THE 25TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 25-27 2013

ATHENS, GREECE

EMSS
2013

EDITED BY

AGOSTINO G. BRUZZONE

EMILIO JIMÉNEZ

FRANCESCO LONGO

YURI MERKURYEV

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2013

ISBN 978-88-97999-16-4 (Paperback)
ISBN 978-88-97999-22-5 (PDF)

ISBN 978-88-97999-16-4 (Paperback)
ISBN 978-88-97999-22-5 (PDF)

The 25th European Modeling & Simulation Symposium

SEPTEMBER 25-27 2013, Athens, Greece

Organized by

DIME – University of Genoa

Liophant Simulation

Simulation Team

IMCS – International Mediterranean & Latin American Council of Simulation

DIMEG, University of Calabria

MSC-LES, Modeling & Simulation Center, Laboratory of Enterprise Solutions

Modeling and Simulation Center of Excellence (MSCOE)

Latvian Simulation Center – Riga Technical University

Logisim

LSIS – Laboratoire des Sciences de l'Information et des Systemes

MIMOS – Movimento Italiano Modellazione e Simulazione

MITIM Perugia Center – University of Perugia

Brasilian Simulation Center, LAMCE-COPPE-UFRJ

MITIM - McLeod Institute of Technology and Interoperable Modeling and Simulation – Genoa Center

M&SNet - McLeod Modeling and Simulation Network

Latvian Simulation Society

Ecole Superieure d'Ingenierie en Sciences Appliquees

Facultad de Ciencias Exactas. Ingegneria y Agrimensura

University of La Laguna

CIFASIS: CONICET-UNR-UPCAM

INSTICC - Institute for Systems and Technologies of Information, Control and Communication

National Russian Simulation Society

CEA - IFAC

AFCEA, HELLENIC CHAPTER

Technically Co-Sponsored

IEEE – Central and South Italy Section Chapter

I3M 2013 Industrial Sponsors

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2013 Media Partners

Inderscience Publishers – International Journal of Simulation and Process Modeling

Inderscience Publishers – International Journal of Critical Infrastructures

Inderscience Publishers – International Journal of Engineering Systems Modelling and Simulation

Inderscience Publishers – International Journal of Service and Computing Oriented Manufacturing

IGI GLOBAL – International Journal of Privacy and Health Information Management

Halldale Media Group: Military Simulation and Training Magazine

Halldale Media Group: The Journal for Healthcare Education, Simulation and Training

Euromerci

EDITORS

AGOSTINO BRUZZONE
MITIM-DIME, UNIVERSITY OF GENOA, ITALY
agostino@itim.unige.it

EMILIO JIMÉNEZ
UNIVERSITY OF LA RIOJA, SPAIN
emilio.jimenez@unirioja.es

FRANCESCO LONGO
DIMEG, UNIVERSITY OF CALABRIA, ITALY
f.longo@unical.it

YURI MERKURYEV
RIGA TECHNICAL UNIVERSITY, LATVIA
merkur@itl.rtu.lv

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2013

GENERAL CO-CHAIRS

AGOSTINO BRUZZONE, *MITIM DIME, UNIVERSITY OF GENOA, ITALY*
YURI MERKURYEV, *RIGA TECHNICAL UNIVERSITY, LATVIA*

PROGRAM CHAIR

FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*

THE 25TH EUROPEAN MODELING & SIMULATION SYMPOSIUM, EMSS 2013

GENERAL CO-CHAIRS

FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA ITALY*
EMILIO JIMÉNEZ, *UNIVERSITY OF LA RIOJA, SPAIN*

EMSS 2013 International Program Committee

Michael Affenzeller, *Upper Austrian Univ. of AS, Austria*
Maja Atanasijevic-Kunc, *University of Ljubljana, Slovenia*
Diego Azofra Rojo, *University of La Rioja, Spain*
Andreas Beham, *Upper Austrian Univ. of AS, Austria*
Julio Blanco Fernández, *University of La Rioja, Spain*
Javier Bermejo, *MTorres, Spain*
Felix Breitenecker*, Technical University of Wien, Austria*
Agostino Bruzzone, *University of Genoa, Italy*
Hipolito Carvajal Fals*, Universityv of Oriente, Cuba*
Priscilla Elfrey, *NASA-KSC, USA*
Maria Pia Fanti, *Polytechnic University of Bari, Italy*
Ernesto Yoel Farinas Wong*, Universidad Central de Las Villas, Cuba*
Idalia Flores*, University of Mexico, Mexico*
Claudia Frydman, *LSIS, France*
Sergio Gallo, Univ. of Modena and Reggio Emilia, Italy
Jorge Luis García Alcaraz, *Universidad Autónoma de Ciudad Juárez, México*
Witold Jacak, *Upper Austrian Univ. of AS, Austria*
Emilio Jiménez, *University of La Rioja, Spain*
Andreas Körner, *Vienna University of Technology, Austria*
Gabriel Kronberger, *Upper Austrian Univ. of AS, Austria*
Carlos Javierre Lardiés, *University of Saragossa, Spain*
Juan Ignacio Latorre Biel, Univ. Pública de Navarra, Spain
Francesco Longo, *MSC-LES, University of Calabria, Italy*
Eduardo Martínez Cámara, *University of La Rioja, Spain*
Marina Massei, *Liophant Simulation, Italy*
Riccardo Melloni, *Univ. of Modena and Reggio Emilia, Italy*
Yuri Merkuryev, *Riga Technical University, Latvia*
Tanya Moreno Coronado, *Uiversidad Nacional Autónoma, Mexico*
Miguel Mújica Mota, *UAB, Spain*
Teresa Murino, *University of Naples Federico II, Italy*
Gasper Music, *University of Ljubljana, Slovenia*
Gaby Neumann, *Tech. Univ. Appl. Sciences Wildau, Germany*
Letizia Nicoletti, *University of Calabria, Italy*
Tudor Niculiu, *University of Bucharest, Romania*
Daniel Niño, *University of La Rioja , Spain*
Tuncer Ören, M&SNet, *University of Ottawa, Canada*
María Otero Prego, *University of La Rioja, Spain*
Mercedes Pérez de la Parte*, University of La Rioja, Spain*
Melchor Gómez Pérez, *University of the Basque Country, Spain*
Miquel Angel Piera, *UAB, Spain*
Simonluca Poggi, *MAST Srl, Italy*
Chumming Rong*, University of Stavanger, Norway*
Juan Carlos Sáenz-Díez Muro, *University of La Rioja, Spain*
Ángel Sanchez Roca, *University of Oriente, Cuba*
Liberatina Santillo, *University of Naples Federico II, Italy*
Boris Sokolov, *Russian Accademy Science, Russia*
Chrysostomos Stylios, *Technological Educational Institute of Epirus, Greece*
Fei Tao, *Beihang University, China*
Alberto Tremori, *University of Genoa, Italy*
Walter Ukovich, *University of Trieste, Italy*
Stefan Wagner, *Upper Austrian Univ. of AS, Austria*
Thomas Wiedemann, *University of Applied Sciences at Dresden, Germany*
Stephan Winkler, *Upper Austrian Univ. of AS, Austria*
Lin Zhang*, Beihang University, China*
Levent Yilmaz, *Auburn University, USA*

Tracks and Workshop Chairs

Discrete and Combined Simulation
Chair: Gasper Music, University of Ljubljana, Slovenia; Thomas Wiedemann, HTW Dresden FB Informatik, Germany

Industrial Processes Modeling & Simulation
Chair: Agostino Bruzzone, DIME, University of Genoa, Italy

Industrial Engineering
Chair: Francesco Longo, DIMEG, University of Calabria, Italy

Agent Directed Simulation
Chairs: Tuncer Ören, University of Ottawa, Canada; Levent Yilmaz, Auburn University, USA

Petri Nets based Modelling & Simulation
Chairs: Emilio Jiménez, University of La Rioja, Spain; Juan Ignacio Latorre, Public University of Navarre, Spain

Simulation and Artificial Intelligence
Chair: Tudor Niculiu, University "Politehnica" of Bucharest, Romania

Workshop on Cloud Manufacturing
Chairs: Lin Zhang, Beihang University ,Beijing, China; Fei Tao, Beihang University, Beijing, China; Simonluca Poggi, MAST Srl, Italy

Simulation Optimization Approaches in industry, services and logistics processes
Chairs: Idalia Flores, University of Mexico, Mexico; Miguel Mújica Mota, Universitat Autonoma de Barcelona, Spain

Human-centred and Human-focused Modelling and Simulation
Chair: Gaby Neumann, Technical University of Applied Sciences Wildau, Germany

Workshop on Soft Computing and Modelling & Simulation
Chairs: Michael Affenzeller, Upper Austrian University of Applied Sciences, Austria; Witold Jacak, Upper Austrian University of Applied Sciences, Austria

Workshop on Cloud Computing
Chairs: Alberto Tremori, Simulation Team, Italy; Chunming Rong, University of Stavanger, Norway

Advanced Simulation for Logistics Systems
Chairs: Maria Pia Fanti, Polytechnic of Bari, Italy; Chrysostomos Stylios,Technological Educational Institute of Epirus, Greece; Walter Ukovich, University of Trieste, Italy

Modelling and Simulation Approaches in and for Education
Chairs: Maja Atanasijevic-Kunc, Univ. Ljubljana, Slovenia; Andreas Körner, Vienna Univ. of Technology, Austria

Modelling and Simulation in Physiology and Medicine (common track EMSS-IWISH)
Chairs: Maja Atanasijevic-Kunc, Univ. Ljubljana, Slovenia; Felix Breitenecker, Vienna Univ. of Technology, Austria

Simulation and Modelling for Occupational Health and Safety
Chairs: Riccardo Melloni, Univ. of Modena and Reggio Emilia, Italy; Sergio Gallo, Univ. of Modena and Reggio Emilia, Italy

Advanced Models and Applications of Logistics & Manufacturing
Chairs: Teresa Murino, University of Naples Federico II, Italy, Liberatina Santillo, University of Naples Federico II, Italy

General Co-Chairs' Message

Welcome To EMSS 2013!

We are sure to speak on behalf of all the people that have provided, along different years and around different countries, their invaluable contributions, when we say that we are glad and proud to be part of the 25th Edition of the European Modeling & Simulation Symposium. EMSS, also known since 1996 as Simulation in Industry, has always involved scientists, researchers and practitioners that, with their effort and enthusiasm, have tremendously contributed to the success of this Symposium creating an ideal forum to share and discuss ideas, propose new concepts and innovative theories, provide evidence on the relevance of Modeling & Simulation as cutting edge technology, bring together people from Industry, Academia, and Agencies.

And this year, again for the 10th time, EMSS 2013 will be co-located with the 10th International Multidisciplinary Modeling & Simulation Multi-conference, I3M 2013, renovating the opportunity to have an International Multi-conference involving different areas and topics related to Modeling & Simulation. This year I3M 2013 includes 7 International Conferences and Workshops (EMSS 2013, HMS 2013, MAS 2013, IMAACA 2013, DHSS 2013, IWISH 2013 and SESDE 2013) providing the attendees with the possibility of exploring thematic Modeling & Simulation areas and joining one of the most important and interesting worldwide events in this field.

We would like to thank for their work and continuous support all the members of the International Program Committee, the Reviewers, the Local Organization Committee, and, above all, the authors that, as every year, have strongly contributed to the success of the Symposium by submitting high quality scientific papers. As follows some of the EMSS 2013 numbers: more than 90 selected papers, 31 countries (including Europe, Latin and North America, Asia, Africa and Australia), 16 tracks, 7 International Journal Special Issues (as part of the I3M Multi-conference) that will publish the best papers of EMSS 2013.

EMSS 2013 is held in Athens, Greece, one of the world's oldest cities, the cradle of western civilization, the birthplace of democracy and a continuous source of science and knowledge, where the organizers wish to all the attendees a fruitful Symposium and a pleasant stay in the city.

Therefore, on behalf of all the people who have made it possible: Welcome to EMSS 2013!

Francesco Longo
DIMEG, University of
Calabria, Italy

Emilio Jiménez
University of La Rioja,
Spain

ACKNOWLEDGEMENTS

The EMSS 2013 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The EMSS 2013 IPC would like to thank all the authors as well as the reviewers for their invaluable work.
A special thank goes to all the organizations, institutions and societies that have supported and technically sponsored the event.

LOCAL ORGANIZATION COMMITTEE

This International Workshop is part of the I3M Multiconference: the Congress leading **Simulation around the World and Along the Years**

Index

MODELLING INTERACTIONS IN A MIXED AGENT WORLD

Wafa JOHAL[(a)], Julie DUGDALE[(b)], Sylvie PESTY[(c)]

[(a) (b) (c)]Grenoble Informatics Laboratory – University of Grenoble - France

[(a)]wafa.johal@imag.fr [(b)]julie.dugdale@imag.fr [(c)]sylvie.pesty@imag.fr

ABSTRACT

The emergence of mixed agent-human societies poses challenges in designing companion agents that are able to form meaningful relationships with humans. This paper describes the first step in developing companion agents. Problem/situations have been identified where companion agents may provide important social contact with humans. Based on a scenario, interactions between human and artificial agents have been modelled in the Brahms modelling and simulation environment. This provides us with a far deeper understanding of the roles that a companion agent should fulfil and how it could switch from one social role to another.

Keywords:
Companion agents, human-agent world, scenarios, human-agent relationship, virtual characters, robots, Brahms

1. INTRODUCTION

We are entering a new era of computing where software agents are becoming increasingly prevalent in our environment; they are found in the technological supports that we use, and are manifested as embodied conversational agents or robots. This is leading to an inevitable increase in interactions between artificial agents and humans resulting in the emergence of mixed human-agent societies. However a common problem of artificial agents is that they fail to establish any meaningful relationship with the user. In order to achieve acceptance and to create value from adopting a new technology, the creation of a meaningful relationship is essential. Our work is conducted in the context of the French ANR funded MOCA project[1] where the overall goal is to construct a mixed agent society, composed of companion agents, such as robots and virtual characters, as well as humans. In this society, human-agent relationships will take the same form as with human-human relationships.

In order to establish a long term relationship, we propose to integrate personality and social concepts into the world of artificial companions. Several studies in the literature (Bickmore 2005, Grandgeorge 2011) note that one of the most important challenges raised by new technologies is to provide a new type of human-machine interfaces that could create and maintain new types of relationship (Pesty 2011) with humans.

Technological advances in robotics have developed what is called 'Service Robots', which assist humans in performing useful service, sometimes in home situations. Such robotic devices interact with the consumer in a homely environment. As robots move beyond just helping us with household chores, we must start to question the nature of our relation with robots. The MOCA project aims to explore how can we design these daily life companions in order for them to improve our quality of life even if we are not expert in new technologies.

The work presented in this paper describes the first step in the development of companion agents. We define two aspects of the companions: the *role* that they play in a problem/situation; and their embodiment in the *device*, such as a robot or virtual agent.

Following the user's choice, the companion will express itself through a device such as a Reeti (Robopec) or Nao Robot (Aldebaran 2009); or Mary (Courgeon et al. 2008) or Greta (Poggi et al., 2005) virtual character (figure 1).

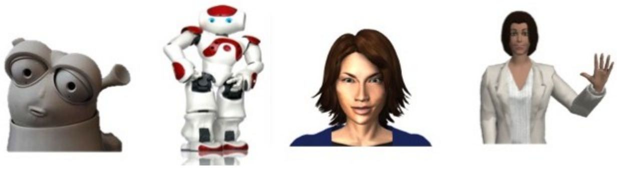


Figure 1: Companion Agents: Reeti, Nao, Mary and Greta

The roles cover the expected behaviours of the companion(s) to respond to a problem or situation encountered by the user. We aim to enrich the interaction between these agents by taking into account the social context. Classically in Human-Machine-Interaction, context-aware technologies take into

[1] "My Little Artificial Companion World" MOCA project, ANR-2012-CORD-019-02.

account 3 variables: the environment of use, the user and the platform used. Research in context-aware technologies focuses mainly on how to characterise the context in order to adapt the service to it. Closer to our work Calvary and her colleagues, speak about the plasticity of interfaces towards a context of use (Calvary 2001). Aiming to have companions (robot or virtual agents) that are socially intelligent, we also want to integrate social rules (family rules, social roles, personality) in the context model.

2. RELATED WORK

Virtual companion agents and especially robots are still very costly. Progress in terms of new features and increased performance of the companions is rapid and new devices, with more features appear on the market everyday. However, as shown by previous unfortunate experiences, the acceptability and adoption of new technologies by users is vital and is not only matter of innovation (Leonardi 2009, Dubois 2009). In the literature, multi-agent scenario-based methods of development are widely described and used (Iglesias 1999). More particularly, agent based modelling and simulation allows us to test and understand at an early stage of those requirements that are difficult to envisage other than by fully developing the robot or the virtual character and testing it in a real situation.

Barreteau (2003) develop the approach of companion modelling dealing with “collective decision making process of stakeholders sharing a common resource”. The principle is to iteratively build models and use mediation process for collective learning. The notion of companion agent as we envisage differs. We use the term of companion to insist on the stability and long term relationship between the virtual agent and the user.

Modelling and simulation involving interactive robots often focus on the motor level, and studies in that vein usually aim to test controllability and limitations of a new technology (Dudenhoeffer 2001). Aiming to develop companion agents that are socially intelligent, we chose a higher level of modelling using the BDI (Belief-Desire-Intention) framework in order to introduce notion of personality and emotions in the reasoning of the agent (Adam 2007).

Role assignment and cooperation are still problematic and are subject to a lot of research in multi-agent systems (Campbell 2010, Kwak 2012). Roles are often used to simplify the problem of cooperation between multiple agents having to respond to a set of tasks. The role is then described as a specification of the agent in accomplishing the task and strategies of assignment of some corresponding task can be chosen.

In our work, we consider the role of the agent as principally being a social role. A same agent can then put on a different role according to the context. Assignment or redundancy will be dependent on the social role, and a pre-design user study will be conducted in order to determine how roles may be the best distributed among our multi-agent system. The social role of a companion will be both the functionality and the set of tasks it can accomplish to respond to a goal, together with its relationship with the user (Clavel 2013). The focus of this study is to identify and model these roles as well as finding triggers that invoke a role.

3. PROBLEM / SITUATION

We have adopted a scenario-based development approach (Rosson and Carroll, 2002), supported by Worth centred design (Cockton, 2004) (Cockton, 2006). Following these, the first step is to identify problem/situations. A problem/situation is a situation in which our system can provide a service that will facilitate or reply to a problem faced by the user. We have chosen to focus on how children, in the 8 to 12 year age bracket, would interact with the companion agents in a family setting. The problem-situation is a context within which the children might need help, and where they could find worthiness, in Cockton’s terminology, by using the companion. The companion then takes a role in order to respond to the problem or situation.

Based on a previous Robofesta[2] survey (Clavel et al. 2013) we elicited 5 problem/situations that a child may encounter and the associated abilities that a companion agent would need to fulfil in that role (Table 1).

Problem/ Situation	Description	Companion abilities
Teaching (Prof)	adapted help and support for *homework* and school matters	• Should have the knowledge and the expertise to help the child in the homework task. • Should motivate and reward good performance; critique and discourage bad performance. • Should be able to interpret the mental state of the child to give him/her appropriate feedback or treatment. • Should be able to track the child’s performance over time to monitor their progress and adjust pedagogic parameters • Should be able to summarize the accomplished daily and communicate this to the parents or teacher.

[2] RoboFesta is an International Organisation established to promote the study and enjoyment of science and technology through hands-on, robot-related events.

Problem/ Situation	Description	Companion abilities
Playing (Buddy)	need a friend to play with	• Should suggest stimulating games • Should suggest and play: game for creativity and imagination • Should make possible group games • Could joke (Simple jokes)
Guarding (Bodyguard)	need to feel more secure	• Should be able to start an alarm • Should be able to call the parents, or emergency services • Should reassure the child • Should provide advice on how to react to the situation
Comforter (Dolly)	need for a cuddle, affection, comfort	• Should be able to perceive child's mood (alert parents if necessary) • Should listen and give advice
Coacher (Coach)	need to be coached to discover extra-curricular activities (learn knowledge, other than that taught in school)	• Should encourage activities • Should give instructions to do activity in security • Should be able to supervise the activity

Table 1: Problem/situations and abilities

Normally such situations would involve a parent or guardian. However, financial pressures on the family mean that parents increasingly have to work and are unavailable for child-care. In this case the companion agent may be used to provide social contact that has been shown to be important for cognitive development (Piaget 1966, Vernon 2011). Other problem/situations are possible but the above were chosen based on interview previously conducted and (Clavel 2013) and because they occur frequently in everyday life.

4. SCENARIO

The next step was to devise a scenario that incorporated the above problem/situations and that highlighted the interactions between the companion agents and the children. It should be noted that a role, e.g. Prof, can be deployed through one or many forms (e.g. virtual agents and/or robots agents) that will collaborate and cooperate in order to accomplish the tasks and to reply to the needs of a specific problem/situation. Below is a natural language description of the scenario.

Ben is 11 years old and in his first year of middle school. His father and mother both work until 8pm. Ben finishes school at 4.30pm. The school is a few streets away from home and Ben usually walks home with some friends every evening. Ben usually has homework to do every evening. His school grades are average but with more help from his parents and teacher they would improve. Although Ben knows that he should do his homework he prefers to watch TV or play video games. In the evening, his neighbour, Alan, usually comes over to play. Ben's parents don't really like him being alone at home, but they have heard about the MOCA system and they already have some devices (virtual characters) at home. Ben would love to have robot companions and so his parents decided to buy him the one he liked from the big city supermarket. They downloaded the MOCA software onto his already existing devices. The MOCA system deploys itself forming a world of companions that can be with Ben in the evenings. The parents configure the world with different roles according to their needs. They download:

- *Playing software, a perfect pal to play with Ben when he is alone (avoiding the video games)*

- *Comforting software, in case Ben feels sad and needs some comfort*
- *Teaching software, which will help Ben with his homework, and to organise and keep track of schoolwork*
- *Coaching software will help Ben in extra-scholar activities (preparing the snack, music lessons).*

- *Finally, in case of problems, Security software*

Ben would love to learn music with 'Coach'. When Ben gets hungry he usually goes to the kitchen and gets a snack He prefers chocolate bars rather than fruit, but the coach usually reminds him that he won't have a dessert after dinner if he didn't have his apple. The activities of the Coach can also be extended by Hip-Hop lessons and Ben would like to be given that for Christmas.

COACH

The house rule is that around 5pm and before playing any game, Ben should have done his homework. Prof proposes to help with the homework and informs the other companions when it is finished. Prof also gets information from the parents and the school agenda. The information is related to the subject that Ben needs to study. Prof makes a synthesis of the work accomplished by Ben and gives a summary to Ben's parents or teacher if they ask for it. The results are added Ben's diary that is managed by the Cloud.

PROF

The Prof encourages Ben to do his homework with care. When the Prof encourages Ben, his motivation increases and he believes more that he can complete the task. Nevertheless Ben can be a bit stubborn, and sometimes Prof needs to threaten Ben with calling his parents in order to try to make him do his work.

Alan and his artificial companions are pretty good at strategic games, and Ben doesn't win

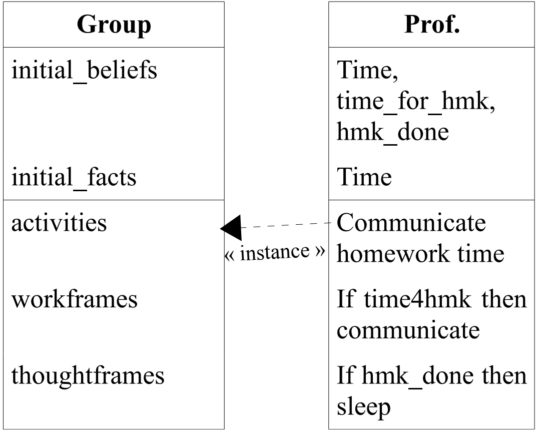
often because Buddy is new in the house and a bit shy. Nevertheless having Buddy and Poto means that there are more 'people' to plays games, and they can all play together.	BUDDY
When Ben looses he is always a bit sad, but Dolly is there to cheer him up and to play some nice songs that take his mind off loosing. Dolly is very sweet, and Ben knows that he can share his secrets with her.	DOLLY
Being home alone, Ben feels reassured when Bodyguard advises him on what to do and check who is at the door before opening it	BODYGUARD

5. SCENARIO MODELLING IN BRAHMS

The aim of modelling the interactions is to frame clearly the companions' roles and their interactions. We have chosen to use BRAHMS (Business Redesign Agent-Based Holistic Modelling System) as a modelling tool (Sierhuis et al. 2003). BRAHMS is an agent oriented language and development environment for modelling and simulation. Brahms is able to represent, people, places, objects, behaviour of people over time and their social behaviours [Sierhuis et al. 2007]. In support, Brahms provides several models with which the developer can specify their world: agent, object, activity and geography. Furthermore BRAHMS has similarities with a BDI (Belief-Desire-and-Intention) approach (Georgeff & al. 1998) in that it allows goal-oriented behaviours and the manipulations of beliefs.

BRAHMS is structured around the following concepts (given in italics) (Brahms tutorial, 2003): *Groups* contain *agents* who are located and have *beliefs* that lead them to engage in *activities*. The activities are specified by *workframes* that consist of preconditions of *beliefs* that lead to *actions* (consisting of *communication actions*, *movement actions*, *primitive actions*) and other *composite activities*, consequences of new beliefs and world facts, *thoughtframes* that consist of preconditions and consequences. Through the use of a time-line we are able to analyse the individual behaviours and interactions of each agent.

In the geography model we model the physical environment of the neigbourhood, including the school and the house, the latter of which is divided into rooms with linking pathways. The children, adults, and artificial companions are all modelled as agents (specified in the agent model), each agent having their own characterising attributes and beliefs. All agents are part of the GroupWorld; this allows us to define general activities, attributes and reasoning process shared by all agents. Groupworld contains three main groups: Adults, Children, and Companions, each group has their own specific needs, locations, actions (the abilities), e.g. adults can be at the office, and children have homework and need to play. Thus the reasoning and abilities of companion agents are dependent on their role. Figure 2 shows how we can instantiate a group in Brahms modelling language into a role, with specific beliefs, activities (tasks), workframes (functionalities) and thoughtframes. In this example, the Prof has a belief about the time for homework. If it is time for homework he can communicate the need to to the homework to another agent.

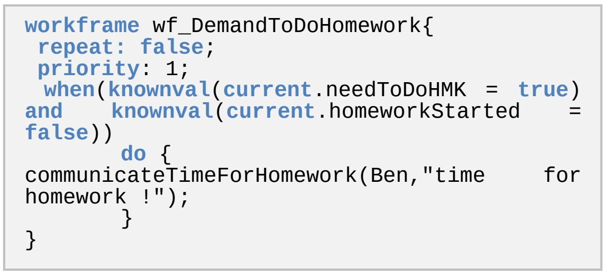


Group	Prof.
initial_beliefs	Time, time_for_hmk, hmk_done
initial_facts	Time
activities	Communicate homework time
workframes	If time4hmk then communicate
thoughtframes	If hmk_done then sleep

Figure 2: Group in Brahms instantiated as the role Prof.

In the scenario we define 4 kinds of activities: primitive activities that are defined by their duration (e.g. the activity of listening to a song); move activities that specify a goal location, such as a particular room, and which uses the geography model; communicate activities, defined by a receiver and a message; broadcasting activities, which allows communication to all agents in the same location as the broadcasting agent.

Workframes contain the actions of the agent with associated preconditions and consequences. For example, the Prof agent may have the workframe symbolising the rule: if there is the need to do the homework and the homework hasn't started yet, Prof should communicate with Ben to tell him that it is time for the homework (figure 3).



```
workframe wf_DemandToDoHomework{
 repeat: false;
 priority: 1;
  when(knownval(current.needToDoHMK  =  true)
and    knownval(current.homeworkStarted    =
false))
             do {
communicateTimeForHomework(Ben,"time     for
homework !");
           }
}
```

Figure 3: A Prof agent Workframe

Thoughtframes allow the manipulation of agents' beliefs and adding uncertainty to a belief (e.g. a belief may only be held 75% of the time. This could be interpreted as an agent 'changing its mind' and ultimately means that each run of the simulation can differ. In figure 4 below we see a Brahms screenshot showing the situation when the prof reminds Ben at 5pm, after he's been watching TV, that it's time to do his homework.

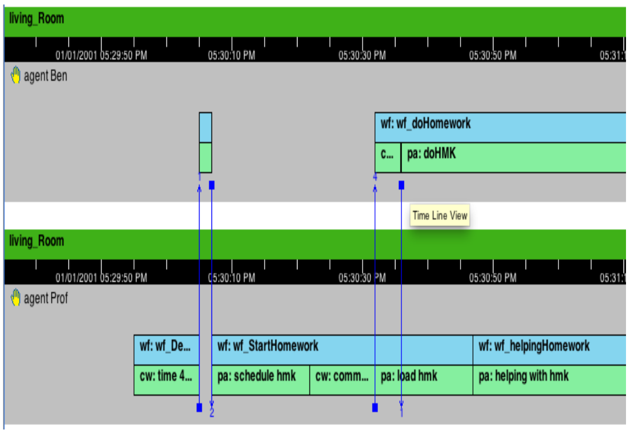


Figure 4: Teaching situation.

The location "living_Room" is shown for each agent (Ben and Prof), together with the date and time, horizontally at the top and middle of the figure. *wf*, *cw*, and *pa* refer to workframe, communicative activity and primitive activity respectively. Blue vertical arrows show the communication between agents; we have made the content of the communications explicit, but they may be seen by clicking on the arrows.

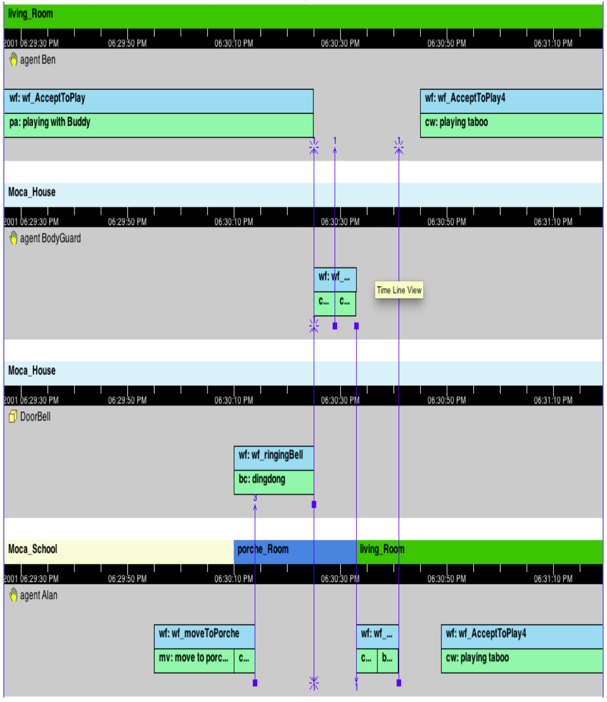


Figure 5: Guarding situation (someone at the door). The Bodyguard agent enters the action, followed by group games with companions and children.

Figure 5 shows a screenshot where, the doorbell, modelled as an object, suddenly rings. Ben is a bit scared. The security agent, called Bodyguard in the figure, sees that it is Alan, who brought a companion with him. Knowing them (modelled as a belief), the Bodyguard lets them in and reassures Ben, telling him that Alan is here with Poto. Buddy and Poto suggest making teams with Alan and Ben to play strategy games, one of Poto's favourite games.

We modelled each role as a group of agent sharing abilities (workframes and thoughtframes). Since in BRAHMS agents are situated, this allowed us to detach the role from the device and also to instantiate the role by more than one situated companion (Poto and Buddy belong to the Playing Group). Indeed, we can imagine one situated agent member of all of the group, being able to accomplish all the roles. Figure 6 below shows how the Buddy agent is member of both the Coach and Playing groups, and hence it can play both roles according to the context.

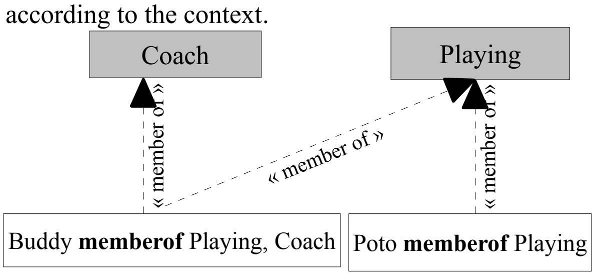


Figure 6: Example of memberships of Buddy and Poto agents

6. DISCUSSION

Modelling and simulation the scenarios in BRAHMS had the advantage of highlighting the behavourial variability that we have to face for the design of the MOCA system. Indeed, variability of context of use (Calvary 2001) composed by the user, the platform, and the environment, shows the importance of the predefined family rules that will help the companion agent to take a decision according to its role. In this study, a family is a mixed group of companions and human that will share some beliefs and throughtframes manipulating these beliefs (figure 7).

```
group Family_One memberof Families {
...
        initial_beliefs:
(current.time_for_homework = 5);
(current.Kids_allow_to_watchTV = false);
...
```

Figure 7: Listing extracted from the group Family_One (Brahms file)

In order to be able to simulate the variability of the device that can play a role, we consider the social role as a group that can be instantiated by several situated agents.

By detaching the role from the device allows us to instantiate several roles in one device. This highlighted the importance of context into the decision of the agent of taking a role. Indeed, as a nanny can help with the homework, she can also be a play buddy after this task is done.

One issue that has been raised by our simulations is the coordination of multiple companion agents.

Indeed, several strategies can be chosen when a role has to be played when several agents are capable of playing this role. We may enrich the model by adding cooperation, assignment, or redundancy decision in the role played by multiple companions. This will depend on the context and some roles will need redundancy (Bodyguard) in order to insure detection whereas other role will offer more benefit with cooperation (Prof).

7. CONCLUSION AND FUTURE WORK

We have shown how interactions in a hybrid society, composed of human and companion agents may be modelled in Brahms, following a definition of problem/situations and a scenario. We are currently in the process of validating the usefulness of these problem/situations via prototyping the companion agents on physical devices (commencing with Reeti and Nao). In order to assess their usefulness, we use the notion of worth, defined as "the value for the user of the system" (Cockton, 2004) (Cockton, 2006). Cockton proposed Worth Centred Design (WCD) framework. WCD focuses the development on increasing the worth of using the developed system. Indeed, some factors such as the appearance, of the system impact on the user experiences with the system. This influences its motivations, its worth in using the system. Where User Centred Design classically considered as factors influencing user experience primarily the functionalities and the ease of use, WCD expends the set of factors to sociologic, emotional or economical factors.

Practically, we are implementing the 'play' (Buddy) situation in a prototype, which will then be tested with sample users; their opinions will then be gained through post-experimentation questionnaires and interviews. The aim of this step is to see if the problem/situations that we have identified are really what users want from companion agents. In parallel we are investigating the worth of adding personality to our companion agents. Personality is a key element of establishing relationships (Mischel et al. 2004), and provides stability in terms of recognizing and relating to the companion. This personality will be expressed when accomplishing the roles and through devices. The addition of personality extends the bipartite composition of the companion agent, from role and device, to one now composed of three parts (figure 8). Ultimately this allows users to choose personalities for their agent.

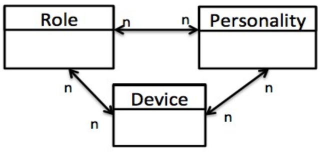


Figure 8: Tripartite aspect of the companions

Once we have established the worth of the system, we will evaluate the interactional capabilities of the agents and the role of personalities through simulation. This may be achieved by extending the Brahms model to run full simulations. This will allow us to design more effective interaction functionalities before further implementation occurs. Thus simulation will be used as an aid to designing companion agent.

ACKNOWLEDGMENTS

This work is conducted as part of the French funded ANR "My Little Artificial Companion World" MOCA project, ANR-2012-CORD-019-02. Julie Dugdale would also like to thank the University of Agder, to which she is affiliated.

REFERENCES

C. Adam. "Emotions: from psychological theories to logical formalization and implementation in a BDI agent". PhD Thesis. Institut National Polytechnique de Toulouse, 2007.

Aldebaran-robotics. N. A. O. http://www.aldebaran-robotics.com/eng

Brahms tutorial, 2003 http://www.agentisolutions.com/documentation/tutorial/tt_title.htm

O. Barreteau & al. (2003). Our companion modelling approach. *Journal of Artificial Societies and Social Simulation*, *6*(2), 1.

T. W. Bickmore, & R. W. Picard, Establishing and maintaining long term human-computer relationships. ACM Transactions on Human Computer Interaction, 12(2), 2005.

A. Campbell, & A. S. Wu, Multi-agent role allocation: issues, approaches, and multiple perspectives. *Autonomous Agents and Multi-Agent Systems*, *22*(2), 317–355, 2010.

C. Clavel, C. Faur, S. Pesty, J-C. Martin, D. Duhaut. Artificial Companions with Personality and Social Role. *Proc. of the 2013 IEEE Symposium on Computational Intelligence for Creativity and Affective Computing (IEEE CICAC 2013)*, Singapour, 2013.

G. Cockton, "From quality in use to value in the world," *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, p. 1287, 2004.

G. Cockton, "Designing worth is worth designing," *Proceedings of the 4th Nordic conference on Human-computer interaction : changing roles* no. October, pp. 14–18, 2006.

M. Courgeon, Jean-Claude Martin et Christian Jacquemin. MARC : a Multimodal Affective and Reactive Character. *In Proc. of The 1st workshop on Affective Interaction in Natural Environments* (AFFINE), Chania, Crete, 2008.

G. Calvary, J. Coutaz, and D. Thevenin, "Supporting context changes for plastic user interfaces: a process and a mechanism," People and Computers, 2001.

M. Dubois, & M.-É. Bobillier-Chaumon, (2009). L'acceptabilité des technologies : bilans et nouvelles perspectives. *Le travail humain*, *72*(4), 305.

D. D. Dudenhoeffer, J. B. David, and L. D. Midge. "Modeling and simulation for exploring human-robot team interaction requirements."*Proceedings of the 33nd conference on Winter simulation*. IEEE Computer Society, 2001.

M. P. Georgeff, B. Pell, M. E. Pollack, M. Tambe, M. Wooldridge. 1998. The Belief-Desire-Intention Model of Agency. In *Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages* (ATAL '98), Jörg P. Müller, Munindar P. Singh, and Anand S. Rao (Eds.). Springer-Verlag, London, UK, UK, 1-10.

M. Grandgeorge, D. Duhaut, (2011). Human Robot: from Interaction to Relationship. *Proccedings of The 14th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines* (CLAWAR2011). Paris.

C. Iglesias, M. Garijo, and J. González, “A survey of agent-oriented methodologies,” framework, pp. 317–330, 1999.

S. Kwak, B. Han, & J. Han, Multi-Agent Event Detection: Localization and Role Assignment. *Computer Vision and Pattern Recognition.* 2013.

P. M. Leonardi, (2009), Why Do People Reject New Technologies and Stymie Organizational Changes of Which They Are in Favor? Exploring Misalignments Between Social Interactions and Materiality. *Human Communication Research*, 35: 407–441.

W. Mischel, Y. Shoda and R. E. Smith (2004). Introduction to Personality: Toward an Integration (7th ed.), Hoboken, NJ: J. Wiley & Sons.

S. Pesty, D. Duhaut, Artificial companion: Building a impacting relation, *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on* , pp.2902,2907, 7-11 Dec. 2011.

J. Piaget, B. Inhelder, *The Psychology of the Child* (New York: Basic Books, 1962) .

I. Poggi, C. Pelachaud, F. de Rosis, V. Carofiglio, B. de Carolis. GRETA. A Believable Embodied Conversational Agent. In Patrizia Paggio et Bart Jongejan, editeurs, Multimodal Communication in Virtual Environments, pages 27–45. Springer, 2005.

Robopec 2010 http://robopec.com/index.html

H. Prendinger., & M. Ishizuka, (2001). Social role awareness in animated agents. *International Conference on Autonomous Agents,* (pp. 270–277).

M.B Rosson & Carroll, J.M., 2002. Usability engineering: Scenario-Based Development of Human-Computer Interaction. San Francisco: Morgan-Kaufmann.

M. Sierhuis, W. J. Clancey, and R. v. Hoof, "Brahms: a multiagent modeling environment for simulating social phenomena", presented at First conference of the European Social Simulation Association (SIMSOC VI), Groningen, The Netherlands, 2003.

M. Sierhuis, W. J. Clancey, R. van Hoof, Brahms - a multiagent modeling environment for simulating work practice in organizations, International Journal of Simulation and Process Modelling 3(3) (2007) 134-152.

D. Vernon, C. Hofsten, and L. Fadiga, *A Roadmap for Cognitive Development in Humanoid Robots*, vol. 11. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

SIMULATION FOR SAFETY ENGINEERING: A COMPARISON BETWEEN EXPERIMENTAL DATA AND FIRE MODELS

Govoni Andrea[(a)], **Davoli Giovanni**[(b)], **Gallo Sergio A.**[(c)], **Melloni Riccardo**[(d)]

Department of Engineering “Enzo Ferrari”, University of Modena and Reggio Emilia, Via Vignolese 905, Modena, Italy

[(a)]andrea.govoni@unimore.it, [(b)]giovanni.davoli@unimore.it, [(c)]sgallo@unimore.it, [(d)]riccardo.melloni@unimore.it

ABSTRACT

A comparison analysis has been conducted between experimental data of a concrete structure and three different models developed with CFAST, by NIST. This work concentrates on the possibility of modeling simplified fire objects. Full-scale experiments are simulated by two-layers zone models. Through a reverse approach on heat release rate estimation, a comparison between different models is reported analyzing time to flash over, maximum temperature, shape of the time-temperature curve. The possible use of these models as an engineering design tool is carried out. Results show that different ventilations lead to an uncertain time-temperature response. The benefits that should come from decomposition of fire load into many fire objects is counterbalanced by reduced accuracy due to increased model complexity and interactions between fire plumes.

Keywords: fire modeling, CFAST, safety engineering, two-zone model.

1. INTRODUCTION

Fires are usually very complex to analyze. The complexity arises because of simultaneous presence of phenomenon which controls fire and smoke development, like combustion, turbulence, radiation, convection, etc. Reduced-scale experiments although provide useful information, yet they alone are not sufficient to reproduce full-scale features. A better understanding can be obtained by carrying out full-scale experiments, but they are expensive. Therefore, mathematical modeling can be used for reduce the needs of experiments. However, mathematical modeling should be validated by full- or reduced-scale experiments, wherever possible, in order to achieve a practical solution. A comparison between two-layers zone models and full-scale experiments is provided in this paper. Like all predictive models, the best predictions come with a clear understanding of the limitations of the model and care in the choice of data provided to the calculations. A number of models have been proposed (Olenick and Carpenter 2003) to assist fire safety scientists and engineers to predict fire growth and smoke movement, considered as essential components in fire risk analysis. These models can be grouped into three basic types: field models or CFD (Computational Fluid Dynamics), zone models and hybrid models (Yao et al., 1999). Compared to zone models, field models are relatively younger and usually require high setup, prohibitively high computational resources or unacceptably long computing time. They divide the space of interest into a large number of control volumes in space and steps in time, and then apply and solve a set of partial differential conservation equations over these volumes and steps to produce results with high temporal and spatial resolution. In contrast, zone models such as CFAST have a longer development history. They consider the regions of interest as a whole or divide the regions into only a few zones, assuming that the gas phase inside each zone is well stirred and uniform. Then, a set of ordinary differential or algebraic conservation equations are solved to obtain average variables for each zone (Zhang and Hadjisophocleous 2012).

The main interest of this research is to verify if the accuracy of model results obtained by simplified fire objects modeling and fast simulation time is suitable to use them in design phase. Zone models can give satisfactory although approximate results at a lower cost, while multi-layer zone models have been proposed (Suzuki et al. 2003, Xiaojun et al. 2005), the most frequently used zone models are still the single-layer and two-layer zone models. While single-layer zone models cannot be used in pre-flashover fires due to the assumption of homogeneous properties in the zone (Luo 1997). Two-layer zone models, such as CFAST (Jones et al. 2009), are often used. Two-layer zones concept was proposed in early 1960s (Cox 1995). However, two-layer zone approach emerged in the mid-1970s when fire development was intensively studied. Since then, many two-zone models have been proposed and reviewed; while they are different in features and details, they are similar in basic treatment, assumptions, and sub-models.

A building in which a fire occurs inside is considered for analyzing fires in growth phase and steady burning period. Our aim is to reproduce its geometry, developing a model within the software, analyzing the fire load, and studying the time-temperature response by comparing it with the experimental data obtained as a result of the full-scale destructive tests performed at BRE (Building Research Establishment) Cardington, Lennon and Moore (2003).

The 8 tests have been undertaken in a compartment with overall dimensions of 12 m x 12 m x height 3.4 m, they have investigated the influence of compartment linings, fire load type and through draft condition on the severity of fully developed, post-flashover fires. By this tests, fire development in the pre-flashover phase and smoke movement are investigated too. This experiment was chosen because the building is very similar not only in the geometry but also in the fire load, to many warehouses of small enterprises or big offices. In Italy, compartments over 100 m^2 and below 200 m^2 are very common. The fire load of 5760 kg wood equivalent (97920 MJ) is typical for small warehouses.

A common limitation of fire models is the estimation of heat release rate (HRR). The HRR is a critical parameter to characterize a fire, it can be estimated with different methods that are usually expensive and destructive. The most widespread techniques are based on mass balance if the heat of combustion of the fuel is known. If the burning material is unidentified, calorimetric principles can be used, relying on oxygen consumption or carbon oxides generation measurements. In the last tests provided in this study a reverse approach was chosen avoiding the need of HRR estimation.

2. THE MODEL

The Consolidated model of Fire growth And Smoke Transport (CFAST), developed by NIST, is a multi-room two-layer zone model, with the capability to model multiple fires and targets. This model divides a space into two layers: an upper, hot layer and a lower, cooler layer, considering obvious temperature gradients and hence buoyancy-induced stratification. It can be used to calculate, during a fire scenario, the evolving temperature and distribution of smoke and fire gases throughout a building. CFAST is a merging of ideas that come out from FAST and CCFM.VENTS projects.

CFAST, as many other two-layer zone models, considers heat and mass transfer from the lower layer to the upper one and even downward heat and mass transfer due to venting flows, nevertheless it ignore the mixing that pass through the interface between the two layers due to the temperature difference between the layers. According to Zhang and Hadjisophocleous (2012), both experimental data and results of numerical simulations of field models showed the existence of mixing at the interface, the lack of which is identified as an important limitation of two-layer zone models. The mixing maybe caused by natural convection as a result of the temperature difference between the two layers and the circulating flow resulted by the plume induced flow. Comparisons between experiments and modeling results (Remesh and Tan 2007, Peacock et al. 2008, Fu and Hadjisophocleous 2000) suggest that it may result in over-prediction of the upper layer temperatures and under-prediction of the lower layer temperatures. All the aspects that were considered in the development process of the model and in the software limitations are presented below.

2.1. Compartment geometry and thermal properties

The building used in the experiment has a regular shape which can be easily recreated in the model. Walls and floor are plain, only the ceiling is built with concrete slabs and its shape is more complex. CFAST is able to manage only plain surfaces for the ceiling, so it was modeled as a plain horizontal surface. Authors agree that is a minor limitation of the software and this should not considerably affect the results. The building is composed of only one compartment, so surface connections are not needed. Surface connections in are a system to consider gas species and energy flux between compartments.

The materials database available in CFAST is not very wide, therefore many materials were added inserting the specific characteristics of the products used in the experiment.

2.2. Flow vents

The building has only horizontal natural flow connections in the front and rear walls. In the model, openings are differently managed if we consider natural or forced ventilation. In this paper mechanical flow vents and vertical flow vents were not tested, because of in the experiment there was not fans, hatches, floor or ceiling holes.

Two cases were analyzed:

- front openings only, 2 openings, height 3.4 m, width 3.6 m
- front and back openings, 4 openings, height 1.7 m, width 3.6 m

In both cases, area remains 24.48 m^2, the opening factor changes because of different height. Even though predictions can be affected because two-layers zone models neglect radiation losses from room openings, which is not entirely insignificant at high temperatures, there are no other significant software limitations for modeling the building we are interested in.

2.3. Fires

Modeling fires is critical, every fire object is generated by a wide set of parameters, many of them are recognized to be very significant in literature. For the plume, McCaffrey model was used, however it was noticed no significant difference by the use of Heskestad in these models. CFAST has built in two types of fire modeling structure.

2.3.1. Fire modeling structures

The first one is to create a new database object with HRR data from a destructive test, it can be very accurate because it is possible to generate the HRR curve with high level of detail. Nevertheless destructive tests are expensive and their cost and complexity is usually too high for a predictive model. Moreover, the uncertainty of the model could be high even if the fire was modeled from a destructive test. The internal database for fires is very limited for using this type of fire modeling in many scenarios, also data from bibliography are not as much as they should be for this aim.

The second fire modeling method let us generate an HRR curve starting from 4 parameters. In real fires, the initial fire development is always accelerating, and a suitable way to describe this is to use the t-squared fire (i.e., HRR= $\alpha \cdot t^2$) (Karlsson and Quintiere 2000). The HRR curve is composed of two parabolic parts during growth and decay period and constant value in steady burning as shown in Fig. 1. This method is less accurate than the first but more easy to use as an engineering design tool. The parameters used to develop t-squared fires are:

- Time to 1 MW
- Maximum HRR
- Steady burning period
- Decay period

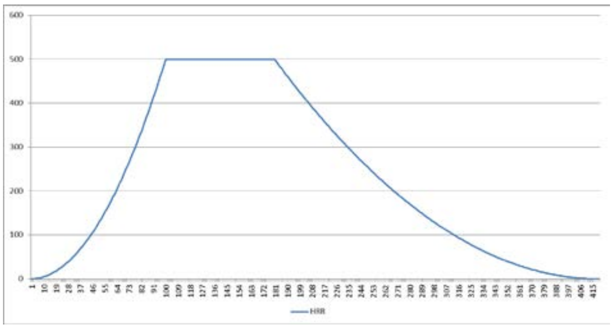


Figure 1: t-squared fire model

Both methods have many limitations in the software and because modeling fires is critical many alternatives have to be evaluated as to pass over software limits. This work concentrates only on the second method. Even if the experiment was conducted to ensure a rapid development of fire, it is possible to recreate this particular condition with CFAST.

2.3.2. Number of fire objects

In all the experiments 49 timber cribs burn in the compartment, but CFAST is limited to a maximum of 31 fire objects, so it is not possible to model every crib as a fire object. The simulation model must be simplified and this situation frequently happens when the building is big enough to contain a lot of furniture, accommodation and stored goods. In the model different fire objects were tested as to compare what solution is more suitable to simulate the experimental fire load. Modifying the number of fire objects, results consistently change.

In the first model only one fire object burns, so that its characteristics are equivalent to all 49 burning wood cribs. In the last model, 30 fire objects are distributed on the floor (with regular pattern) to maintain energy density per floor area equal to the experiments.

2.4. Detection / Suppression

The software provides tools for simulate smoke and fires detectors, sprinklers and suppression systems. None of these systems was used in the model because experiments were conducted without suppression systems.

3. METODOLOGY

A comparison between experimental and simulative results was conducted focusing on time-temperature response.

The accuracy of simulative results was analyzed in this paper. According to Lennon and Moore (2003) the fire growth and steady burning period was chosen as the limit for the comparison. Moreover two-layers zone model are a limiting factor in the prediction of the decay phase. Experimental results are provided every 60 seconds, so, for each step, relative error between experimental and simulative results was calculated. It would also be possible to generate a time-temperature curve by interpolation, using error analysis for continuous functions instead of discrete values, but this approach would create an intrinsic error due to interpolation itself.

For every test, the maximum temperature in the compartment (T_{max}), the time necessary to reach it and the steady burning period, were compared.

3.1. Test A

In test A, a model was created to simulate Cardington experiment 2. The purpose of this test is to evaluate if a fire load composed by only one fire modeled with the $\alpha \cdot t^2$ equation is suitable to fit experimental results. If the fire load is modeled with the generation of HRR in detail it would be possible to create a time-temperature response that fits almost perfectly the experimental one. By the use of the simplified t-squared method it is possible to create a series of fire objects with different values for all parameters (Time to 1 MW, Maximum HRR, Steady burning period) as to generate the most accurate time-temperature curve. Through a set of trials, a model that fits the experimental results with the highest accuracy is provided. The decay period is an useless parameter because we are not considering that phase in the comparison, so the number of trials is reduced.

By this approach, it is possible to minimize model error due to HRR response. According to Au, Wang and Lo (2007) it is found that the characteristic rate of heat release per unit area has the most critical effect on CFAST calculations. The parameters that minimize model error can be used as a baseline in next tests, therefore by this reverse approach there is no need to estimate HRR.

3.2. Test B

One of the more difficult aspects to analyze during the validation of a model is the capability to produce accurate results when the geometry changes. Cardington experiment 2 differs from 4 only for the shape of the building. In the experiment 2, there are only front openings, while in the experiment 4 the building has front and back openings. Openings area over compartment volume ratio remains unchanged.

Considering the fire load calculated by test A, it is possible to compare results of two models that differ only in geometry. Having reduced to the minimum the error on fire load, without need to estimate HRR, the model with the new geometry should fit experimental 4 curve.

3.3. Test C

Two-layers zone models are more accurate when the fire load is decomposed using as many fire objects as the real scenario. Nevertheless CFAST predicts fire development with higher accuracy when the fire object is in the centre of the compartment. The aim of this work is to assess this last statements and define how essential is the decomposition of fire load. A model was developed with a regular pattern of 30 identical fire objects. Using multiple fires let us spread over the compartment all the fire objects in order to recreate a pattern more similar to the experimental one. The energy density per floor area is the same as the previous tests and the experiment. The results of this test compared to Test A and experimental data show how much the prediction is influenced by the number of simplified fire objects.

4. RESULTS

4.1. Test A

A series of trials was conducted to evaluate the α · t^2 method for creating fire objects and the capability of CFAST to simulate accurately the experimental scenario. Each parameter involved in fire modeling was varied (with steps of 10 s or 0.5 MW) as to reduce the gap between model and experiment. The purpose is to check if simplified fire objects could be suitable to be used by designers. According to literature, the time to reach 1 MW in a fire with rapid development is 200 s. The estimated maximum HRR is 30 MW. A detailed description of fire load is provided in Lennon and Moore (2003).

The best combination of factors that minimize the error between model and experiment yields a time-temperature curve very similar for values and shape with experimental data, as shown in Fig 2.

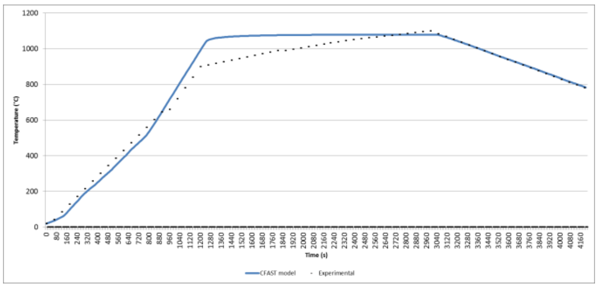


Figure 2: Upper Layer Temperature, Test A

Decay period starts 2990 s after ignition, until that time the error between model and experimental results is 8.68%. The maximum deviation between measurements is about 14%. This is the best result achievable with simplified modeling, is presented in bold in table 1. Only a small number of all other trials are presented in table 1.

Table 1: Fire Load Parameters

Time to 1 MW (s)	Maximum HRR (kW)	Steady Burning (s)	Decay Period (s)	Error %
260	**24000**	**1800**	**5000**	**8.68**
270	24000	1800	5000	8.91
250	24000	1800	5000	8.86
260	24500	1800	5000	8.98
260	23500	1800	5000	8.83
260	24000	1810	5000	8.72
260	24000	1790	5000	8.70

During steady burning the model over-predicts temperatures, as was to be expected considering limitations of two-zone models.

Table 2: Test A Results

Description	Experiment	Simulation	ε %
T_{max} (°C)	1100	1079	1.9
Time at T_{max} (s)	2990	3040	1.6
Steady Burning (s)	1740	1790	2.8

It is remarkable that also in the decay phase the model is very accurate, even if CFAST is not validated to analyze scenarios in this phase. Simulated and experimental curves almost coincide.

4.2. Test B

The aim of test B is to investigate the accuracy of the model when the geometry of the building change. The experiments 2 and 4 in Cardington differs only for the openings. Starting from the model used in Test A, front and back openings were modeled. If the modify does not affect much the result we should find a small error between new model and experiment 4 results. If the model would be able to perfectly manage the modifies in the geometry of the compartment, the error in this case would not be far to 8.68 %, as in Test A. The prediction of a slower fire development is appreciable, as shown in Fig. 3.

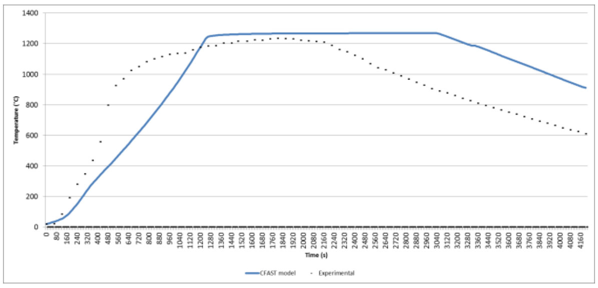


Figure 3: Upper Layer Temperature, Test B

Decay period starts 1800 s after ignition, until that time the error between model and experimental results is 29.12%. The maximum deviation between measurements is about 69%. The variation in the geometry is a critical factor in modeling with this software. However, the model correctly predicts the duration of the burning period and the maximum temperature. Nevertheless it is not able to capture the fire growth rate within the compartment.

If we consider that, during the design, a model is used only for prediction, the high error produced by a slight difference in the geometry is problematic. Even if the designer knows the behavior of the same fire in a different environment, the model is not accurate enough to rely on the forecast without analyzing more aspects in detail.

Table 3: Test B Results

Description	Experiment	Simulation	ε %
T_{max} (°C)	1235	1268	2.6
Time at T_{max} (s)	1800	2960	48.7
Steady Burning (s)	1540	1700	9.9

Comparing experimental results of both tests, a big difference in the quickness of fire development during the heating phase can be noticed. The 800 °C temperature is respectively reached in 1120 s and 480 s. The model is not able to correctly manage this change and it shows a very high sensibility to geometric modifies.

4.3. Test C

Test C investigates the simultaneous presence of 30 fire objects, spreading fire load all over the floor, evaluating if interactions of fire plumes are correctly managed in CFAST. The Cardington experiment 2 was simulated and results can also be compared to Test A.

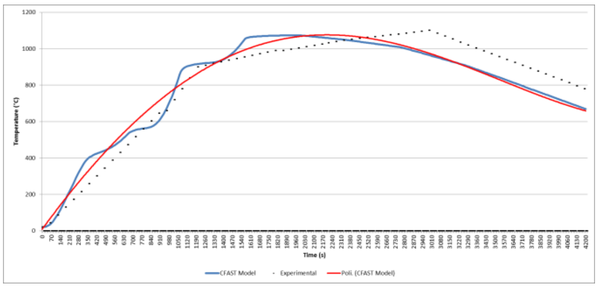


Figure 4: Upper Layer Temperature, Test C

Decay period starts 2990 s after ignition, until that time the minimum mean square error between model and experimental results is 11.2%. The maximum deviation between measurements is about 40%. The decomposition into many fire objects generate a time-temperature response with many peaks. It is difficult to recognize flash-over point and quantify steady burning period. However, the model correctly predicts T_{max}.

Table 3: Test C Results

Description	Experiment	Simulation	ε %
T_{max} (°C)	1100	1074	2.4
Time at T_{max} (s)	2990	1960	41.6
Steady Burning (s)	1740	uncertain	---

To assure that the error is not due to irregular shape of the evolving temperature curve, a polynomial approximation was calculated to reduce the number of peaks. The third grade polynomial is:

$$T= 2.03 \cdot 10^{-8} \cdot t^{3} - 2.93 \cdot 10^{-4} \cdot t^{2} + 10.11 \cdot t + 20$$

This curve, in red in Fig. 4, helps to recognize that in the growth phase the development rate is captured. The error between polynomial curve and experimental results is 10.7%. The accuracy of the model is lower than in the Test A, probably because of increased complexity. Because every crib has an intricate structure, fire growth on it can be quite complex and somewhat variable from one nominally identical sample to the next. In the model, fire objects are identical and according to Jain et al. (2008), fire objects located far from the centre of the compartment and especially those located near vents can lead to incorrect predictions.

5. CONCLUSIONS

A parametric study was performed to examine the effects of different fire characteristics. The use of CFAST in forecasting growth phase and steady burning of a fire is possible under a series of limitations. The need of HRR distribution obtained through a destructive test is a very limiting factor, it can be overcome with the simplified modeling of fire objects. The possibility to create an accurate model, compared to experimental data, was investigated by the use of t-squared fires. The results of using simplified fire objects are accurate and the evolving temperature profile is predictable when all fire load is collapsed in one fire object. In all performed tests T_{max} was predicted with a relative error lower than 2.6%.

During the heating phase, experimental data suggest that the quickness of the development is very influenced by openings. Even if the size of openings is constant, changing from front openings only to front and back highly decrease time to flash-over. The model shows a very high sensibility to geometric modifies and the error is increased of about 20%.

The time-temperature response is less accurate when the fire load is decomposed into many fire objects. During computation by CFAST, location of fire source in the compartment is a critical issue. The increased complexity of the model and the interactions between fire plumes do not lead to a more accurate prediction. The model behavior depends on a complex interaction of the combination of thermal and geometric characteristics of materials and fire objects and one fire object in which all the fire load is collapsed is better managed.

As an engineering design tool, CFAST is suitable to be used with t-squared fires, but a clear understanding of the limitations is necessary.

This study has also identified areas of future research. For example, different methods for generating simplified fire objects can be compared.

REFERENCES

Olenick, S.M., Carpenter, D.J., 2003. An updated international survey of computer models for fire and smoke. *Journal of Fire Protection Engineering* 13: 87–110.

Yao, J., Fan, W., Kohyu, S., Daisuke, K., 1999. Verification and application of field-zone-network model in building fire. *Fire Safety Journal* 33: 35–44.

Zhang, X., Hadjisophocleous, G., 2012.An improved two-layer zone model applicable to both pre- and post-flashover fires. *Fire Safety Journal* 53: 63–71.

Suzuki, K., Harada, K., Tanaka, T., 2003. A multi-layer zone model for predicting fire behavior in a single room. *Fire Safety Science* 7: 851–862.

Xiaojun, C., Lizhong, Y., Zhihua, D., Weicheng, F., 2005. A multi-layer zone model for predicting fire behavior in a fire room. *Fire Safety Journal* 40: 267–281.

Luo, M., 1997. One zone or two zones in the room of fire origin during fires? The effects of the air-handling system. *Journal of Fire Science* 15: 240–260.

Jones, W.W., Peacock, R.D., Forney, G.P., Reneke, P.A., 2009. *CFAST—Consolidated Model of Fire Growth and Smoke Transport (Version 6) Technical Reference Guide*. NIST Special Publication 1026, National Institute of Standards and Technology.

Cox, G. 1995. Chapter 6 Compartment fire modeling, in: Cox G. eds. *Combustion Fundamentals of Fire*, London, Academic Press, pp. 329–404,.

Lennon, T., Moore, D., 2003. The natural fire safety concept—full-scale tests at Cardington. *Fire Safety Journal* 38: 623–643.

Remesh, K., Tan, K.H., 2007. Performance comparison of zone models with compartment fire tests. *Journal of Fire Science* 25: 321–353.

Peacock, R.D., McGrattan, K., Klein, B., Jones, W.W., Reneke, P.A., 2008 *CFAST—Consolidated Model of Fire Growth and Smoke Transport (Version 6) Software Development and Model Evaluation Guide,* NIST Special Publication 1086, National Institute of Standards and Technology.

Fu, Z., Hadjisophocleous, G., 2000. A two-zone fire growth and smoke movement model for multi-compartment buildings. *Fire Safety Journal* 34: 257–285.

Karlsson, B., Quintiere, J.G., 2000. *Enclosure Fire Dynamics*. Boca Raton, Florida: CRC Press.

Au, S.K., Wang, Z.H., Lo, S.M., 2007. Compartment fire risk analysis by advanced Monte Carlo simulation. *Engineering Structures* 29: 2381–2390.

Jain, S., Kumar, S., Kumar, S., Sharma, T.P., 2008. Numerical simulation of fire in a tunnel: Comparative study of CFAST and CFX predictions. *Tunnelling and Underground Space Technology* 23: 160–170.

AUTHORS BIOGRAPHY

Eng. A. Govoni received MS from University of Modena and Reggio Emilia (Italy) in 2008. He is a Lecturer at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy. His research interests include: fire protection engineering, safety science, discrete event simulation, supply-chain, stocks management and logistic problems.

Dr. Eng. G. Davoli received MS and Ph.D. from University of Modena and Reggio Emilia (Italy) in 2005 and 2009, respectively. He is a Lecturer at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy. His research interests include: BPR and lean production practices, discrete event simulation, supply-chain, stocks management and logistic problems.

Dr. Eng. S.A. Gallo received MS and Ph.D. from University of Naples – Federico II (Italy) in 1993 and 1998, respectively. He is a Contract Professor at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena. He had applied as lecturer at the University of Naples - Federico II 1998 to 2005. His research interests include: projects management practices, discrete event simulation, scheduling and logistic problems.

Prof. Eng. R. Melloni received MS and Ph.D. from University of Bologna (Italy) in 1984 and 1991, respectively. He is a Full Professor at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena (Italy) since February 2005. He had applied as associate professor from November 2001 to January 2005. He had applied as lecturer at the University of Parma (Italy) from April 1991 to October 2001. His research interests include: safety system management, projects management, BPR and lean production practices, discrete event simulation, scheduling, supply-chain and logistic problems.

ON AGENTS NEGOTIATION IN A TRADING COMPANY SIMULATION

Roman Šperka[(a)], Marek Spišák[(a)]

[(a)]Silesian University in Opava
School of Business Administration in Karviná
Czech Republic

[(a)]sperka@opf.slu.cz, spisak@opf.slu.cz

ABSTRACT

The aim of this paper is to propose an innovative approach to describe the customer behavior in the trading processes of a virtual company. Agent-based modeling and simulation techniques are used to implement a multi-agent system to serve as a simulation framework. The framework should be a basic part of a management system operating in the integration with real system of a company (e.g. ERP system) to investigate and to predict chosen business metrics of a company. This will ensure the management of a company to support their decision-making processes. The paper firstly presents some of the existing theories about consumer behavior and the types of factors influencing it. Secondly, characterizes multi-agent model of a virtual company, the agents participating in the seller-customer negotiation, and the production function. The production function is used to count the product price while negotiating. Lastly, the simulation results and their validation are described. To sum up, the proposed approach to consumer behavior in an agent-based model could properly contribute to better decision-making process.

Keywords: system modeling, multi-agent systems, agent negotiation, decision support, consumer behavior

1. INTRODUCTION

In the contemporary, dynamic, global and competitive market environment, consumer behavior depends on many different types of factors, which are difficult to grasp. With personal and social factors deals e.g. Enis (1974). With physical factors deals e.g. McCarthy and Perreault (1993). More complex view on the social, economic, geography and culture factors gave Keegan et al. (1992). Schiffman (2007) brought marketing mix and environment into the types of factors mentioned herein above. Previous discussions have so far either relied on an objectivist (complete information of customers, constant decision mechanism, constant consumer preferences) or a constructivist view (consumption discourses, consumption as a crucial aspect in the construction of identity). However, both have failed to integrate the consumers' interactions with their social behavior and physical environment as well as the materiality of consumption (Gregson et al. 2002, Jackson et al. 2006). The complexity of the factors influencing consumer behavior and their changes in the time shows relations between external stimuli, consumer's features, the course of decision-making process and reaction expressed in his choices. As a result, the investigation of consumer behavior seems to be too complicated for traditional analytical approaches (Forrester 1971, Challet and Krause, 2006).

Agent-based modeling and simulation (ABMS) provides some opportunities and benefits resulting from using multi-agent systems as a platform for simulations with the aim to investigate the consumers' behavior. Agent-based models are able to integrate individually differentiated types of consumer behavior. They are characterized by a distributed control and data organisation, which enables to represent complex decision processes with only few specifications. In the recent past there were published many scientific works in this area. They concern in the analysis of companies positioning and the impact on the consumer behavior (e.g. Tay and Lusch 2002, Wilkinson and Young 2002, Casti 1997). Often discussed is the reception of the product by the market (Goldenberg et al. 2010, Heath et al. 2009), innovation difussion (Rahmandad and Sterman 2008, Shaikh et al. 2005, Toubia et al. 2008). More general deliberations on the ABMS in the investigating of consumer behavior shows e.g. (Adjali et al. 2005, Ben 2002, Collings et al. 1999).

The approach introduced in this paper uses an agent-based model in the form of multi-agent system to serve as a simulation platform for the seller-customer negotiation in a virtual trading company. The overall idea comes from the research of Barnett (2003). He proposed the integration of the real system models with the management models to work together in real-time. The real system (e.g. ERP system) outputs proceed to the management system (e.g. simulation framework) to be used to investigate and to predict important company's results (metrics). Actual and simulated metrics are compared and evaluated in a management model that identifies the steps to take to respond in a manner that drives the system metrics towards their desired values. We used a generic control loop model of a company (Wolf 2006) and implemented multi-agent simulation framework, which represents the management system. This task was rather complex, therefore we took only a part of the model – trading processes and the negotiation of seller and customer.

The work described in this paper aims at proposing an approach to describe the customer behavior in the trading processes of a virtual company. Implemented simulation framework will be a basic part of a future management system simulating business metrics – key performance indicators (KPIs) of a real company's system. The paper is structured as follows. In the section 2 the multi-agent model is described. In the section 3 the seller-customer negotiation is introduced. The core of this section is the production function definition. The simulation results are presented in section 4.

2. MULTI-AGENT MODEL

To ensure the outputs of customer behavior simulations a simulation framework was implemented and used to trigger the simulation experiments. The framework covers business processes supporting the selling of goods by company sales representatives to the customers – seller-customer negotiation (Fig. 1). It consists of the following types of agents: sales representative agents (representing sellers, seller agents), customer agents, an informative agent (provides information about the company market share, and company volume), and manager agent (manages the seller agents, calculates KPI). Disturbance agent is responsible for the historical trend analysis of sold amount (using his influence on customer agent). All the agent types are developed according to the multi-agent approach. The interaction between agents is based on the FIPA contract-net protocol (FIPA 2002).

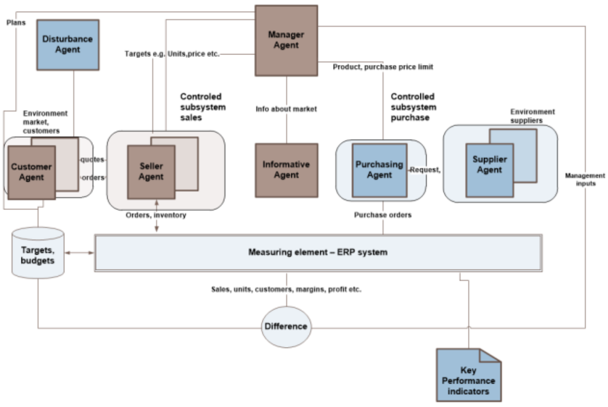

Plans
Disturbance Agent
Targets e.g. Units,price etc.
Manager Agent
Product, purchase price limit
Environment market, customers
Controled subsystem sales
Info about market
Controlled subsystem purchase
Environment suppliers
Customer Agent
quotes
orders
Seller Agent
Informative Agent
Purchasing Agent
Request,
Supplier Agent
Management inputs
Orders, inventory
Purchase orders
Targets, budgets
Measuring element – ERP system
Sales, units, customers, margins, profit etc.
Difference
Key Performance indicators

Figure 1: Generic Model of a Business Company (Source: adapted from Šperka et al. 2013).

The number of customer agents is significantly higher than the number of seller agents in the model because the reality of the market is the same. The behavior of agents is influenced by two randomly generated parameters using the normal distribution (an amount of requested goods and a sellers' ability to sell the goods). In the lack of real information about the business company, there is a possibility to randomly generate different parameters (e.g. company market share for the product, market volume for the product in local currency, or a quality parameter of the seller). The influence of randomly generated parameters on the simulation outputs while using different types of distributions was presented in (Vymetal et al. 2012).

3. SELLER-CUSTOMER NEGOTIATION

In this section, the seller-customer negotiation workflow is described and the mathematical definition of a production function is proposed. Production function is used during the contracting phase of agents' interaction. It serves to set up the limit price of the customer agent as an internal private parameter.

Only one part of the company's generic structure, defined earlier, was implemented. This part consists of the sellers and the customers trading with stock items (e.g. tables, chairs). One stock item simplification is used in the implementation. Participants of the contracting business process in our multi-agent system are represented by the software agents - the seller and customer agents interacting in the course of the quotation, negotiation and contracting. There is an interaction between them. The behavior of the customer agent is characterized in our case by proposed customer production function (Equation 1).

At the beginning disturbance agent analyzes historical data – calculates average of sold amounts for whole historical year as the base for percentage calculation. Each period turn (here we assume a week), the customer agent decides whether to buy something. His decision is defined randomly. If the customer agent decides not to buy anything, his turn is over; otherwise he creates a sales request and sends it to his seller agent. Requested amount (which was generated based on a normal distribution) is multiplied by disturbance percentage. Each turn disturbance agent calculates the percentage based on historical data and sends the average amount values to the customer agent. The seller agent answers with a proposal message (a certain quote starting with his maximal price: *limit price* * *1.25*). This quote can be accepted by the customer agent or not. The customer agents evaluate the quotes according to the production function. The production function was proposed to reflect the enterprise market share for the product quoted (a market share parameter), seller's ability to negotiate, total market volume for the product quoted etc. (in e.g. Vymetal et al. 2012). If the price quoted is lower than the customer's price obtained as a result of the production function, the quote is accepted. In the opposite case, the customer rejects the quote and a negotiation is started. The seller agent decreases the price to the average of the minimal limit price and the current price (in every iteration is getting effectively closer and closer to the minimal limit price), and resends the quote back to the customer. The message exchange repeats until there is an agreement or a reserved time passes.

The customer production function for the m-th seller pertaining to the i-th customer determines the price that the i-th customer accepts (adjusted according to Vymetal et al. 2012).

$$c_n^m = \frac{\tau_n T_n \gamma \rho_m}{O v_n} \qquad (1)$$

c_n^m - price of n-th product offered by m-th seller,
τ_n - market share of the company for n-th product $0 < \tau_n < 1$,
T_n - market volume for n-th product in local currency,
γ - competition coefficient, lowering the success of the sale $0 < \gamma \leq 1$,
ρ_m - m-th sales representative ability to sell, $0.5 \leq \rho_m \leq 2$,
O – number of sales orders for the simulated time,
v_n - average quantity of the n-th product, ordered by i-th customer from m-th seller.

The aforementioned parameters represent global simulation parameters set for each simulation experiment. Other global simulation parameters are: lower limit sales price, number of customers, number of sales representatives, number of iterations, and mean sales request probability. The more exact parameters can be delivered by the real company, the more realistic simulation results can be obtained. In case we would not be able to use the expected number of sales orders O following formula can be used

$$O = ZIp \text{ where}$$

Z - number of customers
I - number of iterations,
p - mean sales request probability in one iteration.

Customer agents are organized in groups and each group is being served by concrete seller agent. Their relationship is given; none of them can change the counterpart. Seller agent is responsible to the manager agent. Each turn, the manager agent gathers data from all seller agents and stores KPIs of the company. The data is the result of the simulation and serves to understand the company behavior in a time – depending on the agents' decisions and behavior. The customer agents need to know some information about the market. This information is given by the informative agent. This agent is also responsible for the turn management and represents outside or controllable phenomena from the agents' perspective.

4. SIMULATION RESULTS

Agent count and their parameterization are listed in Table 1.

Table 1: Multi-agent System Parametrization

AGENT TYPE	AGENT COUNT	PARAMETER NAME	PARAMETER VALUE
Customer Agent	500	Maximum Discussion Turns	10
		Mean Quantity	40 m
		Quantity Standard Deviation	32
Seller Agent	25	Mean Ability	1
		Ability Standard Deviation	0.03
		Minimal Price	0.36 EUR
Manager Agent	1	Purchase Price	0.17 EUR
Market Info	1	Item Market Share	0.15
		Item Market Volume	1 033 535EUR
		Competition coefficient	0.42
		No items sold in one iteration	1 330
		Iterations count	52
Disturbance Agent	1		

Agents were simulating one year – 52 weeks of interactions. As mentioned above – manager agent was calculating the KPIs. Total gross profit was chosen as a representative KPI. Figure 2 contains the month sums of total gross profit for real and generated data. As can be seen from this figure, the result of simulation was quite similar to the real data.

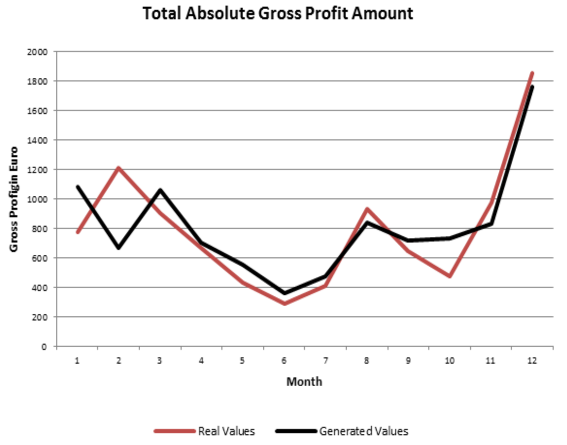


Figure 2: The Generation Values Graph – Monthly (Source: own)

To prove the relationship between the real and generated data – two instruments were chosen – Correlation Analysis to show the correlative relation between them and Chi-Square Test for Independence to show the similarity of distribution for both data series.

Correlation coeficient for total gross profit amount was 0.857, which represents very strong correlation between real and generated data.

Also the Chi-Square Test for Independence has proven that the distribution of real and generated values is very similar. In figure 3, there is a frequency histogram of gross profit for real and generated values.

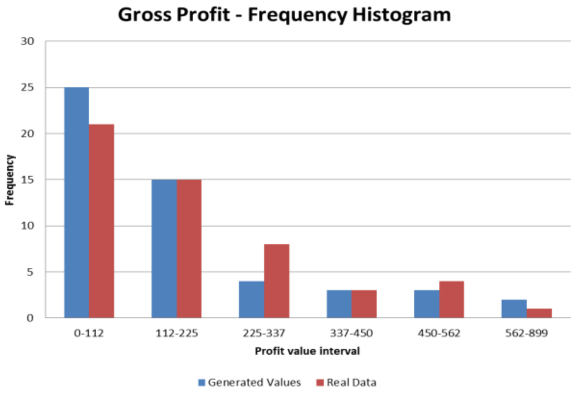


Figure 3: Gross Profit Frequency Histogram (Source: own).

CONCLUSION

Introduction of disturbance agent has caused closer distribution similarity between real and generated data. By its influence on the amounts sold in every turn (even if the price remained as a result of negotiation between seller and customer agent) very strong correlation between reality and generation has risen.

For the future experiments – two improvements shall be made – implement the disturbance agent more sophisticated in history analyzing and also each customer agent shall be more individualistic – have its own targets, beliefs, desires – not only to follow the production function.

ACKNOWLEDGMENTS

This work was supported by grant of Silesian University no. SGS/6/2013 "Advanced Modeling and Simulation of Economic Systems”.

REFERENCES

Adjali, I., Dias, B. and Hurling R. 2005. Agent based modeling of consumer behavior. In: *Proceedings of the North American Association for Computational Social and Organizational Science Annual Conference*. University of Notre Dame. Notre Dame. Indiana. Available from: http://www.casos.cs.cmu.edu/events/conferences/2005/conference. [Accessed 14 March 2012].

Barnett, M. 2003. *Modeling & Simulation in Business Process Management*. Gensym Corporation, pp. 6-7, Available from: http://news.bptrends.com/publicationfiles/1103%20WP%20Mod%20Simulation%20of%20BPM%20-%20Barnett-1.pdf. [Accessed 16 January 2012].

Ben, L., Bouron, T. and Drogoul, A. 2002. Agent-based interaction analysis of consumer behavior. In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*: part 1. ACM. New York, pp. 184-190.

Casti, J. 1997. *Would-be Worlds. How Simulation is Changing the World of Science*. Wiley. New York.

Challet, D. and Krause, A. 2006. What questions to ask in order to validate an agent-based model. In: *Report of the 56th European Study Group with Industry*. pp. J1-J9. Available from: http://www.maths-in-industry.org/miis/107/1/ Unilever-ABM-Report.pdf. [Accessed 28 March 2013].

Collings, D., Reeder A., Adjali, I., Crocker, P. and Lyons, M. 1999. Agent based customer modelling. *Computing in Economics and Finance*. (1352). Available from: http://econpapers.repec.org/paper/scescecf9/1352.htm. [Accessed 28 March 2013].

Enis, B. M. 1974. *Marketing principles: the management process*. Goodyear Pub. Co. (Pacific Palisades, Calif). 608 p., ISBN 0876205503.

Foundation for Intelligent Physical Agents, FIPA. 2002. *FIPA Contract Net Interaction Protocol*. In Specification [online]. Available from: http://www.fipa.org/specs/fipa00029/SC00029H.pdf. [Accessed 13 June 2011].

Forrester, J. 1971. Planung unter dem Einfluss komplexer Sozialer Systeme. In: *Politische Planung in Theorie und Praxis*. Ed by. G. Schmieg. Piper Verlag. München. p. 88.

Goldenberg, J., Libai, B. and Muller, E. 2010. The Chilling effect of network externalities. *International Journal of Research in Marketing*. 27(1): pp. 4-15.

Gregson, N., Crewe, L. and Brooks, K. 2002. Shopping, space, and practice. In: *Environment and Planning* D 20 (5), pp. 597–617. DOI: 10.1068/d270t.

Heath, B., Hill R. and Ciarallo F. 2009. *A survey of agent-based modeling practices* (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation* 12(4): pp. 5-32.

Jackson, P., Perez Del Aguila, R. P., Clarke, I., Hallsworth, A., De Kervenoael, R. and Kirkup, M. 2006. Retail restructuring and consumer choice 2. Understanding consumer choice at the household level. In: *Environment and Planning* A 38 (1),pp. 47–67. DOI:10.1068/a37208

Keegan, W., Moriarty, S. and Duncan, T. 1992. *Marketing*. Prentice-Hall. Englewood Cliffs. New Jersey. 193 p.

McCarthy, E. J. and Perreault, W. D. 1993. *Basic marketing: a global-managerial approach*. Irwin, 792 p., ISBN 025610509X.

Rahmandad, H. and Sterman, J. 2008. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation Models. *Management Science*. 54(5): 998-1014.

Shaikh, N., Ragaswamy, A. and Balakrishnan A. 2005. *Modelling the Diffusion of Innovations Using*

Small World Networks. Working Paper. Penn State University. Philadelphia.

Schiffman, L. G. and Kanuk, L. L. 2007. *Purchasing Behavior* (9th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Šperka, R., Vymětal, D. and Spišák, M. 2013. Validation of Agent-based BPM Simulation. In proceedings: *Agent and Multi-Agent Systems: Technology and Applications 2013*. Hue City, Vietnam. (to be published)

Tay, N. and Lusch, R. 2002. Agent-Based Modeling of Ambidextrous Organizations: Virtualizing Competitive Strategy. *IEEE Transactions on Intelligent Systems* 22(5): 50-57.

Toubia, O., Goldenberg, J. and Garcia, R. 2008. A New approach to modeling the adoption of new products: Aggregated Diffusion Models. *MSI Reports: Working Papers Series*. 8(1): 65-76.

Wilkinson, I; Young, L. 2002. On cooperating: Firms. relations. networks. *Journal of Business Research*. 55: pp. 123-132.

Wolf, P. 2006. *Úspěšný podnik na globálním trhu*. Bratislava: CS Profi-Public. ISBN 80-969546-5-2.

Vymetal, D., Spisak, M. and Sperka, R., 2012. An Influence of Random Number Generation Function to Multiagent Systems. In Proceedings*:* *Agent and Multi-Agent Systems: Technology and Applications*. Dubrovnik, Croatia.

AUTHORS BIOGRAPHY

R. Šperka is an assistant professor at Department of Informatics at the Silesian University in Opava, School of Business Administration in Karviná, Czech Republic. He holds a PhD in Business economics and management. He is author of more than 40 publications.

M. Spišák is an assistant at Department of Informatics and a PhD student at the Silesian University in Opava, School of Business Administration in Karviná, Czech Republic. He is author of more than 15 publications.

QUEUING THEORY SIMULATION MODEL FOR CALCULATING NET PRESENT VALUE CORRECTIVE FACTOR IN INVESTMENT PROJECT APPRAISAL

Zoran Petrovic[a], Sladjana Benkovic[b] Ugljesa Bugaric [c], Dusan Petrovic[d] , Gordana Markovic Petrovic[e]

[a] Mokra Gora School of Management
[b] Faculty of Organizational Sciences, University in Belgrade
[c] Mechanical Faculty, University in Belgrade
[d] Mechanical Faculty, University in Belgrade
[e] Dom Zdravlja. Zemun

[a]zoran@tecon.rs,[b] benko@fon.bg.ac.rs, [c]ubugaric@mas.bg.ac.rs , [d] dpetrovic@mas.bg.ac.rs,[e] gm5rovic@gmail.com

ABSTRACT

In contemporary project management, project life cycle is defined by project development phases. Most important phase for project lifecycle is Opportunity phase in which project profitability is evaluated. On the end of this phase is determined if project will be developed in full life cycle, or rejected as non-profitable. Criteria that is used, for project evaluation, is Net Present Value criteria (NPV).

In order to propose methodology for getting more accurate results for NPV, system which is subject of investment is modeled as queuing theory model with balking and reneging. Input parameters of the system are collected from case study. Based on mentioned combined model, probability of service is calculated. In order to make conclusions more versatile, simulation model is build and validated against results from queuing theory model and case study results.

Probability of service, calculated from validated simulation model, is used as corrective factor for calculation of NPV, based on realistic assumption of serviced units, which are participating in income.

Keywords: Queuing theory, Net Present Value, Investment evaluation, Simulation

1. INTRODUCTION

Project life cycle in the industry is divided in several development phases (Newell and Grashina 2004). First and most important phase is feasibility phase in which project economic gain is evaluated. Based on results of this phase, decision is made – either to develop all project cycle phases, or to mark project as non-profitable, cancel it and find alternative one. If project passes initial phase, then it is possible to continue with all other general phases – Intermediate and Final phase (Figure 1).

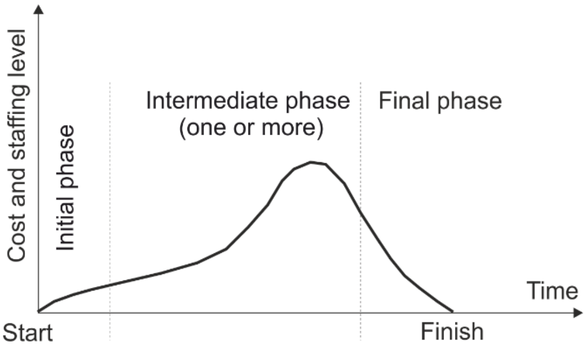


Figure 1: Project life cycle phases (Newell and Grashina 2004)

According to Đuričin and Lončar (Đuričin and Lončar 2012), there is even earlier stage in which project is evaluated – Opportunity phase. In this phase preliminary evaluation of the project is conducted and if project passes this evaluation, Feasibility phase is started. Once again, after finishing of feasibility study evaluation is made, contract is signed and project is officially started (Figure 2).

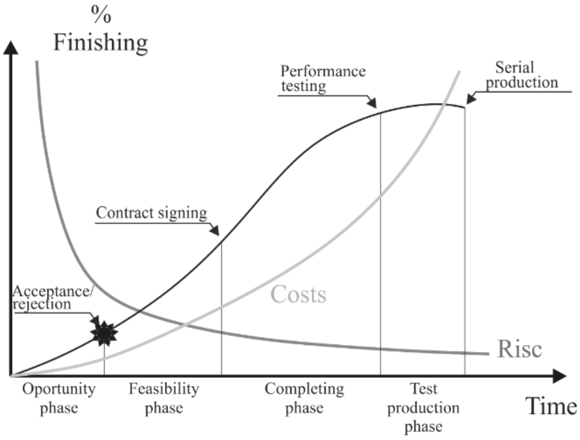


Figure 2: Movement of main parameters through project phases (Đuričin and Lončar 2012)

Regarding project evaluation, lot of different criteria can be used. Some of them are static (Return of Investment, Investment productivity, Employment ratio, etc.), some of them dynamic (Net Present Value, Internal Return Ratio, etc..). There are also lot of different criteria which takes uncertainty into consideration (Break even point analysis, Probability analysis, Game theory analyses, Monte Carlo simulation etc.). One of the most used criteria is Net Present Value (NPV) criteria, which will be also used in this paper.

1.1. Net Present Value criteria

Net Present Value is relatively simple criteria, which takes into consideration Net Income during investment period. In this paper is considered that there is only initial investment in the beginning of the project (I) and that investment period is equal to project life cycle duration (Newnan, D.G., Eschenbach, T.G., Lavelle, J.P., 2004).

$$NPV = \frac{NI_1}{(1+i)^1} + \frac{NI_2}{(1+i)^2} + \cdots + \frac{NI_n}{(1+i)^n} \qquad (1)$$

$$NPV = \sum_{k=0}^{n} \frac{NI_k}{(1+i)^k} - I \qquad (2)$$

Where:

NPV – criteria of Net Present Value

NI_k – Net Income (difference between Income and Cost) in evaluated period k

i – Discount factor

According to Net Present Criteria, investment is approved if NPV > 0.

NPV is chosen since it is very simple criteria which gives accurate preliminary investment evaluation results, satisfactory for opportunity phase analysis.

1.2. Disadvantages of NPV criteria

Problem with Net Present Value (NPV) criteria is anticipation of Net income in evaluated period. This is usually done by observation of Net Income of similar investments, which were already finished in the past. Such methodology is very risky, since it is very hard to find exactly the same equipment, in the same working surrounding, with same dependent and independent costs.

According to found information, there are two potentially risky scenarios.

First one is that Net Income is underestimated, so result of the NPV can be underestimated, also. This would lead to rejecting of potentially profitable project.

Second one in that Net income is overestimated, which would lead to accepting potentially non profitable project.

Idea of this paper is to suggest certain safety factor, which would be used as corrective factor for NPV Criteria, in order to get better estimation of Net Income.

2. METHODOLOGY

2.1. Queuing theory model

First step in the methodology is to represent potential investment in technological equipment as queuing theory model. Equipment is usually observed as part of internal logistic process (Pfohl 2010).

As it can be seen from Figure 3., potential investment is modelled as queuing theory model, which consist of units coming into the system, waiting line for units in front of the servers, and n-servers. After servicing, units are leaving the system. System is generally considered to be with infinite units arrival, without possibility for units to come back for servicing. Units are considered to be intelligent, so they can decide not to enter the system at all (balking), or they can decide to leave waiting line (reneging). This model is chosen, since it can be used in most cases when investment in production lines is considered. Units which are coming can be in form of: material, spare parts, sub-assemblies, assemblies, which are coming on production line. Waiting line can be considered as buffer before entering on production line and servers can be considered as production machines or lines. For various of reasons units can be rejected before entering system (for example quality control before entering into the system, etc.), which was modelled as balking. Also some units can leave waiting line if waiting in the line is taking too much time – for example in the zinc coating process, casting, etc. System state transitions are represented on Figure 4.

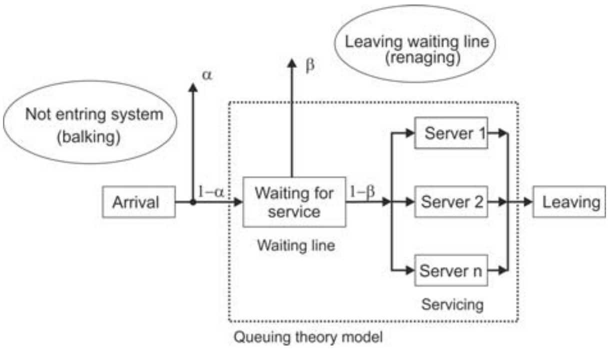


Figure 3: Queuing theory model with balking and reneging

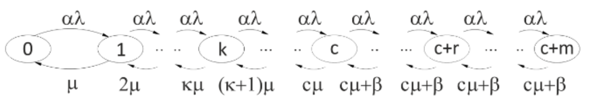


Figure 4: System state transitions

Probabilities of each state of the system are:

$$-\alpha \cdot \lambda \cdot p_0 + \mu \cdot p_1 = 0 \tag{3}$$

$$\alpha \cdot \lambda \cdot p_{k-1} - (\alpha \cdot \lambda + \mu) \cdot p_k + \\ + (k+1)\mu \cdot p_{k+1} = 0 \tag{4}$$

for k= 1,2,…, (c - 1)

$$\alpha \cdot \lambda \cdot p_{c+r-1} - (\alpha \cdot \lambda + c\mu + \beta) \cdot p_{c+r} + (c\mu + \beta) \cdot \\ \cdot p_{c+r+1} = 0 \tag{5}$$

for r = 0,1,2,…, (m -1)

$$\alpha \cdot \lambda \cdot p_{c+m-1} - (c\mu + \beta) \cdot p_{c+m} = 0 \tag{6}$$

Solving system of equations by expressing state probability p₀, gives:

$$p_1 = \frac{\alpha \cdot \lambda}{\mu} \cdot p_0 \tag{7}$$

.

.

$$p_k = \frac{1}{k!} \cdot \left(\frac{\alpha \cdot \lambda}{\mu}\right)^k \cdot p_0,\ k = 1,2,..c \tag{8}$$

$$p_{c+r} = \left(\frac{1}{c \cdot \mu + \beta}\right)^r \cdot \frac{(\alpha \cdot \lambda)^{c+r}}{c! \cdot \mu^c} \cdot p_0 \tag{9}$$

$r = 1,2,,\ldots,m$

Sum of all state probabilities has to be equal one:

$$\sum_{k=0}^{c} p_k + \sum_{r=1}^{m} p_{c+r} = 1 \tag{10}$$

Probability p₀ is calculated:

$$p_0 = \frac{1}{\sum_{k=0}^{c} \frac{1}{k!} \cdot \left(\frac{\alpha \cdot \lambda}{\mu}\right)^k + \sum_{r=1}^{m} \left(\frac{1}{c \cdot \mu + \beta}\right)^r \cdot \frac{(\alpha \cdot \lambda)^{c+r}}{c! \cdot \mu^c}} \tag{11}$$

Or after transformations:

$$p_0 = \frac{1}{\sum_{k=0}^{c} \frac{\left(\frac{\lambda \cdot \alpha}{\mu}\right)^k}{k!} + \frac{\left(\frac{\lambda \cdot \alpha}{\mu}\right)^c}{c!} \cdot \frac{\lambda \cdot \alpha}{c\mu + \beta} \cdot \frac{1 - \left(\frac{\lambda \cdot \alpha}{c\mu + \beta}\right)^m}{1 - \frac{\lambda \cdot \alpha}{c\mu + \beta}}} \tag{12}$$

All other probabilities can be expressed by using expressed probability po.

Probability of serviced with balking and reneging is:

$$P_{srv} = \sum_{i=0}^{c+m+1} p_i = 1 - p_{c+m} = \\ = 1 - \left(\frac{\alpha \cdot \lambda}{c\mu + \beta}\right)^m \cdot p_c \tag{13}$$

Or

$$P_{srv} = 1 - \left(\frac{\alpha \cdot \lambda}{c\mu + \beta}\right) \cdot \frac{1}{c!} \cdot \left(\frac{\alpha \cdot \lambda}{\mu}\right) \cdot p_0 \tag{14}$$

2.2. Case study

In order to validate theoretical queuing theory model, self service car was system was observed. System is consisted of 2 washing bays (servers) and 3 places in waiting line. Layout of observed system is presented on Figure 5.

Number of observed events was chosen based on research of Barlett et al. (Barlett, Kotrlik, Higgins, and Chadwick 2001). Mentioned authors determined minimum sample size, for given population size for categorical and continuous data. Based on their research, for analysis of one year (365 days – population size), for margin of error of 0.05, p= 0.50, t=1.96, sample should be 180. According to this fact, system was observed for 180 randomly chosen days, from opening to it's closing - 12 working hours.

During that time, following data were written in study protocol: exact time of first unit coming in the wide system aria, time between arrival of the next unit – until last one.

Number of units which entered wide system area, but didn't enter system itself, number of units left waiting line and finally time from starting of service, until end of the service and leaving system.

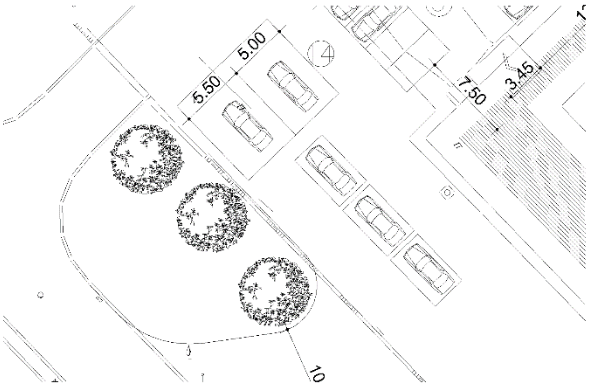


Figure 5: System layout

Based on mentioned data from protocol, following statistical analyses were done. First of all, mean time between units arrival was calculated for each day. Mean time of units servicing was calculated for each day. Both variables were tested by using Kolmogorov-Smirnov test for goodness of fit to exponential theoretical distribution, in order to confirm preliminary assumption, of Markov process of birth and death.

Results from statistical analysis are given in separate chapter - 3. Results.

If both mean time between arrival of the units in the system and mean time of servicing are exponentially distributed, then these data can be used as input values for queuing theory model and probability of servicing can be calculated, using theoretical model.

2.3. Simulation

Theoretical model of queuing theory can be used if assumption of Markov process is fulfilled, so probability of servicing can be analytically calculated. For some special cases, with some assumptions, calculation of probability of servicing is also possible, but for most of the cases, only way to calculate probability of servicing is by using simulation.

Simulation model for calculating probability of servicing was made in Mathlab Simulink, according to general methodology proposed by Mitroff et al. (Mitroff, Betz, Pondy and Sagasti 1974) and more detail methodology proposed by Lopatenok and Merkuryev (Lopatenok, Merkuryev 2000) .

In the model, units are generated according signal from random number generator. After being generated, units are coming to first junction, in which they can go in one of two direction based on the entrance signal. Entrance signal is probability of balking, which is entered as input data in the model. First direction is continue to the waiting line and second one is to the system exit through balking. If unit is leaving the system based on balking, it is noted as balked unit. Algorithm of simulation model is represented on Figure 6.

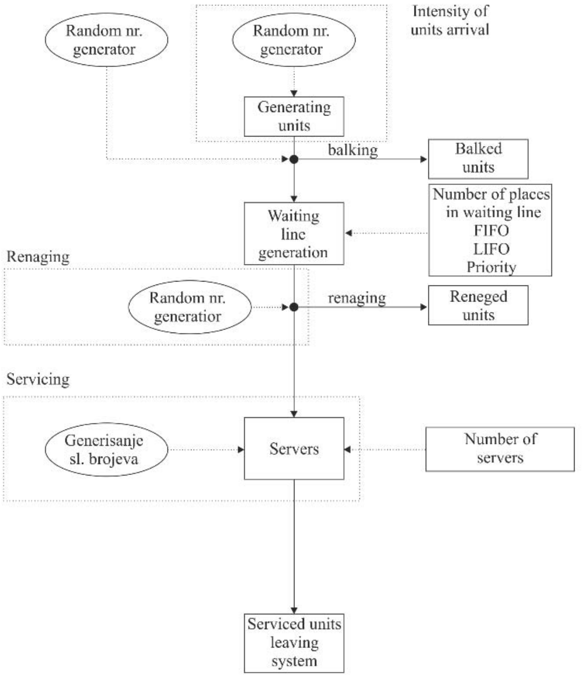


Figure 6: Algorithm of simulation model

Unit which continues goes to the waiting line. Discipline of waiting line can be FIFO, LIFO and Priority. In this paper only FIFO discipline was observed. From the waiting line unit is going into the second junction in which it can go again to the one of two potential directions. Signal which is determining in which direction will unit go, is probability of reneging. If unit goes to the reneging direction it goes out of the system and it is noted as reneging unit. If unit is not going in reneging direction, it goes to the servers. After finishing servicing, unit is leaving system and it is noted as serviced unit. Number of simulation events will be 180, according to research of Barlett et al.

2.4. Probability of service as corrective factor

Observing the system, general behavior of units was noted. Not all units that come in wide system area are entering the system. Also, not all units that enter into the system go to servicing – some of them leave waiting line and go out of the system, without service.

According to proposed methodology, probability of service can be used as corrective factor for calculation of Net Present Value criteria:

$$NPV^{re} = P_{srv}^{sim} \cdot NPV \qquad (15)$$

Where:

NPV^{re} – Realistic NPV

P_{srv}^{sim} - Simulated probability of service

3. RESULTS

Based on proposed methodology, according to Kolmogorov – Smirnov test, for distribution fit, with value α =0,05, all samples have exponential distribution, with different mean value. On the Figure 7 – mean time between arrivals distribution is shown and on the Figure 8 – mean time of servicing.

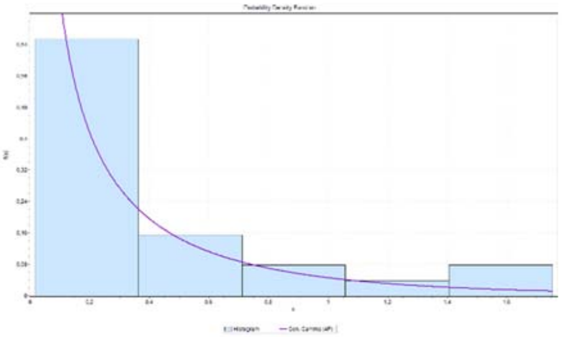


Figure 7: Mean time between arrivals

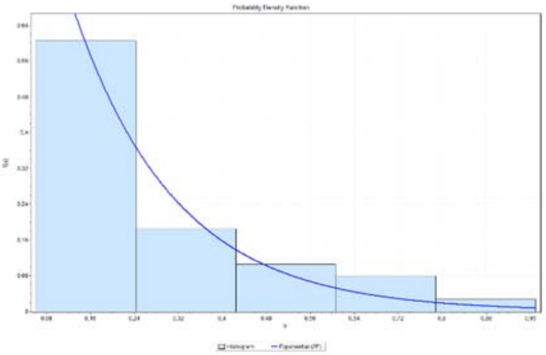


Figure 8: Mean time of servicing

Based on that fact, it is possible to conclude that assumption of Markov process, both in arrival and service is confirmed. From this conclusion, comes second conclusion that it is possible to use theoretical expression for probability of service, for observed case study.

From all gathered samples, mean time between arrival of the units along with mean time of service, were calculated.

Mean time between arrival of the units is $\bar{t}_d = 5{,}15$ min., which gives intensity of arrival of λ = 11,65u/h. Mean time of servicing is $\bar{t}_o = 5{,}05$ min., which gives intensity of service μ = 11,88 u/h.

Also, it was noted from results in the protocol, that average balking probability is 10% and intensity of reneging was 0,1μ.

According to proposed model from queuing theory with g intensity of arrival of λ = 11,65 u/h and intensity of service μ = 11,88 u/, gives that probability of service is Psrv = 0,999.

Based on this value, 180 simulating events were started. Average result for probability of servicing was: P_{srv}^{sim} = 0,9892.

In order to validate simulation model, methodology proposed by Bugarić and Petrović was used (Bugarić and Petrović 2011).

Table 1: Validation of simulation model

Value	**Psrv**
Theoretical	0,999
Simulation	0,989
Average deviation $\sigma = \sqrt{\frac{\sum_{1}^{br.sim}\left(Vel_{teor} - Vel_{exp}\right)^2}{br.sim}}$	0,0108
Estimated error $\frac{\sigma}{\sqrt{br.sim}}$	0,00108
Relative error $\left\|Vel_{teor} - Vel_{exp}\right\|$	0.01

4. DISCUSION

Proposed queuing model is used since it can be analytically calculated. Similar model was proposed by Whitt (Whitt 1999) in his study about informing customers about anticipated delays. Boots and Tijms (Boots and Tijms 1999.) were exploring general M/M/M/c queue with impatient customers, with similar model. Model which was proposed in this paper was chosen for two reasons.

First one is that lot of production lines can be very well described by it. Second one is that it gives good bases for validation of simulation model, since it can be solved analytically.

Based on results from combined theoretical queuing model and case study, probability of service was calculated. This value was compared to calculated value from simulation model and according to Table 1. , difference between those two values is not statistically important, meaning that simulation model is validated.

5. CONCLUSION

Methodology that was proposed in this paper helps to calculate Net Present Value of investment project more accurately, which gives certain safety factor in decision making. Simulation model takes into consideration only units that are being served, excluding ones that balked or reneged from system. This model can be used for any investment in production line that can be represented with proposed queuing model.

Benefit from validated simulation model is that it is not sensitive to distribution of time between arrival of the units, or servicing time. It can be used without need for starting assumptions of Markov processes of birth and death. Simulation model can also be used for analyzing worst case scenario of equipment capacity in different theoretical distribution of units arrival and servicing and bulk units arrival, or it can be used for optimization of system configuration based on existing or predicted parameters.

REFERENCES

Barlett, J.E., Kotrlik, J.W., Higgins, M., Chadwick C., 2001. Organizational research: Determining appropriate sample size in surway research. *Informational Technology, Learning and Performance Journal,* 19(1), 43-50.

Boots, N.K., Tijms, H., 1999. An M/M/M/c queue with impatient customers. *Sociedad de Estadistica e Invesigacion Operativa,* 7(2), 213-220.

Bugarić, U., Petrović, D., 2011. *Modeliranje sistema opsluživanja.* Beograd: Mašinski fakultet, Univerzitet u Beogradu.

Đuričin, N.D., Lončar, D.M., 2012. *Menadžment pomoću projekata.* Beograd: Centar za istraživačku delatnost Ekonomskog fakultata u Beogradu

Lopatenok, V., Merkuryev, V., 2000. Simulation and Analysis of Production Facility Operation. *IFAC Symposium on Manufacturing, Modeling, Management and Control*, pp 467-471. July 12-14, University of Patras (Patras, Greece)

Mitroff, I.I., Betz, F., Pondy, L.R., Sagasti, F., 1974. On managing science in the systems age: two schemas for the study of science as whole systems phenomenon. *Interfaces ,* 4(3), 46-58.

Newell, M.W., Grashina, M.N., 2004. *The Project Management Question and Answer Book.* New York: AMACOM.

Newnan, D.G., Eschenbach, T.G., Lavelle, J.P., 2004. *Engineering Economic Analysis.* 9th ed. London: Oxford Press.

Pfohl, H., 2010. *Logistiksysteme Betriebswirtschaftliche Grundlagen.* 8th ed. Heidelberg: Springer.

Whitt, W., 1999. Improving Service by Informing Customers About Anticipated Delays. *Management Science,* 45(2), 192-206.

AUTHORS BIOGRAPHY

Zoran Petrović finished Mechanical Faculty, University in Belgrade in 2003. In past ten years developed his career in different professional and scientific fields. He is working as executive manager in company Tecon Sistem from 2006. He finished his doctoral thesis on Mechanical Faculty in Belgrade in July 2013 and at the moment he is studying EMBA studies at Mokra Gora School of Management.

Slađana Benković has been an associate professor at the Faculty of Organizational Sciences, University in Belgrade for the last 15 years. During 2007/2009 she spent time at the George Washington University, Washington D.C. as a visiting professor. She is Deputy President of the Management Board of the “Endowment of Milivoje Jovanović and Luka Ćelović”, as well as a member of the Management Board of the “Endowment of Đoko Vlajković”. Her teaching and research fields are financial management with a research focus on project finance, modalities of financing development projects of companies, technical evaluation of investment profitability and determination of corporate capital structure.

Uglješa Bugarić is professor at the Mechanical Faculty, University in Belgrade and chairman of Industrial engineering department. His research is mainly oriented to logistics, operations research problems and optimization.

Dušan Petrović has been an associate professor at the Mechanical Faculty, University in Belgrade, since 1999. His research is mainly oriented to logistics and warehouse management problems.

Gordana Marković Petrović finished Medical Faculty, University in Belgrade and Master in Emergency Surgery and Management in Healthcare. At the moment she is working as executive manager in Dom Zdravlja, Zemun and finishing Master studies in traditional medicine at Medical Faculty, University in Belgrade.

RETSIM: A SHOE STORE AGENT-BASED SIMULATION FOR FRAUD DETECTION

Edgar Alonso Lopez-Rojas[a], **Stefan Axelsson**[b], and **Dan Gorton**[c]

[a],[b]Blekinge Institute of Technology , School of Computing
[c]KTH Royal Institute of Technology , Department of Transport Science , Center for Safety Research

[a]edgar.lopez@bth.se, [b]stefan.axelsson@bth.se, [c]dan.gorton@abe.kth.se

ABSTRACT

RetSim is an agent-based simulator of a shoe store based on the transactional data of one of the largest retail shoe sellers in Sweden. The aim of RetSim is the generation of synthetic data that can be used for fraud detection research. Statistical and a Social Network Analysis (SNA) of relations between staff and customers was used to develop and calibrate the model. Our ultimate goal is for RetSim to be usable to model relevant scenarios to generate realistic data sets that can be used by academia, and others, to develop and reason about fraud detection methods without leaking any sensitive information about the underlying data. Synthetic data has the added benefit of being easier to acquire, faster and at less cost, for experimentation even for those that *have* access to their own data. We argue that RetSim generates data that usefully approximates the relevant aspects of the real data.

Keywords: Multi-Agent Based Simulation, Retail Store, Fraud Detection, Synthetic Data.

1. INTRODUCTION

In this paper we introduce *RetSim*, a **Ret**ail shoe store **Sim**ulation, built on the concept of Multi Agent-Based Simulation (MABS). RetSim is based on the historical transaction data provided by one of the largest Nordic shoe retailers. This data contains several hundred million records of diverse transactional data from a few years ago, and covering several years. That is, this data is recent enough to reflect current conditions, but old enough to not pose a risk from a competitor analysis standpoint.

The defence against fraud is an important topic that has seen some study. In the retail store the cost of fraud are of course ultimately transferred to the consumer, and finally impacts the overall economy. Our aim with RetSim is to learn the relevant parameters that governs the behaviour in a retail store to simulate *normal* behaviour, which is our focus in this paper.

The main contribution and focus of this paper is a method to generate anonymous synthetic data of a retail store, that can then be used as part of the necessary data for the development of fraud detection techniques. Even so, the data set generated could also be the basis for research in other fields, such as demand prediction, logistics and demand/supply research.

Later we plan to address the actual fraud and develop techniques to develop malicious agents to inject fraudulent and anomalous behaviour, and then develop and test different strategies for detecting these instances of fraud. Even though we do not address these issues in this paper, we describe some typical scenarios of fraud in a retail store. As this is our ultimate goal, fraud heavily influenced the design of RetSim.

The main goal of developing this simulation is that it enables us to share realistic fraud data, without exposing potentially business or personally sensitive information about the actual source. As data relevant for computer security research often is sensitive due to a multitude of reasons, i.e. financial, privacy related, legal, contractual and other, research has historically been hampered by a lack of publicly available relevant data sets. Our aim with this work is to address that situation. However, simulation also have other benefits, it can be much faster and less expensive than trying different scenarios of fraud, detection algorithms, and personnel and security policy approaches in an actual store. The latter also risks incurring e.g. unhappiness amongst the staff, due to trying e.g. an ill advised policy, which leads to even greater expense and unwanted problems.

Outline: The rest of this paper is organized as follows: Section 2. introduce the topic of fraud detection for retail stores and present related work. Sections 3. describes the problem, which is the generation of synthetic data of a retail store. Section 4. shows a data analysis of the current data. Section 5. presents an implementation of a MABS for our domain and shows the description of some retail fraud scenarios. We present our results and verification of the simulation in section 7. and finish with a discussion and conclusions, including future work in section 8..

2. BACKGROUND AND RELATED WORK

Simulations in the domain of retail stores have traditionally been focused on finding answers to logistics problems such as inventory management, supply management, staff scheduling and for customer queue reductions (Chaczko and Chiu, 2008; Schwaiger and Stahmer, 2003; Bovinet, 1993).

There is currently a lack of research in the area of simulation of the retail environment for fraud detection and here is where we focus in this work.

We have previously analysed the implications of using machine learning techniques for fraud detection using a synthetic dataset (Lopez-Rojas and Axelsson, 2012a). We then built a simple simulation of a financial transaction system based on these assumptions, in order to overcome our limitations and lack of real data (Lopez-Rojas and Axelsson, 2012b). However, this work was not based on any underlying data, but rather on assumptions of what such data could contain. Here we continue and build a realistic simulation based on a real data set that in the future can be used to test diverse fraud detection techniques.

Data mining based methods have been used to detect fraud (Phua et al., 2010). This lead to the result that machine learning algorithms can identify novel methods of fraud by detecting those transactions that are different (anomalous) in comparison with the benign transactions. This problem in machine learning is known as novelty detection. Supervised learning algorithms have previously been used on a synthetic data set to prove the performance of outliers detection (Abe et al., 2006), however this has not been done over transactional data. There are tools such as IDSG (IDAS Data and Scenario Generator (Lin et al., 2006)) which was developed with the purpose of generating synthetic data based on the relationship between attributes and their statistical distributions. IDSG was created to support data mining systems during their test phase and it has been used to test fraud detection systems.

Nowadays with the popularity of social networks, such as *Facebook*, the topic of Social Network Analysis (SNA) has been given special interest in the research community (Alam and Geller, 2012). Social Network Analysis is a topic that is currently being combined with Social Simulation. Both topics support each other for the benefit of representing the interactions and behaviour of agents in the specific context of social networks.

Our approach aims to fill the gap between existing methods and provide researchers with a tool that generates reliable data to experiment with different fraud detection techniques and compare them with other approaches.

3. PROBLEM

Fraud and fraud detection is an important problem that has a number of applications in diverse domains. However, in order to investigate, develop, test and improve fraud detection techniques one needs detailed information about the domain and its specific problems.

There is a lack of data sets available for research in fields such as money laundering, financial fraud and illegal payments. Disclosure of personal or private information is only one of the many concerns that those that own relevant data have. This leads to in-house solutions that are not shared with the research community and hence there can be no mutual benefit from free exchange of ideas between the many worlds of the data owners and the research community.

After describing the problem we formulated the main research question that we address on this paper:

RQ *How could we model and simulate a retail shoe store and obtaining a realistic synthetic data set for the purpose of fraud detection?*

4. Data Analysis

To better understand the problem domain we began by performing a data analysis over the historical data provided by the retailer. We are interested in finding the necessary and sufficient attributes to enable us to simulate a realistic scenario in which we could reason about and detect interesting cases of fraud.

We initially started by selecting five stores that represent different sizes of store in the company. We selected two big stores, one medium and two small. We extracted statistical information from the data set, presented in table 1. All prices given are in a fictitious currency.

Due to a lack of space we will focus our presentation of the analysis on one of the big stores by sales volume, store one. Store one is relatively richer in data than the smaller stores. This is specially interesting, since we are more likely to find actual cases of fraud in a big store. We took a sample that comprises the sales during a year. We selected the transaction tables that detail cash flow and the articles inventory, which give us a good idea of how many transactions a big store can produce in a year, and how many different types of articles and their quantities that are sold in a year.

Table 1: Statistical analysis of five stores during one year

Stat-Store	1	2	3	4	5
Transactions	147037	180626	44446	37776	28456
Receipts	43406	38376	10094	8595	7619
Returns	9,25%	9,67%	11,43%	9,89%	9,33%
Members	5509	6381	1375	1152	16
Mem. Rec	16,02%	14,14%	18.12%	22,33%	0,56%
Avg. Price	762,49	772,32	665,2	575,93	409,62
Std. Price	494,52	514,51	459,05	616,74	416,36

4.1. Statistical Analysis

The store one sample contains 147 037 records of transactions. Note that this does not mean receipts, as a single receipt can produce several records. The retailer runs a fidelity program that allows customers to register their purchases. From this one store we identified 5509 unique members that made at least one purchase during the period resulting in 16,02% of the receipts. This means that the majority of receipts belongs to unidentified customers. However in all these records we can identify the item(s), sales price and the sales clerk.

We extracted statistical information, presented in table 1 and plotted in figure 1 which represents the sales summary per day and figure 2 which shows the number of customers per day.

Some observations that stand out in the data set:

- There were 67 receipts where the customer did not pay anything for the item, it means that the discount was 100% without returning any other article to the store. This could possible be due to a fraud, and when investigated could be used for injecting malicious behaviour.
- It was very rare for a customer to buy the same article more than once in the same purchase, this happened only three times during the year.

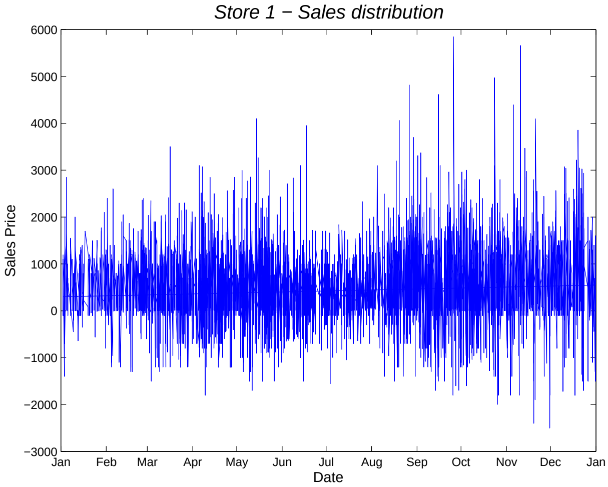


Figure 1: Store one - sales distribution

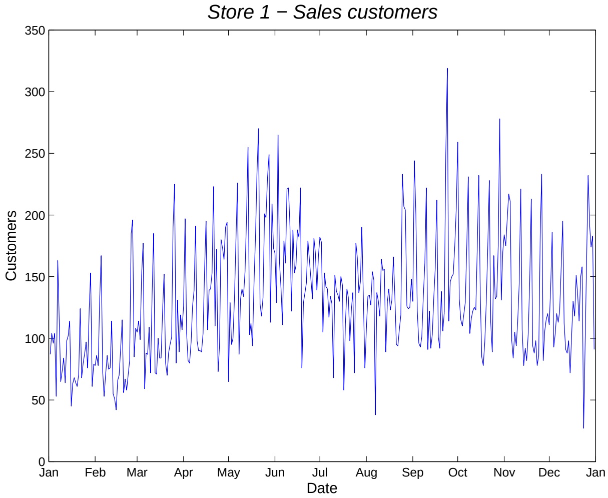


Figure 2: Store one - number of customers per day

We then investigated the performance of the staff. We divided the sales staff into three categories: *top*, *medium* and *low*. *Top* refers to staff that works regularly at the store. *Medium* refers to seasonal staff that works usually for a period between one and three months. Finally *Low* refers to staff that worked for less than one month. Table 2 shows the distribution of frequencies found in the data. Top sale clerks work an average of 66% of the time at the store, and they are only 22% of the total number of sales staff.

Table 2: Sales clerk frequency

Type	Avg. Days	Avg. Cust	Std. Cust	Quantiy
Top	155,75	45,43	28,17	22,22%
Med	63,20	38,97	23,83	11,11%
Low	13,57	33,93	16,68	66,67%

Table 3: Article categories

Category	Probability	Rank
Top	0,2705	+1000
High	0,2122	100-999
Medium	0,1109	20-99
Low	0,3495	3-19
Unfreq	0,0569	1-2

4.2. Network Analysis

Fraud has traditionally had a strong association to network analysis. Due to the possibility of several actors participating in a specific fraud in order to confuse the investigators and dilute the evidence. Another advantage of a network analysis is the ability to visualize the network by using different layout algorithms such as *Force Atlas* or *Yifan Hu* (Hu, 2005). In this project we used the *Gephi* software, that does network analysis and allows the use of different layout algorithms for the visualization of the network (Bastian et al., 2009).

We can create a network based on the interactions between each of the sales clerks and their respective customers. For the weight of the edges we use the total sales price with respect to each customer. Figure 3 shows one way to visualize the sample data extracted from the database using *Yifan Hu* layout.

The network topology resembles a hub topology, where the sales clerks are the central nodes of the hubs, and a few customers that have been helped by more than one sales clerk act as bridges between the hubs.

The store one sample contains 5545 nodes where 36 of them are sales staff, with the rest being customers. The network contains 6120 edges that connects the sales staff and customers. Each edge weight represents the total amount of purchases per customer. Table 4 show more information about the network used for calibrating the simulation.

Figure 3 shows a visualization of the network for the store, the size of the nodes is determined by the outdegree of the sales clerks. The number inside the nodes also represent the number of customers that were helped

Table 4: Network Analysis

Statistic	**Store one**
Nodes	5.545
Sales Clerks	36
Customers	5.509
Avg. Degree	1.104
Diameter Undirected	10
Avg. Path Undirected	3.98

by the sales clerk. The In-degree distribution can be better visualized in figure 4.

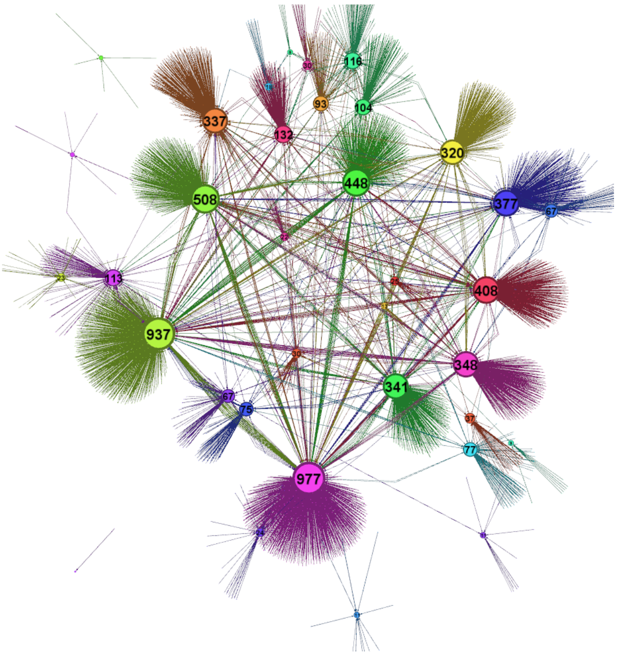


Figure 3: Store one - Network of customers and sales clerks

From the network analysis there is a lot of data we can use for our model, e.g. that 90.26% of the members have been helped by only one sales clerk, as described by the out-degree distribution.

5. MODEL AND METHOD

The design of RetSim was based on the ODD model introduced by Grimm et al. (2006). ODD contains 3 main parts: *Overview*, *Design Concepts* and *Details*.

5.1. Overview

5.1.1. Purpose

We aim to produce a simulation that resembles a real retail store. Our main purpose is to generate a synthetic data set of business transactions that can be used for the development and testing of different fraud detection techniques. It is important due to the difficulty to find diverse and enough cases of fraud in a real data set. However this is not the case of a simulated environment, where fraud can be injected following known patterns of fraud.

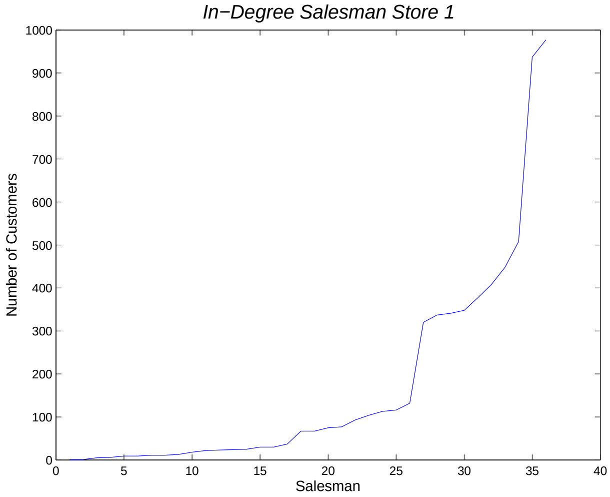


Figure 4: Store one - Customers per sales clerks

5.1.2. Entities, state variables and scales

There are three agents in this simulation: *Manager*, *Sales clerk* and *Customer*.

Manager This agent decides the price, check inventory and order new items.

Sales clerk Is in charge of promoting the items and issues the receipt after each sale. A sales clerk can be in state busy when the clerk is serving its maximum amount of customers.

Customer The behaviour is determined by the goal of purchasing one or several items. A customer is in an active *need-help* state, when no sales clerk is assisting with shopping.

5.1.3. Process overview and scheduling

During a normal step of the simulation a customer enters the simulation, and a sales clerk sense nearby customers in the *need-help* state and offers help. There are two different outcomes: Either a transaction takes place, with probability p, or no transaction takes place with, trivially, probability $1 - p$.

The time granularity of the simulation is that each step represents a day of sales. So a normal week has seven steps and a month will consist of around 30 steps. We do not make any explicit distinction between specific days of the week. Instead we handle differences between days by using a different distribution of the customers per day (see figure 2).

5.2. Design Concepts

The *basic principle* of this model is the concept of a commercial transactions. We can observe an *emergent* social network from the relation between the customers and the sales clerks. Each of the customers have the *objective* of purchasing articles from the store. The sales clerks *objective* is to aid the customers and produce the receipt necessary for the generation of the data set. Managers play a special role in the simulation. They serve as the schedulers for the next step of the simulation. Given the specific step of the simulation the manager generate a supply of customers for the next day and activate or deactivate specific sales clerks in the store. In our virtual environment the *interaction* between agents is always between sales clerk and customer. Purchase articles from another customer or selling articles to a sales clerk is not permitted.

Customers and sales clerks can scout the store in any radial direction from their current position and search or offer help, respectively.

The agents do not perform any specific learning activities. Their behaviour is given by probabilistic Markov models where the probabilities are extracted from the real data set.

5.3. Details

5.3.1. Initialization

The simulation starts with a number of sales clerks that serve the customers, an initial number of customers and one manager that does the scheduling.

The In-degree distribution is used as an indication of how good a sales clerk can be. Each sales clerk is assigned an in-degree value in each step of the simulation when the sales clerk searches for customers in need of assistance. The bigger their in-degree the more customers they can help.

5.3.2. Input Data

RetSim has different inputs needed in order to run a simulation. The input data concerns the distributions of probabilities for scheduling the sales clerks, the items that can be purchased and different statistic measures for the customers. A CSV file which contains an identifier, description, price, quantity sold and total sales specify these inputs. For setting the parameters, including the name of the CSV-file, we use a parameter file that is loaded as the simulation starts or the can also be set manually in the GUI.

5.3.3. Submodels

Figure 5 shows the different use cases of the agents. This model represent the different actions that an agent can take inside the system.

Manager scheduler This agent is in charge of scheduling the next step of the simulation. There is only one manager per store. This agent creates the new customers that are going to arrive to the store according to a distribution function extracted from the original data set. The manager also allocate the sales clerks that are going to be active during the this step of the simulation.

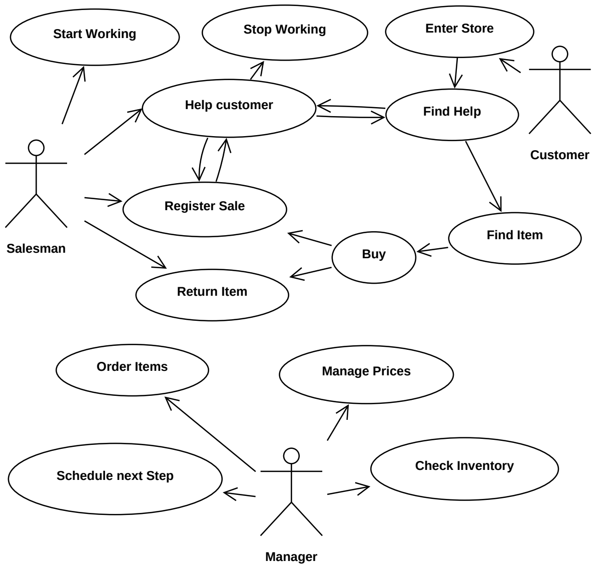


Figure 5: RetSim Use Case Diagram

Customer finder Is performed by the sales clerk and it starts with the agent searching nearby for a customer that is not being helped by an other sales clerk. Once the contact is established a sale is likely to occur with a certain probability.

Sales clerk finder Customers that are still in need for help can also look for nearby sales clerks. This again could lead to a sale.

Network generation Every time a transaction is performed between a customer and a sales clerk, an edge is created in the network composed of the customers and the sales clerks in attendance. The weight of the edge represent the sales price. The network grows by the inclusion of new customers or sales clerks.

Item selection for purchasing Items are classified into 5 different categories according to their quantity or units sold (see table 3). From the original data we extracted the probabilities of each of the categories and quantities. A customer can also purchase more than one item.

Item return after purchasing A customer can also decide to return a purchased item with a certain probability *p*.

Log of receipt transactions Each time an item is purchased a receipt is created. A receipt contains the information about the customer, sales clerk, item(s), quantities, sales price, date and discount if any.

6. Fraud Scenarios in a Retail Store

In this section we describe how three examples of retail fraud can be implemented in RetSim. These fraud scenarios are based on selected cases from Thornton (2009) report. As can be seen in section 5., the different scenarios can be implemented in almost the same way. Furthermore, a fraudulent sales clerk will probably use several different methods of fraud, which means that RetSim needs to be able to model combinations of all fraud scenarios implemented. Although the implementation of these scenarios are out of the scope of this paper, we include a description and explain how to implement them in RetSim.

6.1. Sales cancellations

This scenario includes cases where the sales clerk cancels some of the items in the sale without telling the customer, i.e., the customer pays the full sales price, and the sales clerk keeps the difference. In terms of the object model used in RetSim the sales cancellation scenario can be implemented by the following setting: Estimate the average number of cancellations per sale and the corresponding standard deviation. Use these statistics for simulating normal cancellations in the RetSim model. Fraudulent sales clerks will perform normal cancellations, as well as fraudulent once. The volume of fraudulent cancellations can be modelled using a sales clerk specific parameter. The "red flag” for detection will in this case be a high number of cancellations for a sales clerk with a low number of average sales.

6.2. Refunds

This scenario includes cases where the sales clerk creates fraudulent refund slips, keeping the cash refund for him-or herself. In terms of the object model used in RetSim the refund scenario can be implemented by the following setting: Estimate the average number of refunds per sale and the corresponding standard deviation. Use these statistics for simulating refunds in the RetSim model. Fraudulent sales clerks will perform normal refunds, as well as fraudulent once. The volume of fraudulent refunds can be modelled using a sales clerk specific parameter. The "red flag” for detection will in this case be a high number of refunds for a sales clerk.

6.3. Coupon reductions/discounts

This scenario includes cases where the sales clerk registers a discount on the sale without telling the customer, i.e., the customer pays the full sales price, and the sales clerk keeps the difference. In terms of the object model used in RetSim the coupon reduction/discounts scenario can be implemented by the following setting: Estimate the average number of cancellations per sale and the corresponding standard deviation. Use these statistics for simulating discounts in the RetSim model. Fraudulent sales clerks will perform normal discounts, as well as fraudulent ones. The volume of fraudulent discounts can be modelled using a sales clerk specific parameter. The "red flag" for detection will in this case be a high number of discounts for a sales clerk with a low number of average sales.

7. RESULTS

RetSim uses the Multi-Agent Based Simulation toolkit MASON which is implemented in Java (Luke, 2005). MASON offers several tools that aid the development of a MABS. We justified our choice mainly for the benefits of supporting multi-platform, parallellization, good execution speed in comparison with other agent frameworks; which is specially important for computationally intensive simulations such as RetSim (Railsback et al., 2006). RetSim can be run with GUI, that helps the user see the states and relations between the sales clerks (bigger circles) and customers, as can be seen in the example in figure 6.

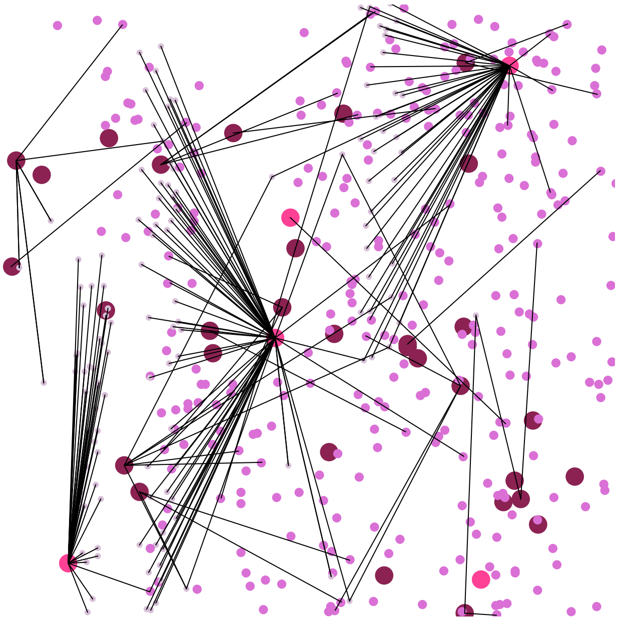


Figure 6: Screenshot of RetSim during a step

In RetSim we do not make any distinction between customers that are part of the membership programme or not. RetSim assumes that all the customers are members. This give us a way to track individual behaviours of all customers, which is beneficial.

The output of RetSim is a CSV file that contains the fields: *Step*, *Type* of *Transaction* (e.g. one sale, three returns), *Customer Id*, *Sales Clerk Id*, *Sales Price*, *Item Id* and *Item Description*.

7.1. Scenarios simulated

We aimed to perform a simulation that would produce a comparable data set to our sample data set which contained 36 sales clerks and around 45000 receipts and 81500 articles sold. The simulation was loaded with a subset of about 11000 articles from the real store.

We ran RetSim for 361 steps (working days of the store), several times and calibrated the parameters given in order to obtain a distribution that get closer enough to be reliable for testing. We collected several log files and selected three from the latest runs. Table 5 compares three runs of RetSim against the original data. Since this is a randomised simulation the values are of course not identical.

Table 5: Statistical Analysis Store one vs RetSim Simulations

Statistic	Store 1	RetSim1	RetSim2	RetSim3
Articles sold	81441	103716	95847	96492
Avg. Sales Price	372.3	405.5	405.2	407.1
Std. Sales Price	510.9	555.1	550.7	552.2

7.2. Social Network Calibration

We experimented with calibrating our results and aim to simulate the network presented in section 4.2.. Our aim was to obtain approximately the same amount of nodes and edges. We used the out-degree distribution to associate sales clerks with customers. So each sales clerk is capable to handle more or less customers during each step of the simulation and this creates the difference between nodes. This difference is interpreted in the real world by two parameters. The first is how many days a sales clerk work and the second is how good sales clerks they are. Accordingly, we only allow sales clerks with a high *in-degree* to be active during most of the steps. It means that we deactivate some sales clerks during any one specific step.

After several experimental runs and around 180 steps, keeping the most of the parameters from the original simulation, we selected one of the simulation runs to show in table 6.

Table 6: Network Simulated

Statistic	RetSim
Nodes	4948
Edges	5339
Sales Clerks	36
Customers	5303
Avg. Degree	1.079
Avg. Weighted Degree	499.1
Modularity Undirected	0.845
Diameter Undirected	8
Avg. Path Undirected	4.19

7.3. Evaluation of the model

We start the evaluation of our model with the verification and validation of the generated simulation data (Ormerod and Rosewell, 2009). The verification ensures that the simulation correspond to the described model presented by the chosen scenarios. We described RetSim in section 5.. In our model, we have included several characteristics from a real store, and successfully generated a distribution of sales that involved the interaction of sales clerks and customers. However, there are a few characteristics left from the real model such as discounts.

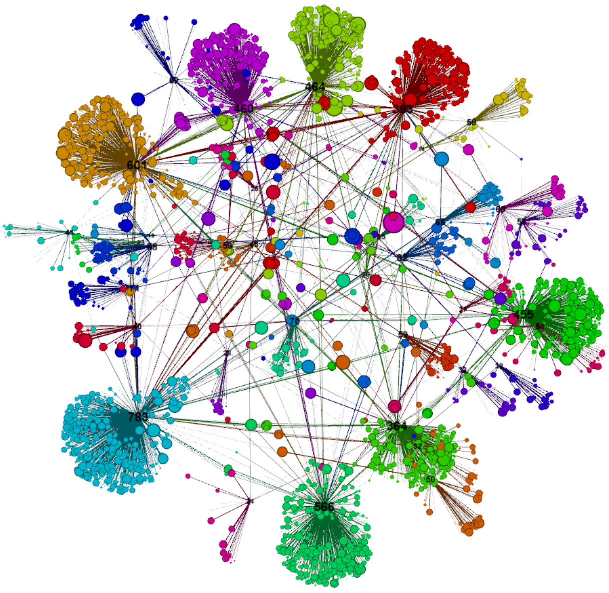


Figure 7: Small Simulated network

The validation of the model answer the question: *Is the model a realistic model of the real problem we are addressing?* After several runs of the simulation to calibrate it, we are able to answer that question affirmatively. We present some generated distributions of sales that are comparable visually in figure 8, 9 and 10.

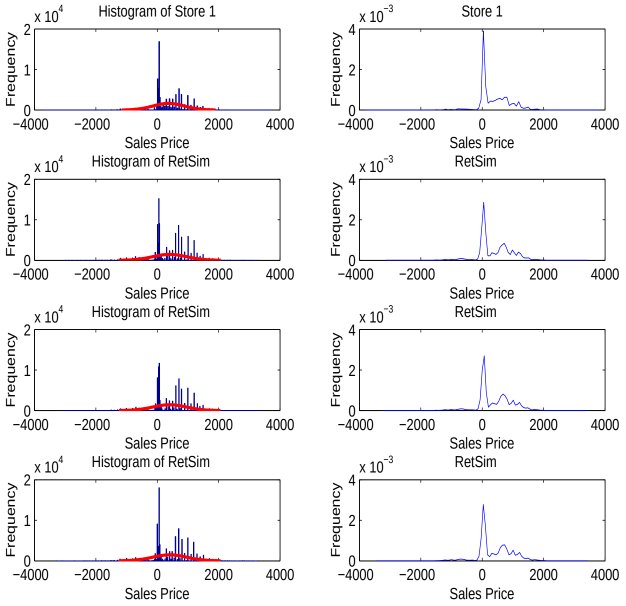


Figure 8: Comparison of simulated vs real data

Figure 8 shows a comparison of RetSim and the real sample data extracted from store one. We note several things: first the shape of the distributions look similar. Before zero are all the returns with a shape of a flat normal distribution. Between zero and 100 are the most frequently sold items such as shoe laces or accessories, which produces a peak. After 100 and before 2000 is the most common rank for shoes, so it presents another part of the distribution that contains the mean.

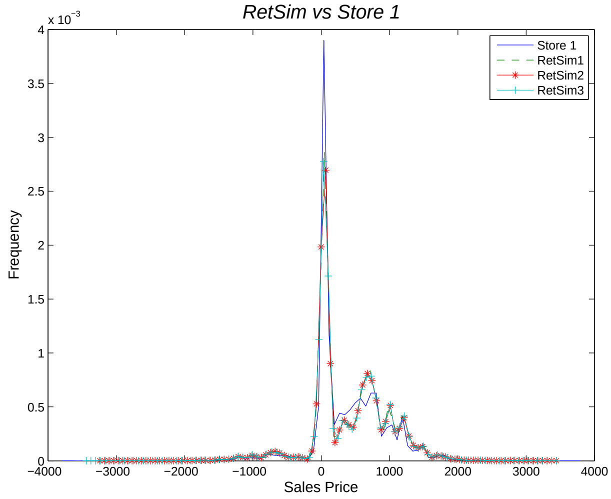


Figure 9: Comparison of distribution of simulated vs real data

Figure 9 shows an overlap of our sample store with different simulation runs by RetSim. Visually the distributions look similar. However there are several differences in the small shapes.

In figure 10 we can see a box plot comparison of store one with the RetSim runs. We can visually identify that the five statistical measures provided by the box plot are similar without being identical.

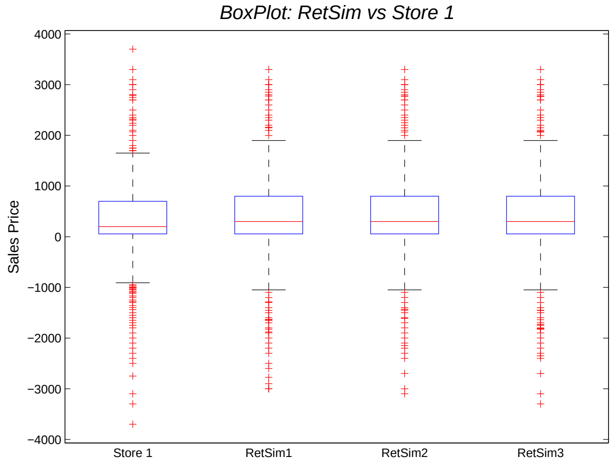


Figure 10: Box plot of simulated vs real data

Now we will focus on evaluating the simulated network presented in section 7.2.. The simulation in comparison with the original data seems visually very similar. There are similarities between the hub topology, number of nodes, and sales clerks. However we also find some dissimilarities between the weighted average degree, which in the simulation was below the original data.

There is more homogeneity between the purchases of the customers in the original data than in the simulated data. This could be due to the random nature of the selection of items in the simulation. Notice the visual differences between figure 3 and 7.

Another difference that we found is that the simulated network generates one single giant component. In the original data we could perceive a few sales clerks that perhaps just worked there for a single/few days and only served few customers. Those sales clerks are identified as islands and separated components. The analysis of these islands might be of interest for fraud detection.

We can also look at the modularity of the simulated network as an emerging behaviour of the customers. Both, the original and the simulated network are very similar and build their communities around the sales clerks. This can be clearly visualized by the different colours used in all the visualizations.

So in summary, our agent model with its programmed micro behaviour, produces the same type of overall interaction network that we can observe in the original data, and furthermore, this interaction network give rise to the same macro behaviour for the whole store as for the real store as well.

Since we are running a simulation we argue that the differences are not significant for our purpose, which is to use this distribution to simulate the normal behaviour of a store, and later combine this with injected anomalies and known patterns of fraud.

8. CONCLUSIONS

RetSim is a simulation of a retail shoe store with the objective to generate a sales data set that can be used for research into fraud detection. Synthetic data sets generated with RetSim can aid academia, companies and governmental agencies to test their methods or to compare the performance of different methods under similar conditions on the same test data set.

In section 3. we formulated our research question for this paper: *How could we model and simulate a retail shoe store and obtaining a realistic synthetic data set for the purpose of fraud detection?* In section 5. we presented the RetSim model, which is based on the ODD methodology. In order to better support our claim and answer our research question we analysed the type of data needed to generate and output as a CVS file (see section 7.) and we evaluated and verified our model in section 7.3..

It is important to know how much information from the real data set is contained in the generated synthetic data. First we do not keep any record of who is purchasing anything in the store, we based our simulation purely on statistical measures and network measures that give us an approximate description of how the individual agents behave. This means that the retail store can be sure that the privacy from the customers is preserved when using RetSim.

We argue that RetSim is ready to be used as a generator of synthetic data sets of commercial activity of a retail store. Data sets generated by RetSim can be used to implement fraud detection scenarios and malicious behaviour scenarios such as a sales clerk returning stolen shoes or unusually low productivity of a sales clerk during a specific day which could mean that the clerk is not entering some of the receipts into the system. We will make a stable released of RetSim available to the research community together with standard data sets developed for this article and further research.

For future work we plan several improvements of and additions to the current model. RetSim can be calibrated to improve the results presented in section 7. and make the data set more realistic.

In order to generate records with malicious behaviour we plan to extend RetSim to also generate malicious activity that can come from the sales clerk, customer or even the managers, or combinations of these.

Among the additions we consider are: inventory control, discounts and promotions that affect the demand of certain products. We can also add hidden parameters to sales clerks such as skills in sales, which will increase the number of customers and the average cost of items purchased. Another possible inclusion in future versions is an interesting behaviour, the self transaction, where a sales clerk can play the role of a customer and a sales clerk at the same time. This behaviour can play a key role in order to find cases of fraud.

ACKNOWLEDGMENTS

The first author wants to thank his wife Helena Lopez for bringing life to his daughter Linnea just a couple of weeks after this work was submitted. Both wife and daughter helped as inspiration and motivation for this work.

REFERENCES

Naoki Abe, Bianca Zadrozny, and John Langford. Outlier detection by active learning. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 504, 2006. doi: 10.1145/1150402.1150459.

SJ Alam and Armando Geller. Networks in agent-based social simulation. *Agent-based models of geographical systems*, pages 77--79, 2012.

Mathieu Bastian, Sebastien Heymann, and M Jacomy. Gephi: An open source software for exploring and manipulating networks. *International AAAI conference on …*, 2009.

JW Bovinet. *RETSIM: A Retail Simulation with a Small Business Perspective*. West Pub. Co., Minneapolis/St. Paul, 1993. ISBN 0314016708.

Z. Chaczko and C.C. Chiu. A smart-shop system - Multi-agent simulation system for monitoring retail activities. pages 20--26, 2008. ISBN 8890073268;978-889007326-7.

Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K. Heinz, Geir Huse, Andreas Huth, Jane U. Jepsen, Christian Jø rgensen, Wolf M. Mooij, Birgit Müller, Guy Pe'er, Cyril Piou, Steven F. Railsback, Andrew M. Robbins, Martha M. Robbins, Eva Rossmanith, Nadja Rüger, Espen Strand, Sami Souissi, Richard a. Stillman, Rune Vabø, Ute Visser, and Donald L. DeAngelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115--126, September 2006. ISSN 03043800. doi: 10.1016/j.ecolmodel.2006.04.023.

Yifan Hu. Efficient and High Quality Force-Directed Graph. *The Mathematical Journal*, 10:37--71, 2005.

P.J. Lin, B. Samadi, and Alan Cipolone. Development of a synthetic data set generator for building and testing information discovery systems. In *ITNG 2006.*, pages 707--712. IEEE, 2006. ISBN 0769524974.

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Money Laundering Detection using Synthetic Data. In Julien Karlsson, Lars ; Bidot, editor, *The 27th workshop of (SAIS)*, pages 33--40, Örebro, 2012a. Linköping University Electronic Press.

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Multi Agent Based Simulation (MABS) of Financial Transactions for Anti Money Laundering (AML). In Audun Josang and Bengt Carlsson, editors, *Nordic Conference on Secure IT Systems*, pages 25--32, Karlskrona, 2012b.

S. Luke. MASON: A Multiagent Simulation Environment. *Simulation*, 81(7):517--527, July 2005. ISSN 0037-5497. doi: 10.1177/0037549705058073.

Paul Ormerod and Bridget Rosewell. Validation and Verification of Agent-Based Models in the Social Sciences. In Flaminio Squazzoni, editor, *LNCS*, pages 130--140. Springer Berlin / Heidelberg, 2009. ISBN 978-3-642-01108-5.

Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler. A comprehensive survey of data mining-based fraud detection research. *Arxiv preprint arXiv:1009.6119*, 2010.

S. F. Railsback, S. L. Lytinen, and S. K. Jackson. Agent-based Simulation Platforms: Review and Development Recommendations. *Simulation*, 82(9):609--623, September 2006. ISSN 0037-5497. doi: 10.1177/0037549706073695.

Arndt Schwaiger and B Stahmer. SimMarket: Multiagent-based customer simulation and decision support for category management. *Multiagent System Technologies*, pages 74--84, 2003.

Grant Thornton. Reviving retail Strategies for growth in 2009 Executive summary. Technical report, Grant Thornton, 2009.

AUTHORS BIOGRAPHY

MSc. Edgar A. Lopez-Rojas

Edgar Lopez is a PhD student in Computer Science and his research area is related with Multi-Agent Based Simulation, Machine Learning techniques with applied Visualization for fraud detection and Anti Money Laundering (AML) in the domains of retail stores, payment systems and financial transactions. He obtained a Bachelors degree in Computer Science from EAFIT University in Colombia (2004). After that he worked for 5 more years at EAFIT University as a System Analysis and Developer and partially as a lecturer. In 2011 he obtained a Masters degree in Computer Science from Linköping University in Sweden.

Dr. Stefan Axelsson
Stefan Axelsson is a senior lecturer at Blekinge Institute of Technology. He received his M.Sc in computer science and engineering in 1993, and his Ph.D. in computer science in 2005, both from Chalmers University of Technology, in Gothenburg, Sweden. His research interests revolve around computer security, especially the detection of anomalous behaviour in computer networks, financial transactions and ship/cargo movements to name a few. He is also interested in how to combine the application of machine learning and information visualization to better aid the operator in understanding how the system classifies a certain behaviour as anomalous. Stefan has ten years of industry experience, most of it working with systems security issues at Ericsson.

Dan Gorton, Licentiate of Engineering
Dan Gorton is a Ph.D. candidate at KTH Royal Institute of Technology. He received his M.Sc. in computer science in 1997 at KTH Royal Institute of Technology, in Stockholm, Sweden, and a Licentiate of Engineering in computer engineering in 2003 at Chalmers University of Technology, in Gothenburg, Sweden. His current research focuses on risk management of online financial services, including fraud detection. Previous research has focused on extending intrusion detection with alert correlation and intrusion tolerance. Dan has 15 years of industry experience, working with different security and risk issues primarily within the banking, defense, and telecom sectors.

A STOCHASTIC APPROACH TO SECURITY SOFTWARE QUALITY MANAGEMENT

Vojo Bubevski
Legal & General Account
TATA Consultancy Services Ltd
London, United Kingdom
vojo.bubevski@landg.com

ABSTRACT

The conventional approach to security software quality management specifically for ongoing projects has two major limits: (1) Six Sigma is not applied; and (2) analytic risk models are used. This paper proposes a stochastic method, which applies Six Sigma Define, Measure, Analyze, Improve and Control (DMAIC), Monte Carlo Simulation and Orthogonal Security Defect Classification (OSDC). DMAIC is tactically applied to assess and improve quality. Simulation predicts quality (reliability) and identifies and quantifies the quality risk. OSDC allows qualitative analysis. DMAIC is a verified structured methodology for systematic process and quality improvements. Simulation is superior to analytic risk models. OSDC offers qualitative improvements. This synergetic method eliminates observed deficiencies gaining important benefits including savings, quality and customer satisfaction. It is CMMI® (Capability Maturity Model Integration) compliant. The method is simplistically elaborated on a published third-party project.

Keywords: Six Sigma; DMAIC; Simulation; Security Software; Quality Management;

1. INTRODUCTION

Software quality is a multidimensional attribute including reliability, functionality, usability, performance, etc. It is a direct consequence of software processes. Software processes are inherently variable and uncertain, thus involving substantial risks. A key factor in Software Quality is *Software Reliability* as it is one of the the quality attributes most exposed to customer observation. In this paper, “reliability” and “quality” are used interchangeably.

Software quality and customer satisfaction are very important. Managing the software quality, particularly for security software, is a critical factor for software projects.

Six Sigma is used across industries for improving processes, quality and customer satisfaction. One of the principal Six Sigma methodologies is *DMAIC (Define, Measure, Analyze, Improve, Control)*. In Software Engineering, Six Sigma is compatible with *Capability Maturity Model Integration (CMMI®)*. Applications of Six Sigma methodologies in Software Development are discussed in published works (Tayntor 2002; Mandl 1985; Tatsumi 1987; Brownlie, Prowse and Phadke 1992; Bernstein and Yuhas 1993; Siviy, Penn and Stoddard 2007; Nanda and Robinson 2011).

Monte Carlo simulation is a methodology which iteratively evaluates a deterministic model by applying a distribution of random numbers as inputs, which allows to use probability and statistical tools to analyze the results. It is used for modeling phenomena with significant uncertainty, such as software development processes (Bratley, Fox, and Schrage 1983; Rubinstein and Kroese 2008). The term “simulation” is generically used in this paper to refer to “Monte Carlo simulation”.

Software Reliability is a main subject in *Software Reliability Engineering (SRE)* (Lyu 1996). The software reliability analytic models have been available since the early 1970s (Lyu 1996; Kan 2002; Xie 1991). The need for a simulation approach to software reliability was recognized in 1993 by *Von Mayrhauser et al.* (Von Mayrhauser et al., 1993). Subsequently, substantial work on simulation was published (Gokhale, Lyu and Trivedi 1997; Gokhale, Lyu and Trivedi 1998; Tausworthe and Lyu 1996; Bubevski 2009; Bubevski 2010).

Six Sigma Software practitioners usually employ conventional analytic models. It has been reported that for Six Sigma in general, simulation models are superior to conventional analytic models (Ferrin, Miller and Muthler 2002).

The Orthogonal Security Defect Classification (OSDC) was established and used by *Hunny* to improve the quality of security software (Hunny 2012). OSDC is based on the Orthogonal Defect Classification (ODC), which was elaborated by *Chillarege* and implemented by IBM™ (Lyu 1996, Chapter 9). OSDC provides for applying qualitative analysis of security software.

1.1. Problem Statement and Proposal

The conventional approach to manage the quality of security software, specifically for an ongoing software projct, has two major limitations: (1) it doesn’t apply Six Sigma methods on a current project, but uses the previous release’s data to improve the quality of the next release; and (2) it uses analytic risk models.

The paper proposes a stochastic approach to Security Software Quality Management. This approach is based on the new method published by *Bubevski* (Bubevski, 2013). However, the approach presented herein applies the DMAIC and Simulation methodologies specifically to Security Software by using OSDC.

The synergy of DMAIC, Simulation and OSDC eliminates the limitations identified above. By using this method, substantial savings and quality improvements can be achieved, increasing customer satisfaction.

1.2. Related Work

Hunny used OSDC in order to improve the quality of security software. He presented how the failure data collected for past releases of two software systems, which are OSDC-classified, can be used to improve the quality of the next/future release of the systems by applying analytic models (Hunny 2012).

Bubevski elaborated stochastic approaches to software quality management, which applied Six Sigma and Simulation. The methods were demonstrated and verified on real software projects using published data (Bubevski 2009; Bubevski 2010).

Bubevski also devised and elaborated a new approach to Software Quality Management of ongoing software projects by applying the Six Sigma DMAIC, Simulation and ODC methodologies. The nw method was proven in practice achieving savings, quality imrovments and high customer satisfaction (Bubevski, 2013).

Chillarege applied ODC and the Inflection S-shaped Software Reliability Growth Model for relative risk assessment of the final testing stage. The Inflection S-shaped Software Reliability Growth Model is analytic; it is used to predict the future course of the software reliability growth curve. This helped the project to reduce risk, meet the schedule and assure good field reliability gaining significant benefits (Lyu 1996, Sec. 9.5, Sec. 3.3.6).

Tausworthe & Lyu applied simulation for software reliability assessment on the Galileo spacecraft software project at the Jet Propulsion Laboratory™. The reliability simulation results were substantially better than the reliability predictions obtained by the analytic models such as Jelinski-Moranda, Musa-Okumoto and Littlewood-Verrall models (Tausworthe and Lyu 1996).

Gokhale, Lyu & Trivedi developed simulation models for failure behaviour of the most commonly used fault tolerance architectures. They demonstrated the ability to simulate very complex failure scenarios with various non-trivial dependences (Gokhale, Lyu and Trivedi 1997).

Gokhale, Lyu & Trivedi simulated the reliability of component based software. Discrete event simulation was applied to analyze complex systems, i.e. a terminating application, and a real time application with feedback control. The simulation models applied were superior to the conventional analytic models including Prevalent Markovian and Semi Markovian methods (Gokhale, Lyu and Trivedi 1998).

Simulation was applied by *Gokhale & Lyu* for structure-based analysis of software reliability. Simulation provided for tailoring the testing and repair strategies, and achieving the desired reliability cost-effectively (Gokhale and Lyu 2005).

Siviy, Penn and *Stoddard* used Six Sigma to reduce defects and improve quality. Conventional Six Sigma tools were used such as Rayleigh Fitted Histogram Defect Model and Cause-and-Effect Model, including Computer Aided Software Reliablity Estimation (Siviy, Penn and Stoddard 2007, Sec. 9.1).

Murugappan & Keeni combined Six Sigma with CMMI® to create a quality management system. The aim was to improve software processes and achieve CMMI® Level 4 compliance, which provides for quantitative and qualitative software quality management (Murugappan and Keeni 2003).

An application of CMMI® and Six Sigma in software processes improvement was elaborated by *Xiaosong et al.* The software process management was considered and Six Sigma and CMMI® integration was implemented achieving quality improvements (Xiaosong, Zhen, ZhangMin and Dainuan 2008).

Nanda and *Robinson* published a book including two case studies, which use DMAIC and conventional Six Sigma statistical tools for software defect reduction purposes. The book demonstrates how Six Sigma is applicable to the IT industry, with compelling success stories from today's leading IT companies (Nanda and Robinson 2011, Chapter 5).

Galinac & Car elaborated an application of Six Sigma in the continuous improvement of software verification process. Appling Six Sigma, change management, and statistical tools and techniques, solved the problem of fault slippage through the verification phases (Galinac and Car 2007).

Macke & Galinac presented experiences and results of applying Six Sigma for process improvements in a global software development organization including process definition, awareness for different levels of expectations in globally distributed teams, and introduction of regular scanning mechanisms. Success indicators were defined connecting process capability to business value in order to measure the improvement success (Macke and Galinac 2008).

A six sigma DMAIC approach to software quality improvement was presented by *Redzic & Baik.* Tactical changes were identified and established, which substantially increased the software quality of all software products (Redzic and Baik 2006).

Xiaosong et al used Six Sigma DMAIC and accomplished continuous quality improvement throughout the software development process for high-quality software product. . The software process deficiencies were identified and eliminated to ensure the desired software quality (Xiaosong, Zhen, Fangfang and Shenqing 2008).

2. THE METHOD DEMONSTRATION (HYPOTHETICAL SCENARIO)

The elaboration is based on a real software project using published data (Lyu 1996, Dataset ODC4). The project is finished so this case is hypothetical. The failure data are available for the entire life cycle but only the data from the last 15 months are used (Table 1). To emulate the scenario of an ongoing project, the data from the first 12 months are used. The last three months' data are used to verify the results. We also pretend that the failure data are for a security software project. Thus, the original data use ODC, but they are mapped to OSDC (Hunny 2012, Table 3.1) because OSDC is specifically applicable to sequrity software. The OSDC Defect Types considered are: Security Functionality (SF), Security Logic (SL) and Miscellaneous (Misc.).

Table 1: Actual Failure Count Data

Month	SF	SL	Misc.
1	16	20	32
2	11	8	6
3	203	36	22
4	37	20	7
5	107	43	13
6	240	43	21
7	27	64	18
8	30	112	23
9	147	98	23
10	24	93	23
11	24	106	28
12	24	33	23
13	6	14	8
14	7	7	3
15	4	15	1

2.1. Assumptions

The method involves quantitative analysis of the metrics data, so the results are data driven. Consequently, the metrics data must be verified and reliable. Also, the organisation and the software project must have capabilities and experiences with quantitative analysis in order to use the method and provide for good and consistent results. Therefore, the fundamental assumption for the method feasibility is that the software organization and the software project are compliant with CMMI® Level 4.

CMMI® Level 4 requires quantitative management of software processes and products within an organization. Thus, the criteria are as follows: (i) detailed measures of the software process and product quality are collected; and (ii) both the software process and products are quantitatively understood and controlled.

Also, there was no information about testing profile, defect relationship, fix (removal) rate or the rate of introducing new defects in the published data. Therefore, for the purpose of the demonstration only and to simplify the simulation models, at least until the experimental results are adequately verified, it is assumed that (i) testing operation profile is uniform, (ii) failures occur independently, (iii) defects are removed in the same time interval as they are encountered and (iv) no new defects are introduced with the fix. It should be considered that the demonstration simulation model's results were satisfactorily verified, so the assumptions were proven to be acceptable.

2.2. Software Quality Risk Management Using Six Sigma and Simulation

The method follows the DMAIC methodology as a tactical framework.

2.2.1. The Project Definition (Define)

We assume that the project is within the final testing stage at the end of Month 12 (TI(12)), which is three months from the targeted delivery date of the product.

Project Objective: Complete final test phase by the end of Month 15 (TI(15)) as planned and deliver the system on time, whilst achieving the quality goal. The delivery date is at the beginning of Month 16.

Project Quality Goal: The aim is to ensure that the system is stable and ready for delivery. All detected defects should be fixed and re-tested before the end of testing. Also, the final month of testing (Month 15) should have one defect per Defect-Type and three defects in total. Maximum two defects per Defect-Type and six defects in total are allowed.

Problem Statement: Assess and mitigate the risk to deliver the system on time, whilst achieving the quality goal. Critical to Quality (CTQ) for the project is the sequrity software reliability.

2.2.2. The Project Metrics (Measure)

In order to define the Failure Intensity Function (FIF) deterministic models by Defect-Type, which are required for simulation, we need to (1) transform the actual data by applying Rank Transformation (Conover and Iman 1981) to get the transformed FIF by Defect Type; and (2) approximate the transformed FIF by Defect Type.

Logarithmic and exponential approximations were tried. The R-square values by Defect Type were (1) Logarithmic: i) SF: $R^2 = 0.9254$; ii) SL: $R^2 = 0.8981$; iii) Misc.: $R^2 = 0.7385$; and iv) Total: $R^2 = 0.9604$; and (2) Exponential: i) SF: $R^2 = 0.8999$; ii) SL: $R^2 = 0.9276$; iii) Misc.: $R^2 = 0.7642$; and iv) Total: $R^2 = 0.9665$.

Comparing the R-square values, the exponential approximation is more accurate. Thus, the exponential approximation of the FIF is selected, which is used by the Musa's Basic Execution Time software reliability model (Lyu 19969, Sec. 3.3.4). The aproximations, i.e. the deterministic models of the FIF by Defect Type are as follows:

$$FIFf(k) = 262.33 \exp(-0.274\ k) \quad (1)$$
$$FIFl(k) = 179.17 \exp(-0.217\ k) \quad (2)$$
$$FIFm(k) = 41.138 \exp(-0.127\ k) \quad (3)$$
$$FIFt(k) = FIFf(k) + FIFa(k) + FIFm(k) \quad (4)$$

Where, FIFf, FIFl, FIFm and FIFt are the FIF for SF, SL & Misc. Defect-Type and the Total respectively, and k is the time interval (k = 1,2,…, n).

2.2.3. Six Sigma Process Simulation (Analyze)

To analyze the process, we simulate the FIFs for the future three months , i.e. from TI(13) to TI(15) inclusive. The simulation is based on the Musa's Basic Execution Time deterministic model (Lyu 19969, Sec. 3.3.4), which applies exponential FIFs. The Poisson distribution is used for the simulation.

To define the quality targets for Month 15 we use the Six Sigma *Target Value*, *Lower Specified Limit (LSL)* and *Upper Specified Limit (USL)*: a) Target Value is one for all defect types and three defects for the Total; b) USL is two for all defect types and six for the Total; b) LSL should be zero, but it will be set to a very small negative number to prevent an error in the Six Sigma metrics calculations, i.e. LSL is -0.0001 for all defect types including the Total. The Six Sigma process simulation results follows.

Figure 1 shows that the Total's distribution in Month 15 of testing totally deviates from the target specifications (i.e. LSL, Target Value and USL). Also, there is a 0.90 probability that the Total would be in the range 11 – 25; 0.05 probability that the Total would be more than 25; and 0.05 probability that the Total would be less than 11.

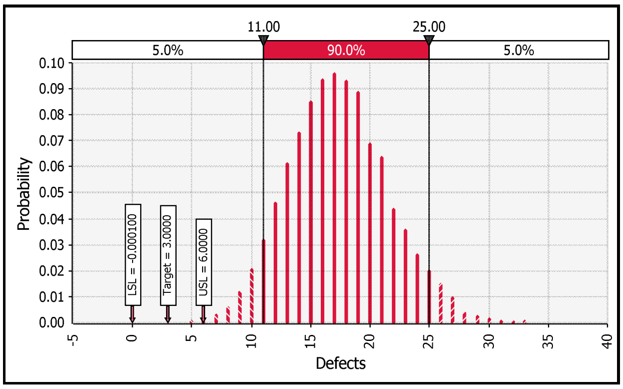


Figure 1: Total Defects Probability Distribution Month 15

Table 2 shows the predicted mean (μ), Standard Deviation (σ) and Minimum and Maximum Values for total number of defects by Defect-Type in the final month of testing TI(15) including the Total.

Table 1: Predicted FIF for Month 15

Process	μ	σ	Min	Max
SF	4	2.06	0	14
SL	7	2.62	0	19
Misc.	6	2.47	0	19
Total	17	4.19	5	36

The predicted Total in TI(15) is 17, with Standard Deviation of 4.19 defects. This indicates that the product will not be stable for delivery at the end of Month 15.

The Six Sigma metrics used to measure the performance are: a) *Process Capability (Cp)* ; b) *Sigma Level*; and c) *Probability of Non-Compliance (PNC).* The Sigma metrics by Defect-Type for Month 15 is given in Table 3. For example, the SF type has the lowest PNC equal to 0.8057, which is 80.57% deviation from the desired target range. The PNC for the Total is equal to 0.9988, which is 99.88% deviation from the specified target. All three Six Sigma metrics strongly suggest that the process would not perform well, so it would not deliver the desired quality at the end of Month 15.

Table 3: Process Six Sigma Metrics for Month 15

Process	Cp	PNC	Sigma Level
SF	0.1618	0.8057	0.2460
SL	0.1272	0.9687	0.0392
Misc.	0.1349	0.9461	0.0676
Total	0.2389	0.9988	0.0015

2.2.4. Six Sigma Simulation Sensitivity Analysis: CTQs Identification (Analyze)

The simulation sensitivity analysis is used to determine the influence of the change of a particular Defect-Type to the change of Total Defects for all Defect Types.

The correlation sensitivity shows that the most influential defect type, i.e. the top risk CTQ, is SL Defect-Type with correlation coefficient to the Total of 0.63. The Misc. and SF Defect-Type are less influential as their correlation coefficients are 0.58 and 0.49 respectively.

The regression sensitivity shows the quantitative parameters of the influence of the defect types to the total if they change by one Standard Deviation. That is, if the SL defects increase by one Standard Deviation, the Total will increase by 2.62 defects, which is the top risk defect type.

The regression coefficients for Misc. and SF defects are 2.47 and 2.06 defects respectively, so they are less influential. These results are consistent with the correlation sensitivity results.

2.2.5. Six Sigma Analysis Conclusions and Recommendation (Analyze)

The following are the conclusions from this Six Sigma analysis: a) The testing process would not perform well as shown by the considered Six Sigma metrics. Therefore, the system would not be ready for delivery as the quality goals would not be met at the beginning of Month 16 if the project maintains the current situation; and b) The CTQ to deliver the system is the software reliability, i.e. the predicted Total in TI(15) is 17 defects, versus the target value of three defects.

Analysis Recommendation: In order to deliver the system on time and achieve the quality goal, immediately undertake an improvement project to improve the process and enhance the software reliability, which is the CTQ.

2.2.6. Improvement Six Sigma Simulation (Improve)

The purpose of this Six Sigma simulation is to quantitatively determine the solution for improvement, i.e. to predict all the escaped defects (i.e. the defects that are believed to be in the system but they are not captured). Therefore, the software reliability for the future period will be simulated to predict when the reliability goal will be achieved.

It was analyzed and identified that this target could be met in Month 24. Thus, FIF by Defect-Type was simulated for the future period of 12 months, i.e. from Month 13 to Month 24. All the parameters for this simulation were exactly the same as for the previous simulation.

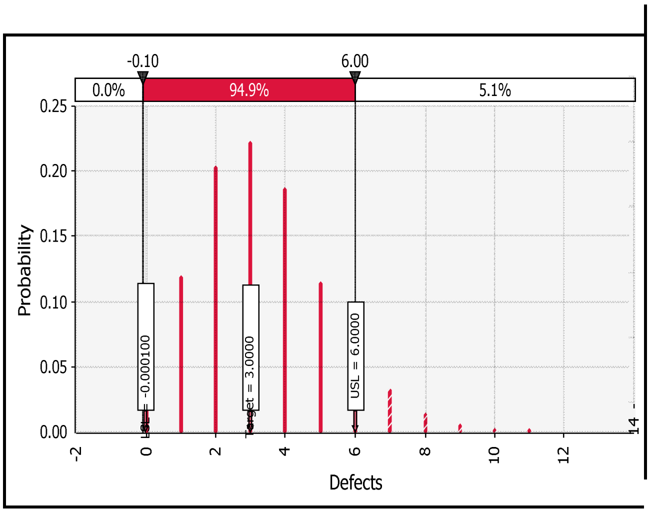


Figure 2: Total Defects Probability Distribution Month 24

As Figure 2 shows, the Total's distribution in Month 24 of testing fits in the process target specifications (LSL, Target Value and USL are marked on the graph). Also, there is a 0.949 (94.9%) probability that the Total in TI(24) would be in the specified target range 0-6 defects; and 0.051 (5.1%) probability that the Total would be more than six. The probability that there would be three defects in total is approximately 0.22 (22%).

According to this prediction, the process could achieve the reliability goal in Month 24 if the project maintains the current situation.

Table 4: Predicted FIF for Month 24

Process	μ	σ	Min	Max
SF	0	0.61	0	4
SL	1	0.98	0	7
Misc.	2	1.39	0	8
Total	3	1.81	0	12

Table 4 shows that predicted number of defects for all types, including the Total, is within the specified target range. The Standard Deviation for all types including the Total, however, is relatively high.

The process Six Sigma metrics at the end of Month 24 are given in Table 5. For example, PNC for Misc. defects is 0.3112 (i.e. 31.12% deviation). The Total however shows only 5.11% deviation.

Table 5. Process Six Sigma Metrics for Month 24

Process	Cp	PNC	Sigma Level
SF	0.5468	0.0074	2.6783
SL	0.3397	0.0739	1.7872
Misc.	0.2391	0.3112	1.0127
Total	0.5510	0.0511	1.9506

All three Six Sigma metrics suggest that there are realistic chances that the process could perform and deliver the desired quality at the end of Month 24.

2.2.7. Improvement Recommendations (Improve)

The following defines and quantifies the solution for the improvement. The predicted total numbers of defects by Defect-Type including the Total for the future periods are shown in Table 6.

The predicted defects for Month 13 – 15 are expected to be detected and removed by the current project until the end of Month 15. The predicted defects expected to be found in the system from Month 16 to Month 24 are unaccounted for. These defects need to be detected and removed until the end of Month 15 in order to achieve the quality goal.

Table 6: Predicted Defects per Defect-Type for Future Periods

Time Period	SF	SL	Misc	Total
TI(13) – TI(15)	17	27	21	65
TI(16) – TI(24)	11	25	31	67
TI(13) – TI(24)	28	52	52	132

Therefore, the process improvement recommendation is: Immediately undertake an improvement project to deliver the system quality improvements as required to achieve the quality goals. The objectives of this project are:

1. Reanalyze the unstable defects applying *Casual Analysis and Resolution (CAR)*;
2. Determine the quality improvement action plan, establishing an additional tactical test plan;
3. Execute the tactical test plan to additionally test the system and detect and repair the escaped defects, i.e. the defects that is believed are in the system but have not been detected. According to the simulation above, there are 67 predicted escaped defects in total (TI(16) – TI(24), Table VI);
4. The additional testing, detection and correction of the escaped defects should be completed by the end of Month 15 to achieve the quality goal.

2.2.8. The Improvement Project and Employment of Additional Resources (Improve)

To undertake the improvement project, additional resources with special skills are needed. The current project is not behind schedule and is running according to plan. The problem is the quality of the product.

To minimize the Brooks' Law effect in employing the additional resources, a "surgical team" should be assigned to the project (Brooks 1995). The objective of the "surgical team" is to deliver the required quality improvement only. The current team working on the project should continue their work according to plan. The "surgical team" will not share any work with the current team.

2.2.9. Improvement Definition (Improve)

The process improvement is a new testing project, which is totally independent of the current testing in progress. There are only three months available to accomplish the improvement, as the quality goal needs to be met at the end of testing (i.e. at the ond of Month 15).

Keeping one month as the time interval for observation is not good because it provides for only two future check points. Thus, the time interval for observation will be reduced to one week. Thus, the proposed schedule for the testing improvement project during the next 13 weeks is: a) one week to start the project and appoint the staff; b) three weeks to complete the required analysis and test plans; and c) nine weeks of testing where the escaped defects will be detected and fixed.

The predicted distribution of the escaped defects by Defect-Type including the Total, which need to be detected and fixed during the testing period of nine weeks, i.e. TI(1) – TI(9), is: SF: 11 Defects; b) SL: 25 Defects; c) Misc.: 31 Defects; and d) Total: 67 Defects.

2.2.10. Six Sigma Simulation for Monitoring (Control)

It is imperative to establish continuous monitoring in order to discover any variances in the process performance, and determine and implement the appropriate corrective actions to eliminate the deviations. This will ultimately mitigate the risk and allow for the delivery of the product on time and the achievement of the quality goals.

In order to deliver the product on time and meet the quality goals, the control phase should be applied to both the current and the improvement testing process. It is recommended to create two additional Six Sigma simulation models and to apply them regularly on a weekly basis to both processes until the end of the projects.

A Six Sigma simulation model for monitoring of the improvement testing process will be demonstrated now. It is assumed that the improvement testing project is at the end of Week 3. An actual defect distribution by Defect-Type for the first three weeks of testing is also assumed, which is given in Table 7.

We need first to transform the assumed actual failures over the three weeks period, to determine the FIF. Also, we need to approximate the FIF by Defect-Type as required for the simulation. The logarithmic and exponential approximations were tried.

Table 7: Assumed Actual Failure Count Data

TI (Week)	SF	SL	Misc.	Total
1	2	3	5	10
2	3	3	4	10
3	2	4	4	10
Total:	7	10	13	30

The R-square values are:

(1) Logarithmic: i) SF: $R^2 = 0.8668$; ii) SL: $R^2 = 0.8668$; iii) Misc.: $R^2 = 0.8668$; and iv) Total: $R^2 = 0.8668$.

(2) Exponential: i) SF: $R^2 = 0.75$; ii) SL: $R^2 = 0.75$; iii) Misc.: $R^2 = 0.75$; and iv) Total: $R^2 = 0.75$.

Comparing the R-square values, the logarithmic approximation is more precise. Thus, we will select the Logarithmic Poisson Reliability Model for the simulation. The aproximation of the FIF by Defect Type is as follows:

$$FIFf(k) = -0.968 \ln(k) + 2.9112 \quad (5)$$

$$\mathrm{FIFl}(k) = -0.968 \ln(k) + 3.9112 \tag{6}$$

$$\mathrm{FIFm}(k) = -0.968 \ln(k) + 4.9112 \tag{7}$$

$$\mathrm{FIFt}(k) = \mathrm{FIFf}(k) + \mathrm{FIFa}(k) + \mathrm{FIFm}(k) \tag{8}$$

Where, FIFf, FIFl, FIFm and FIFt are the FIF for SF, SL & Misc. Defect-Type and the Total respectively, and k is the time interval (k = 1,2,…, n).

The software reliability for the future period of six weeks will be predicted (simulated), i.e. from TI(4) to TI(9) inclusive. For this prediction, the discrete event simulation is used applying the Poisson distribution on the formulas above (5 - 8).

The major objective of the improvement project is to capture and fix the escaped defects. The escaped defect distibution by Defect-Type is a) SF: 11 Defects; b) SL: 25 Defects; c) Misc.: 31 Defects; and d) Total: 67 Defects. Thus, the Six Sigma Target Value, LSL and USL are: a) SF: Target Value is 11, LSL is 9 and USL is 13; b) SL: Target Value is 25, LSL is 22 and USL is 28; c) Misc.: Target Value is 31, LSL is 28 and USL is 35; and d) Total: Target Value is 67, LSL is 60 and USL is 74.

The process Six Sigma simulation results (Figure 3) show that the Total's distribution for the final week of testing TI(9) fits well within the process target specifications. For example, there is a 0.742 (74.2%) probability that the Total in Week 9 would be in the specified target range 60-74 defects. This indicates that the improvement project could achieve the quality target.

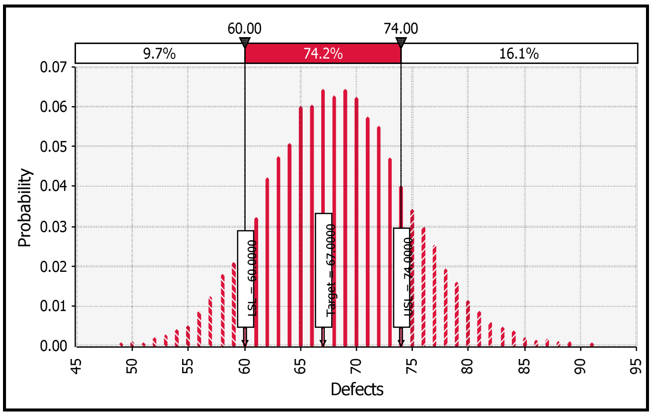


Figure 3: Total Defects Probability Distribution for Week 9

Table 8: Predicted FIF for Week 9

Process	μ	σ	Min	Max
SF	14	2.61	7	26
SL	23	3.56	12	39
Misc.	32	4.32	18	50
Total	69	6.16	48	91

Also, for Week 9 (Table 8) the predicted Total is 69 defects with Standard Deviation of 6.16 defects (8.93%), which is acceptable.

The process Six Sigma metrics (Table IX) shows that for the Total, the PNC metric is 0.2582, i.e. 25.82% deviation from the desired target range, which is acceptable. Therefore, the chances that the improvement testing process could perform as expected are high.

Table IX. Improvement Testing Process Six Sigma Metrics

Process	Cp	PNC	Sigma Level
SF	0.2554	0.5609	0.5815
SL	0.2812	0.5520	0.5948
Misc.	0.2700	0.4230	0.8012
Total	0.3789	0.2582	1.1307

All three Six Sigma metrics strongly suggest that the improvement testing process performed well during the first three weeks of testing. Therefore, there is no need for any corrective action as at the end of Week 3. However, it is required to continue to analyze the process performance by applying the above DMAIC-Simulation analysis regularly, i.e. at the end of every week, until the end of the project.

Similarly, a Six Sigma simulation model can be easily created to monitor the current testing process regularly on a weekly basis until the end of the project. For this purpose, the predicted defect distribution for the period TI(13) – TI(15) should be transformed in a desired weekly defect distribution.

2.2.11. Verification of Results

The experimental results, i.e. the predictions, are compared with the actual available data for verification. It should be underlined that there are no data available from System's Operation. Thus, it is impossible to verify the predictions for improvments and predictions for control.

Two comparisons are performed as presented below: a) Partial Data Comparison; and b) Overall Data Comparison.

Partial Data Comparison:

Table 10: Partial Data Comparison

Process	Defects		
	Actual	Pred.	Error %
SF	17	17	0
SL	36	27	-25
Misc.	12	21	75
Total	65	65	0

The results (Table 10) are verified by comparing the predicted total number of defects by Defect-Type including the Total for the three months period TI(13) – TI(15), versus the corresponding actual defects.

The SF defects and the Total are accurately predicted. The SL defects are underestimatedand Misc. defects are overestimated. These prediction results are acceptable.

Overall Data Comparison:

The overall data comparison is shown in Table XI.

Table 11: Overall Data Comparison

	Defects		
Process	Actual	Pred.	Error %
SF	907	907	0
SL	712	703	-1.2640
Misc.	251	260	3.5857
Total	1870	1870	0

The results are verified by comparing the actual and predicted total number of defects by Defect-Type including the Total for the entire period TI(1) – TI(15), with the corresponding actual defects. Again, the SF defects and the Total are accurately predicted. The SL defects are underestimated with a minimal error. The Misc. defects are slightly overestimated. Thus, these prediction results are very good.

Considering the calculated errors in Table 10 and Table 11, the experimental results are satisfactorily verified.

3. CONCLUSION

The conventional security software quality management of ongoing projects has two major weaknesses: i) analytic risk models are used; and ii) structured methodologies for process and quality improvements are not systematically applied. The proposed novel practical method applies Six Sigma DMAIC, Monte Carlo Simulation and OSDC methodologies. Simulation is superior to analytic risk models and DMAIC is a proven and recognized methodology for systematic process and quality improvements. OSDC provides for qualitative analysis offering qualitative improvements. This synergetic method eliminates the observed limitations of the conventional approach.

The method fully follows the DMAIC framework including the five phases: define, control, analyse, improve and control. It is compatible with CMMI® and can substantially help software projects to deliver the product on time and achieve the quality goals.

The method tactically uses the synergy of the three applied methodologies, i.e. Six Sigma DMAIC, Monte Carlo Simulation and OSDC, which provides for strong performance-driven software process improvements and achieves important benefits including savings, quality and customer satisfaction.

In comparison with the conventional methods, the stochastic approach is more reliable and comprehensive as the inherent variability and uncertainty are accounted for, allowing for probability analysis of the risk. Therefore, the confidence in the method's decision support is substantial, which is of mission-critical importance for software projects.

The simulation models used to demonstrate the method are simple for practical reasons in order to facilitate the elaboration. The models could be easily enhanced to provide for more complex analysis of the ongoing software projects.

ACKNOWLEDGMENTS

I acknowledge *Lyu* (Lyu 1996) for published data used in this work. Also, I would like to thank my daughter, Ivana Bubevska, for reviewing the paper and suggesting relevant improvements.

REFERENCES

Tayntor, C.B., 2002. *Six Sigma Software Development.* Auerbach: Boca Raton, Florida, US.

Mandl, R., 1985. Orthogonal Latin Squares: An Application of Experiment Design to Compiler Testing. *Communications of the ACM,* Vol. 128, No. 10, pp. 1054-1058.

Tatsumi, K., 1987. Test Case Design Support System. *Proceedings of ICQC,* Tokyo.

Brownlie, R., Prowse, J., and Phadke, M.S., 1992. Robust Testing of AT&T PMX/StarMAIL Using OATS. *AT&T Technical Journal,* Vol. 71. No. 3, pp. 41- 47.

Bernstein, L., and Yuhas, C. M., 1993. Testing Network Management Software. *Journal of Network and System Management*, Vol. 1, No. 1.

Siviy, J.M., Penn, L.M., and Stoddard, R.W., 2007. *CMMI® and Six Sigma: Partners in Process Improvement (SEI Series in Software Engineering).* Addison-Wesley Professional: Boston, Massachusetts, US.

Bratley, P., Fox, B.L., and Schrage, L.E., 1983. *A Guide to Simulation.* Springer-Verlag: New York.

Rubinstein, R.Y., and Kroese, D.P., 2008. *Simulation and the Monte Carlo Method.* John Wiley & Sons: New Jersey.

Lyu, M.R., 1996. *Handbook of Software Reliability Engineering.* IEEE Computer Society Press: Los Alamitos, CA, US.

Kan, S.H., 2002. *Metrics and Models in Software Quality Engineering.* Addison-Wesley Professional: Los Alamitos, CA, US.

Von Mayrhauser, A., et al., 1993. *On the need for simulation for better characterization of software reliability.* Proceedings of Fourth International Symposium on Software Reliability Engineering, pp. 264-273. Denver, Colorado, US.

Gokhale, S.S., Lyu, M.R., and Trivedi, K.S., 1997. Reliability Simulation of Fault-Tolerant Software and Systems. *Proceedings of Pacific Rim International Symposium on Fault-Tolerant Systems*, Taipei, Taiwan.

Gokhale, S.S., Lyu, M.R., and Trivedi, K.S., 1998. Reliability Simulation of Component-Based Software Systems. *Proceedings of Ninth International Symposium on Software Reliability Engineering*, Paderborn, Germany.

Tausworthe, R.C., and Lyu, M.R., 1996. Software Reliability Simulation. In: Lyu, M.R., ed. *Handbook of Software Reliability Engineering,* Chapter 16. IEEE Computer Society Press: Los Alamitos, CA, US.

Bubevski, V., 2009. A Simulation Approach to Six Sigma in Software Development. *Proceedings of the 2009 Summer Computer Simulation Conference*. pp. 125-132. Istanbul, Turkey.

Bubevski, V., 2010. An Application of Six Sigma and Simulation in Software Testing Risk Assessment. *Proceedings of the 2010 Third International Conference on Software Testing, Verification and Validation*. pp. 295-302. Paris, France.

Ferrin, D.M., Miller, M.J., and Muthler, D., 2002. Six Sigma and simulation, so what's the correlation?. *Proceedings of the 2002 Winter Simulation Conference*, December 2002, San Diego, California, US.

Lakey, P.B., 2002. Software Reliability Prediction is not a Science… Yet. Cognitive Concepts, St. Louis, US.

Brooks, F.P. Jr., 1995. The Mythical Man-Month (Essays on Software Engineering, Anniversary Edition). Addison-Wesley: Boston, Massachusetts, US.

Nanda, V., and Robinson, J.A., 2011. *Six Sigma Software Quality Imrovment.* McGraw-Hill Professional: New York City , NY, US.

Xie, M., 1991. *Software Reliability Modelling*, World Scientific: Singapore.

Conover, W.J., and Iman, R.L., 1981. Rank Transformations as a Bridge Parametric and Nonparametric Statistics. *The American Statistician*, Vol. 35. No. 3, August 1981, pp. 124.

Gokhale, S.S., and Lyu, M.R, 2005. A simulation approach to structure-based software reliability analysis. *Software Engineering, IEEE Transactions on*, Vol. 31, Issue 8, August 2005, pp. 643-656.

Murugappan, M., and Keeni, G., 2003. Blending CMM and Six Sigma to meet business goals. *Software, IEEE*, Vol. 20, Issue 2, Mar/Apr 2003, pp. 42 – 48.

Xiaosong, Z., Zhen, H., ZhangMin, Y.W., and Dainuan, Y., 2008. Process integration of six sigma and CMMI. *Proceedings of 6th International Conference on Industrial Informatics (INDIN*), pp. 1650-1653. 2008, Daejeon, Korea.

Galinac, T, and Car, Z., 2007. Software verification improvement proposal using Six Sigma. *LNCS*, Vol. 4589, pp. 51-64.

Macke, D., and Galinac, T., 2008. Optimized software process for fault handling in global software development. *LNCS*, Vol. 5007, pp. 395-406.

Redzic, C., and Baik, J., 2006. Six Sigma approach in software quality improvement. *Proceedings of 4th International Conference on Software Engineering Research, Management and Applications (SERA),* pp. 396-406. 2006, Seattle, Washington, US.

Xiaosong, Z., Zhen, H., Fangfang, G., and Shenqing, Z., 2008. Research on the application of six sigma in software process improvement. *Proceedings of 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP),* pp. 937-940. 2008, Harbin, China.

Hunny, U., 2012. *Orthogonal Security Defect Classification for Secure Software Development*. Thesis (PhD) Queen's University, Kingston, Ontario, Canada.

Bubevski, V., 2013. A Novel Approach to Software Quality Risk Management", *Software Testing, Verification & Reliability – STVR.* In Early View.

AUTHORS BIOGRAPHY

Vojo Bubevski comes from Berovo, Macedonia. He graduated from the University of Zagreb, Croatia in 1977, with a degree in Electrical Engineering - Computer Science. He started his professional career in 1978 as an Analyst Programmer in Alkaloid Pharmaceuticals, Skopje, Macedonia. At Alkaloid, he worked on applying Operations Research methods to solve commercial and pharmaceutical technology problems from 1982 to 1986.

In 1987 Vojo immigrated to Australia. He worked for IBM™ Australia from 1988 to 1997. For the first five years he worked in IBM™ Australia Programming Center developing systems software. The rest of his IBM™ career was spent working in IBM™ Core Banking Solution Centre.

In 1997, he immigrated to the United Kingdom where his IT consulting career started. As an IT consultant, Vojo has worked for Lloyds TSB Bank in London, Svenska Handelsbanken in Stockholm, and Legal & General Insurance in London. In June 2008, he joined TATA Consultancy Services Ltd.

Vojo has a very strong background in Mathematics, Operations Research, Modeling and Simulation, Risk & Decision Analysis, Six Sigma and Software Engineering, and a proven track record of delivered solutions applying these methodologies in practice. He is also a specialist in Business Systems Analysis & Design (Banking & Insurance) and has delivered major business solutions across several organizations. He has received several formal awards and published a number of written works, including a couple of textbooks. Vojo has also been featured as a guest speaker at several prominent conferences internationally.

PRODUCTION CAPACITY AND INVESTMENT POLICY FOR A MANUFACTURING COMPANY USING SIMULATION AND INTEGER PROGRAMMING

Jorge A. García-Hernández

Engineering Faculty, UNAM (Mexico)

artaban@comunidad.unam.mx

ABSTRACT

In this paper we constructed a simulation model to determine production capacity in a manufacturing company. The main objective was to define a range of expected production for a particular footwear model. This information was necessary to ensure that the factory was able to meet the new costumers demand prior to set an agreement with bigger orders. The results confirmed that the factory did have sufficient capacity for these new orders; nonetheless the time dedicated to produce other orders was quite narrow. We also detected the need for increasing production capacity; therefore an integer program was constructed in order to explore two goals: 1) Maximize production given a fixed investment budget, and 2) Minimize the investment cost to obtain a certain production. The results of the integer program were tested in the simulation model to obtain new production capacity.

Keywords: production capacity, simulation, integer programming, investment policy

1. INTRODUCTION

Simulation has been used as a tool to explore complex systems due to its capacity to incorporate elements with stochastic behavior and logical interactions between them.

Production systems are exposed to many and different inputs, each of them have an impact in the overall outcome; therefore simulation technique may be useful to generate information that allows us to describe and predict the behavior of the production system, and moreover, generate insights for decision-making.

The company studied belongs to the leather and footwear sector. It had had a wide spectrum of product mix; nevertheless due to commercial reasons, the company decided to focus on footwear products.

Before addressing costumers with bigger orders, it decided to determine its production capacity if the whole factory were dedicated to shoe-manufacturing. Therefore, the research question was: How many shoes can be produced by the factory given current conditions?

Simulation was selected because of the reasons given above and because of the flexibility for adding relevant elements and for integrating or disintegrating objects attributes. In Figure 1 we show a picture of the footwear product.

Figure 1. Picture of One Product

2. LITERATURE REVIEW

There are analytical formulations for production process that optimize certain attributes of the system (e.g. maximize production, minimize total cost), these formulations give an optimal or almost optimal solution (Pinedo 2005); nonetheless they are very time-consuming and impractical for real world problems, therefore different approaches have been tried such as network modeling, simulation, and hybrid approaches.

Network formulations can easily grow fast with the addition of elements and become impractical (Argoneto 2008).

Carvalho et al. (2012) used simulation to analyze different scenarios for a three-echelon supply chain and improve the overall system. Heilala (2008) constructed a simulation model to design a sustainable manufacturing system while optimizing different subsystems.

Lee et al. (2002a) tried an analytical-simulation approach; the simulation model is used to deal with operation times. Lee et al. (2002b) also combined discrete and continuous simulation models to represent the supply chain.

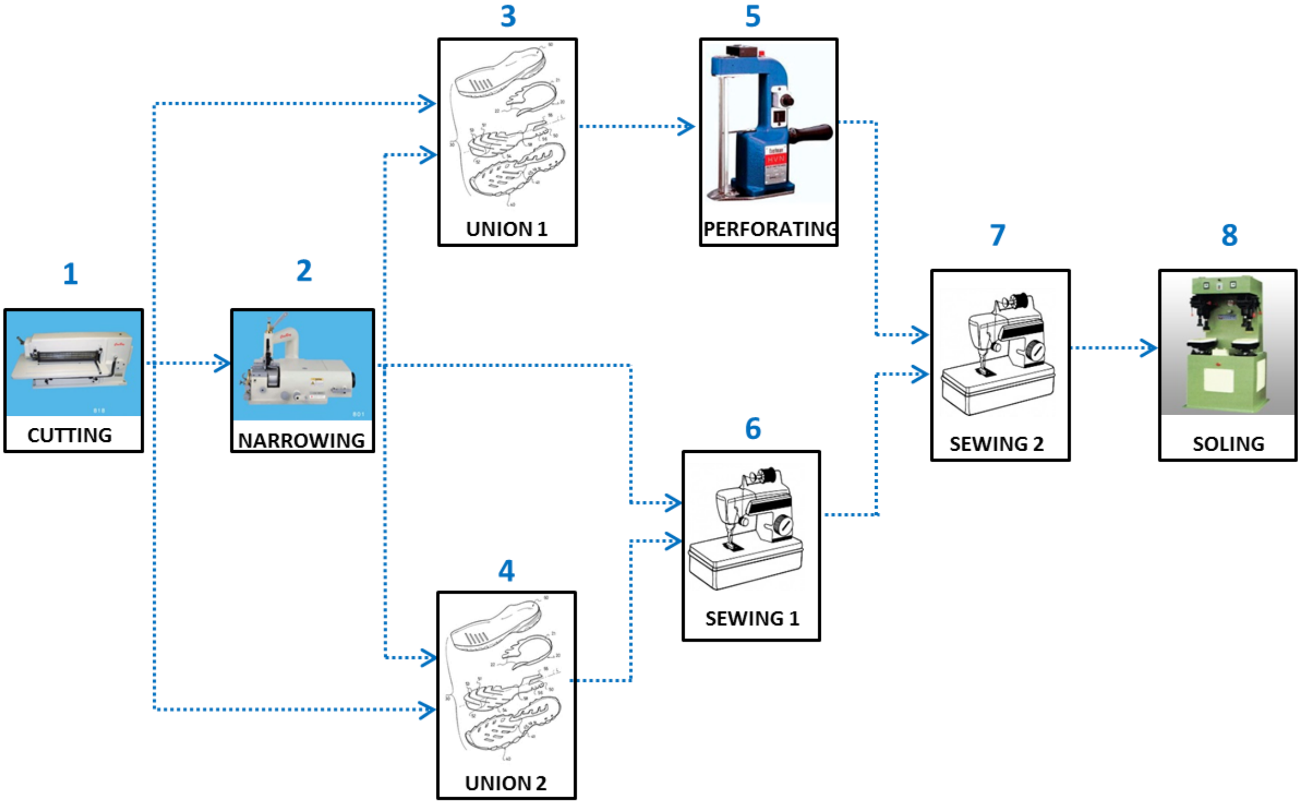


Figure 2. Shoe-Manufacturing Process

3. MANUFACTURING PROCESS

The shoe-manufacturing process is shown in Figure 2.

3.1 Cutting (Op. 1)

Raw material consists of three different kinds of leather: lamb, pork, and veal. All of them pass through the cutter machine where standard molds produce pieces of different sizes and colors.

3.2 Narrowing (Op. 2)

Pieces from lamb leather must have the same thickness; therefore they are process by the narrower machine.

3.3 Union 1 (Op. 3)

Here the insole pieces from different leathers are attached into one.

3.4 Union 2 (Op. 4)

In this operation an attachment is performed between pieces that will be located in the upper part of the shoe.

3.5 Perforating (Op. 5)

The pieces from operation 3 are perforated along the edge to guide the next sewing operation. This is made by a hammer machine.

3.6 Sewing 1 (Op. 6)

A first sewing is performed by a sewing machine to keep the pieces all together.

3.7 Sewing 2 (Op. 7)

The second sewing is performed to give the shoe a hand-made artistic appearance.

3.8 Shoe soling (Op. 8)

Finally plastic or leather soling sheets are attached to the piece from operation 7. This machine processes groups of exactly three shoes; once it finishes with one group then receives another. So the production is always a multiple of 3.

3.9 Times

Times associated with operations are shown in Table 1; they depend on the size of the shoe and the skills of the worker. The table shows the minimum, mode and maximum of the data.

The first operation has different times depending on the raw material type. Pieces *b*, *c*, *d* are from the lamb leather, and *a*, *d* are from pork and veal respectively.

Table 1. Operation Times

Operation	Name	Time (minutes)		
		Min	Mode	Max
1.*a*	Cutting *a*	2	2.5	4
1.*b*	Cutting *b*	2	2.5	4
1.*c*	Cutting *c*	1	1	1
1.*d*	Cutting *d*	2	3	4
1.*e*	Cutting *e*	2.5	3	4.5
2	Narrowing	1	1.5	2
3	Union 1	4	5.5	7
4	Union 2	2.5	3.5	6
5	Perforating	4	5	6
6	Sewing 1	3	3.5	4
7	Sewing 2	6	7	9
8	Shoe soling	20	20	20

4. SIMULATION MODEL

4.1 Data collecting

Operation and transport times were taken. Triangular distributions were selected to model operation times; Kolmogorov-Smirnov and Anderson-Darling tests were applied to ensure its vality. We consider transport times as negligible.

4.2 Assumptions

1. Shoe manufacturing was the only type of production consider. Other products were discarded.
2. Different sizes and colors were modeled by the probability distribution functions of each machine.
3. Transport times between operations were negligible.
4. There were always sufficient raw materials for production.
5. Machines are always operational.

4.3 Software

Simio Simulation Software was selected to carry out the simulation due to its robustness and the flexibility to represent industrial environment.

4.4 Model

Operations were represented with objects from Simio library, and data tables and add-in processes were used.

Simulation runs started with no semi-finished product. We show a schematic procedure of the simulation in Figure 3, operations are shown with rectangles.

4.5 Results

A total of 2500 replications were made, the results of the model allowed us to calculate the Expected Weekly Production and a Confidence Interval of size 96%. This is shown in table 2.

Table 2. Simulation Weekly Production

Mean	Standard deviation	Min	Max	Confidence interval 96%
124	15.75	102	147	[111, 135]

The distribution of the production is shown in figure 4. The results are multiple of 3 because of the last machine constraint commented above.

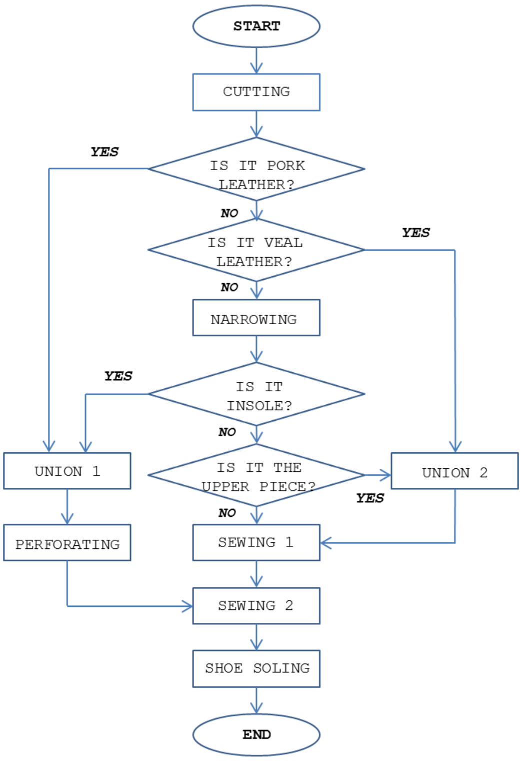


Figure 3. General Simulation Diagram

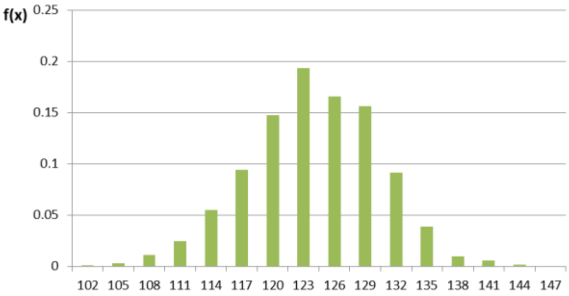


Figure 4. Weekly Production Distribution

To ensure the validity of the model and the results, we took the daily production of one month and tested a statistical hypothesis to determine whether population mean was different from the one obtained from simulation. We concluded that the data analyzed did not provide any information for supporting the statement that they were different.

5. INTEGER PROGRAM

We constructed an integer program to study two goals: 1) Maximize production given a fixed investment budget, and 2) Minimize the investment cost to obtain a certain production. Moreover, the integer program also balances the production line, so a measure of efficiency can be decided.

We only used one model to explore both goals, by setting the objective function in the first as a constraint in the second and conversely.

5.1 Data Collecting

The output of the simulation model was used to determine production parameters. Machine prices, investment budget and minimal expected production were provided by the company.

5.2 Assumptions

1. The investment policy considered machines and tools; the cost of hiring and training workers was excluded.
2. Machines had the same capacity than the current ones in each operation.

5.3 Software

Lingo 13 was selected because it provides easiness to introduce short instructions and the capacity to link with *.txt*, *.xls*, *.dll*, and other data files.

5.4 Model

The integer program is the following:

$$Max\, z = x_N \tag{1}$$

$$\sum_{i=1}^{N} c_i w_i \leq b \tag{2}$$

$$x_i \leq a_i(1 + w_i) \quad \forall i \tag{3}$$

$$(1 - d) f_i x_N \leq x_i \quad \forall i \neq N \tag{4}$$

$$x_i \leq (1 + d) f_i x_N \quad \forall i \neq N \tag{5}$$

$$x_i \geq 0 \quad \forall i \tag{6}$$

$$w_i \in \mathbb{Z}^+ \quad \forall i \tag{7}$$

x_N represents the rate of production of the last operation (Op. 8). c_i is the cost to buy and to install a machine in operation i. b is the total investment money. a_i is the rate of production of the machine i, this was obtained from the simulation model. d is the allowed proportion deviation (% of efficiency), f_i is a coefficient that represents the number of units from operation i needed to produce one final product. x_i and w_i are decisional variables.

In order to explore the second goal we changed equation (1) to (1.a) and equation (2) to (2.a) as follows:

$$Min\, z = \sum_{i=1}^{N} c_i\, w_i \tag{1.a}$$

$$p \leq x_N \tag{2.a}$$

The rest of the equations remained the same. p is the minimal desired production.

5.5 Results

First we will review the results associated with the first model (maximize production) and then those associated with the second model (minimize investment cost).

In the first model we tried with several values of d, finally we decided along with the manager, to set d=*0.10*, which implies 90% of efficiency.

Table 3. Optimal Solution of the First Model with d=*0.10*

Op.	RP (current)	New mach.	RP (new)	Balanced RP
1	50.69	1	101.38	88.20
2	28.23	2	84.69	72.17
3	29.00	2	87	87.00
4	28.36	2	85.08	85.08
5	24.05	2	72.16	72.16
6	28.80	2	86.41	86.41
7	22.80	3	91.20	88.20
8	20.66	3	82.64	80.18

In table 3, *Op.* is the operation; *RP (current)* is the maximal rate of production in current conditions; *New mach.* is the number of new machines to buy according to the optimal solution; *RP (new)* the maximal rate of production considering new machines; *Balanced RP* is the optimal balanced rate of production considering new machines.

This solution consumes the entire investment budget and the production goes from 124 to 480 shoes weekly, this represent an increase of 287%.

For the second model (or goal), we decided to explore a set of minimal weekly productions p in order to associate production to different investment levels. Again 90% of efficiency was selected, i.e. d=*0.10*.

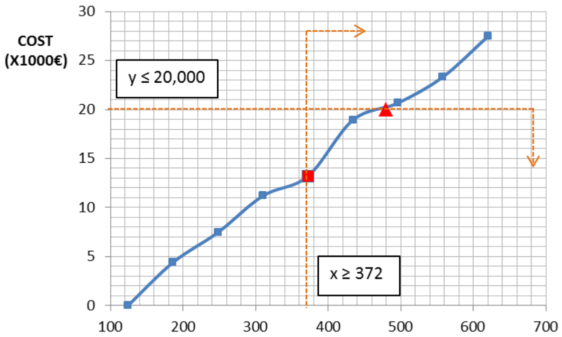


Figure 5. Weekly Production Associated with Investment Levels.

In figure 5 we can see that the investment cost function is approximately linear. The distance between blue dots is constant in x-axis. The production and financial constraints are shown with dotted lines.

The red triangle shows the optimal solution found for the first integer model, the red square shows the optimal for the second.

The solutions selected to be tested in the simulation model are the following:

Table 4. Optimal Solutions Selected

	Op.	Solution	
		A	B
Investment policy (new machines)	1	1	1
	2	2	1
	3	2	1
	4	2	1
	5	2	2
	6	2	1
	7	3	2
	8	3	2
Weekly production		480	372
Investment cost (€)		20,000	13,200

Solution A is the optimal obtained from the first integer program, solution B was selected because it provides the higher ratio (production/investment cost), it is above the weekly production objective set by the manager, and below the investment budget.

6. SIMULATION OF OPTIMAL SOLUTIONS

We made a simulation experiment of 2500 replicates for both solutions and obtained weekly productions about five percent higher than predicted from the integer program, this was because the simulation model considered several other but it was still a good prediction.

Table 5. Statistic Measures of the Optimal Solutions.

	Solution	
	A	B
Mean	501.7	390.4
Range	87	33
SD	12.5	5.7
CV	2%	1%
Average Efficiency	90.6%	93%

The range of results of B solution was narrower than A's and its standard deviation was smaller, but both solutions had a low coefficient of variation. Average efficiency was higher in B solution, which means that the rate of production implies a better use of resources.

Both solutions represent an intended full capacity of the system useful for providing insights for decision-making. With any of these levels of production the company is capable to meet the new customers demand.

Also, the production obtained can be used as an objective production considering that the assumptions of the model did not involve any production line stopping.

The decision between the optimal solutions has to consider the preference of the manager about installed capacity.

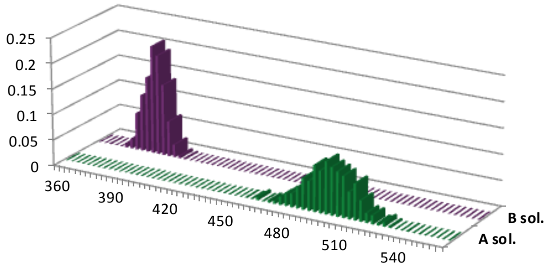


Figure 6. Weekly Production Distribution for A (green) and B (purple) solutions.

In Figure 7 we show how the mean production was approaching to its statistic regularity value as more replicates were made.

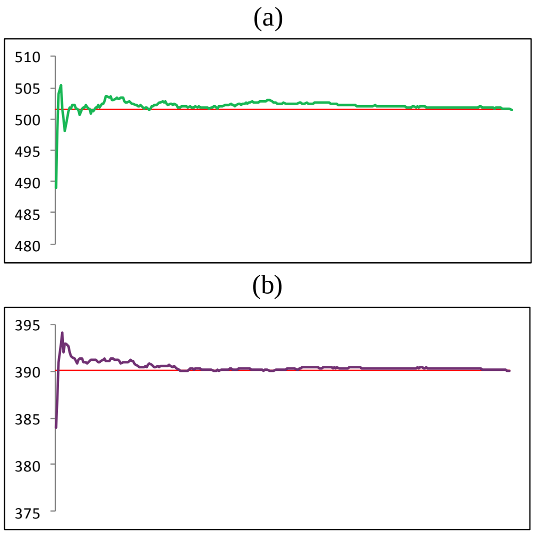


Figure 7. Mean´s Statistic Regularity for A *(a)*, and B *(b)* solutions.

7. CONCLUSIONS

With the conditions the company had, it would be completely dedicated to shoe-manufacturing if it wanted to meet new customers demand. It would not be capable to meet other orders.

We recommended increasing the capacity of the factory according to A solution only if the company is capable to sell such levels of production; otherwise it is advisable to increase the capacity according to B solution.

We used the integer program to find 'promissory solutions', then we used simulation for a deeper research in those solutions. We believe this is a good approach to face situations in which time and resources are scarce. Trying to explore through simulation every possible investment configuration would lead to a combinatory problem that would be very expensive and time-consuming.

Also we must say that the simulation model was a good fund-finding tool for the company. It generated tangible information about levels of production associated with investment budgets that the investors can rely on.

REFERENCES

Argoneto, P and Perrone, G, 2008. *Production planning in production networks. Models for medium and short-term planning.* London; Springer.

Carvalho H. et al., 2012. Supply chain redesign for resilience using simulation. *Computers & Industrial Engineering*, Vol. 62: 329-341.

Heilala, J. et al. 2008. Simulation-based sustainable manufacturing system design. *Proceedings of the 2008 Winter Simulation Conference*, 1922-1930. December, San Diego, CA.

Lee, Y. H. and Kim, S. H., 2002a. Production-distribution planning in supply chain considering capacity constraints. *Computers & Industrial Engineering*, Vol. 43: 169-190.

Lee, Y. H. et al., 2002b. Supply chain simulation with discrete-continuous combined modeling. *Computers & Industrial Engineering*, Vol. 43: 375-392.

Pinedo, M. L., 2005. *Planning and scheduling in manufacturing and services*. New York; Springer.

Papoulis, A. and Pillai, S. U., 2002. *Probability, Random Variables, and Stochastic Processes*. Fourth edition, New York; McGrawHill.

Ríos Insúa, David et. al., 2009. *Simulación. Métodos y Aplicaciones*. Second edition, México; AlfaOmega.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Idalia Flores for her support and Pablo I. Alcántara for the facilities provided by the company.

AUTHOR´S BIOGRAFY

Jorge Andrés García studied Industrial Administration at the National Polytechnic Institute (IPN) and then a Master in Operations Research at the National Autonomous University of Mexico (UNAM). His research lines are Optimization in Production Planning and Simulation.

SIMULATION OPTIMIZER AND OPTIMIZATION METHODS TESTING ON DISCRETE EVENT SIMULATIONS MODELS AND TESTING FUNCTIONS

Pavel Raska[a], Zdenek Ulrych[b], Petr Horejsi[c]

[a] Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia,
Univerzitni 22, 306 14 Pilsen
[b] Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia,
Univerzitni 22, 306 14 Pilsen
[c] Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia,
Univerzitni 22, 306 14 Pilsen

[a]praska@kpv.zcu.cz, [b] ulrychz@kpv.zcu.cz, [c] tucnak@kpv.zcu.cz

ABSTRACT

The paper deals with testing of selected optimization methods used for optimization of specified objective functions of three discrete event simulation models and four selected testing functions. The developed simulation optimizer uses modified optimization methods which automatically adapt input parameters of discrete event simulation models. Random Search, Hill Climbing, Tabu Search, Local Search, Downhill Simplex, Simulated Annealing, Differential Evolution and Evolution Strategy were modified in such a way that they are applicable for discrete event simulation optimization purposes. The other part of the application is focused on testing the implemented optimization methods. We have proposed some evaluation techniques which express the success of the optimization method in different ways. These techniques use calculated box plot characteristics from the series of optimization experiments.

Keywords: simulation optimization, heuristic optimization methods, discrete event simulation models, testing function

1. INTRODUCTION

Many of today´s industrial companies try to design their own production system as effectively as possible. The problem is that this intention is affected by many factors. We can say the problem is NP-a hard problem in most cases. A possible answer to the problem is using discrete event simulation and simulation optimization. The use of discrete event simulation focuses on the invisible problems in the production system many times and also avoids bad decisions made by the human factor.

The next question is effectively finding a suitable solution to the modelled problem. We can use a simulation optimizer to find an optimal/suboptimal feasible solution respecting the defined model constraints. The basic problem of global optimization can be formulated as follows:

$$\breve{\mathbf{X}} = \arg\min_{\mathbf{X}\in\tilde{X}} F(\mathbf{X}) = \left\{\breve{\mathbf{X}} \in \tilde{X} : F\left(\breve{\mathbf{X}}\right) \leq F(\mathbf{X}) \forall \mathbf{X} \in \tilde{X}\right\} \quad (1)$$

where $\breve{\mathbf{X}}$ denotes the global minimum of the objective function; $F(\mathbf{X})$ denotes the objective function value of the candidate solution – the range includes real numbers; $\tilde{X}$ denotes the Search space. This optimal solution is represented by the best configuration (input parameters values) of the simulation model.

Current simulation software (Arena, Witness, PlantSimulation etc.) uses its own simulation optimizers. These integrated optimization modules are black-boxes but many of them use similar optimization methods. We have tested the following optimization methods: Random Search, Hill Climbing, Tabu Search, Local Search, Downhill Simplex, Simulated Annealing, Differential Evolution and Evolution Strategy. These methods were modified in such a way that they are applicable for discrete event simulation optimization purposes. The goal of our research is to compare some of these widely used optimization methods. Hence we have designed our own simulation optimizer. We have to say that it is not possible to implement exactly the same optimization methods which are used in these simulation optimizers.

Another reason for testing the optimization methods and designing our own simulation optimizer was that our department focuses on modelling and optimizing production and non-production processes in industrial companies (Kopecek, 2012; Votava, Ulrych, Edl, Korecky and Trkovsky, 2008). Some projects have to be solved with difficulty without the use of special simulation optimization tools because of the large complexity of the discrete event simulation model. The problem is that some integrated simulation optimizers cannot affect all the parameter types of the designed simulation model.

2. SELECTED OPTIMIZATION METHODS

We have transformed some of the selected optimization methods to use the principle of evolutionary algorithms. These optimization methods generate a whole population (instead of one possible solution) in order to avoid getting stuck on a local optimum. Previous testing of optimization methods confirms that generating one solution leads to premature convergence in most cases (depending on objective function type). Different variants of selected optimization methods obtained from a literature review were united into the algorithm. The user can combine different variants of optimization methods by setting the optimization method parameters.

2.1. Random Search

A new candidate solution is generated in the search space with uniform distribution (Monte Carlo method). This method is suitable for cases where the user has no information about the objective function type. The user is able to perform a number of simulation experiments.

2.2. Downhill Simplex

This method uses a set of n + 1 linearly independent candidate solutions (n denotes search space dimension) - Simplex. The method uses four basic phases – Reflection, Expansion, Contraction and Reduction. (Tvrdík 2004; Weise 2009)

2.3. Stochastic Hill Climbing

Candidate solutions are generated (populated) in the neighbourhood of the best candidate solution from the previous population. Generating new possible solutions is performed by mutation. This method belongs to the family of local search methods.

2.4. Stochastic Tabu Search

The newly generated candidate solution is an element of the Tabu List during the optimization process. This candidate solution cannot be visited again if the aspiration criterion is not satisfied (this feature prevents the method from becoming stuck at a local optimum). The method uses the FIFO method of removing the candidate solution from the Tabu List. The user can set whether the new candidate solution is generated using mutation of the best candidate solution from the previous population or the new solution is generated using mutation of the best found candidate solution. (Monticelli, Romero and Asada 2008; Weise 2009)

2.5. Stochastic Simulated Annealing

A candidate solution is generated in the neighbourhood of the candidate solution from the previous iteration. This generating could be performed through the mutation of a randomly selected gene or through the mutation of all genes. Acceptance of the worse candidate solution depends on the temperature. Temperature is reduced if the random number is smaller than the acceptance probability or the temperature is reduced if and only if a worse candidate solution is generated. If the temperature falls below the specified minimum temperature, temperature is set to the initial temperature. (Monticelli, Romero and Asada 2008; Weise 2009)

2.6. Stochastic Local Search

A candidate solution is generated in the neighbourhood of the best candidate solution.

2.7. Evolution Strategy

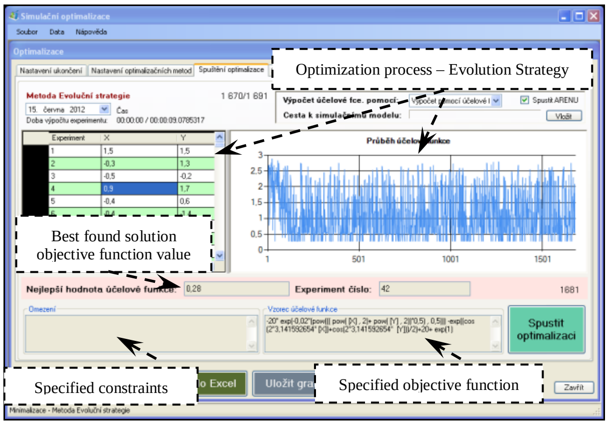
This optimization method uses Steady State Evolution – population consists of children and parents with good fitness. A candidate solution (child) is generated in the neighbourhood of the candidate solution (parent) and it is based on the Rechenberg 1/5th-rule. The population is sorted according to the objective values (Rank-Based Fitness Assignment). The optimization method uses Tournament selection. (Koblasa, Manlig and Vavruska 2013; Miranda 2008; Tvrdík 2004)

2.8. Differential Evolution

Selection is carried out between the parent and its offspring. The offspring is created through a crossover between the parent and the new candidate solution (individual) which was created through the mutation of four selected individuals and the best one selected from the population – BEST method. The optimization method uses General Evolution and the Ali and Törn adaptive rule. The user can define the probability of a crossover between the new candidate solution and the parent. (Tvrdík 2004; Wong, Dong, 2008)

3. DEVELOPED APPLICATION

We have developed our own simulation optimization application which addresses the problems listed in the first chapter. The application contains seven different global optimization methods. This application contains two modules. The first module is a simulation optimizer which enables optimization of developed simulation models in ARENA or PlantSimulation simulation software. The objective function of the models is specified within the discrete simulation models. The user can also test a specified objective function without the need of creating the simulation model – Figure 1.



Metoda Evoluční strategie 1 670/1 691

15. června 2012 Čas

Doba výpočtu experimentu: 00:00:00 / 00:00:09.0785317

Výpočet účelové fce. pomocí: Výpočet pomocí účelové f

☑ Spustit ARENU

Cesta k simulačnímu modelu:

Experiment	X	Y
1	1,5	1,5
2	-0,3	1,3
3	-0,5	-0,2
4	0,9	1,7
5	-0,4	0,6
6	-0,4	1,4

Nejlepší hodnota účelové funkce: 0,28 Experiment číslo: 42 1681

Omezení

Vzorec účelové funkce: -20* exp(-0,02*(pow(((pow([X] , 2)+ pow([Y] , 2))*0,5) , 0,5))) -exp((cos (2*3,141592654* [X])+cos(2*3,141592654* [Y]))/2)+20+ exp(1)

Minimalizace - Metoda Evoluční strategie

Figure 1: Graphical user interface of simulation optimizer - first module

The application was created in Visual Basic 2010. This programming language was used for connection to ARENA simulation software and Microsoft Access database. The data from simulation experiments results and settings are stored in this database. The file contains information about:

1. Controls – identification, names, low and high boundaries, type (discrete vs. continuous), initial values of controls and comments.
2. Constraints – specification of the constraint function through using the mathematical operator buttons and the list of controls. User can validate the built expression.
3. Objective function - specification of the objective function without the need of a simulation software tool. Objective function is composed of mathematical operators and selected controls from the list of all controls. User can validate the built expression.
4. Optimization experiment setting – minimization vs. maximization of objective function, Termination criterion (Value to reach, number of simulation experiments, specified time, sub-optimum improvement ratio etc.), parameters settings of selected optimization method, low and high boundaries of selected optimization method parameters, number of replications, creation of a knowledge base of a simulation model, etc.

The second module is designed for testing the behaviour of the implemented optimization method in terms of setting the parameters for the optimization method. The user can specify the range of optimization method parameters. After finishing the number of optimization experiments replications (series of concrete optimization method setting) the data are exported to MS Excel workbook.

We have also developed an application which enables 3D visualization of simulation experiments when there are two controls and one objective function. Simulation experiments are represented by the points in the 3D chart of the objective function. The objective function surface is generated from the data obtained from simulation experiments. Another possibility is to generate a whole 3D chart from the data obtained from the simulation runs of all possible settings of the simulation model input parameters – complete search space.

4. DISCRETE EVENT SIMULATION MODELS AND OBJECTIVE FUNCTIONS

The testing of optimization methods which search for global optima was applied to three discrete event simulation models. These models reflect real production systems of industrial companies. Discrete event simulation models were built in Arena simulation software. We specified different objective functions considering the simulated system. All possible solutions and their objective function values were mapped to find the global optimum in the search space.

4.1. The Manufacturing System and Logistics

This discrete event simulation model represents the production of different types of car lights in a whole production system. The complex simulation model describes many processes; for example, logistics in three warehouses, production lines, 28 assembly lines, painting, etc. The objective function is affected by the sum of the average utilization of all assembly lines and average transport utilization. The objective function is maximized. Controls are the number of forklifts responsible for: transport of small parts from the warehouse to the production lines and assembly lines, transport of large parts from the warehouse to the assembly lines and the transport of the final product from the assembly lines to the warehouse. The objective function landscape of this model when the number of forklifts for transport of large parts = 14 is shown in Figure 2.

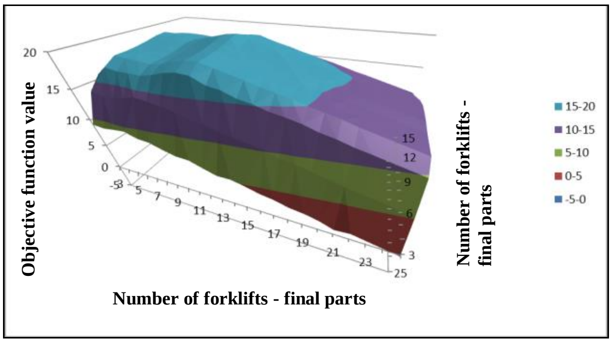


Figure 2: Objective Function - The Manufacturing System And Logistics Discrete Event Simulation Model - Number of Forklifts for Large Parts = 14

4.2. The Penalty

This simulation model represents a production line which consists of eight workstations. Each workstation contains a different number of machines. Each product has a specific sequence of manufacturing processes and machining times. The product is penalized if the product exceeds the specified production time. A penalty also occurs if the production time value is smaller than the specified constant. The penalty function is shown in Figure 3 where T denotes production time; $T_{\min}$ denotes required minimum production time; $T_{\max}$ denotes required maximum production time; T_{crit} denotes critical production time; k_{α} denotes the penalty for early production (slope of the line - constant); k_{α_1} denotes the penalty for exceeding the specified production time (slope of the line - constant); P_1 denotes the penalty for exceeding the specified production time (constant); P_2 denotes the penalty for exceeding the specified critical production time (constant); P denotes the penalty of the product.

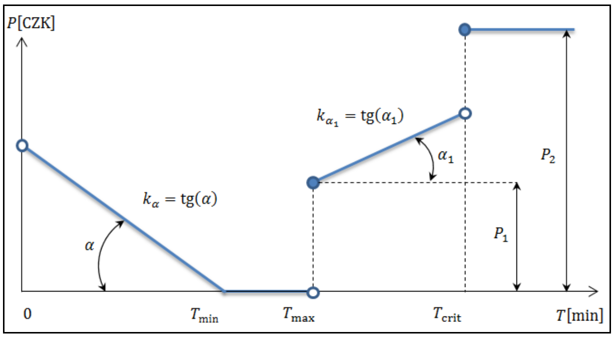


Figure 3: Penalty Function

This rule is defined because premature production leads to increasing storage costs – the JIT product. The objective function is affected by the total time spent by the product in the manufacturing system. The objective function is minimized. Controls of the production line simulation model are the arrival times of each product in the system. The objective function is shown in Figure 4.

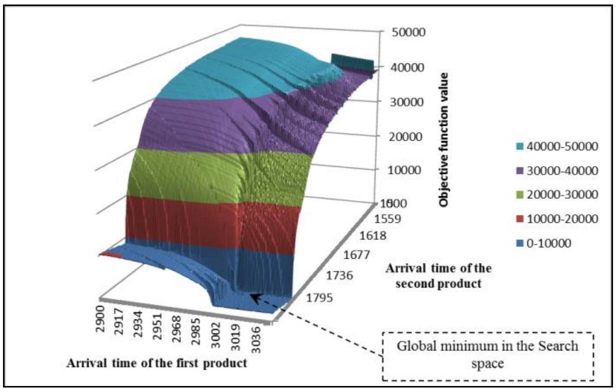


Figure 4: Objective Function – The Penalty Discrete Event Simulation Model

4.3. The Assembly Line

This model represents an assembly line. Products are conveyed by conveyor belt. The assembly line consists of eleven assembly workplaces. Six of these workplaces have their own machine operator. The rest of the workplaces are automated. A specific scrap rate is defined for each workplace. At the end of the production line is a sorting process for defective products. The objective function reflects the penalty which is affected by the number of defective products and the palettes in the system. The objective function is maximized. The objective function is shown in Figure 5. The input simulation model parameters (controls) are the numbers of fixtures in the system and the number of fixtures when the operator has to move from the first workplace to the eleventh workplace to assemble waiting parts on the conveyor belt.

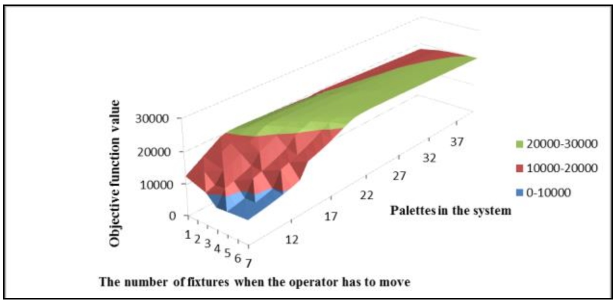


Figure 5: Objective function - The Assembly Line discrete event simulation model

5. TESTING FUNCTIONS

We also tested implemented optimization methods on four standard testing functions. All testing functions are minimized.

5.1. De Jong´s Function

It is a continuous, convex and unimodal testing function. The function definition:

$$F(\mathbf{X}) = \sum_{j=1}^{n} x_j^2 \qquad (2)$$

where $F(\mathbf{X})$ denotes the objective function; *j* denotes index of control; *n* denotes the dimension of the search space; x_j denotes the value of control. The objective function is shown in Figure 6.

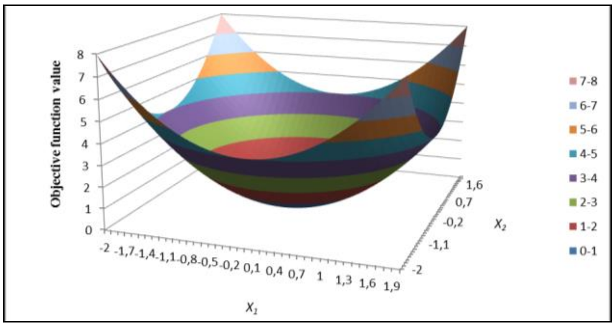


Figure 6: Objective Function - De Jong´s Function

5.2. Rosenbrock´s Function

Rosenbrock´s (Rosenbrock's valley, Rosenbrock's banana) function is a continuous, unimodal and non-convex testing function. The function definition:

$$F(\mathbf{X}) = \sum_{j=1}^{n-1} 100 \cdot (x_j^2 - x_{j+1})^2 + (1 - x_j)^2 \qquad (3)$$

The objective function is shown in Figure 7.

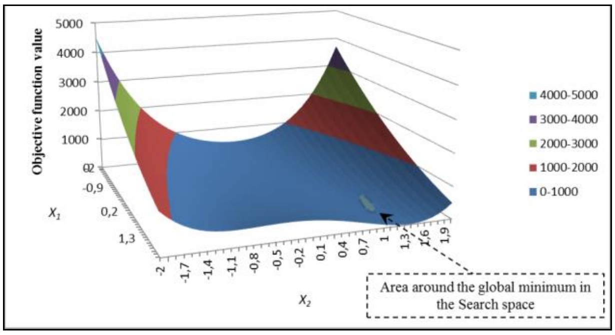


Figure 7: Objective Function - Rosenbrock´s Function

5.3. Michalewicz´s Function

Michalewicz´s function is a multimodal test function (n! local optima). The parameter *m* defines the "steepness" of the valleys or edges. Larger *m* leads to a more difficult search. For very large *m* the function behaves like a needle in a haystack (the function values for points in the space outside the narrow peaks give very little information on the location of the global optimum). (Pohlheim 2006)

$$F(\mathbf{X}) = -\sum_{j=1}^{n} \sin(x_j) \cdot \left(\sin\left(\frac{j \cdot x_j^2}{\pi} \right) \right)^{2 \cdot m} \tag{4}$$

$$j = 1 : n, 0 \le x_j \le \pi \tag{5}$$

We selected $m = 5$ in our simulation model. The objective function is shown in Figure 8.

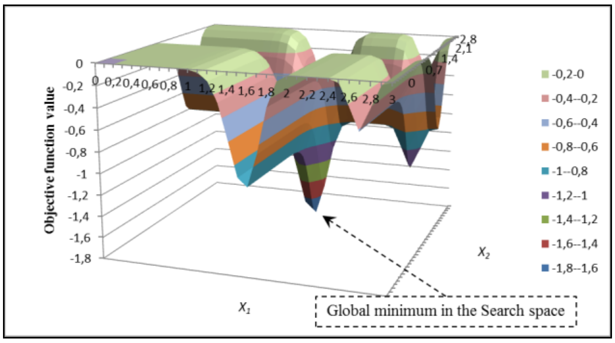


Figure 8: Objective Function - Michalewicz´s Function

5.4. Ackley´s Functions

Ackley´s function is a multimodal test function. This function is a widely used testing function for premature convergence. (Tvrdík 2004)

$$F(\mathbf{X}) = -20 \cdot \exp\left(-0.02 \cdot \sqrt{\frac{1}{n} \cdot \sum_{j=1}^{n} x_j^2} \right) - \exp\left(\frac{1}{n} \cdot \sum_{j=1}^{n} \cos 2 \cdot \pi \cdot x_j \right) + 20 + \exp(1) \tag{6}$$

$$j = 1 : n, -30 \le x_j \le 30 \tag{7}$$

The objective function is shown in Figure 9.

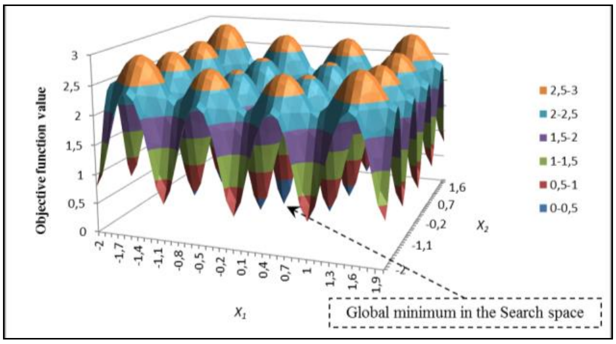


Figure 9: Objective Function - Ackley´s Function

6. EVALUATION METHOD

Simulation experiments results are saved to a database file during simulation experiments if the user uses a simulation optimizer. Simulation experiments results are visualized in the objective function chart and stored in the table placed in the application. The graphical user interface of the first module is shown in Figure 1.

If the second module is used the simulation experiments data are exported to MS Excel workbook after finishing the series (series - replications of optimization experiments with concrete optimization method setting). Excel was selected because of its wide usage

Considering the number of simulation experiments we can divide the number of simulation experiments – Figure 10:

1. Simulation experiment – simulation run of simulation model.
2. Optimization experiment – performed with concrete optimization method setting to find optimum of objective function.
3. Series – replication of optimization experiments with concrete optimization method setting.

The second module focuses on testing the behaviour of the implemented optimization method in terms of setting the parameters for the optimization method. The user can set up the parameters of a selected optimization method, low and high boundaries of the selected optimization method parameters, number of replications, and export the objective function chart to image – Figure 11.

The same conditions had to be satisfied for each optimization method, e.g. the same termination criteria, the same search space. If the optimization method has the same parameters as another optimization method, we set up both parameters with the same boundaries (same step, low and high boundaries).

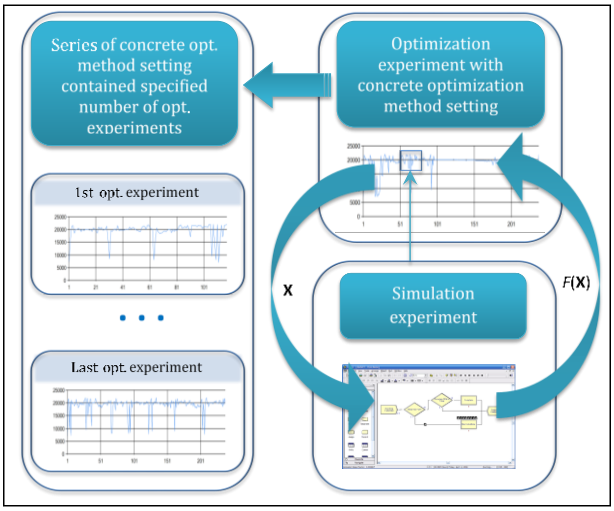
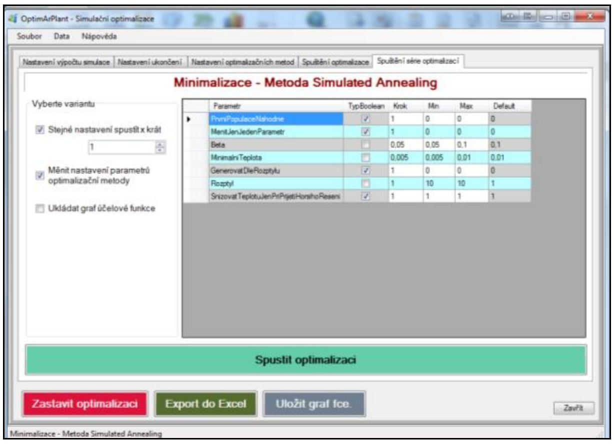

Series of concrete opt. method setting contained specified number of opt. experiments
Optimization experiment with concrete optimization method setting
1st opt. experiment
Last opt. experiment
Simulation experiment
X
F(X)

Figure 10: The Number of Simulation Experiments



Minimalizace - Metoda Simulated Annealing

Vyberte variantu

☑ Stejné nastavení spustit x krát

1

☑ Měnit nastavení parametrů optimalizační metody

☐ Ukládat graf účelové funkce

Parametr	TypBoolean	Krok	Min	Max	Default
PrvniPopulaceNahodne	☑	1	0	0	0
MenitJenJedenParametr	☑	1	0	0	0
Beta	☐	0,05	0,05	0,1	0,1
MinimalniTeplota	☐	0,005	0,005	0,01	0,01
GenerovatDleRozptylu	☑	1	0	0	0
Rozptyl	☐	1	10	10	1
SnizovatTeplotuJenPriPrijetiHorsihoReseni	☑	1	1	1	1

Minimalizace - Metoda Simulated Annealing

Figure 11: GUI of the Second Module

Box plot characteristics (the smallest observation – sample minimum Q_1, lower quartile Q_2, median Q_3, upper quartile Q_4, and largest observation - sample maximum Q_5) are calculated for each performed setting of the optimization method parameters – Figure 12.

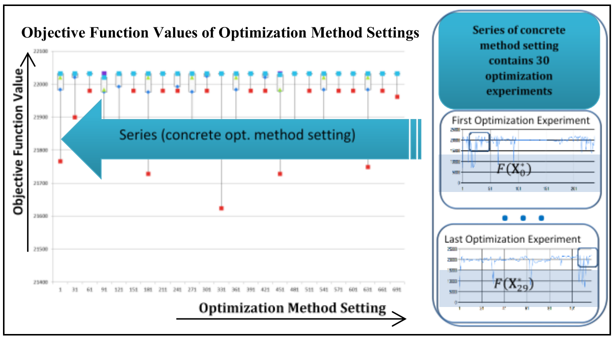


Figure 12: Example of Results from Simulation Optimization Experiments Provided by Evolution Strategy Displayed in Box Plot Chart – The Assembly Line Simulation Model

$F(\mathbf{X}_0^*)$ denotes the found optimum (local in this case). These characteristics are visualized in the box plot chart – Figure 12.

Three box plot charts are generated - Best objective function value, Range of provided function objective values during the simulation experiments, and Number of experiments required to find global (local) optimum. Visualization can help the user to find a suitable setting of optimization method more quickly.

Due to the large volume of data (over 4 billion simulation experiments) we have to propose evaluation techniques (criteria) which express the failure of the optimization method in different ways. Each criterion value is between [0, 1]. If the failure is 100[%] the criterion equals 1 therefore we try to minimize all specified criteria. We implemented the graphical user interface to MS Excel workbook which enables the user to set up the weights of each criterion and other parameters of the evaluation. These parameters are automatically loaded from the simulation experiments results. We used the VBA for MS Excel.

6.1. Optimization Method Success

The first criterion f_1 is the value of not finding the known VTR (value to reach). This value is expressed by:

$$f_1 = \frac{s - n_{succ}}{s} \tag{8}$$

where s denotes the number of performed series, n_{succ} denotes the series where the VTR was found. Simulation runs of all possible settings of simulation model input parameters were performed. This means that we have evaluated all possible solutions of the search space hence we can determine the global optimum (VTR) in the search space. Average Method Success of Finding Optimum can be formulated as follows:

$$f_{avg} = \left(1 - \frac{\sum_{i=1}^{s} f_{1_i}}{s}\right) \cdot 100[\%] \tag{9}$$

where i denotes the index of one series, f_{1_i} denotes the value of the first criterion (Optimization method success – the best value is zero), s denotes the number of performed series. The average optimization method success of finding the optimum of testing functions is shown in Figure 13.

We can see that the Evolution Strategy and Simulated Annealing are successful optimization methods. Random Search also achieves good results. It was affected by doing many simulation experiments by this method. The probability of finding the optimum increases with a high number of simulation experiments. This strategy is simply random and if the search space is huge (NP-hard) we can say it is lucky to find the optimum. This method is usable when the user has no information about the objective function type. We have to evaluate each possible solution in the search space to obtain the optimum hence the search space cannot be too huge.

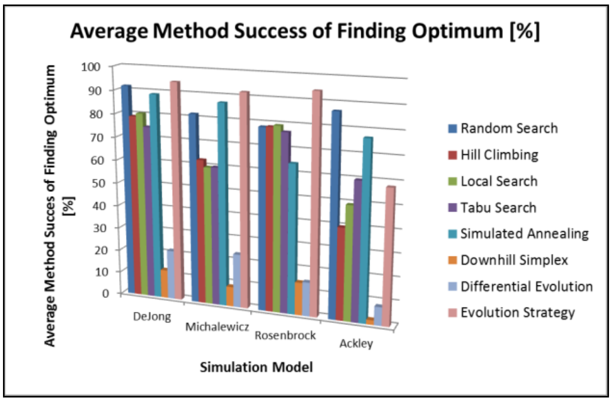


Figure 13: Average Optimization Method Success – Simulation Optimization Results of Testing Functions

Average optimization method success of finding the optimum of discrete event simulation models is shown in Figure 14. We can say that Simulated Annealing and Evolution Strategy are quite successful optimization methods again. Random Search was not successful in the case of the Penalty model because of the larger search space. The Penalty discrete event simulation model has a complicated objective function landscape. The area around the optimum is straight and the method could not obtain information about rising or decreasing the objective function terrain.

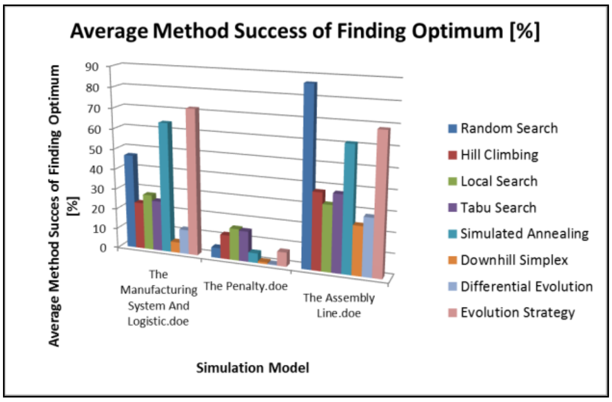


Figure 14: Average Method Success – Simulation Optimization Results of Discrete Event Simulation Models

Previous charts express the average success of optimization methods of all optimization methods settings. These charts also contain bad settings therefore we separated the bad series from the good series. The next chart contains the filtered series with the best found first criterion value only (in this case $f_1 = 0$ so the optimum was found in each optimization experiment). The percentage of absolutely successful series compared to all performed series is shown in Figure 15. It is obvious that the favourite, Evolution Strategy, has problems with the multimodal Ackley function. The success of this method was affected by the number of individuals randomly chosen from the population for the tournament – exploration vs. exploitation of the search space.

The first approach is to generate other new solutions which have not been investigated before - exploration. Since computers have only limited memory, the already evaluated solution candidates usually have to be discarded in order to accommodate new ones. Exploration is a metaphor for the procedure which allows search operations to find new and maybe better solution structures. Exploitation, on the other hand, is the process of improving and combining the traits of the currently known solutions, as done by the crossover operator in evolutionary algorithms, for instance. Exploitation operations often incorporate small changes into already tested individuals leading to new, very similar solution candidates or try to merge building blocks of different, promising individuals. They usually have the disadvantage that other, possibly better, solutions located in distant areas of the problem space will not be discovered. (Michalewicz 2004)

The behaviour of Hill Climbing, Local Search and Tabu Search is similar considering the similar pseudo gradient principle.

Substandard results were achieved with the Downhill Simplex method. This optimization method works by calculating the points of the centroid (center of gravity of the simplex). We have to modify this optimization method in such a way that it is applicable for discrete event simulation optimization purposes where the step in the search space is defined. We use the rounding of coordinates of the vector (new calculated point) to the nearest feasible coordinates in the search space and this leads to deviation from the original direction. We performed other simulation experiments with smaller steps and the success of finding the optimum was higher than before. This problem can be solved by using a calculation with the original points and the objective function value will be calculated by the approximations of the objective value of the nearest feasible points in the search.

Differential Evolution uses the elitism strategy in our case. This leads to copying of identical individuals which suppresses the diversity of new promising individuals. Random Search looks successful, but there were only two possible settings – generating the same individual possibility. This evaluation can be modified by using the coefficient which recalculates the value of success depending on the number of performed series. The termination criterion was the number of possible solutions in the search space when there is little search space. This led to increasing the probability of success of this optimization method.

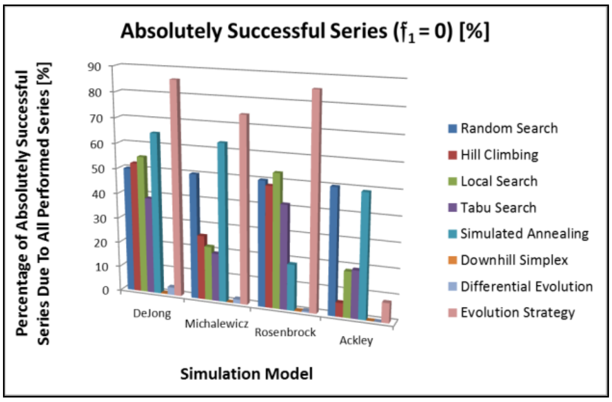


Figure 15: Percentage of Absolutely Successful Series Compared To All Performed Series - Testing Functions

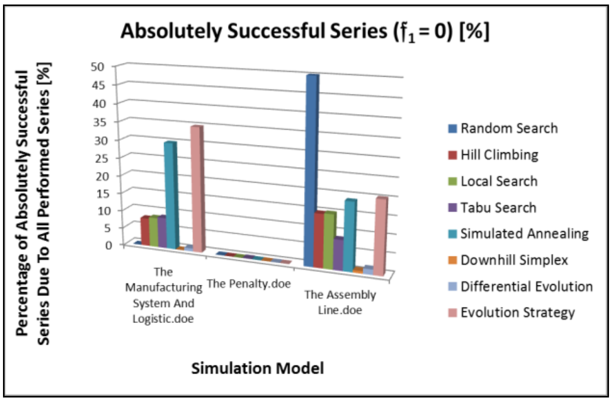


Figure 16: Percentage of Absolutely Successful Series Considering All Performed Series - Discrete Event Simulation Models

6.2. The Difference between Optimum and Local Extreme

The second criterion f_2 is useful when there is no series which contains any optimum or the solution whose objective function value is within the tolerance of optimum objective function value. The first criterion f_1 equals zero in this case. The function where the output of the function can take value $f_2 \in [0,1]$. This function evaluates the difference between the objective function value of the best solution found in the series and the optimum objective function value. The effort is to minimize f_2. The list of found optimums considering objective function value using the comparator function is sorted in ascending order. After that the value of the second criterion is calculated using the formula:

$$f_2 = \frac{\left|F(\mathbf{X}^*) - F(X_{\text{Best}})\right|}{\left|F(\mathbf{X}^*) - F(X_{\text{Worst}})\right|} \qquad (10)$$

where $F(\mathbf{X}^*)$ denotes the objective function value of the global optimum of the search space; $F(X_{\text{Best}})$ denotes the objective function value of the best solution found in concrete series; $F(X_{\text{Worst}})$ denotes objective function value of the worst solution (element) of the search space.

The difference between the optimum and the local extreme is shown in Figure 17 (testing functions) and Figure 18 (discrete event simulation models). The charts contain only series where the $f_1 = 0$ (no optimum was found in the series). The average of second criterion f_2 is shown for each optimization method – these values express the failure of the optimization method. Output of function can take value $f_2 \in [0,1]$.

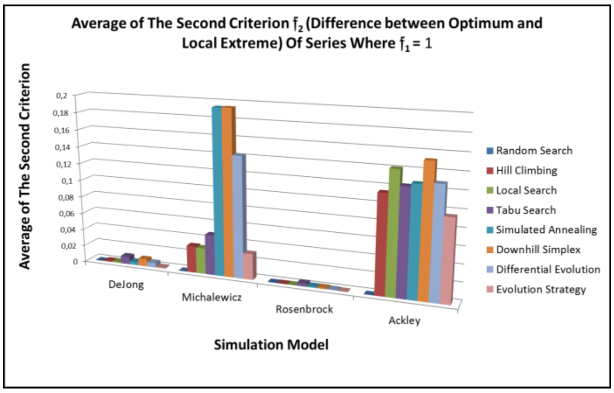


Figure 17: Average of the Second Criterion f_2 - Difference between Optimum and Local Extreme - Testing Functions

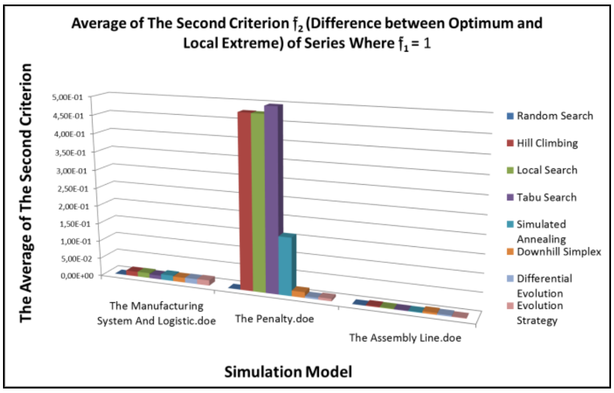


Figure 18: Average of the Second Criterion f_2 - Difference between Optimum and Local Extreme - Discrete Event Simulation Models

6.3. The Distances of Quartiles

Third criterion f_3 expresses the distance between quartiles of a concrete series. Weights are used for evaluation purposes. These weights penalize the solutions) placed in quartiles. Values of the weights were defined based on the results of the simulation experiments. The user can define the weight value. The sum of weights equals one. The third criterion when the objective function is minimized can be formulated as follows:

$$f_3 = \frac{\left|Q_1 - F\left(\mathbf{X}^*\right)\right| + w_{4 f_3}\left|Q_1 - Q_2\right| + w_{3 f_3}\left|Q_2 - Q_3\right| + w_{2 f_3}\left|Q_3 - Q_4\right| + w_{1 f_3}\left|Q_4 - Q_5\right|}{\left|F\left(\mathbf{X}^*\right) - F\left(X_{\mathrm{Worst}}\right)\right|} \quad (10)$$

where $F\left(\mathbf{X}^*\right)$ denotes the objective function value of the global optimum of the search space; $w_{4 f_3}$ denotes the weight (penalty) of objective function values between sample minimum Q_1 and lower quartile Q_2; $w_{3 f_3}$ denotes the weight of objective function values between lower quartile Q_2 and median Q_3; $w_{2 f_3}$ denotes the weight of objective function values between median Q_3 and upper quartile Q_4; $w_{1 f_3}$ denotes the weight of objective function values between upper quartile Q_4 and largest observation - sample maximum Q_5; $F\left(X_{\mathrm{Worst}}\right)$ denotes objective function value of the worst solution (element) of the search space. The evaluation of optimization experiments using the third criterion is shown in Figure 19 and in Figure 20.

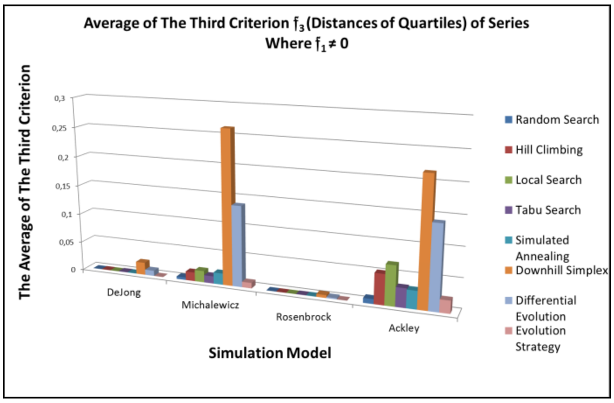


Figure 19: Average of the Third Criterion f_3 - Distances of Quartiles - Testing Functions

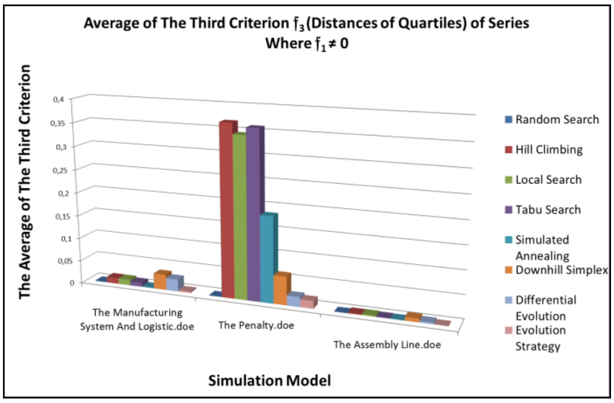


Figure 20: Average of the Third Criterion f_3 - Distances of Quartiles - Discrete Event Simulation Models

The effort is to minimize f_3 ($f_3 \in [0,1]$). If the first criterion equals zero $f_2 = 1$ then the third criterion equals zero $f_3 = 0$ (absolutely successful series). The Downhill Simplex optimization method provided the worst optimization results of all tested optimization methods due to rounding the coordinates. Pseudo gradient optimization methods found solutions of similar quality. Simulated Annealing provides a worse solution than the Evolution Strategy.

6.4. The Number of Simulation Experiments Until the Optimum Was Found

The fourth criterion f_4 evaluates the speed of finding the optimum – the number of performed simulation experiments until the optimum/best solution was found in each series. The effort is to minimize f_4 ($f_4 \in [0,1]$). The fourth criterion when the objective function is minimized can be formulated as follows:

$$f_4 = \frac{\left|Q_1 - 1\right| + w_{4 f_4}\left|Q_1 - Q_2\right| + w_{3 f_4}\left|Q_2 - Q_3\right| + w_{2 f_4}\left|Q_3 - Q_4\right| + w_{1 f_4}\left|Q_4 - Q_5\right|}{m_{\tilde{X}}} \quad (11)$$

where $w_{4 f_4}$ denotes the weight (penalty) of number of simulation experiments until the optimum was found between sample minimum Q_1 and lower quartile Q_2; $w_{3 f_4}$ denotes the weight of number of simulation experiments until the optimum was found between lower quartile Q_2 and median Q_3; $w_{2 f_4}$ denotes the weight of number of simulation experiments until the optimum was found between median Q_3 and upper quartile Q_4; $w_{1 f_4}$ denotes the weight of number of simulation experiments until the optimum was found between upper quartile Q_4 and largest observation - sample maximum Q_5; $m_{\tilde{X}}$ denotes the number of feasible solutions in the search space. The evaluation of optimization experiments using the third criterion is shown in Figure 21 and in Figure 22.

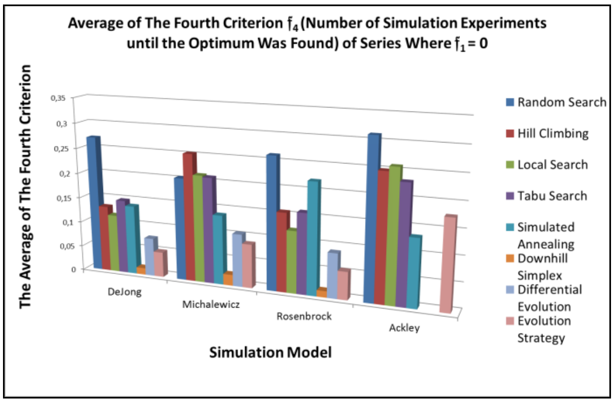


Figure 21: Average of the Fourth Criterion f_4 - Number of Simulation Experiments until the Optimum Was Found - Testing Functions

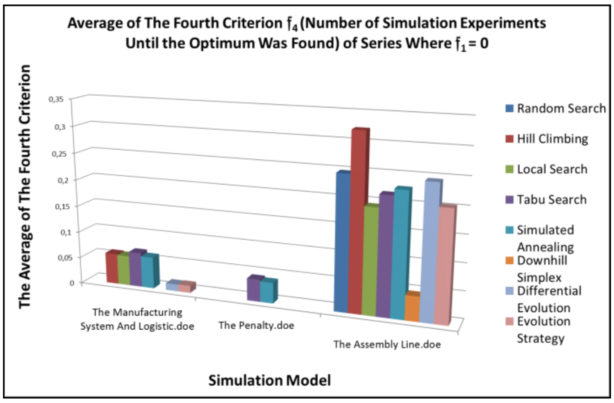


Figure 22: Average of The Fourth Criterion f_4 - Discrete Event Simulation Models

7. CONCLUSION

The goal of our research is to compare selected modified optimization methods (Random Search, Hill Climbing, Tabu Search, Local Search, Downhill Simplex, Simulated Annealing, Differential Evolution and Evolution Strategy) used in the developed simulation optimizer and used in the second module which is focused on testing the implemented optimization methods. Optimization methods generate whole populations instead of one possible solution which prevents premature convergence. The success of optimization methods depends on the objective function landscape. Evolution Strategy is a suitable optimization method for all the tested objective functions (a little propensity to bad tuning of the method parameters). This optimization method achieves good values for specified criteria. The alternative to Evolution Strategy optimization methods is Simulated Annealing. Simulated Annealing has the ability to escape from the local extreme thanks to the implemented approach of setting the temperature to the initial temperature. We can expect to find good results using Random Search if there is a small search space. If the dimension of the search space is bigger, there is little probability of success. Optimization methods based on pseudo-gradient searching such as Hill-Climbing, Local Search, Tabu Search achieve almost the same results for the simple objective function landscape due to their similar nature. Differential Evolution avoids repressing the diversity of solutions (elitism - an advantage of this approach is the faster finding of a feasible solution but not the finding of the global optimum). The range of provided simulation optimization results using this optimization method is better than the optimization methods based on pseudo-gradient searching.

ACKNOWLEDGEMENTS

This paper was created with the subsidy of the project SGS-2012-063 “Integrated production system design as a meta product with use of a multidisciplinary approach and virtual reality” carried out with the support of the Internal Grant Agency of University of West Bohemia and with the subsidy of the Motivation System (POSTDOC) of University of West Bohemia. The paper uses the results of the project CZ.1.07/2.3.00/09.0163 carried out with the support of Ministry of Education, Youth and Sports.

REFERENCES

Koblasa, F., Manlig, F., Vavruska, J., 2013. Evolution Algorithm for Job Shop Scheduling Problem Constrained bythe Optimization Timespan. *Applied Mechanics and Materials*, 309, 350-357.

Kopecek, P, 2012. Heuristic Approach to Job Shop Scheduling. *DAAAM International Scientific Book 2012*, pp. 573-584. October 24-27, Zadar (Croatia).

Michalewicz, Z., Fogel, D. B., 2004. *How to Solve It: Modern Heuristics*, Berlin: Springer,

Miranda, V., 2008. Fundamentals of Evolution Strategies and Evolutionary Programming. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 43–60.

Monticelli, A.J., Romero, R., Asada, E., 2008. Fundamentals of Tabu Search. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 101–120.

Monticelli, A.J., Romero, R., Asada, E., 2008. Fundamentals of Simulated Annealing. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 123–144.

Pohlheim, H., 2006. *Genetic and Evolutionary Algorithm Toolbox for use with MATLAB*. GEATbx. Available from: http://www.geatbx.com/docu/fcnindex-01.html [accessed 20 November 2011]

Tvrdik, J., 2004. *Evolutionary Algorithms - textbook*. Virtual information center for Ph.D. students, Ostrava University. Available from: http://prf.osu.cz/doktorske_studium/dokumenty/Evolutionary_Algorithms.pdf [accessed 6 February, 2011]

Votava, V., Ulrych, Z., Edl, M., Korecky, M., Trkovsky, V., 2008. Analysis and Optimization of Complex Small-lot Production in new Manufacturing Facilities Based on Discrete Simulation. *Proceedings of 20th European Modeling & Simulation Symposium EMSS 2008*, pp. 198-203. September 17-19, Campora San Giovanni (Amantea, Italy).

Weise, T., 2009. *Global Optimization Algorithms - Theory and Application 2nd Edition*. Thomas Weise - Projects. Available from: http://www.it-weise.de/projects/book.pdf [accessed 2 February, 2012]

Wong, K.P., Dong, Z.Y., 2008. Differential Evolution. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 171–186.

SIMULATION AND OPTIMIZATION OF PRODUCTION SYSTEM BASED ON FUZZY LOGIC AND QUICK RESPONSE MANUFACTURING

Centobelli P. [(a)]**, Murino T.** [(b)]**, D'Addona D.** [(c)]**, Naviglio G.** [(d)]

[(a), (b), (c), (d)] *Department of Materials and Production Engineering*
University of Naples Federico II, Naples, Italy
P.le Tecchio 80, 80125, Naples, Italy

[(a)] piera.centobelli@unina.it, [(b)] murino@unina.it, [(c)] daddona@unina.it, [(d)] giuseppe.naviglio@unina.it

ABSTRACT

Since 1979, european and american clients had benchmarked the performance of theirs factories with those of Japanese competitors. The differences included substantially higher productivity, better quality, significantly less inventory, less space, more flexibility and much faster throughput times. Everyone knows that time is money and mangers understand the importance of quick response to customers. Lean Manufacturing techniques can be powerful in several situations, but for companies making a large variety of products with variable demand or companies making highly engineered products, Lean Manufacturing has several drawbacks. Quick Response Manufacturing (QRM) can be more effective competitive strategy for companies targeting such markets, which focuses on lead time reduction. The importance of define the lead time required in an engineer-to-order company is critical in particular during the New Product Development (NPD) process. This paper presents how to apply Quick Response Manufacturing to a manufacturing industry through the previous calculation of product components Run Time using a Fuzzy Logic approach, in order to predict whether a decision will improve lead times.

Keywords: Fuzzy Logic, quick response manufacturing, processing time, new product development

1. INTRODUCTION

In the past few years the world attend a rapid growth in the number of options provided by manufacturers to their customers. Even beyond providing pre-specified options though, is the fact that the modern technology has given companies the ability to custom-engineer and then manufacture products for individual clients without incurring the high additional costs that such customization would have required two decades ago.

Along with this has come the power of internet, which allows customers to easily evaluate many different options and select from them. All of these development mean that there will be increasing demand for customized products in the 21st century (D'Addona and Teti, 2008). Today customers expect products to be delivered with a much shorter lead time than was acceptable in the past (Converso, Santillo and De Vito, 2013), (Chiocca, Guizzi, Murino, Revetria and Romano, 2012), (Converso, Aveta, Santillo and Gallo, 2012).

The improvement of flexibility has become increasingly important as a method to achieve competitive advantage in manufacturing (Beckman, 1990), (DeMeyer et al., 1989§), (Holusha, 1989), (Goldhar and Jelinek, 1983), (Zelenovic, 1982).

Flexibility may be seen as both a set of capabilities (internal) and a source of competitive advantage in a particular environment (external). It is important to distinguish the (internal) capability of being flexible from the (external) competitive need it is intended to match or the advantage derived from it (Shaouta and Al-Shammari, 1998). A possibility is to build capabilities which allow the manufacturing system to switch effortlessly and quickly between products, avoiding the carrying cost of the inventory and facilitating "just-in-time" production. This internal form of flexibility has been termed mobility, see (Upton, 1994),(Murino, Romano and Santillo, 2011),(Guizzi, Chiocca and Romano 2012).

Recently, many manufacturing companies affected by the economic slump due to challenge of competition by low-wage countries in the globalized market have looked inward, struggling to find ways to reduce response time, improve quality and costs (Suri, 2003).

The ability to change the product being manufactured quickly, on an on-going basis is the capability which most frequently supports the ability to provide quick response (Danny J. Johnson, 2003). And this is where Quick Response Manufacturing (QRM) comes in. These strategy enables companies to dramatically shorten their lead times to deliver products faster and, at the same time, improving their quality.

Factories lead times, work-in-process and actual capacities are all the result of complex dynamics and interactions on the manufacturing shop floor. A powerful methodology, called Rapid Modeling Technology (RMT), describes factory floor dynamics particularly well. An easy-to-use software tool (based on RMT) to assist companies in achieving and sustaining quick response in their manufacturing is the

MPX®. MPX® can assist engineers and managers analyze their operations to find opportunities for improvements related capacity, work-in-process, labour allocation, new product introduction and many other manufacturing issues. The key issue here is that lead time depends on both processing time and queue time that is the time parts spend waiting to get their turn at machines when the machines are busy.

While processing time may be known based on the machining parameters, queue time depends on many "dynamic" factors such as, which other parts are already in queue to use the machine, whether the machine has broken down, whether an operator is available, and so on. In order to predict whether a decision will improve lead times, it is thus necessary to be able to predict these queue times, which means any lead time reduction tool must model these dynamics and interactions. The RMT technology in MPX© models these complex dynamics of the manufacturing facility in terms of mathematical equations. Until a few years ago, these equations couldn't be solved. However, with the progress that has been made in queuing theory in recent years, very good estimates can now be obtained for system performance with amazingly little computer time, often just seconds on a personal computer (MPX user manual).

In this paper the MPX© has been used to simulate 'what if' scenarios which impact a variety of manufacturing parameters, including parts routing, labour, equipment, equipment failure/repair, set-up, run time and lot size.

The software has help the experts to evaluate the effects of alternative management decisions during the new product development. It helps in obtaining an insight into the factors that influence the lead time performance of cells and establish what would be the ideal cell configuration. The components processing time has calculated using a fuzzy approach.

The process followed is shown in Figure 1.

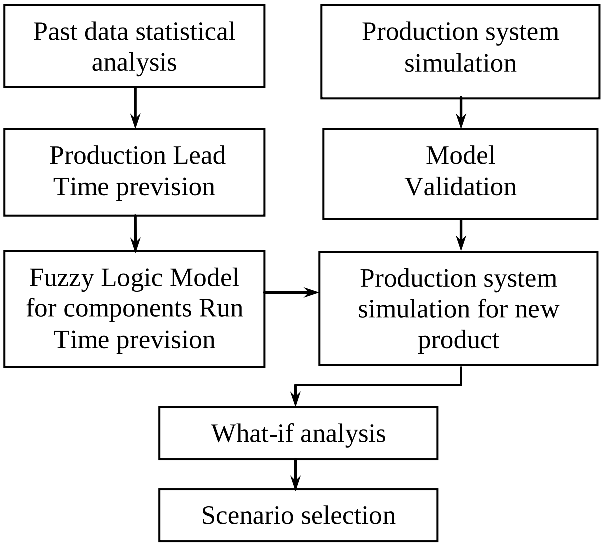

Past data statistical analysis
Production system simulation
Production Lead Time prevision
Model Validation
Fuzzy Logic Model for components Run Time prevision
Production system simulation for new product
What-if analysis
Scenario selection

Figure 1: Application Architecture

2. STATISTICAL ANALYSIS

The product at issue is a gas turbine coating for thermal and sound insulation. The gas turbine can be divided in six different macroareas similar for geometric structure. However every component runs the same operations routing shown in Figure 2.

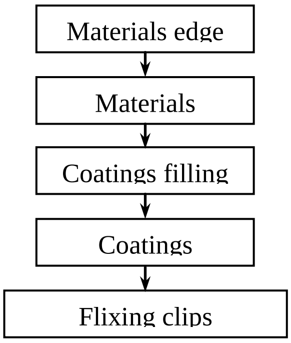


Figure 2: Components operations routing

First were analyzed labours daily work schedules related to time spent by each for every commission and for each operations. Time has been plotted choosing a significant feature as allocation base for each operation routing as shown in Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7 dividing two-dimensional (2D) and three-dimensional (3D) case for 2D and 3D components.

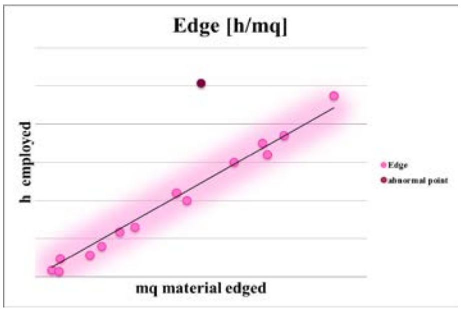


Figure 3: Materials edge

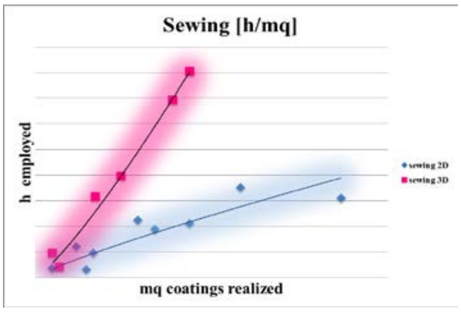


Figure 4: Materials sewing

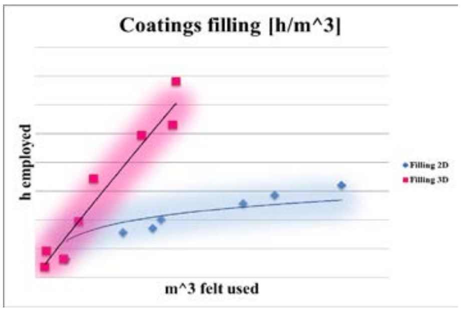


Figure 5: Coatings filling

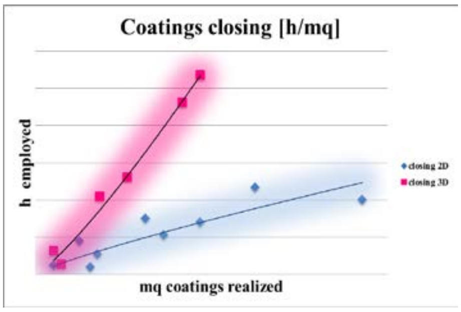


Figure 6: Coatings closing

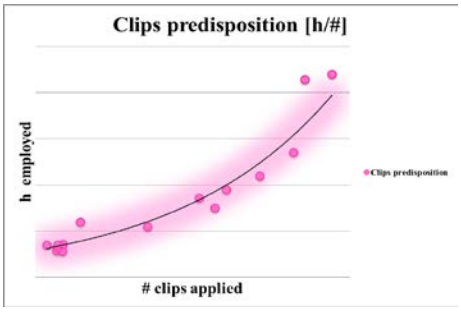


Figure 7: Number of clips applied

Near each trend line a cloud has been represented with amplitude ±10%, because mathematical methods are purely quantitative and they don't consider marketing strategies, structural problems and details, particular clients demands and other qualitative aspects.

This trend line model has been validate monitoring other processed commissions and the result is that the error of prevision never exceeds the 10%.

3. FUZZY LOGIC

The Fuzzy Set Theory allows us to represent the ambiguity contained in linguistic information (Zadeh, 1965). The first original paper on fuzzy logic encountered skepticism and hostility. Fourty years later many international journals have been published papers which include the word "fuzzy" in their title and thousands of patents have been applied.

By 1973, Zadeh had stated the principle of incompatibility on which the fuzzy approach is based: "As the complexity of a system increases, our ability almost mutually exclusive characteristics. It is in this sense that precise quantitative analyses of the behavior of humanistic systems are not likely to have much relevance to the real world societal, political, economic, and other types of problems which involve humans either as individuals or in groups" (Zadeh, 1973).

Given a universe of the discourse U, a fuzzy set A in U is defined by a membership function that assigns to each element in U a value between 0 and 1 (Figure 8). When a value 0 is assigned to an element 'u', 'u' doesn't belong to A; if instead it assumes the value 1 then it completely belongs to set A. But differently to what happens in the traditional set theory, in the fuzzy set theory a generic value can be assume an intermediate value between 0 and 1 then the element will partially belong to A with a specific membership degree (Iandoli and Zollo, 2007).

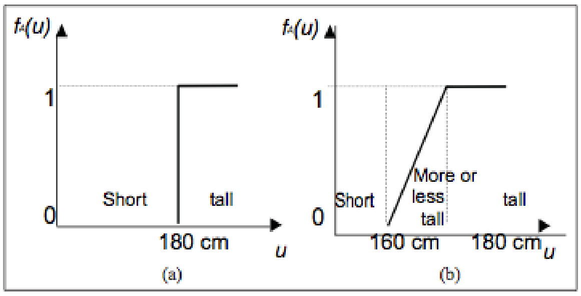


Figure 8: Example of membership function for: (a)traditional set and (b) fuzzy set

3.1. Dual Truth Model

The 'dual truth model' is a model proposed to represent verbal judgments through fuzzy logic. One of the factors which makes natural language such a flexible and efficient tool is its inherent vagueness. It is surprising, in fact, how even a fairly limited vocabulary is enough to enable a person to carry out even very complex tasks.

This model has been introduced in 1996 for the evaluation of competencies of professional workers within a large organization (Zollo, Cannavacciuolo, Capaldo, Ventre, and Volpe, 1996). In 2002 the model was used for methodological approach for the evaluation of innovation capabilities in small software firms (Capaldo, Iandoli, Raffa and Zollo, 2002).

In this paper the dual truth model has been used to define the processing time of each component of product analyzed. Dual truth model appears considerably suitable for decision making processes because, generally, decision making involves uncertainty. When the firm decides to introduce a new product there aren't historical data about time production for each operation to be processed on each component.

In this kind of problems it's very important to take in count expert's judgments on the bases of their ability and experience in a particular operation. The main task is the ability to handle these imprecise, incomplete and vague informations.

The work starts with the elaboration of a questionnaire showing component specifics and 3D figure of CAD project of new product, a new coating for gas turbine. It was necessary explain face to face to each worker of each industrial operation, every question presented in the questionnaire, asking their opinion about complexity of each component.

In agreement with dual truth model, has been used a Fuzzy Term Set (FTS) standard of five functions (Figure 9):

FTS = { ANC, PC, C, MC, EC }

Where:
ANC means "absolutely not complex" PC means not "much complex"
C means "complex"
MC means "very complex"
EC means "extremely complex"

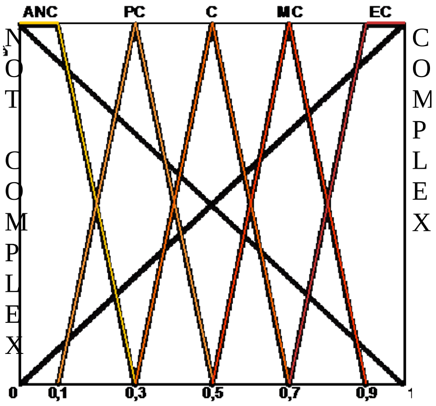


Figure 9: Term Set Fuzzy

The two diagonals of the square, represent the function COMPLEX and NOT COMPLEX. Simply they mean that the truth-value of COMPLEX increases linearly from 0 to 1, and, vice-versa, the truth value of NOT COMPLEX decreases.

Each component has a different complexity so it has been calculated a complexity rate in order to allocate the total lead time calculated with statistical and historical data.

The rate has been calculated doing the fuzzy average and then defuzzificating workers judgments. For example, when judgments are different, a new membership function, triangular or trapezoidal, must be calculated, which is obtained by the convolution of workers judgments (Figure 10).

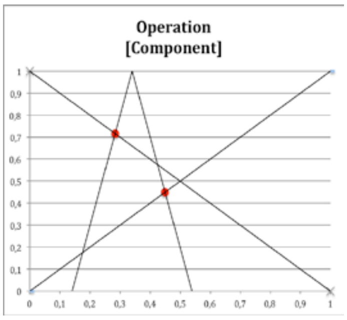


Figure 10: Judgments average represented using Dual Truth Model

Using the dual truth model a truth couple (y_{NC} , y_C) can be immediately associated to the fuzzy representation of the judgment which is obtained as shown in Figure 10 in red. The following formula can be used to get a non fuzzy reading:

$$I_{c_i} = \frac{(y_{C_i}+(1-y_{NC_i}))}{2} \tag{1}$$

The red point on the left (y_{NC}) expresses the probability that the judgment is "not complex" so (1 - y_{NC}) is its complementary, whereas (y_C) is the possibility that the judgment is "complex". So, Ic is a kind of average of two possibilities.

This process must be repeated for each component *i* of product and each operation routing.

In order to associate the real processing time to each component, starting from the total time necessary for each operation, must be calculated the I#, a rate that take in count the number of different kind components must be processed.

$$I_{\#_i} = \frac{n_i}{\sum_i n_i} \tag{2}$$

where
n_i is the number of same components
$\sum_i n_i$ is the total number of components

The total index for time allocation is:

$$I_{tot_i} = \frac{(I_{c_i} * I_{\#_i})}{\sum_i (I_{c_i} * I_{\#_i})} \tag{3}$$

4. MODELING MANUFACTURING FLOOR IN MPX®

When a production isn't automated but mainly manual it will be very difficult simulate through software the industrial plant (Di Franco, Gallo, Guizzi, and Zoppoli, 2009); in particular it will be very difficult define the processing time of each component for each operation. The main objective of the MPX® utilization is establish if the manufacturing system analysed can produce an upgrade of the product actually realized at the same time. Otherwise find solutions to optimize the performances of the manufacturing system (Grisi, Guerra and Naviglio, 2010). In the software mentioned, due to complete the simulation, is necessary insert several input data. These inputs regards:

- General Data: project name, time units, time unit conversions, maxim utilization admitted (in %)
- Labor Data: labor group name, number of workers that are present at one time, overtime (%), time unavailable (%).
- Equipment Data: equipment name and type, number of individual equipment in the group, Mean Time to Failure (MTTF), Mean Time to Repair (MTTR), overtime (%), labor group assigned to work in the group.
- Product Data and Product Operation Data: product name, end use demand for each product, average lot size; operation name and number of sequence, equipment group where this operation will be performed, % assigned. In these step it's necessary insert equipment and labor run time and setup time.

This is the real problem of manufacturing industry analyzed which has no idea of processing time of products never processed before. In this contest the fuzzy logic, and in particular the dual truth model, solved this kind of problem.

4.1. Production System Simulation

After calculated the processing time of each component it's possible insert this data in the software.

The first step is simulate the actual production system employed with the actual production, with data well-known and detectable, and with results comparable with real cases.

After simulating the production system the results related to the utilization of labors and machines (Figure 11 and Figure 12) in each work center have been compared with real cases.

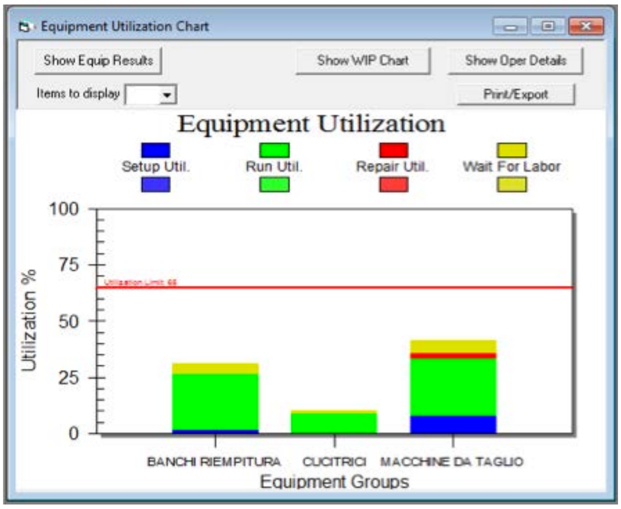


Figure 11: Report Labor Utilization Chart production system "as is"

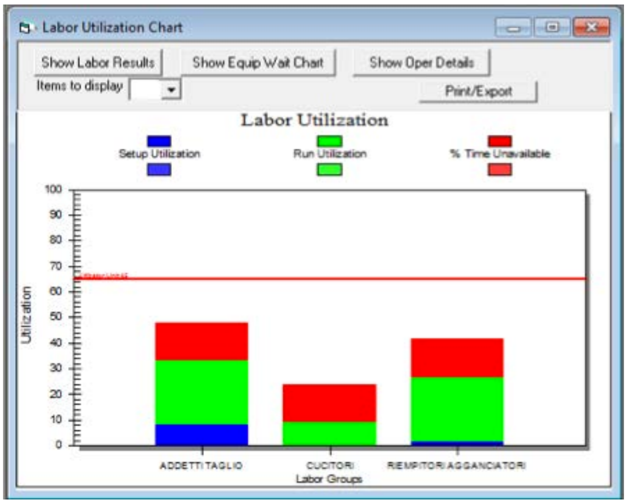


Figure 12: Report Equipment Utilization Chart production system "as is"

The sources most used are labors. Each rate time obtained (setup in Figure 13, run in Figure 14 and unavailable Figure 15) has been compared with real cases results.

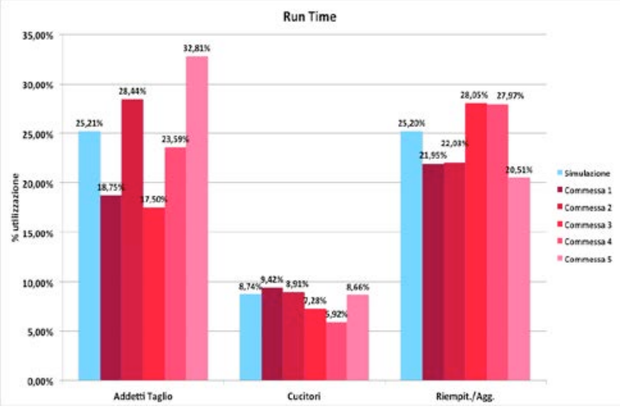


Figure 13: Labor Utilization for Run Time

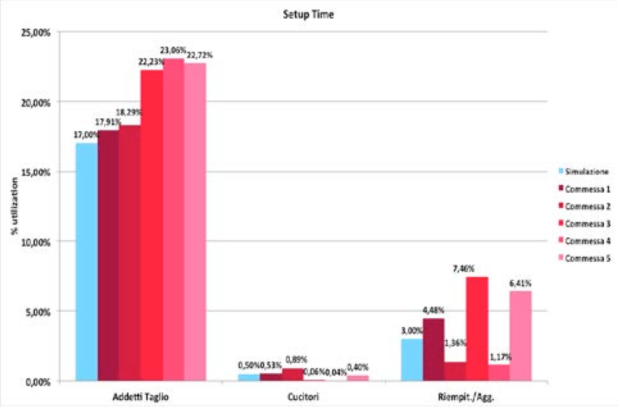


Figure 14: Labor Utilization for Setup Time

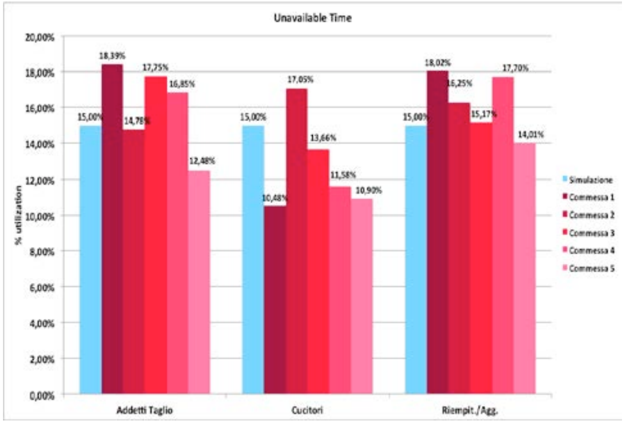


Figure 15: Labor Utilization for Unavailable Time

The difference between simulation results and real data are always under the 10% and this an acceptable value of tolerance specially for a manual manufacturing industrial system, not automated. The simulation software results a good simulator of product flow so seems consistent use it for a prevision analysis.

4.2. Simulation and optimization of industrial system employed with the new product

The new product follows the same operations routing of actual one. Obviously it is required to insert all data related to the new product, included the processing time founded with fuzzy model, and start the simulation.

In the case study analyzed software shows an error message means that production system can't realize the new product in the time established because a group of labor is overused (Figure 16). So it was necessary start a what-if analysis in order to rebalance resources.

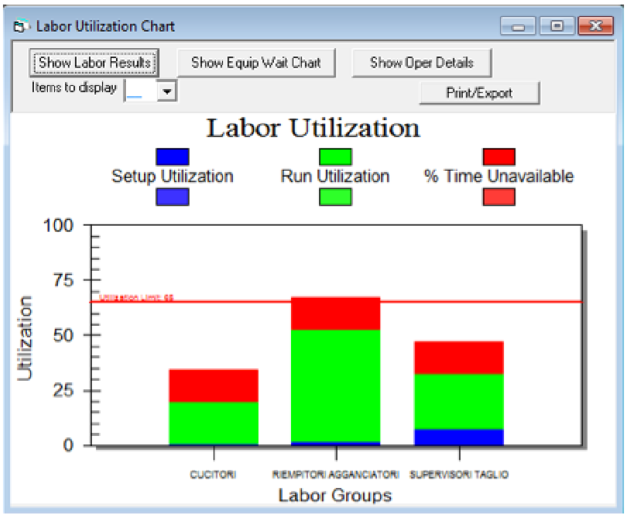


Figure 16: Labor Utilization Chart production system "to be"

4.3 What-if analysis

What-if analysis was used to compare different scenarios and their potential outcomes based on changing conditions. The what-if analysis was performed about some aspects and in particular on:

1. Increase lot-size;
2. Reduce lot-size;
3. Increase labors number in overused work center;
4. Move one (or more than one) labors from a work center to overused one;
5. Improve setup times.

Alternative 1. conduces to a growing of flow times spent waiting for lot while alternative 2. conduces to a growing setup times. Alternative 5. appears not feasible because setup times are not referred to machines.

The only liable way, in order to not increase the cost of the product, is move a labor from the first workcenter represented in Figure 16 named 'cucitori' to the second one, overused, named 'riempitori/agganciatori'. Now the software displays a message telling us the calculations are complete.

This change did enable us to make our production targets.

5. CONCLUSION

The RMT technology that models the complex dynamics of manufacturing facilities in terms of queuing theory mathematical equations was applied, through the use of the MPX© software code, for the evaluation and optimisation of a manufacturing firm during a new product development. Using a fuzzy logic model for time prevision the MPX© utilisation allowed for the analysis and verification of the manufacturing system capability to meet predefined lead time reduction goals and the finding of opportunities for performance improvements in the production system.

REFERENCES

Beckman, S.L., 1990. "Manufacturing flexibility: The next source of competitive advantage", in: P.E. Moody (Ed.), Strategic Manufacturing, Dow Jones-Irwin, pp. 107-132.

Capaldo, G., Iandoli, L., Raffa, M., Zollo, G., 2003. The Evaluation of Innovation Capabilities in Small Software Firms: A Methodological Approach, Small business economics : an internat. journal. - Dordrecht [u.a.] : Springer, ISSN 0921-898X, ZDB-ID 10245601. - Vol. 21.2003, 4, p. 343-354.

Chiocca, D., Guizzi, G., Murino, T., Revetria, R., Romano, E., 2012. A methodology for supporting lean healthcare. Studies in Computational Intelligence, 431, pp. 93-99.

Converso, G., Aveta, P., Santillo, L.C., Gallo, M., 2012. Planning of supply risks in a make-to-stock context through a system dynamics approach. IOS PRESS.

Converso, G, Santillo, L.C., De Vito, L., 2013. Sustainability of global Supply Chain network: the role of research and innovation. WSEAS Press.

D'Addona D., Teti R., Carbone R., 2008. Quick Response Manufacturing Approach for Lead Time Reduction in PCB Fabrication.

Danny J. Johnson, 2003. A framework for reducing manufacturing throughput time, Journal of Manufacturing Systems, Volume 22, Issue 4, 2003.

DeMeyer, A., Nakane, J., Miller, J.G., Ferdows, K., 1989. "Flexibility: The next competitive battle the manufacturing futures survey". Strategic Management J., vol. 10, pp. 135- 144.

Di Franco, R., Gallo, M., Guizzi, G., Zoppoli, P., 2009. Project risk management: A quantitative approach through simulation tecniques. *Proceedings of the 8th WSEAS International Conference on System Science and Simulation in Engineering, ICOSSSE '09*, pp. 326-333.

Gallo, M., Grisi, R.M., Guizzi, G., 2010. A vendor rating model resulting from AHP and the linear model. *International conference on System Science and Simulation in Engineering - Proceedings*, pp. 370-377.

Gallo, M., Guerra, L., Guizzi, G., 2010. Some considerations on inventory-based capacity scalability policies in RMSs. *International conference on System Science and Simulation in Engineering - Proceedings*, pp. 342-347.

Goldhar, J.D., Jelinek, M., 1983. "Plan for economies of scope". Harvard Business Rev., November-December, pp. 141-148.

Grisi, R.M., Guerra, L., Naviglio, G., 2010. "Supplier Performance Evaluation for Green Supply Chain Management. Business Performance Measurement and Management, Springer, Gennaio 2010. ISBN: 978-3-642-04799-2.

Guizzi, G., Chiocca, D., Romano, E., 2012. System dynamics approach to model a hybrid manufacturing system, Frontiers in Artificial Intelligence and Applications, 246, pp. 499-517.

Holusha, J., 1989. "Beating Japan at its own game". The New York Times, 16 July.

Iandoli L., Zollo G., 2007. Organizational Cognition and Learning, Building System for the Learning Organization. Information Science Pub, 2007.

MPX® User Manual, 1999, Network Dynamics Inc., MA, 01803, USA.

Murino, T., Romano, E., Santillo, L.C., 2011. Supply chain performance sustainability through resilience function. *Proceedings – Winter Simulation Conference*, art. no. 6147877, pp. 1600-1611.

Shaouta, A., Al-Shammari, M., 1998. Fuzzy logic modeling for performance appraisal systems A framework for empirical evaluation. Expert Systems With Applications 14, 323–328.

Suri R., 2003. QRM and POLCA: A Winning Combination for Manufacturing Enterprises in the 21st Century, Technical Report, Center for Quick Response Manufacturing, May 2003.

Upton, D.M., 1994. "The management of manufacturing flexibility", California Management Rev., Hass School of Business, University of California, Berkeley, Winter.

Zadeh L., 1965. Fuzzy sets, Information and Control, 8, 338-353.

Zadeh L., 1973. Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems*, Man and Cybernetics, 3(1), 28-44.

Zelenovic, D.M., 1982. "Flexibility - A condition for effective production systems", Internat. J. Prod. Res., vol. 20, no. 3, pp, 319-337.

Zollo, G., Cannavacciuolo, A., Capaldo, G., Ventre, A., Volpe, A., 1996. "The organizational evaluation process: a fuzzy model", Fuzzy Economic Review, Vol. 1, n° 1, pp. 3÷30, 1996.

AUTHORS BIOGRAPHY

Piera Centobelli studied at the University of Naples Federico II where she received with honors the M.Sc. Degree in Management Engineering in 2013. Currently she is a PhD student at the Department of Materials and Production Engineering, University of Naples Federico II. The PhD programme is entitled 'Production Technology and Systems', with a particular focus on Innovative Materials, Processes and Systems.

Teresa Murino is researcher in Department of Materials Engineering and Operations Management University of Naples "Federico II". She teaches "Industrial Logistics" at the University for students in Logistics&Production Engineering (3rd year) and "Manufacturing System Management" for students in Mechanical Engineering (4th year) degree courses. The research interest include Advanced Model and Applications of Logistics&Manufacturing, simulation with system dynamics and optimization problems, data analysis.

Doriana D'Addona works as Assistant Professor at the Dept. of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II. She has taken part in a significant number of regional, national and international research projects. She is member of national and international scientific associations in the field of production engineering: (CIRP) and (AITEM). Her main research interests are focused on manufacturing processes and automation, intelligent computation for manufacturing technology and systems.

Giuseppe Naviglio is a PhD in Production Systems and Technologies at the Department of Materials Engineering and Operations Management of the University of Naples Federico II. He has a master degree in management engineering from the University of Naples"Federico II. His research field is related to modeling and simulation of production systems and on Lean Manufacturing . Now he is a project manager of a Research & Development radar laboratory.

INFORMATION SECURITY INCIDENTS SIMULATION FOR RISK ANALYSIS IN THE NATIONAL AUTONOMOUS UNIVERSITY OF MEXICO

Israel Andrade Canales[(a)]

[(a)]National Autonomous University of Mexico

[(a)]iandradec@unam.mx

ABSTRACT

The National Autonomous University of Mexico (UNAM) offers different types of services to support academic activities. All of these services use valuable information for the achievement of their objectives and goals; consequently, information is one of the most important assets that the University has. However, thousands of security incidents affect these assets every year; for instance, in 2012 the university network suffered about 16,000 incidents provoked by botnets, spam and brute force attacks. Until now, this problem has been confronted by qualitative risk analysis methodologies in order to select counter-measures that mitigate these dangerous events. Nevertheless, these approaches lack either an optimization point of view or accurate results. Because the institution needs to treat risk not only precisely but also plausibly in financial and technical terms, this paper tries to shed light on a mixed model that combines simulation and linear optimization for the prediction and treatment of security incidents.

Keywords: information security, risk analysis, simulation, optimization

1. INTRODUCTION

Nowadays, information is a valuable asset that is used by people and organizations for decision making, communicating ideas, offering services and creating a variety of products. Therefore, Information Technologies (IT) have been developed for processing, storing and transmitting information in a practical manner. However, different risks affect information seriously; for instance, software bugs that generate vulnerabilities, and risky user habits that damage it.

The National Autonomous University of Mexico (UNAM) uses information technologies to offer different types of services that support academic activities, but thousands of security incidents affect these assets every year. For instance, in 2012 the university network suffered about 16,000 incidents provoked by botnets, spam and brute force attacks (UNAM-CERT 2012).

However, this kind of incidents frequently occurs in world-wide organizations; for example, the survey Information Security Branches (Price Waterhouse Coopers 2012) shows that nine of ten large enterprises in the UK reported an information security incident whose impact amounted to somewhere between 110 and 250 thousand pounds.

Until now, this problem has been dealt with risk analysis, i.e. the methodical use of information in order to identify and evaluate risk (ISO 2009). In this sense, there are two manners to assess risks: qualitatively and quantitatively (Gollman 2011, Buchanan 2011).

Qualitative risk methodologies use techniques such as manual inspections, staff interviews and information provided by experts in accordance to structured methods like OCTAVE (Caralli 2007). Although these techniques allow to carry out risk analysis in a coherent, repeatable and documented way, they may be subjective because they lack mathematical models that can give more accurate information to identify and analyze risks.

On the other hand, quantitative risk methodologies are model-based techniques that use mathematical tools like decision trees (Sahinoglu 2005), and simulation, such as those proposed by Winkelvos et al. (Winkelvos 2011). These methodologies give more precise information about risks. Even though these models are more precise than the qualitatively ones for risk analysis, most of them lack an optimization focus that helps to minimize risks for decision making and risk treatment.

Since UNAM needs to treat risk not only precisely but also plausibly in financial and technical terms, this paper tries to shed light on a quantitative risk analysis with two principal purposes: (1) to analyze risks with a model-based technique; and then, (2) to design a feasible risk treatment plan. Therefore, the proposed model combines simulation and linear optimization for the prediction and treatment of risks based on incident reports. First, the methodology used to analyze risk with the model proposed is described. Next, details about the model are given. Then, the results obtained are discussed. And finally, the conclusions that can be drawn from this research are presented.

2. METHODOLOGY

To carry out the information security risk analysis proposed in this paper, a combined model of simulation and optimization was proposed. The simulation model was built to perform a risk analysis in some scenarios of interest. On the other hand, the optimization model was formulated to treat efficiently the simulated risks in order to minimize the negative effects of information security incidents.

Figure 1 shows the connection between the simulation and the optimization model; the arrows represent the variables used to formulate each model, and the variables used to link both models; for example, the predicted incidents, the simulation output, are used as the objective function in the optimization model.

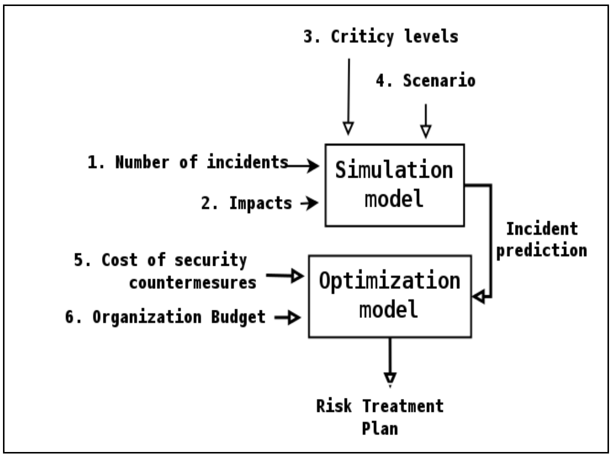


Figure 1: The Model Proposed

2.1. Simulation Model

The simulation model was used to analyze risks. This tool uses two random variables that represent the number of incidents and their impact. These variables were obtained through the data analysis of the information security incident reports of a university office.

2.1.1. Variables

The first variable (number of incidents) expresses the frequency of four types of incidents classified according to their sources: external entities (H), e.g. hackers; configuration errors (C); policy violations (P); and lack maintenance (M). However, because some organization activities affect the probabilities of the security incidents, four main scenarios were analyzed. Table 1 shows the probabilities for each incident in the four considered scenarios: External Projects (1), “normal” days (2), auditing procedures (3) and public events (4).

For example, when the academic office organizes a public event such as a congress, it is more likely to suffer an external attack (0.42) than when the organization has a normal period of activities. On the other hand, when the university office administers external projects, it is more common for configuration errors in the equipment to occur than when the organization is running a public event.

Table 1: Incident Probability Table

	Scenario			
Incident	**1**	**2**	**3**	**4**
H	0.10	0.01	0.11	0.42
C	0.36	0.97	0.77	0.14
M	0.36	0.01	0.01	0.42
P	0.18	0.01	0.11	0.02
Min	2	0	2	3
Max	4	1	6	4

The second variable (Impacts) corresponds to the number of idle hours caused by each incident. This information is the impact of each incident on the organization. Other kind of impacts may be: the cost of each incident, the damage to reputation, etc.

To estimate the productivity hours lost, the incidents reported by the academic institution in 2011 were analyzed. A security incidents histogram was built and adjusted by a probability distribution function; for instance, Figure 2 shows how the histogram of external attacks incidents (H) was adjusted to a beta probability density function; the beta function expresses that the organization loses between 0.5 to 20 productivity hours in an external attack, but the most likely ranges are between 0.5 to 7, and between 14 to 20 hours. Table 2 presents the parameters of the probability density function for each incident.

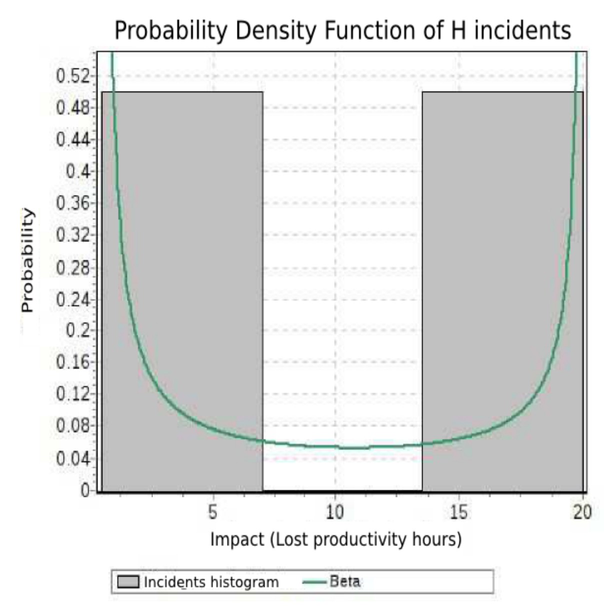


Figure 2: Probability Density Function of External Attacks Incidents (H)

Table 2: Impact Functions

Incident	Function	Parameters

H	Beta	α = 0.063, β = 0.159
		a = 0, b = 20

C	Beta	α = 0.071, β = 0.337
		a = 0.5, b = 24

P	Uniform	Min = 0.3
		max = 4.09

M	Uniform	Min = 0
		max = 2.62

2.1.2. Constants

On the other hand, two constants were used to run the simulation: a critical level of incidents and the experiment scenario. The first constant indicates the ranges of hours for three level categories: high, medium and low. Table 3 shows the ranges defined in the simulation.

The second constant indicates a series of activities that the organization may perform in the analyzed period. This information indicates which probability function to use in the model.

Table 3: Critical Levels of Impact

Levels	Low	Medium	High
Range (hrs)	0-5	5-10	10+

2.1.3. Simulation

Finally, a Monte-Carlo technique was used as a numeric method to relate the behavior of all the variables and constants in order to run the simulation. The Monte-Carlo simulation was programmed in the statistical language *R* (R Core Team 2012). Hundreds of simulation cycles were necessary to obtain stable results. These results acted as input data of the optimization model.

2.2. Optimization Model

The optimization model was formulated in order perform the risk treatment plan after the risk analysis; in other words, it was used to obtain a combination of activities that minimize the information security risks. This model takes into account that each activity has a cost in financial and work time terms. In this sense, it allowed us to obtain a security treatment plan, which is totally feasible for the organization.

The optimization model was derived from an Integer Binary Program which specifies a list of variables that represents the activities to be implemented, *i.e.* the countermeasures such as firewalls or information policies designed by the information security team; this approach has been suggested by Caulkins J. (Caulkins 2007) and Garvey (Garvey 2009) . The formal model is described as follows.

$$Min: \sum_{j=1}^{m} \sum_{i=1}^{n} -r_{ij} x_{ij} \qquad (1)$$

s.t

$$\sum_{j=1}^{m} \sum_{i=1}^{n} c_{ij} x_{ij} \leq B$$
$$x \in \{0,1\}, c \geq 0, B \geq 0 \quad i, j \in \mathbb{N} \qquad (2)$$

The objective function (Equation 1) represents the risks ***r*** to be mitigated by the activity ***x*** in order to minimize the total risk. In the equation, ***n*** is the number of levels registered in Table 3, and ***m*** represents the number of different kinds of incidents, *i.e*. H, C, M and P.

Equation 2 expresses the restriction of cost ***c*** of each activity ***x***, which must be less or equal to the organization's budget ***B***. Table 4 indicates those restrictions and the budget estimated for the academic office.

Once these equations have been solved, the results can be used as a decision-making aid to establish a Risk Treatment Plan that the organization can follow in order to obtain a reasonable state of security.

Table 4: Restrictions Used in the Optimization Model

Restrictions		
Countermeasure	**Time of Implementation**	**Cost**
1	1	1
2	5	3
3	10	10
4	1	5
5	5	10
6	15	50
7	2	10
8	5	50
9	10	100
10	1	10
11	5	20
12	10	50
Budget	32	150

3. RESULTS AND DISCUSSION

In this section, we show and discuss the results of the risk analysis. First we present the simulation results; then, we present a brief comparison between the simulation results and the information reported in 2012 by UNAM. Next, we show how the results of the simulation were transformed to formulate the optimization model. And finally, we report the results of the optimization model that represents a feasible risk treatment plan.

3.1. Simulation results

The simulation reported 21 incidents according to the organization's activities report in 2012. These incidents were distributed as follows (Table 5), 11 configuration errors; 2 external attacks; 3 lack-of-maintenance related incidents; and 5 policy violation incidents. To reach stable results, a hundred simulations were run. Figure 3 shows a graphic with the number of incidents obtained in these tests.

Table 5: Results of the Simulation

Simulation results		
Incident	**Number of Incidents**	**Impact (Idle hours)**
C	11	49.4
H	2	10.4
M	3	4
P	5	10
Sum	**21**	**73.8**

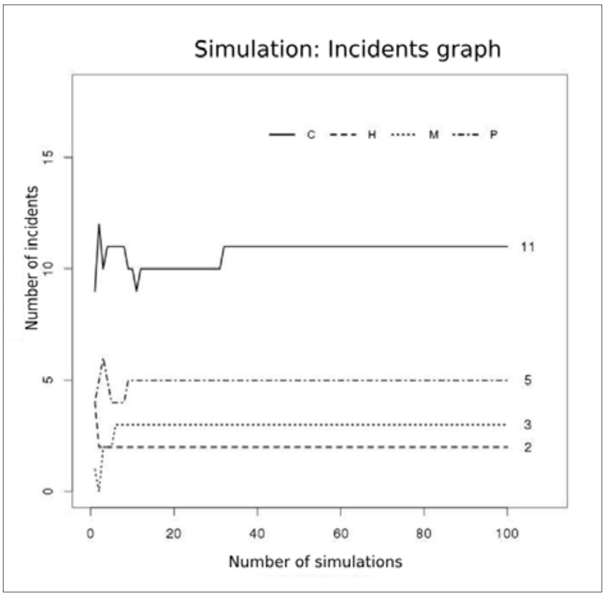


Figure 3: Number of Incidents Obtained by the Simulation Tests

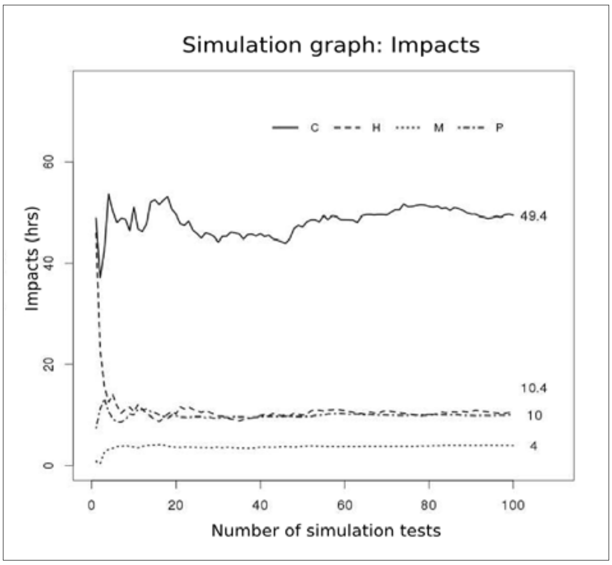


Figure 4: Productive Hours Lost Due to Information Security Incidents Obtained by the Simulation

In addition, the simulation reported 73.8 productive hours lost due to incidents. Figure 4 shows how the impact of each kind of incident, a random variable in the simulation, fluctuated at the beginning of the tests, and then leveled out at the end of the simulations.

3.2. Comparison between simulated and reported results

The simulation shows results consistent with the number and type of incidents reported in 2012 by the academic institution (Table 6). Figure 5 highlights the comparison among the incidents simulated and the incidents reported in 2012. As can be seen, the number of incidents was very similar to the actual number reported by the institution.

Table 6: Results Reported by the Organization in 2012

Simulation results		
Incident	**Number of incidents**	**Impact (Idle hours)**
C	12	52.2
H	3	2
M	2	6
P	6	9.9
Sum	**23**	**70.1**

Furthermore, Figure 6 exhibits a comparison of impacts per incident between the simulation and the incidents reported in 2012. It is important to notice that incidents occurred due to external attacks (H) were considerably fewer than the ones reported in 2012. The probability function does not imitate the real-system variable because external attacks are significantly random in the problem analyzed.

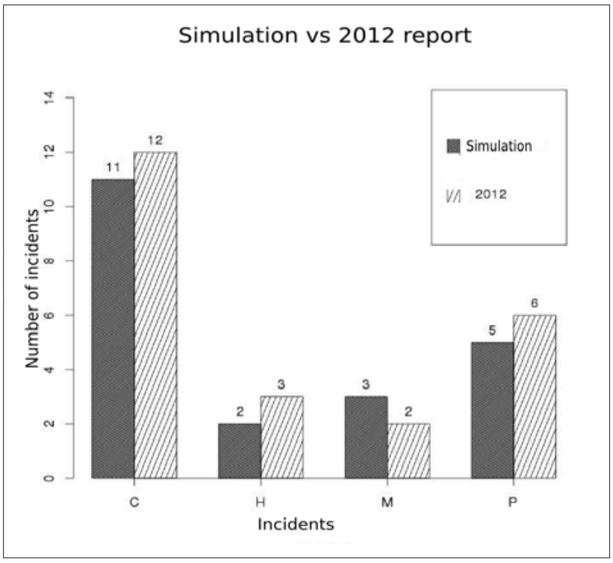


Figure 5: Comparison between Incidents Obtained by Simulation and Incidents Obtained by 2012 Reports

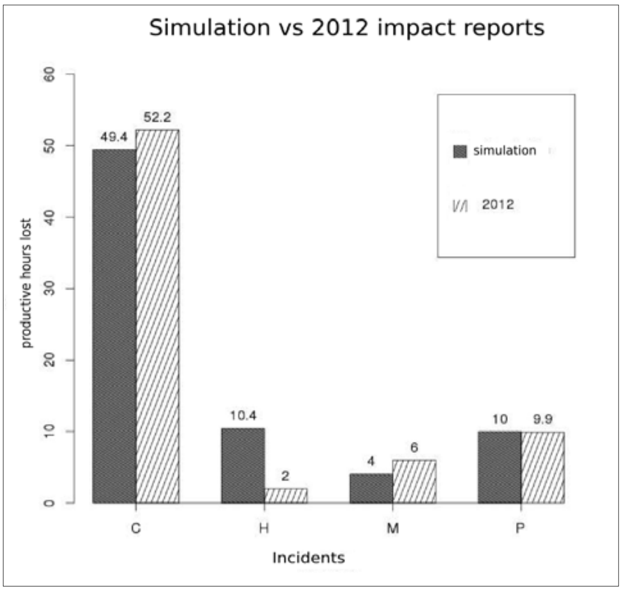


Figure 6: Comparison Between Incident Impacts Obtained by Simulation and Incidents Impacts Obtained by 2012 Reports

3.3. Simulation results as optimization model input data

Because the second objective of this paper was to formulate a risk treatment plan based on the simulation results, we grouped the incidents obtained according to the impacts critical levels (Table 3). These groups were used to establish the objective function (Equation 1) described in Section 2.2

Table 7 highlights the groups of incidents derived by the simulation and the impact critical levels established by the organization; moreover, Equation 3 indicates how this information is used to formulate the objective function.

Table 7: Incident Groups According to the Impact Critical Level

	Incidents			
Critical level	**C**	**H**	**M**	**P**
Low	6.3	1.1	4	10
Medium	3.7	0.8	0	0
High	39.4	8.5	0	0
Total	49.4	10.4	4	10

On the other hand, the restrictions shown in Equation 4 and Equation 5 were derived from the information described in Table 4.

$$Min\!:\!-6.3x_1-3.7x_2-39.4x_3-1.1x_4-0.8x_5 \\ -8.5x_6-4x_7-10x_{10} \tag{3}$$

subject to

$$x_1+5x_2+10x_3+x_4+5x_5+15x_6+2x_7+5x_8+10x_9+\ldots \\ \ldots+x_{10}+5x_{11}+10x_{12}\leq 32 \tag{4}$$

$$x_1+3x_2+10x_3+5x_4+10x_5+50x_6+10x_7+50x_8+\ldots \\ \ldots+100_9+10x_{10}+20x_{11}+50x_{12}\leq 150 \tag{5}$$

3.4. Results of the optimization model

The integer program, used to obtain a risk treatment plan, was solved through the *lpsolve software* (Berkelaar 2004). This software could be integrated with the simulation code to automate either the simulation tests or the optimization model.

The optimization model gave a risk treatment plan that allows to mitigate 69 of the 74 productive hours lost. Table 8 shows what activities should be implemented in order to reach these results. However, because both models are formulated with random data, the results should be used as a decision-making aid tool.

Table 8: Risk Treatment Plan Obtained from the Optimization Model

Risk treatment plan				
Activity	x_1	x_2	x_3	x_4
Plan	1	0	1	1
Activity	x_5	x_6	x_7	x_8
Plan	0	1	1	0
Activity	x_9	x_{10}	x_{11}	x_{12}
plan	0	1	0	0

4. CONCLUSIONS

Information security is a relative new field that can be explored through models like simulation and analytical models. In the approximation presented in this paper, the combination of the two models allowed both to analyze information security risks and treat them efficiently. This can be useful in an organization that deals with some restrictions on security investment.

On the other hand, the comparison between the results obtained and the information reported in the case studied suggests that systematic incidents such as policy violations and human errors caused by ambiguous procedures can be estimated successfully. Nevertheless, the same validation highlighted that some random events, like hacker attacks, are less precisely estimated. However, the two main objectives presented in this paper, not only to perform a risk analysis, but also to treat the impacts of incidents, were reached.

Therefore, this case of study showed that simulation and linear optimization are a powerful technique for a better decision-making in information security.

ACKNOWLEDGMENTS

This research was supported by UNAM-PAPIIT grant IN116012, and the author would like to thank to Dra. Idalia Flores de la Mota and Dr. Juan Manuel Estrada Medina for their valuable and constructive suggestions during the planning and development of this research work, to Martha Mora Barreto for the editing and proofreading help, and to “The Computer Emergency

Response Team (UNAM-CERT)" for making available the data used in the case of study.

REFERENCES

Berkelaar, M., Eikland, K. and Notebaert, P., 2004. *Open source (Mixed-Integer) Linear Programming system*. Poole, GNU lpsolve. Available from: http://lpsolve.sourceforge.net [accessed 1 July 2012]

Buchanan, W. J., 2011. *Introduction to Security and Network Forensics*. 1st ed. The U.S.: CRC Press.

Callari, R.A., Stevens, J.F., Young, L.R. and Wilson W.R., 2007. *Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process*. SEI, Carnegie Mellon University. Available from: http:/*/www.cert.org/archive/pdf/07tr012.pdf* [accessed 1 July 2012]

Caulkins, J. P., Hough, E. D., Mead, N. R., Osman, H., 2007. Optimizing Investments in Security Countermeasures: A Practical Tool for Fixed Budgets. *IEEE Security and Privacy*, 5 (5), 57-60.

Garvey, P. R., 2009. *Analytical methods for risk management: a systems engineering perspective*. 1st ed. Massachusetts, U.S.A: CRC Press.

Gollman, D., 2011. *Computer Security*. 3th ed. The U.K.: Wiley and sons.

Price Waterhouse Coopers, 2012. *Information Security Branches 2012*. Price Weterhouse Coopers. Available from: http://www.pwc.co.uk/en_UK/uk/assets/pdf/olpap/uk-information-security-breaches-survey-technical-report.pdf [accessed 1 November 2012]

R Core Team, 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/ [accessed 1 July 2012]

Sahinoglu M., 2005. Security Meter: A Practical Decision-Tree Model to Quantify Risk. *IEEE Security and Privacy*, 3 (3), 18-24.

UNAM-CERT, 2012. Estadísticas. Universidad Nacional Autónoma de México. Available from: http://www.cert.unam.mx/estadisticas.dsc [accessed 1 November 2012]

Winkelvos, T., Rudolph, C. and Repp, J., 2011. A property based security risk analysis through weighted simulation. *Information Security South Africa (ISSA)*, 1 (8), 15–17.

ISRAEL ANDRADE CANALES

Israel studied Computer Engineering at Facultad de Estudios Superiores Aragón, UNAM. He worked 3 years at the Computer Emergency Response Team (UNAM-CERT) of the National Autonomous University of Mexico in the information security auditing and risk analysis area. He is currently studying a master in Operations Research, and his line of research is optimization in information security.

MODELLING COMPLEX AND FLEXIBLE PROCESSES FOR SMART CYBER-PHYSICAL ENVIRONMENTS

Ronny Seiger[(a)], Christine Keller[(a)], Florian Niebling[(a)], Thomas Schlegel[(a)]

[(a)] Junior Professorship in Software Engineering of Ubiquitous Systems, Technische Universität Dresden, Germany

[(a)] {ronny.seiger, christine.keller, florian.niebling, thomas.schlegel}@tu-dresden.de

ABSTRACT

Cyber-physical Systems introduce several new requirements for modelling and executing autonomous processes. Current workflow languages are not able to completely fulfil these requirements, as they mostly lack expressiveness and flexibility. In this paper, we therefore present a new workflow language for formalizing processes within heterogeneous and dynamic environments. Our approach is highly model-based and uses aspects of component-based software engineering. We present an object-oriented meta-model for describing processes, which enables the hierarchical composition of process components and leverages reusability. In addition, a domain-specific model is used for typification of process elements. Due to the object-orientation, we are able to easily extend our models and create variants of processes. Type-based modelling and polymorphism enable the dynamic selection of appropriate process steps at runtime, creating flexible processes. We present a graphical editor and a distributed execution engine for our meta-model. In addition, we discuss the use of semantic technologies for smart workflows.

Keywords: workflow language, process modelling, cyber-physical systems, meta-modelling, smart factory, automation

1. INTRODUCTION

Business processes have gained an increasing importance in describing complex correlations between distributed systems and executing composite workflows. Especially in the field of online trading and manufacturing, modelling and execution languages for business processes, e. g, BPMN and BPEL, have proven to be well suited to formalize high-level sequences of tasks and activities involving web service invokes and human interaction.

However, the on-going integration and combination of embedded systems and distributed cloud-based services into cyber-physical systems (CPS) and smart environments, lead to a number of new requirements for process modelling and execution. Most current workflow languages lack structure, expressiveness, and flexibility to meet these requirements.

Some of the drawbacks of state of the art process modelling languages include: only weak means for typing of process components and data, mostly static calls to a fixed set of service types, and reduced flexibility considering runtime modelling and adaptation. Modelling tools often produce code, which is incompatible with execution environments, and only a subset of the model elements is supported.

In addition, many long-established workflow modelling languages have been extended and evolved over time, mostly by adding new components and modifying the respective meta-models in order to meet new requirements and provide new functionality. This has led to complex and ambiguous process modelling languages containing special solutions for specific problems and domains.

In this paper, we present a new meta-model for processes designed to meet the requirements of current and future ubiquitous systems. We believe that by using model-based approaches, we can create a modular and extensible workflow language. With the help of this language, we will then be able to model flexible and dynamic processes for the automation of workflows. Current semantic technologies will help us with developing a smart and context-adaptive process engine and modelling environment. We focus on adhering to simple structures for the core of the process meta-model and at the same time being able to easily extend this model by means of object-orientation. Nevertheless, we are able to map process models of other workflow languages to models compatible with our system.

The paper is structured as follows: Section 2 presents some basic terms and explanations. Section 3 lists requirements that are introduced with the emergence of ubiquitous systems. Section 4 gives a brief overview of related work and evaluates state-of-the-art workflow languages with respect to their suitability for cyber-physical systems. Section 5 describes our own process model for complex and flexible business processes in detail. Section 6 demonstrates practical aspects with respect to implementing the model, a modelling tool, and a process execution engine. Section 7 discusses our approach and shows some aspects to further extend our research. Section 8 concludes the paper.

2. BASIC CONCEPTS

We will start with clarifying basic terms and concepts that are used within the context of this paper. As our focus lies on the scope of ubiquitous computing and cyber-physical systems, we will introduce these concepts first, as well as, our understanding of processes. Second, the paradigms of model-driven architecture and meta-modelling will be presented, as our own approach is based on these concepts.

2.1. Ubiquitous Computing

In his article "The Computer for the 21s Century", published in 1991 (Weiser, 1991), Mark Weiser introduced his vision of "the age of calm technology, when technology recedes into the background of our lives" and thereby coined the term "ubiquitous computing". Ubiquitous computing can be found at the intersection of pervasive computing, mobile computing, and ambient intelligence, and stands for systems that are unobtrusively integrated into everyday objects and activities.

2.2. Cyber-physical Systems

Cyber-physical systems (CPS) can be regarded as a major step towards Weiser's vision. CPS comprise networks of embedded, heterogeneous sensors and actuators into complex distributed systems, that are often linked to cloud-based services and cross-boundary systems. A closed loop between local sensing, remote processing, and local controlling can often be found within cyber-physical systems. Real-world objects are represented digitally and taken into consideration when planning and executing processes in a cyber-physical system. In addition, CPS are highly dynamic with respect to their components, i.e. devices and services can be added and removed at any time. By constantly collecting context information (Abowd et al., 1999), cyber-physical systems are able to adapt themselves to the current users and environment, thus evolving into so-called "smart spaces", e.g. smart homes, smart offices, and smart factories. CPS intend to create a strong link between the physical world and the cyber world, and to support their users with performing their daily tasks.

2.3. Processes

Processes (workflows) have been used to describe complex sequences of tasks and function calls in order to model the high-level behaviour of so-called systems of systems. Due to the large increase of distributed and loosely coupled systems over the last decades, the need for an additional layer describing workflows between multiple entities has been generated. With traditional approaches, it is not possible any more to implement all algorithms and cross-boundary interactions within the software application shipped with one product. The usage of processes helps with creating autonomous environment and the automation of repeating tasks.

We therefore define a process for the scope of our work as follows:

Processes represent a set of actions (process steps), which are connected with each other by a unidirectional order relation describing the order of execution of the steps (Schlegel, 2008).

2.4. Model-driven Architecture & Meta-Modelling

Using models throughout the development process of a software system incorporates several advantages with respect to modularization, reusability, extensibility, automatic code generation, and maintenance. The process layer on top of software products and systems should also be highly model-based and described by a platform independent model (PIM).

With the Meta-Object Facility (MOF), the OMG (http://www.omg.org/mof/) introduced a de facto standard for model-driven engineering (Aßmann et al., 2006)), describing several (meta-) levels of abstraction for modelling various kinds of systems. As we will also be dealing with models and meta-models throughout this paper, we want to clarify our understanding of these terms and their use within the context of process modelling at this point.

- *Process Meta-Meta-Model:* A process meta-meta-model (MOF-M2) defines the semantic and syntactic elements and structures used in the process meta-model.
- *Process Meta-Model:* A process meta-model defines all elements, types, and relations that can be used for modelling processes as well as their structural combinations. The process meta-model (MOF-M1) is an instance of the process meta-meta-model.
- *Process Model:* A process model is the abstract description of an actual process, which can be instantiated and executed at runtime. The process model (MOF-M0) is an instance of the process meta-model.
- *Process Instance:* A process instance represents a concrete process at execution time, having a runtime state. The process instance is an instance of the process model.

In the main part of this work, we will put our focus on presenting a new process meta-model, but we will also briefly describe the underlying process meta-meta-model.

3. REQUIREMENTS FOR MODELLING UBIQUITOUS PROCESSES

In order to evaluate current workflow languages with respect to their suitability for being used within ubiquitous systems (UbiSys), we will first outline some special requirements that come along with developing ubiquitous systems. Some of the following requirements are already predominant within current system architectures. However, UbiSys combine them to a large degree.

- **Dynamics:** Ubiquitous systems, as well as cyber-physical systems (CPS), are characterized as being highly dynamic with respect to the number and availability of its components, devices, and services. Therefore, modelling service invocations within processes on the instance level, i.e. the invocation of a concrete service, may not be suitable due to its possible unavailability. Hence, we also need to be able to model process steps and service calls on the type level, i.e. a certain type of service should be invoked. This way, we do not necessarily need to know at modelling time, which concrete service or device will be executing the process step.
- **Heterogeneity:** In a CPS there are usually numerous heterogeneous services and devices integrated into a so-called system of systems. However, when modelling workflows, a unified view on these components would be helpful. In addition, we would like to support a wide range of different services types and be able to easily extend this set. Complementary to the aforementioned requirement, there should also be a way of assigning an activity to a certain handling entity (resource) on the instance or the type level.
- **Complexity:** Processes within CPS can be very complex and contain a large number of process steps, both, composite and atomic, as well as further process elements. This makes means for hierarchical structuring and aggregating process components necessary in order to master high levels of complexity. A modular meta-model can also leverage exchangeability and reusability of process components.
- **Parallelism:** Numerous process instances may exist in parallel in a CPS and their execution times and cycles can also vary considerably. As processes often influence other processes indirectly, i.e. without an explicit specification of the interrelations within the process model, there should be means for supporting this way of process interaction and intercommunication available. This will leverage the integration of loosely-coupled systems and processes.
- **Evolution:** With respect to long-lasting processes, there may also be a need for changing the underlying process model during instance execution, due to a change of conditions or within the context of the process environment. It is often necessary to generate variants of models in line with new requirements or needs, e.g. within custom industrial production processes.
- **Distribution:** An important additional aspect concerns the execution of the process models. Engines for executing business processes are usually designed to be a central orchestration entity calling the services corresponding to the business process model. However, in a cyber-physical environment, e.g. a smart home, there often is no central high-performance server available. Instead, several small low-powered devices are distributed and embedded into the environment. Our future aim is to also use these resources for executing parts of a process in a distributed way.

4. RELATED WORK

During the last years, a lot of special purpose and domain specific modelling languages for processes have evolved. These languages formalize which process elements exist and how they can be composed into a workflow (cf. Meta-Process modelling).

The most well-known and de facto standard graphical notation in the domain of business processes is the Business Process Model and Notation (BPMN 2.0) (http://www.omg.org/spec/BPMN/), which has been under on-going development and extension since 2001. It includes concepts for supporting the process elements mentioned in section 3 and integrates a large variety of additional modelling entities, e.g. conversations and an extensive set of event types. BPMN descends from event-driven process chains (EPC) (Dumas et al., 2005) - another form of process modelling. EPCs are not suitable for modelling complex process structures, though.

The complexity of BPMN has made this workflow language hard to use for non-experts. Due to the large variety of language elements, processes with the same "meaning" can be modelled in a lot of different ways (Wohed et al., 2006). Many of BMPN's elements were introduced in order to fulfil requirements from modelling processes in the business domain, which usually resorts to web services and static calls on the instance level. BPMN does not support the creation of variants of processes or partial processes. As it is our goal to model autonomous workflows for more dynamic and complex heterogeneous environments, we need a more structured and flexible language, which also allows the evolution and extension of models.

The Business Process Execution Language (BPEL) (https://www.oasis-open.org/committees/wsbpel) is similar to BPMN, but it incorporates a more formal way to describe business processes resulting in a stronger execution semantic and therefore better engine support. However, most of the aforementioned drawbacks of BPMN can also be observed with BPEL (Wohed et al., 2006), which therefore does not proof to be suitable for our purpose, as well.

XPDL (http://www.wfmc.org/xpdl.html) is a workflow language intended for interchanging process definitions between different process notations, especially for serializing graphical BPMN models. It is extensible and provides strong execution semantics. However, it is based on BPMN and therefore has similar properties, which makes its application within UbiSys not feasible, too.

There also exists a large variety of further workflow modelling languages developed within an academic or an industrial context. Petri net based languages provide formal execution semantics and verification. Modelling Petri net based workflows is very complex, though, and none of the existing languages provides means for the dynamic allocation of process invocations or the creation of process variants. YAWL (van der Aalst and ter Hofstede, 2005), which is a famous Petri net based language, only supports the modelling of control flow. However, within CPS, we also need to be able to express data flow and assign resources to process steps.

In (Ranganathan and McFaddin, 2004) a workflow execution system based on BPEL is proposed, which is intended to facilitate user interaction with web services in a pervasive environment. This work is mostly concerned with automatic discovery and integration of pervasive web services. It is suitable for solving aspects of the requirements concerning the dynamic integration and deletion of components in a ubiquitous system, but does not consider further requirements in detail.

(Montagut and Molva, 2005) present a workflow management system supporting the distributed execution and dynamic assignment of resources and tasks for pervasive environments. The concept based on BPEL covers several of our requirements, but still does not allow the extension of models and the creation of process variants due to its BPEL-based nature.

When evaluating current workflow languages and management systems with respect to the aforementioned requirements, one finds that the majority of these languages fulfil the requirements listed in section 3 only partially. We therefore propose a new language for modelling workflows within ubiquitous environments.

In designing our own workflow language, we will try to adhere to the basic principles of BPMN and Petri nets. Therefore, we will still be able to support BPMN models and map our process models to higher-level Petri nets. We will also integrate concepts that have been presented within related research to solve some aspects with respect to the requirements presented in section 3.

5. MODELLING COMPLEX AND FLEXIBLE PROCESSES

5.1.1. Process Elements and Information

In order to develop a language for modelling processes within ubiquitous systems, we need to identify the most important elements necessary for formalizing workflows in these environments (Van der Aalst et al., 2003).

- *Process step:* A process step represents an activity or task to be executed.
- *Transition:* A transition represents a unidirectional connection between process steps, creating an ordered workflow.
- *Data:* A data element represents actual data of a specific type being consumed or produced by a process step.
- *Event:* An event represents a certain occurrence of a special happening and can lead to other events or trigger new processes.
- *Logic step:* A logic step is a special type of a process step containing logic for controlling the activation flow of other process steps.
- *Process:* A process contains one or more process steps, transitions, data, events, and logic steps, and can be regarded as the description of a closed sequence of actions.
- *Handling Entities:* A handling entity (resource) is responsible for performing one or more process steps. An entity can be a certain device, a service, and also a human being.

In addition, the process elements mentioned above need to have a certain set of attributes. We will detail this information later on when we describe our concept.

5.2. Meta-Meta-Model for Processes

First of all, we will present the underlying meta-meta-model for our process meta-model. We find that we only need ***Components*** and ***Relations*** as elements for describing the meta-model (Schlegel, 2008).

Components are a well-established concept, e.g. in the field of software engineering, for describing a closed entity providing a defined functionality (Szyperski, 1998). They can be accessed via their interfaces, which describe requirements for using the components and the result of their usage (pre-/postconditions). Components can be composed to larger components and also split into smaller ones up to the point of atomic components. As components provide several positive properties, we will apply this central concept on process steps and processes, which therefore represent instances of components, and use the term “process components” in the meta-meta-model.

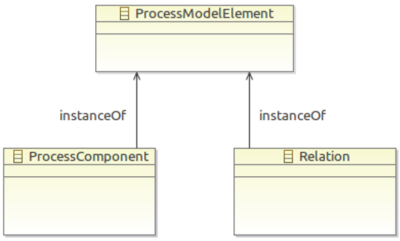


Figure 1: Process Meta-Meta-Model

Relations are used for describing the formal structure of the process meta-model. Using the Unified Modelling Language (UML) as a basis for modelling connections between components, we transfer the object-oriented concepts of inheritance, association, and composition into our meta-model. These meta-model elements therefore represent instances of relations, i.e.

connections between components. Figure 1 depicts this meta-meta-model.

5.3. Meta-Model for Processes

Based on the meta-meta-model presented in the previous section, we can now define the meta-model for processes, which is an instance of the meta-meta-model.

5.3.1. Process Components

The central element of the meta-model is the ***Process Step***, which is the basic component for modelling processes. In our object-oriented model, we differentiate between ***Composite Steps*** containing one or more process steps (depicted by the composition relation *subSteps* in Figure 2), and ***Atomic Steps***. Composite and atomic steps are seen as specializations of a process step (depicted by the inheritance relations in Figure 2). A ***Process*** is regarded as a set of one or more process steps that form a self-contained workflow. This way, processes can themselves contain processes consisting of one or more process steps (cf. composite design pattern) and at the same time, a process can be seen as one step of a super ordinate process (depicted by the *parentStep* association). This modular design leverages extensibility and reusability when modelling complex processes.

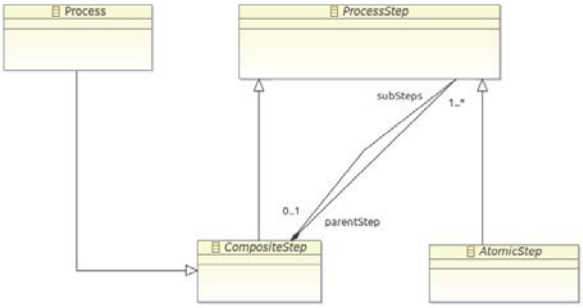


Figure 2: Process Components of the Meta-Model

5.3.2. Component Ports and Flow Relations

In order to describe processes as an ordered control flow and data flow graph of process steps, the meta-model provides transitions between process components. As depicted in Figure 3, we introduce the concept of ***Ports*** as parts of a process step. A port represents an entry or exit point for data or control flow concerning a process step.

At runtime, ports will have an activation state, which will be used to decide on the point of execution of the according process step. The process step will only be executed if all of its start ports are in an activated state.

In general, we differentiate between ***Data Ports*** and ***Control Ports***, which are both specializations of port objects (inheritance relation). Data ports are used for modelling data that are consumed by process steps at their start ports, or that are produced by process steps at their end ports. This concept can be compared to a simple function call within a common programming language specifying in-going data necessary for executing the function, and out-going data as a result of executing the function. Data ports represent data of a certain **Data Type** of a possibly external data type model (*type* association). To support the use of data elements of different types, multiple entry ports (*startDataPorts*) and out-going ports (*endDataPorts*) can be contained within a process step. Data ports will be activated after the successful execution of a process step.

Control ports are used for connecting process steps that do not require a passing of data. Similarly to data ports, diverging control flow can be modelled by using multiple *endControlPorts*. A process step can also contain multiple start ports, which may be connected to multiple preceding process steps. In the end, all data and control ports at the start of a process step need to be activated in order to start the process step execution.

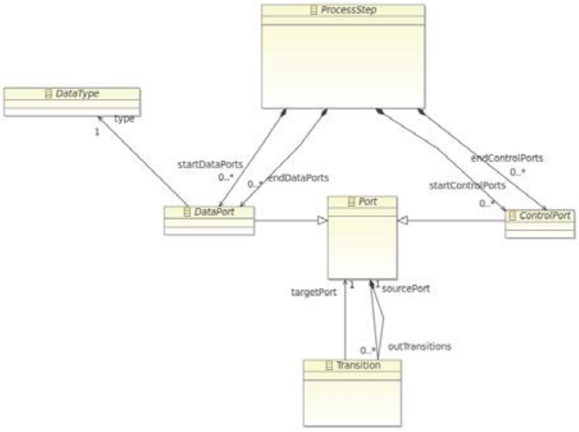


Figure 3: Component Ports and Transition Relation

Connections between process steps are modelled by using ***Transition*** objects, which can be viewed as a relation between exactly one port of a process step (*sourcePort* association) and exactly one port of another process step (*targetPort* association). A transition is defined as part of the port it originates from (composition relation *outTransitions*). This way, a port contains all of its out-going transitions. As the modelling of loops requires additional attention, we will introduce a special process component concept for loops later on. Therefore, transitions are only allowed between the ports of distinct process steps.

The connections between the set of elements of the meta-meta-model and the elements of the meta-model are presented in Figure 4.

When a process step has been executed successfully, all of its end ports become active, which also activates the transitions connected to the respective end ports. In a succeeding step, the transitions' target ports are activated.

When modelling composite process steps, there also needs to be a transition created between the start ports of parent step and its child step, as well as between their respective end ports. Transitions are only allowed between process steps on the same hierarchical level and their direct parents. This cannot be enforced structurally by the model, but it has to be formalized by using additional constraints, for example by using the Object Constraint Language for object-oriented models (http://www.omg.org/spec/OCL/). Consequentially, we eliminate dependencies between process steps that are not adjacent to each other or in a direct parent-child-relation.

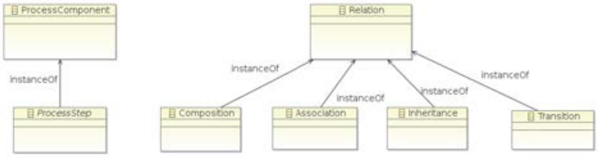


Figure 4: The Meta-Model Elements as Instances of the Meta-Meta-Model

In describing relations between process components in such a manner, we are able to model a flow of process step executions and we can leverage the encapsulation of a closed sub-workflow and its reuse in another process. Figure 5 shows an example of a process model including process steps, which again contain other encapsulated process steps.

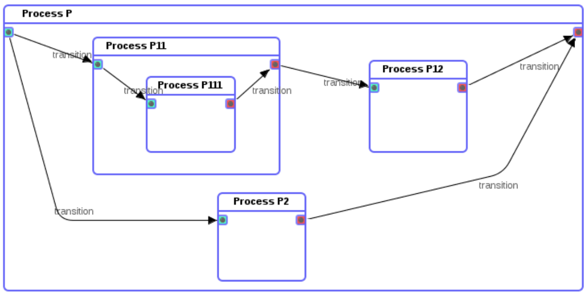


Figure 5: Exemplary Process Model with Nested Processes

5.3.3. Component Specializations

Until now, we mostly described the basic structure of our process meta-model, focussing on process components and process steps respectively, and showing how to compose them. Yet, we need more specific forms of process steps in order to have a comprehensive set of modelling elements. Thanks to the object-oriented approach of modelling the process elements, we can easily extend the previously presented concepts of atomic and composite process steps by inheritance and thereby introduce specializations of process steps. Figure 6 depicts a small set of possible extensions for data and control flow often used within other workflow languages, e.g. BPMN and BPEL.

An extension of the composite step may be used in order to represent ***Loops*** within a process. A loop could again be extended by specialized loop type, e.g. a do-while-loop, containing a loop condition and a loop counter.

Several logic elements for controlling the activation flow within a process are modelled as specializations of an atomic step. In general, a process step will be executed if all of its start ports are in an activated state. This can be seen as the logical AND connective. Other logical connectors for joining the control flow and formalizing a more special activation pattern (e.g. ***OR*** and ***XOR*** connectives) need to be modelled explicitly. In the same way, we can define conditional join operations based on data at the start port of the respective process step (e.g. ***IF***). The forking of an activation flow is modelled by creating multiple transitions from the corresponding out-going port of a process step to the eligible target process steps (see Figure 5).

We also introduced process steps for ***Data Manipulation***. The ***Data Explosion*** component analyses a complex data type and breaks it down into primitive data types. The ***Data Implosion*** step combines primitive data types into a complex type.

For calling external functionality, we added the ***Service Invoke*** component into our model. By further specializing this type of process step, we can support various kinds of service calls, e.g. to REST or SOAP based web services, or to OSGi services, via their respective services addresses or interfaces.

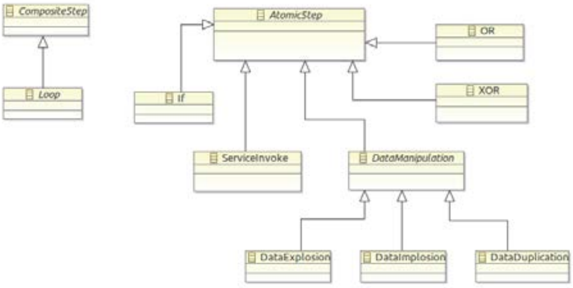


Figure 6: Possible Extensions of Process Steps

Thanks to the object-oriented modelling approach, we can easily extend and further refine the types of process steps via inheritance within the meta-model. In the same way, we can extend the (external) data type model used for defining types of data ports.

5.4. Events and Process Slots

Now we have an extensive set of elements for modelling processes. However, additional means for representing special model elements in ubiquitous processes need to be available.

We introduce ***Events*** as a special type of process step to the model (see Figure 7) to allow for the representation of loosely coupled architectures predominant in cyber-physical systems (Talcott, 2008). Events are viewed and modelled as process steps. The triggering of an event as a consequence of an action within a process is described by creating an event process step and connecting its in-going control port to the control port of the process step responsible for triggering the event. This event can have a special payload and be handled by an event processing engine (cf. complex event processing). The consumption of an event by another process step, as well as the triggering of other events or processes by the event, can be modelled accordingly. Interaction of and communication between processes is therefore event-based.

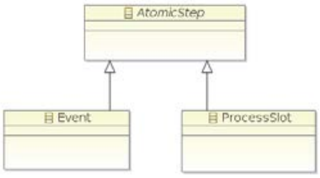


Figure 7: Event and Process Slot Extensions

To allow for the runtime usage of a process step whose implementation is not available at modelling time, we introduce a place holder for process steps called ***Process Slot***. Using this concept, the interface of a process step can be defined without providing a specific functionality. The execution engine is then able to bind the process slot to a specific process step by matching the existing process steps to the slot's ports and name at runtime.

5.5. Creating Domain-specific Processes

Thus far, we defined a domain-independent vocabulary of elements for modelling processes within ubiquitous systems. However, we also need to add semantics to the process components in order to describe the functionality of a process component, and to have more sophisticated means to model and select processes appropriately. Therefore, a *type* attribute for process components is introduced, which represents a domain-specific semantic description of the actions a process actually performs. This could be, for example, a user interaction step or a fetching step in the smart home domain. Backed up by a domain-specific object-oriented model or ontology, we can leverage the properties of this domain-knowledge and create flexible and adaptive process models. Based on the domain model, an execution engine could search from a repository of available process steps for a process step with a matching type attribute.

In using a structured domain-model for the typification of process components, we gain several advantages when choosing an appropriate process step. On the one hand, process step types can be refined via inheritance, e.g. a data input process can be specialized to a speech input, text input, and gesture input process. A fetching process can be specialized to a paper fetching process, as depicted in Figure 8.

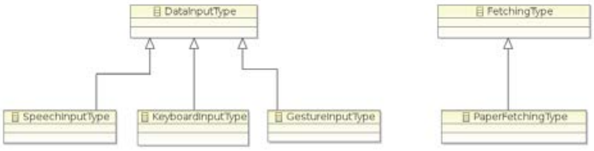


Figure 8: Component Type Refinements via Inheritance

On the other hand, we can make use of polymorphism of process steps, i.e. a process step can be of a certain type and at the same time also of its parent types, which can be continued transitively. At runtime of a process instance, the process engine could walk through the inheritance structures and search for a process step with a matching type or one of its specializations that is suitable and available for execution. For example, a process step providing data input may be required but not specified any further and therefore a speech input process step is used, as it is also of type data input and therefore has the same general properties.

A more sophisticated method of creating and using a domain-specific model would be to use semantic technologies. This would result in more advanced mechanisms for process modelling and selection by using verification and deduction based on logic.

Regardless of which method for modelling the domain-knowledge is applied, creating a comprehensive and structured model for describing the application domain is an important requirement for achieving flexibility in process execution.

5.6. Component Attributes and Roles

In order to meet the requirements described in Section 3 and to complete the meta-model, we introduced a set of further attributes for the process components.

Besides attributes for naming and identifying components on the model and instance level, an optional role-based *handling entity* for a process step can be defined (Montagut and Molva, 2005), (List and Korherr, 2006). This corresponds to the swim lane and pool concepts of BPMN. We can define an entity (resource), again on the instance or type level, that is responsible for executing the respective process step. By using roles of an underlying model for this allocation of process step handler, we are able to orchestrate multiple devices and classes of devices, and also achieve a basic form of access control (Sandhu, 1998). This concept also supports our future goal of distributing process execution across multiple devices in ubiquitous systems.

In addition to the aforementioned attributes considering properties at modelling time, we also need to have component attributes with respect to runtime properties. These include, amongst others, activation states for ports and transitions, as well as an execution state for process steps.

6. MODELLING ENVIRONMENT AND AUTHORING TOOLS

6.1. Technical Realization

The implementation of the introduced process meta-model is based on the Eclipse Modeling Framework (EMF) (http://www.eclipse.org/modeling/emf/), which provides an extensive set of applications and tools for modelling and creating domain-specific languages. This open source framework is based on Java and supports mechanisms for automated source code generation, model verification, and persisting model information with the help of the XMI (XML Metadata Interchange) format. Thanks to its object-oriented design, we can map our models and concepts directly to objects that can be used for process execution by a corresponding process engine.

6.2. Process Authoring

Apart from implementing the meta-model, we have started developing a toolchain for supporting the computer-based authoring of processes.

6.2.1. Process Editor

Using built-in tools of EMF, table-based editors for Ecore models can be generated automatically. Unfortunately, the complexity and low lucidity of these editors requires the user to have in-depth knowledge of the underlying model. To improve usability in terms of consistency, conciseness, and comprehensibility, we have developed a graphical process editor based on the Graphiti tooling infrastructure for EMF (http://www.eclipse.org/graphiti/).

Figure 9 displays a screenshot of the process editor, which can be divided into three areas. (1) shows the main drawing area for the process model, (2) shows the set of modelling elements available, and (3) shows the components' attributes.

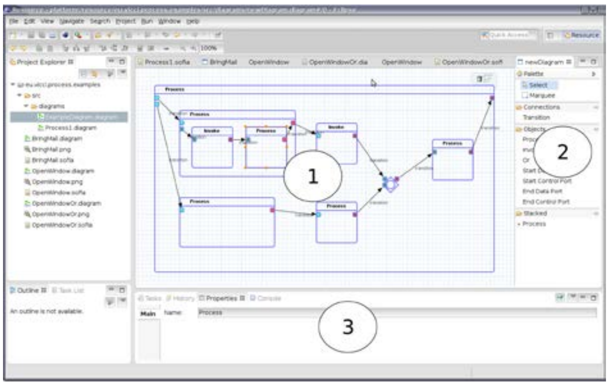


Figure 9: Eclipse-based Process Model Editor

At some points, however, enforcing additional rules for dealing with exceptional combinations of components and relations is necessary during modelling. Formalizing these restrictions inside the meta-model would usually lead to a large increase of its complexity. Constraints that cannot be applied structurally by the model are defined separately using the EMF Validation Framework. After creating a process model using drag and drop from the element list to the main drawing area, a check of the model's validity according to the meta-model and to the separate external constraints is performed.

The result of creating a model is an XML-based representation of the process model including graphical information for visualization of the process model inside the editor and additional process monitoring tools.

6.2.2. Process Repository

In order to model and execute domain-specific processes, we are currently planning on developing a repository for processes and process steps that can be accessed by the editor and the execution engine.

Thanks to the model-based design and modularity, we will be able to use the graphical process editor for the initial creation of processes and process steps which can be submitted to the repository, and to further extend and adapt the processes inside the repository.

At this point, we will also be using a semantic description for processes, their ports, and their domain-specific types to have additional means for checking compatibility of process steps, recommending suitable process steps, and verifying a modelled process. To do so, we are able to draw upon an extensive set of methods from the field of semantic web technologies.

6.3. Process Execution

We have also started implementing a process engine for executing instances of process models based on the meta-model presented before.

The XML-based description of a process generated by the process editor is loaded and validated by the process engine. Afterwards, the engine creates a process instance and walks through the objects defined in the respective process model, calling the methods implemented for handling the specific type of process element.

In order to represent and persist the runtime state of a process, we extended our meta-model with runtime information. Consequently, we also have an extended version of our meta-model for representing the state of process instances (Lehmann et al., 2010).

Process instances are currently executed on a central orchestration server supporting the invocation of web services and OSGi services via remote procedure calls. However, as part of our future work, we will be able to distribute the execution of process steps and complex processes across several servers based on a peer-super-peer network infrastructure (Schlegel, 2009).

7. DISCUSSION

Our aim in designing a process modelling language was to be able to cope with new requirements that come along with the emerging new form of complex systems of systems, called cyber-physical systems.

Despite the complexity of CPS, we tried to adhere to simple structures with respect to the meta-models. Using components as basis for describing process steps, we are able to have modular entities representing one process step, which can be combined into larger, complex process building blocks and reused for modelling. These hierarchical structures help with modelling and visualizing complex processes.

Principles of object-orientation help us with defining connections and relations between process steps on the syntactical level, but also on the semantic level. Based on this structured domain-specific model, the meaning process components, as well as their relations with each other, can be described and used for further semantic processing. However, we have to investigate the feasibility of using semantic technologies, as processes can contain time-critical components and performance may be an issue with their execution.

Due to the object-oriented approach, we are able to extend our process meta-model very easily without changing the model's core structures. The creation of specializations and variants of processes is also possible via object-orientation (Schlegel, 2009). However, the domain model for describing process components needs to be developed and available before modelling processes.

By using the concept of process slots and defining handling entities, as well as, process steps on the type level at modelling time, we achieve a high level of flexibility when executing process instances. Thanks to the model-based description and to polymorphism, the execution engine can search for suitable process steps and handlers at runtime, walking through the inheritance tree. In case a process handler of a certain type is unavailable or a certain process step cannot be executed, the engine could find a matching replacement within one of its specializations or generalizations. In order to make use of these ad hoc replacement mechanisms, we also need a model for describing process handling entities and to define their capabilities, as well as, the matching requirements for the execution of the respective process steps. A basic form of this validation can already be achieved by checking the process steps interface, i.e. its type, its ports, and its name, with respect to compatibility.

The introduction of events for intercommunication and interaction of processes leverages the integration of loosely-coupled systems and supports flexibility in process execution, as process components can trigger other processes without an explicit representation of this relation in the process model. However, we need additional rules for describing correlations among events and between events and process steps, which have to be handled by an auxiliary event processing engine.

Due to the process component composition, we are able to regard every process component as a self-contained process, which can be executed on a set of distributed process engines. Though, in order to achieve this distribution of process execution, there need to be communication and synchronization mechanisms among the process engines.

We based our design on the core concepts and elements of common workflow languages, e.g. BPMN, and therefore are able to map processes created with similar workflow languages to our model and execution engine. Furthermore, our workflow language facilitates the modelling and execution of more complex and dynamic processes within heterogeneous environments, achieving a high level of autonomy of processes.

However, there are several additional models and rules necessary in order to describe all aspects regarding process types, process execution, and process handlers.

Evaluating our concepts with respect to the requirements presented in section 3, we find that we can meet all the requirements that were introduced as being novel with respect to ubiquitous systems. Related research within an academic and industrial context is currently able to satisfy only a subset of the requirements, as the workflow languages are often too static and lack expressiveness, as well as, flexibility.

In order to evaluate our approach and to show its feasibility, we implemented the process meta-model, as well as, started to implement an execution engine and a graphical editor for process models as first elements of a our toolchain for ubiquitous processes. We will extend and improve our tools and models in the near future.

8. CONCLUSION & FUTURE WORK

In this paper, we presented a new meta-model for formalizing workflows within cyber-physical environments developed from a software engineering perspective. State-of-the-art workflow languages often only support parts of the new requirements introduced by cyber-phyiscal systems. Therefore, we developed a new meta-model for processes, which is mainly based on concepts of object-orientation and of component-based systems. We adhered to the paradigm of model-driven architectures, which yielded several benefits with respect to modularity, reusability, and extensibility of process components. By adding domain-specific descriptions to process components and using semantic models, we achieve a high flexibility when executing processes via a dynamic allocation of process handlers. Thus, workflows become more intelligent and autonomous.

However, there are still several open issues that need to be resolved in order to develop an extensive process toolchain for current and future cyber-physical environments. We believe that with our process meta-model, we laid the foundation for smart autonomous workflows within complex heterogeneous environments.

Our future work will include the development of a process component repository and a semantic domain-model for the classification of process components in the area of smart homes. We will also model capabilities of process execution entities and requirements necessary for executing process steps. In doing so, the process engine will be able to select appropriate process steps at runtime. One step further, we will investigate the usage of agent-based technology in order to find appropriate process steps more intelligently during execution.

We will also investigate the mapping of our process meta-model to concepts used within Petri nets. The advantages of formal verification may prove helpful when constructing and analysing safety critical workflows as they may be required within cyber-physical systems. Hierarchical Petri nets may be suitable for our meta-model to be mapped to.

In order to better adapt workflows to the current situation and environment, we are going to use context information collected by the sensors within the ubiquitous systems and thereby make the processes context-aware and more intelligent, (Wieland et al., 2008). These adaptations can be directly incorporated into the process models (Yongyun et al., 2007).

The decentralized execution of workflows will also be one of our main focuses with respect to further research activities (Hens et al., 2010). Distributing workflows across several orchestration peers may increase the availability of the workflow system and make the workflows more resilient against failures and outages (Friese et al., 2005).

ACKNOWLEDGMENTS

This research has been partially funded within the VICCI project under the grant number 100098171 by the European Social Fund (ESF) and the German Federal State of Saxony.

REFERENCES

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., Steggles, P., 1999. Towards a Better Understanding of Context and Context-Awareness. *HUC '99 Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, pp. 304-307. London: Springer-Verlag UK.

Aßmann, U., Zschaler, S., Wagner, G., 2006. Ontologies, Meta-models, and the Model-Driven Paradigm. *Ontologies for Software Engineering and Software Technology*, pp. 249-273. Heidelberg: Springer-Verlag Berlin.

Scheer, A.-W., Thomas, O., Adam, O., 2005. Process Modeling using Event-Driven Process Chains. In: Dumas, M., van der Aalst, W. M. P., ter Hofstede, A. H. M., eds. *Process-Aware Information Systems: Bridging People and Software through Process Technology*. Hoboken, NJ: John Wiley & Sons.

Friese T., Müller, J. P., Freisleben, B., 2005. Self-healing Execution of Business Processes Based on a Peer-to-Peer Service Architecture. In: Beigel, M., Lukowicz, P., eds. *Systems Aspects in Organic and Pervasive Computing, Lecture Notes in Computer Science*. Heidelberg: Springer Berlin Heidelberg, pp. 108-123.

Hens, P., Snoeck, M., De Backer, M., Poels, G., 2010. Transforming standard process models to decentralized autonomous entities. *5th SIKS/BENAIS Conference on Enterprise Information Systems (EIS 2010)*, pp. 97–106. November 16, Eindhoven, The Netherlands.

Lehmann, G., Blumendorf, M., Trollmann, F., Albayrak, S., 2010. Meta-modeling runtime models. *MODELS'10 Proceedings of the 2010 international conference on Models in software engineering*, pp. 209–223. October 2-8, Oslo, Norway.

List, B., Korherr, B., 2006. An evaluation of conceptual business process modelling languages. *SAC '06 Proceedings of the 2006 ACM symposium on Applied computing*, pp. 1532-1539. April 23-27, Dijon, France.

Montagut, F., Molva, R., 2005. Enabling pervasive execution of workflows. *International Conference on Collaborative Computing: Networking, Applications and Worksharing*. December 19-21, San Jose, USA.

Ranganathan, A., McFaddin, S., 2004. Using workflows to coordinate Web services in pervasive computing environments. *Proceedings IEEE International Conference on Web Services, 2004*, pp. 288-295. July 6-9, San Diego, USA.

Sandhu, R. S., 1998. Role-based Access Control. In: Zelkowitz M. V., eds. *Advances in Computers*. Elsevier, 237-286.

Schlegel, T., 2008. *Laufzeit-Modellierung objektorientierter interaktiver Prozesse in der Produktion*. PhD Thesis (in German). Universität Stuttgart.

Schlegel, T., 2009. Object-Oriented Interactive Processes in Decentralized Production Systems. *Human Interface and the Management of Information. Designing Information Environments, Lecture Notes in Computer Science* . Vol. 5617: pp. 296–305.

Szyperski, C., 1998. *Component Software - Beyond Object-Oriented Programming*. Addison-Wesley / ACM Press.

Talcott, C., 2008. Cyber-Physical Systems and Events. *Software-Intensive Systems and New Computing Paradigms*, pp. 101-115. Heidelberg: Springer-Verlag Berlin.

Van der Aalst, W. M. P., Ter Hofstede, A. H. M., Kiepuszewski, B., Barros, A. P., 2003. Workflow Patterns. *Distributed and Parallel Databases* Vol. 14: pp. 5-51.

Van der Aalst, W. M. P., Ter Hofstede, A. H. M., 2005. YAWL: yet another workflow language. *Information Systems* Vol. 30: pp. 245-275.

Weiser, M., 1991. The computer for the 21st century. *Scientific American*, Vol. 265: pp. 94-104.

Wieland, M., Kaczmarczyk, P., Nicklas, D., 2008. Context Integration for Smart Workflows. *Percom 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, pp. 239-242. March 17-21, Hong Kong.

Wohed, P., van der Aalst, W. M. P., Dumas, M., ter Hofstede, A. H. M., Russel, N. 2006. On the Suitability of BPMN for Business Process Modelling. *Lecture Notes in Computer Science* Vol. 4102/2006: pp. 161-176.

Yongyun, C., Kyoungho, S., Jongsun, C., Jaeyoung, C., 2007. A Context-Adaptive Workflow Language for Ubiquitous Computing Environments. Computational Science and Its Applications – ICCSA 2007, *Lecture Notes in Computer Science* Vol. 4706: pp. 829-838.

URBAN TRANSPORT INFRAESTRUCTURE: A STATE OF THE ART

Idalia Flores[a], Ioannis Chatziioannou[b]Esther Segura[c]Salvador Hernández[d]

[a] Faculty of Engineering, UNAM
[b] Faculty of Engineering, UNAM
[c]Institute of Engineering, UNAM
[d]Technological Institute of Celaya

[a]idalia@.unam.mx, [b]j.xatzhiwannou@yahoo.gr.
[c] esegurap@iingen.unam.mx.[d]salvador.hernandez@itcelaya.edu.mx

ABSTRACT

The urban transport infrastructure is one of the most important problems for the cities, and involves many aspects that concern to citizens, governments and the economical growth of the countries. The objective of this paper is to show how this issue has been studied in recent years, with emphasis in the new technologies, the use of simulation and optimization at the whole planning process. Some study cases are shown in order to clarify the concepts presented.

Key words: Urban transport, planning, simulation, geotechnologies

1. INTRODUCTION

Nowadays one of the bigger problems in cities is the transportation system and its infrastructure. There have been lots of studies and research in recent decades trying to find solutions. In general there is an economic impact when countries make an investment in this sector. Most of the studies on transport infrastructure, in particular, focus on its impact on growth. In the past two decades the analytical literature has grown enormously, with studies carried out using different approaches, such as a production function (or cost) and growth regressions, as well as different variants of these models (using different data, methods and methodologies), the majority of these studies found that transportation infrastructure has a positive effect on output, productivity or growth rate Calderon & Serven (2008). One of the pioneers was Aschauer (1991) who, in his empirical study, provided substantial evidence that transport is an important determinant of economic performance. Another example is the study of Alminas, Vasiliauskas and Jakubauskas (2009), who found that transport has contributed to growth in the Baltic region. Another study on the Spanish plan to extend roads and railways that connect Spain with other countries concludes that these have a positive impact in terms of Gross Domestic Product (GDP) Alvarez-Herranz & Martínez-Ruiz (2012). In a study of the railroad in the United States, it is mentioned that many economists believe that the project costs exceed the benefits Balaker (2006). However, the traditional model of cost-benefit assessment does not include the impact of development projects De Rus (2008). In these studies focused on growth, we see there is a bias towards economic rather than social goals. That is why it is important to emphasize the impact of transport infrastructure on development and not just growth.

In order to show the subject clearly, we will use a systems approach, dividing urban transport infrastructure according to The City of Calgary (2009)
Urban Transport infrastructure:

- Transportation Planning
- Transportation Optimization
- Intelligent Transportation Systems

According to this system paradigm, this paper is focused on the description of the research made in the last five years, mainly considering optimization, simulation and the intelligent systems. The structure of this paper is as follows. Section 2 shows the state of the art for the general transportation planning issue. Section 3 is devoted to transportation optimization techniques that have been used in different ways in accordance with the problems they are meant to solve. Section 4 is about intelligent transportation systems and how the development of new technologies interacts with the whole system and where they are being used. Section 5 is about some study cases. These cases are important because of the new technologies used and their successes and failures. Section 6 gives some conclusions and future research on this approach.

The impact that transport infrastructure has in increase of the quality of life can be seen in the next figure:

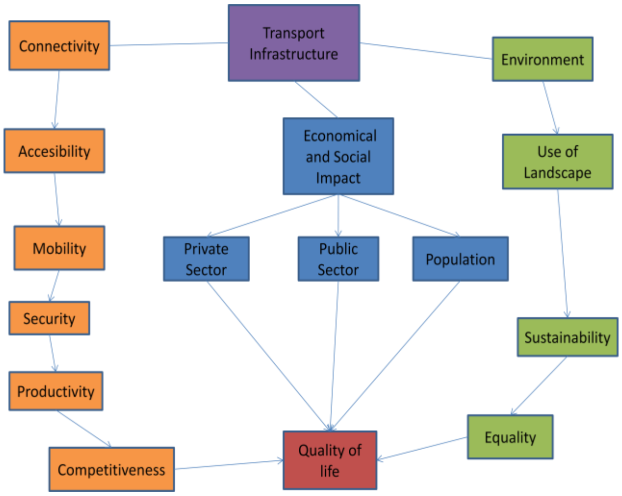


Figure 1 Impact of the transport infrastructure.

2. TRANSPORTATION PLANNING

Transportation planning covers a lot of different aspects and is an essential part of the system. According with Levy (2011), "Most regional transport planners employ what is called the rational model of planning. The model views planning as a logical and technical process that uses the analysis of quantitative data to decide how to best invest resources in new and existing transport infrastructure."

Phases for transportation planning

There are three phases: The first, preanalysis, considers what problems and issues the region faces and what goals and objectives it can set to help address those issues. The second phase is technical analysis. The process basically involves the development of the models that are going to be used later. The post-analysis phase involves plan evaluation, program, implementation and monitoring of the results, Johnston (2004).

Transportation planning involves the following steps:

- Monitoring existing conditions;
- Forecasting future population and employment growth, including assessing projected land uses in the region and identifying major growth corridors;
- Identifying current and projected future transportation problems and needs and analyzing, through detailed planning studies, various transportation improvement strategies to address those needs;
- Developing long-range plans and short-range programs of alternative capital improvement and operational strategies for moving people and goods;
- Estimating the impact of recommended future improvements to the transportation system on environmental issues, including air quality; and
- Developing a financial plan for securing sufficient revenues to cover the costs of implementing strategies.

Transportation planning process. Figure 2 shows the process briefly and clearly.

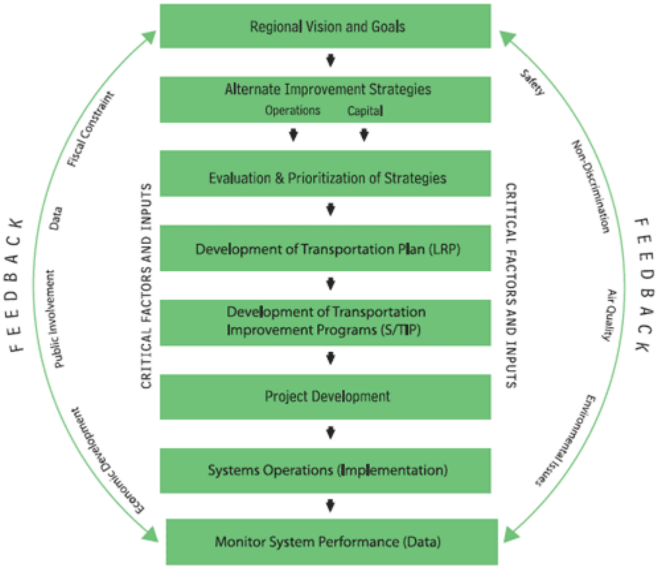

Regional Vision and Goals
Alternate Improvement Strategies
Operations
Capital
Evaluation & Prioritization of Strategies
Development of Transportation Plan (LRP)
Development of Transportation Improvement Programs (S/TIP)
Project Development
Systems Operations (Implementation)
Monitor System Performance (Data)
FEEDBACK
FEEDBACK
CRITICAL FACTORS AND INPUTS
CRITICAL FACTORS AND INPUTS
Fiscal Constraint
Data
Public Involvement
Economic Development
Safety
Non-Discrimination
Air Quality
Environmental Issues

Figure 2 Transportation planning process. Source (FHWA, 2007)

Urban Infrastructure

Urban infrastructure, a human creation, is designed and directed by architects, civil engineers, urban planners among others. These professionals design, develop and implement projects (involved with the structural organization of cities and companies) for the proper operation of important sectors of society. When governments are responsible for construction, maintenance, operation and costs, the term "urban infrastructure" is a synonym for public works.

Road infrastructure is the set of facilities and equipment used for roads, including road networks, parking spaces, traffic lights, stop signs laybys, drainage systems, bridges and sidewalks.

Urban infrastructure includes transportation infrastructure, which can, in turn, be divided into three categories: land, sea, and air, that can found in the following modalities:

• Street networks, high or low-speed railway lines, together with such as bus stops, road signs, traffic lights, parking bays, and so forth. This applies to all the cases cited below:

• Infrastructure for mass transit or bike paths and footpaths

• Canals, bridges

• Ports, airports and lighthouses, etc.

Figure 3 shows a systemic approach to the relationships between transport infrastructures, considering the main elements that require analysis.

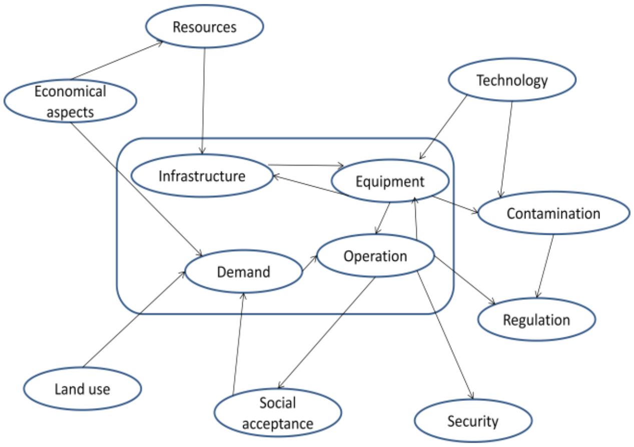


Figure 3 A systemic approach to transport infrastructure

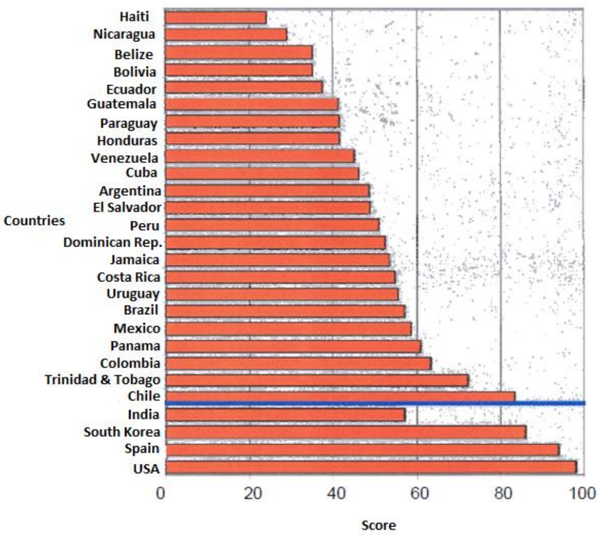
3. TRANSPORTATION OPTIMIZATION

The goal of transportation optimization is to identify, evaluate and plan enhancements that optimize the operation of a transportation system. With this in mind many countries have specific policies for this and a lot of research has been developed over recent years to this end. Optimization deals mainly with the maintenance costs and management of the infrastructure that requires a balance between the performance of the structure and the total cost accrued over the entire life-cycle. There are a series of new technologies that, using GIS, GPS, sensors and cameras, have been used in visual inspections.

The proper and efficient management of urban transport infrastructure includes many technological, political, and social aspects. So it is necessary to use an interdisciplinary approach such as geotechnology, with which digital technologies can be integrated for a spatial analysis of reality.

The maintenance of urban infrastructure consists of a series of actions that require knowledge and experience about the needs of different types of infrastructure (bus stops, signage, benches etc.) to be done in the optimum manner. To achieve this, infrastructure can be changed, expanded and/or replaced in an efficient manner in order to meet the needs of the users of a city.

Urban transport infrastructure has a direct impact on people's daily lives, which can, in a positive or negative way (depending on its condition), affect the competitiveness of people in general and the country at large (depending on the competitiveness of its urban infrastructure on a global level), as can be seen in the following figure 4:



Source: Division of Sustainable Development and Human Settlements (2007) with data of American Economics, January 2007 according to the http://www.cg-la.com database.

Figure 4 Worldwide competitiveness of urban infrastructure

There are many factors that lead to the growth of urban transportation, but we must not forget other important factors such as rural development, use of the countryside or urban development.

Mexico, in particular, has changed from a predominantly rural country to a being mainly urban culture, so the term urban development has a very important role to play in Mexico's sustainability, where urban development involves the growth and quality of new housing, bringing greater wellbeing, as a result of urban expansion, planning and access to credit for housing.

That is why forecasting and transportation data are two other important topics in this section, considering forecasting as an important tool for designing, building, operating, and maintaining models for forecasting the demand for transport. These models are built using optimization algorithms as well as simulation software.

Following with the systemic approach, the transport issue can be seen as a transportation network, and in this way the relationships among the main elements to analyze and study, figure 5 shows these ideas.

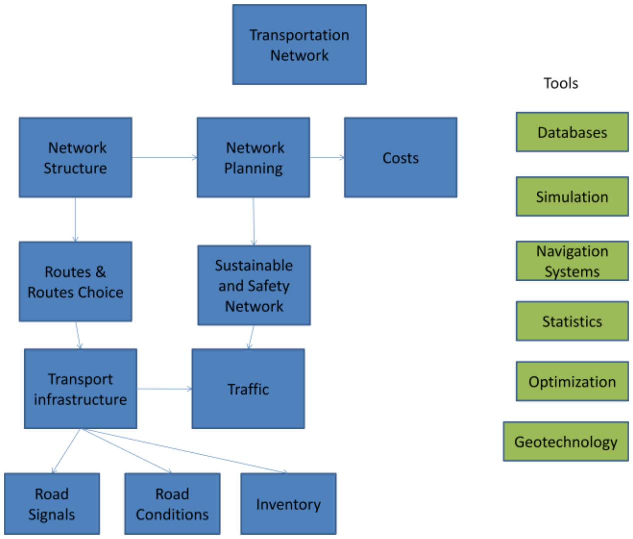


Figure 5 Transportation network

Some of the tools used to diagnose, analyze and solve problems relating to urban transport infrastructure are shown in table 1

Table 1 Methodologies used for urban transport problems

	Methodologies				
Issues	Simulation	Optimization	Statistics	Geotechnologies	Databases
Crashes	✓		✓		✓
Traffic for a future network	✓		✓	✓	✓
Traffic for a modified network	✓		✓	✓	✓
Routes design	✓	✓		✓	✓
Routes selection	✓	✓		✓	✓
Regional accessibility	✓			✓	✓
Level of service	✓	✓	✓	✓	✓
Cost	✓	✓			✓
Supply and demand	✓	✓		✓	✓
Physical condition of the streets, avenues and roads	✓		✓	✓	✓

As can be seen in table 1, simulation is one of several methodologies that can be used for a transportation network, when we want to plan and optimize transportation problems in the all aspects that are open to being solved. There are general-purpose discrete simulation software packages such as SIMIO, Promodel or Flexim, as well as other more specific packages, such as S Paramics, or Simleader.

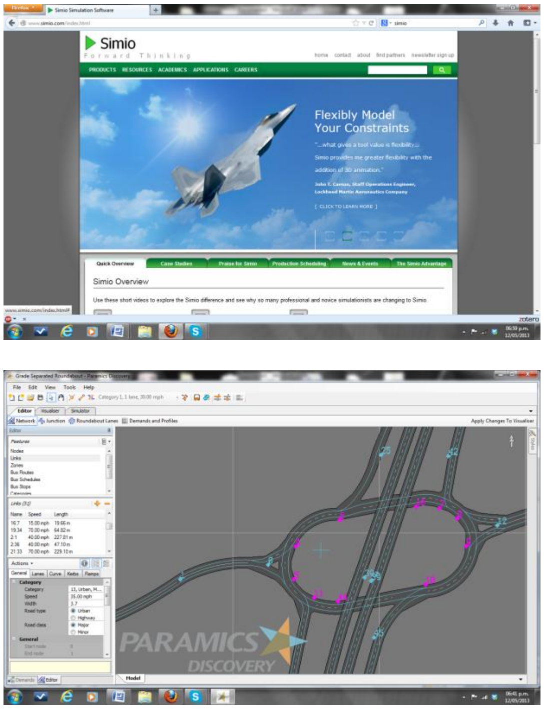

Simio
Forward Thinking
PRODUCTS RESOURCES ACADEMICS APPLICATIONS CAREERS
Flexibly Model Your Constraints
Simio Overview
PARAMICS DISCOVERY
File Edit View Tools Help
Editor Visualizer Simulator

Figure 6 SIMIO and S-Paramics simulation software

However, there is also the need to optimize an entire transportation network. There are two main methods for this, the exact and heuristics, though a hybrid that that combines them both can also be used. In most cases, though, some heuristic algorithms are used because of the size of the problems involved. This subject, according to the vast amount of literature Laporte (2007) Laporte (2009), Toth (2002), Daskin (2008) focuses particularly on optimization from the perspective of logistics and distribution, and most especially on route optimization.
This is largely because distribution is one of the functions that has evolved the most over the last few years in organizations. And this evolution has inevitably resulted in increasingly complex transport and distribution operations that, combined with factors such as the need to lower production costs, constantly rising transport costs or the increasing demands in customer-supplier relations, have made logistics management as a key element of companies' strategies.

In this scenario, the capacity of companies for optimizing their transport and distribution routes appears as a key element of logistics management; however, not all companies approach this problem in a suitable and systematic way.

Therefore companies are interested in route optimization, which, in general, can be understood to mean all those actions that contribute to improving the distribution function, either in terms of level of service, the improvement of quality, lowering costs, etc.

One of the tools that have been proposed for the optimization is the use of the Design of Distribution Networks to include a set of highly combinatory problems, Chekuri (2007), Vishv (2009) and Ambrosino(2009).

Related problems are the inventory distribution problem as well as the vehicle routing problem Cohelo et al (2012) that, in turn, has as a precedent the vendor management inventory problems and the traveling salesman problem. These problems respond to optimization decisions such as: inventory, location and routing. Each one of these problems has its own origin and solution methodologies, all of great complexity, that enable the optimization in matters to do with transport, particularly route optimization. Laporte (2007), Laporte (2009), Toth (2002) and Wu (2002).

The key to approaching a route optimization problem lies in understanding that you deal with it depending on the particularities of each organization and, as such, there are no global solutions capable of solving all the existing distribution models.

The objective of the optimization must be clearly defined: i.e., the scope of the problem you want to solve must be clearly defined as must be the variables that are most critical when measuring the success of the optimization (level of service, cost, etc.)

Clearly demarcating the current service in terms of the characteristics of the product, characteristics of the routes and characteristics of the organization (processes and means it has)

The type of result desired for the project has to be established. That is to say whether one is looking for a system that makes it possible to control numerous routes even at the cost of losing flexibility or, on the contrary, if one wants a more flexible system with more limited scope.

Once all these questions have been analyzed, one is then in a position to tackle the project; its scope and the complexity shall determine how this is done.
Logistics in general and transport in particular has progressively undergone a transformation over the last few years that is directly related to a massive growth in trade that has meant the traders in the supply chain to need to be constantly adapting. This transformation is based on two major trends:

- The growing integration of logistics chains**.**
- Growing attention to **intermodality** and **multimodality** in the distribution chain.

As it was mentioned before according to the solution methods some routing studies have been developed using either genetic algorithm hybridized with Dijkstra algorithms to find the shortest routes, or just some advanced label algorithms as the one shown in Klunder, (2006).

As has already been mentioned, metaheuristics are used because they provide very good solutions in a short time, like the neural networks that are used by Yu et al (2011)
The exact methods we are referring to include branch and bound, branch and cut, dynamic programming. The location and routing problem presented by Belenguer et al (2011) uses branch and cut for the design of logistic networks. In this case the overall distribution cost may be excessive if routing decisions are ignored when locating depots. In order to overcome this problem they propose a branch and cut algorithm for solving it. The proposed method is based on a zero-one linear model reinforced by valid inequalities.
Berman et al (2011) gives us an example of search paths for service. In this paper an optimal search path is found for a service problem that is stated as follows: “A customer residing at a node of a network needs to obtain service from one of the facilities; facility locations are known and fixed. Facilities may become inoperational with certain probability; the state of the facility only becomes known when the facility is visited. Customer travel stops when the first operational facility is found. The objective is to minimize the expected total travel distance”. This problem is NP-hard and a forward dynamic programming procedure is developed.
Communications technologies and IT have been successfully used for years in this scenario, thus permitting the development of freight transport management. However, its growing development under the global umbrella of ITS (Intelligent Transport Systems) has made it possible to more efficiently mold transport operations that, in intermodality environments, are getting ever more complex to manage.
The most significant of these different technologies are:

- Geographic Information Systems.
- Geolocation Systems (GPS, for example)
- Computer applications capable of calculating mathematical route optimization models based on a series of intrinsic constraints on the logistic process (fleet availability, geographic location of the distribution and delivery points, loading, reception and delivery time slots, variable distribution cost, etc).

Nowadays these technologies tend to form part of global solutions that have given rise to a multitude of IT programs offering companies the possibility of managing their transport operations more efficiently and effectively.

In this sense, the IT programs include simulation programs with built-in optimization for the construction of scenarios that, apart from solving the problems already mentioned, are used for aspects such as:

- Measuring the traffic on main streets and how it can be solved by constructing scenarios that make it possible to look for alternatives to lighten it at certain peak hours.
- Measuring traffic when streets have to be closed for repairs, in which case scenarios are also constructed to assess alternatives and alternate routes
- Assessing routes in transport networks to find the safest routes with less traffic.
- Distributing transport in critical routes, taking into account safety, sustainability and efficiency.

4. INTELLIGENT TRANSPORTATION SYSTEMS.

The issue of the proper and efficient administration of urban transport infrastructure contains many technological, political, and social aspects. So it is necessary to use an interdisciplinary approach such as geotechnology, whereby digital technologies can be integrated for a spatial analysis of reality.

Geotechnology, in other words, is presented as a new vision of geographic space that enhances the field of computer systems using cybernetic human and electronic systems for the analysis of physical and social Buzai (2012) and its scope is ever expanding geoBlog (2007).

Some geotechnological tools are:

- Geographic Information systems (GIS).
- Global Positioning System (GPS).
- Aerial Photos.

GIS integrates hardware, software and data for capturing, managing, analyzing and displaying geographically referenced information. GIS allows us to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns and trends in the form of maps, globes, reports, and charts. A GIS helps us to answer questions and solve problems by looking at all the available data in a way that is quickly understood and easily shared. Some of the top 5 benefits that GIS has to offer are the following:

- Cost saving and increased efficiency
- Better decision making
- Improved communication
- Better record keeping
- Geographical management

A GIS system can help its users, by mapping out where things are, allowing them to find places with the particular features they are looking for and letting them see patterns. GIS also allows users to map quantities and by mapping quantities people can find places that meet their criteria and take the necessary actions. Public health officials might want to map the numbers of physicians per 1,000 people in each census tract to identify which areas are adequately served, and which are not. In the case of map quantities, the user can, with a density map, measure the number of features using a uniform areal unit so the distribution can be clearly seen. This is especially useful when mapping areas, such as census tracts or counties that vary greatly in size. On maps showing the number of people per census tract, the larger tracts might have more people than smaller ones. But some smaller tracts might have more people per square mile—a higher density. To find what is inside this way, a person can use GIS to monitor what is happening and take specific action by mapping what is inside a specific area and then, finally map the change in an area to anticipate future conditions, decide on a course of action, or to evaluate the results of an action or policy. By mapping where and how things move over a period of time, you can gain insight into how they behave.

Nowadays we are living in an era characterized by technological advances, mobile devices are much stronger, more efficient and capable than they used to be and for this reason a new type of commerce has been created, called Mobile Commerce, where people can make transactions through their mobile device. A subcategory of Mobile Commerce is Location Based Commerce whereby a mobile device can inform its user through a GPS system certain information that can make the user’s life easier; for example a user can be informed whether he is near a gas station, hospital or restaurant. Thus we can see that geography and the technologies associated with it are connected with humans to such a degree that they can help us in our daily round. Location based m-commerce, according to Turban et al. (2008), can be divided into the following 5 categories:

i. **Location**: the service that can determine the place of a person.
ii. **Mapping**: the service relating to the creation of maps for specific locations.
iii. **Tracking:** the surveillance of a person through his/her route.
iv. **Navigation:** the creation of the ideal route between two locations.
v. **Timing:** the calculation of the time that a vehicle needs to cover a specific route.

One system that is well known nowadays is the GPS (Global Positioning System). This was first used by the United States Department of Defense but its applicability to civilian life was recognized from the first moment. Its working principle, according to Turban et al. (2008), is based on satellites (originally 24 satellites). Each satellite’s position is identified by a signal sent by the satellite from its highly sensitive and accurate onboard clock. According to Pike (2009), GPS consists of three segments: the space segment, the control segment and the user segment. The space segment consists of the satellites. The second GPS segment consists of the equipment on land (antennas, monitoring stations etc). The third segment of GPS consists of the receivers the users need in order to know their location. GPS receivers can be stand-alone devices or built into mobile devices. The receivers receive information (position) from the satellites in terms of latitude and longitude but use the GIS (Geographical Information System) software to change this information into a form that the average user can recognize (addresses), according to Turban et al. (2008). GIS is an electronic system that processes location-related data. Thus GPS/GIS can provide drivers with valuable information about how to get to their destination, such as how to take the shortest route. Moreover the head office of a company will be able, at any one time, to know the position of each of its vehicles. This provides transport companies with a higher level of security as they can immediately inform the police if a vehicle changes leaves its designated route. The connection to the GPS is something simple nowadays, because everybody with mobile device will be able to connect and to depict the geographical form of the route that he is going to follow and the vehicle on it, moreover the device will be able to provide the user with several statistical elements. Every request the user makes (for example, the shortest alternative route because the ideal route is closed for some reason) is received by the company’s server, which will give the user the best solution.

Another type of geotechnology is the use of aerial photos (Orthophotography). These photos were made for the first time in 1960 Smith (1995). The technology of the early 1970s brought to this data source a good commercial application and its use began to expand. This technique has the advantages of a conventional map, but in contrast to this, is able to display the up-to-date details of the land, rather than a cartographic representation. There is software (GIS) that allows you to display the aerial photo and lets the user process it with annotations or geographic symbols such as schools, hospitals, police departments and stations, gas stations, etc. An example of an aerial photo can be seen in the following figure.

Figure 7 Example of aerial photo.

In today's world, the role of official agencies and departments of transport has evolved significantly; the main job of these agencies has expanded over and above the mere construction and maintenance of transport infrastructure. The agencies also have to be responsible for the operation of networks, to achieve improvements in safety, fluidity, reliability, comfort and efficiency. Improving mobility and safety, lowering fuel consumption and the emission of pollutants, and providing travelers with dynamic and effective information, are the main goals on the market today.

The need for efficient transport networks means that their operation has become a primary focus, so intelligent transport systems have been used as tools to make this efficiency possible. ITS systems apply transport systems technology to solve problems and achieve optimum performance.

Some example of ITS are given below:
• Video cameras to detect accidents
• Dynamic message boards of road information boards
•Vehicle detectors to calculate travel times
• Electronic payment of collection systems
• Traffic management centers
• Systems integration software
• Intelligent traffic lights

The iRAP project uses technologies such as GIS and GPS with cameras in vans for the purpose of preventing road accidents and, when combined with ITS, can prove to be a really useful tool in the hands of the agencies responsible for the urban infrastructure, helping them to obtain in-depth knowledge of the state of the urban infrastructure in order to be able to meet future needs through a well-defined planning strategy. The International Road Assessment Programme (iRAP) is a registered charity dedicated to preventing the more than 3,500 road deaths that occur every day worldwide.

As they say “Our vision is a world free of high risk roads.”
iRAP works in partnership with government and non-government organizations to:

- Inspect high-risk roads and develop Star Ratings and Safer Roads Investment Plans

- Provide training, technology and support that will build and sustain national, regional and local capability
- Track road safety performance so that funding agencies can assess the benefits of their investments

So, for the intelligent transportation system, it is important to use all these tools in order to support the provision of a safe, efficient and environmentally-friendly multi-modal transportation system through the application of the best technologies, practices, and partnerships. The City of Calgary (2009)

This section shall present and discuss how ITS are used and, more precisely, the next section shall examine the case studies.

5. EXAMPLES AND CASE STUDIES

This section gives some examples to illustrate the importance this issue has around the world. Some of them are taken from the report, *Integrating Australia's Transport Systems: A Strategy For An Efficient Transport Future* Booz Allen (2012) that describes cities with integrated transport planning and the use of ITS.

London

London's overall public transport network is characterized by a well-established rail network complemented by an extensive bus network together with a ferry network. These networks are integrated by multi-modal stations designed for ease of interchange for high volumes of passengers. At major stations, purpose-built bus interchanges have been developed to be within walking distance of the railway and underground stations, often manned by bus station staff and furbished with real time information systems (e.g. Countdown – which shows the number of minutes until the next bus is due to arrive).

Hong Kong

Hong Kong public transport services include railways, trams, buses, minibuses, taxis and ferries. This results in very high public transit mode share (90%) and very low vehicle ownership rates (50 vehicles per 1000 population). Hong Kong transport services are provided by several operators.

Singapore

Singapore is considered an international leader in integrated multi-modal transport planning. It established the world's first area licensing and electronic road pricing systems, and uses a quota system to limit vehicle ownership. The government makes continued investments in transport infrastructure.

Colombia

In 2011 in Colombia, an attempt was made to create an inventory of the streets and roads Quintero (2011) in terms of an inventory that includes: road infrastructure, infrastructure, signaling and control devices, parking infrastructure, the whereabouts infrastructure, and the infrastructure of public transport, routes and urban passenger transport. The case of the *Transmilenio* buses system was also studied.

6. CONCLUSIONS AND FURTHER RESEARCH

Big cities are experiencing and will continue to experience significant growth, so it is very important to be able to deliver the urban transportation infrastructure in a successful and sustainable manner. Experience with cases and projects in the past have given us insight into the failures, and, according to KPMG (2010), these experiences show that we need to take some points into consideration if we want to succeed

- Project environment and turbulence
- Political control and sponsorship
- Role of national government
- Effectiveness of planning
- Effectiveness of procurement and financing
- Organizing for operations

The next step is to consider each one of these factors and to analyze projects from this point of view.

ACKNOWLEDGMENT

This research was supported by UNAM-PAPIIT grant IN116012. The second author specially acknowledges the support from CONACYT scholarship program.

7. REFERENCES

Alminas, M. Vasiliauskas, A. V. Jakubauskas, G. 2009. The Impact of transport on the competitiveness of national economy. *Department of Transport Management*, 24(2): 93-99.

Álvarez-Herranz, A. Martínez-Ruíz, M.P. 2012. Evaluating the economic and regional impact on national transport and infrastructure policies with accesibility variables. *Transport, 27 (4),* 414-427.

Ambrosino, D., Sciomachen, A., Scutellà, M.G., 2009. "A heuristic approach based on multi-exchange techniques for a regional fleet assignment location-routing problem", *Computers & Operations Research,* 36 (2), 442 - 460, Scheduling for Modern Manufacturing, Logistics, and Supply Chains.

Aschauer DA. 1991. *Transportation spending and economic growth: the effects of transit and highway expenditures*. (Report). Washington, D.C: American Transit Association.

Balaker, T. 2006. Do economists reach a conclusion on rail transit? *Econ Journal Watch,* 3(3): 551.

Belenguer, J.M, Benavent, E., Prins,C., Prodhon, C., WolflerCalvo, R. 2011, A Branch-and-Cut method for the Capacitated Location-Routing Problem. *Computers &OperationsResearch* 38, 931–941

Berman, O., Ianovsky,E., Krass, D. 2011, Optimal search path for service in the presence of disruptions, *Computers & Operations Research* 38, 1562–1571, Elsevier.

Booz Allen 2012, *Integrating Australia's Transport Systems: A Strategy For An Efficient Transport Future, Infrastructure Partnership Australia* . Available from www.infrastructure.org.au/DisplayFile.aspx FileID=812 . [accessed 12 April 2013]

Buzai, G., 2012. *El ciberespacio desde la Geografía. Nuevos espacios de vigilancia y control global*, Meridiano, Revista de Geografía No.1. Available from http://www.gesig-proeg.com.ar/documentos/articulos/2012-Buzai-Meridiano1.pdf. [accessed 10 November 2012]

Calderón, C. Servén, L. 2008. Infrastructure and economic development in Sub-Saharan Africa. *The World Bank Policy Research* Working Paper, 4712.

Cal & Mayor *Intelligent transportation Systems (ITS)*, PDF format, Available from http://www.calymayor.com.mx/website/documentos/its.pdf. [accessed at 10 May 2013].

Chekuri, C., Shepherd, F.B., Oriolo, G and Scutellá, M.G., 2007. Hardness of Robust Network Design, *Networks* 50, 50-54.

Coelho, J.L.C., Cordeau,J.F., Laporte, G., 2012.*Thirty Years of Inventory-Routing, Transportation Science* accepted for publication. (Technical Report) September 2012. CIRRELT

Daskin, M.S. 2008. What you should know about location modeling. *Naval Research Logistics (NRL)* Volume 55, Issue 4, pages 283–294, June.

De Rus, G. 2008. *The Economic effects of high speed rail investment.* University of Las Palmas, Spain, Discussion Paper, 2008-16.

Dijkstra, A 2011, *EN ROUTE TO SAFER ROADS. How road structure and road classification can affect road safety*. SWOV. Available from http://www.swov.nl/rapport/Proefschriften/Atze_Dijkstra.pdf. [accessed 3 March 2013].

GEOblog 2007. Available from http://www.antoniofraga.com.[accessed on 11 July 2012].

GIS Mapping Software 2013. Available from http://www.esri.com. [accessed 12 March 2013].

Johnston, R. A. 2004. The Urban Transportation Planning Process. In S. Hansen, & G. Guliano (Eds.), *The Geography of Urban Transportation* (pp. 115-138). The Guilford Press.Mult-Modal Transportation Planning.

Klunder,G.A. Post, H.N. 2006, The Shortest Path Problem on Large-Scale Real-Road Networks, *Networks* 182-194.

KPMG 2010. *International, Success and failure in urban transport infrastructure projects*. A study by Glaister, Allport, Brown and Travers KPMG's Infrastructure Spotlight Report.

Laporte, G., 2007 What You Should Know about the Vehicle Routing Problem. *Naval Research Logistics* vol n 54(8) pp.811-819.

Laporte,G., 2009. Fifty Years of Vehicle Routing, *Transportation Science*, 43 408-416. Published online before print October 21, 2009, doi: 10.1287/trsc.1090.0301.

Levy, J. M. 2011. *Contemporary Urban Planning*. Boston: Longman.

Pike, 2009 GPS III and Operational Control Segment. Available from http://www.globalsecurity.org/space/systems/gps.htm. [accessed 17 March 2011].

Quintero, J.R., 2011. Road Inventories and the RoadNet Categorization in the Traffic and Transport Engineering Studies, *Revista Facultad de Ingeniería, UPTC*, I., vol. 20, No.30., pp 65-77.

Smith, G., 1995. Digital Orthophotography and GIS. Smith, *Proceedings of the 1995 ESRI User Conference, 22-26 May*, Palm Springs, California. Available from http://www.esri.com/library/userconf/proc95/to150/p124.html. [accessed 8 February 2013].

The Transportation Planning Process: Key Issues A Briefing Book for Transportation Decision makers, Officials, and Staff. *A Publication of the Transportation Planning Capacity Building Program Federal Highway Administration Federal Transit Administration* Updated September 2007 Publication Number: FHWA-HEP-07-039

The City of Calgary, *Transportation Department*. Available from http://www.calgary.ca/Transportation/Pages/Transportation-Department.aspx. [accessed 8 march 2013].

The International Road Assessment Programme (IRAP). Available from http://www.irap.net. [accessed 7 January 2013].

Toth, P., Vigo, D., 2002. The Vehicle Routing Problem. *SIAM.Monographs on Discrete Mathematics and Applications*. Society for Industrial and Applied Mathematics, Philadelphia. 2002

Turban. et al., 2008. *Electronic Commerce A Managerial Perspective*, Pearson education.

Vishv, J., Erhan, K., Amit, P., 2009. Logistics network design with inventory stocking for low-demand parts: modeling and optimization. (Report) *IIE Transactions May1*.

Wu, T.H., Low,Ch., Jiunn-Wei Bai, 2002. Heuristic solutions to multi-depot location-routing problems *Computers & Operations Research 29, 1393-1415.*

Yu, B, Lam W, Lam Tam, M. 2011, Bus arrival time prediction at bus stop with multiple routes, *Transportation Research Part C*. Elsevier.

AUTHORS BIOGRAPHY

Idalia Flores received a Master with honors, being awarded the Gabino Barreda Medal for the best average of her generation, in the Faculty of Engineering of the UNAM, where she also obtained her Ph.D. in Operations Research. Dr. Flores is a referee and a member of various Academic Committees at CONACYT as well as being a referee for journals such as Journal of Applied Research and Technology, the Center of Applied Sciences and Technological Development, UNAM and the Transactions of the Society for Modeling and Simulation International. She is a full time professor at the Posgraduate Program at UNAM and her research interests lie in simulation and optimization of production and service systems.

Ioannis Chatziioannou did his undergraduate studies in Automation and Control Engineering at the Technological Institute of Piraeus, Greece, and his Master's Degree in Information Technology at the University of the West of Scotland and is currently studying his PhD in Systems Engineering at the UNAM. He has twice participated as co – author of an article in the International Congress.

Esther Segura is a professor in the Department of Industrial Engineering of the Universidad Nacional Autónoma de México (UNAM). She did her Masters and Ph.D. in Operations Research in the Postgraduate Program of the Faculty of Engineering UNAM and is currently in post doctorate studies in the Institute of Engineering of the same University. Her research and teaching interests include developing models and algorithms for production and transportation problems.

Salvador Hernández is a Professor of Operations Research at the Instituto Tecnológico de Celaya (ITC). He received his BEng in Chemical Engineering from the Universidad La Salle and a M. Eng and a PhD in Engineering (Operations Research) from the Universidad Nacional Autónoma de México. Prior to joining to ITC he worked as a planner in ceramics and chemical industries. Dr Hernández teaches courses in Operations Research and Optimization. His current research interests include modeling and analysis techniques for production and manufacturing systems

APPLICATION OF THREE-DIMENSIONAL VISUALIZATION TECHNIQUES AND DECISION MAKING USED IN THE DETECTION OF MOVEMENTS OF THE GROUND AND UNDERGROUND PIPELINES IN UNSTABLE AREAS

Robson da Cunha Santos[a], Marcelo Silva[b], Gerson Gomes Cunha[c]

[a] Fluminense Federal Institute, Campus Macaé, Engineering and Automation Control
164 km Amaral Peixoto Road, Brazil and Estácio de Sá University, General Alfredo Bruno Gomes Martins Highway, s/n - Braga - Cabo Frio / RJ, Brazil
[b] Estácio de Sá University, General Alfredo Bruno Gomes Martins Highway, s/n - Braga - Cabo Frio / RJ, Brazil
[c] Federal University of Rio de Janeiro, Alberto Luiz Coimbra Institute Graduate Engineering and Research, Civil Engineering Program/COPPE/UFRJ

[a]profbsonso@yahoo.com.br, [b]msc.marcelosilva@gmail.com, [c]gerson@lamce.ufrj.br mail

ABSTRACT

Techniques for three-dimensional visualization and simulation of construction are applied in regions where companies install pipes for oil and gas. These regions can present aspects of geological and geotechnical risks, which can compromise the structural integrity of the pipes, by movement of the solids and interaction between soil and pipe. A system capable of evaluate these regions was created based on digitalized data of aerial photographs, coordinates of control instruments and on ground topography. The system consists of a three-dimensional environment (3D) using technics on virtual reality (VR) developed to support the analysis of existing problems. The system is completely geo-referenced, which permits definig adequate solutions for the projects as well as avoiding problems for some specific areas. In addition, the system permits interaction with specialists, enabling them to indicate directly on the 3D images, the risk aspects for further evaluation *in situ* and for taking adequate decisions.

Keywords: three-dimensional, visualization and simulation, virtual reality, unstable areas

1. INTRODUCTION

New studies on pipelines demonstrated that Brazil has about 22000 km of pipelines underground, and many of these areas are considered at risk (Monitor Mercantil 2010). Many of these pipelines cross regions with distinct geologic features, presenting several unstable areas as the region of Serra do Mar. In these regions the pipelines are submitted to additional loadings imposed by ground movements.

Structural integrity of the ducts should be in perfect condition installed in these areas. It becomes necessary to survey and map all areas unstable and stud the soil mass movements. Creep movements usually involve extensive areas and present slow speed. In general they are difficult to detect through visual inspection.

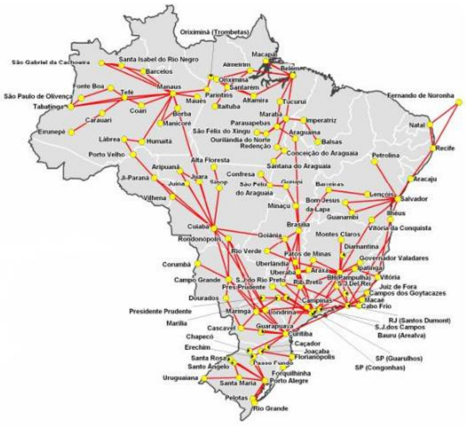


Figure 1: Brazilian Pipeline Map (Calhambequi 2013)

New studies have been implemented to improve operational safety pipeline, new technologies are being developed to detect unstable areas and estimate their effect on the pipelines. A complete set of visualization and numerical simulation software platform is available and it is being used to build a three-dimensional model of all the geotechnical risky areas in Serra do Mar. The installation and operation of a pilot monitoring system, including piezometers and inclinometers on the slope and strain gauges on the pipeline, at three different pipelines crossing Serra do Mar, with data acquisition in real time is also being undertaken.

Techniques for three-dimensional visualization of areas around the pipeline have demonstrated to be a great tool to detect geologic and geotechnical risky features. It allows a detailed analysis of the adjacent slopes, being useful to guide the implementation of monitoring systems and stabilization works.

2. REGION EVALUATED

The region studied and analyzed consists of a mountainous range Brazilian relief that extends for about

1500 km along the east / south coast, ranging from the state of Rio de Janeiro to the north of the state of Santa Catarina known as Serra do Mar. The saw is characterized by the presence of thick deposits talus colluvium. Due to its formation process, these deposits are very heterogeneous, presenting all kinds of grain sizes. It is located in a region of high precipitation what in association with the geologic and geotechnical features creates a potential condition for the occurrence of creep and soil sliding movement.

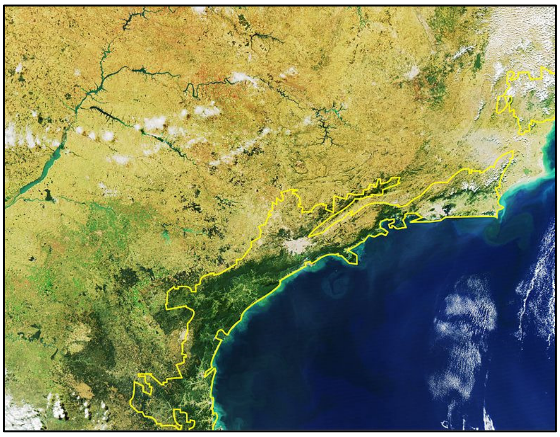


Figure 2: Satellite Image of the Serra do Mar (GoogleEarth 2013)

The best definition for the Creep can be a slow and continuous soil mass movement most of the times with no clear limits. It is caused by the action of the gravity, being activated or increased by variation of temperature and pore pressure in the soil mass. Depending on the season, the speed of soil movements can increase or decrease. It can even cease during the drought period.

Sometimes these types of movements can be noticed on the surface by the change in the verticality of trees, fences or other structures, being observed through visual inspections. When movements are very slow these evidences can escape to the eyes. In these cases it is necessary to use other ways to detect and follow the evolution of the process like the installation of field monitoring.

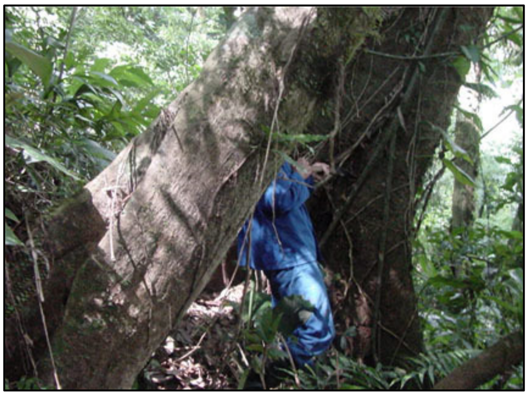


Figure 3: Change of Verticality of Trees

3. DATA ENTRY

The 3D visualization system of geological features of risk depends on various information and input data so that they can do analysis and make some decisions. The main input data are: the aerial photographs, contour lines to generate digital terrain model (DTM), the coordinates of the pipeline sections and coordinates the instruments.

3.1. Aerial photography

Aerial photography is considered to be the basic tool for mapping and the reconnaissance of terrain. Since the discovery of photography and its application in mapping until today, their contribution has been remarkable in the context of Cartography (Hohl 1990).

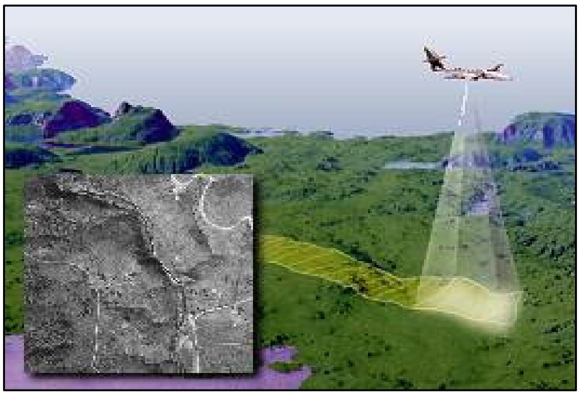


Figure 4: Model of Capture Aerial photography (Esteio Engenharia2013)

Aerial photography was solidified as an essential element for mapping with the creation of science called Aerophotogrametry and its further evolution happened in the period of the World Wars with its constant use for military purposes. With the end of periods of conflict and the discovery of new processes, equipment and materials, aerial photography has become an invaluable product for the planner, researcher and entrepreneur, besides being the raw material for the work of the cartographer (Plewe 1997).

It is, in technical terms, an aerial photograph such as that obtained by strictly air chamber (calibrated focal length, lens distortion parameters and known frame size) mounted with the optical axis of the camera in a near vertical aircraft properly prepared and approved to receive this system. The set of operations required to obtain these photos or set of pictures that superimposed and represent the area flown. This area is called the coverage aero photogrammetric. Aerial photographs are usually obtained in a sequential and overlapping longitudinal and lateral imaging allowing the entire region of interest is covered (Andrade, 1999).

3.2. Level Contours

As result of longitudinal superposition between two consecutive aerial photographs stereoscopic images were obtained. These images are placed on devices restorative semi-analytical and analytical and a basis for the model definition. These images are projected on an optical system so as to form a three dimensional model in the eyepiece (Hohl 1990).

The stereoscopic model should be oriented so as to reproduce all the characteristics of the terrain, without presenting deformations and dislocations. The orientation must also assign to the stereoscopic model, the coordinate system used for mapping, that is, it is the placement of the scale model and appropriate coordinates. The coordinates of these points are coming from aerial triangulation (Wolfgram 1993). By means of a suitable system, all of the recorded points are transmitted to equipment that composes the original cartographic. The stereoscopic mark should always be kept in contact with the surface of the feature in order to maintain the correct coordinate. For each point registers values of the X, Y and Z represent the spatial location.

The catchment of altimetry is performed with the mark stereoscopic always with a certain altitude, where the operator must search of the terrain visualized the points that have this same altitude, thus materializing the lines of constant altitude or contours, which in practice we call level contours .

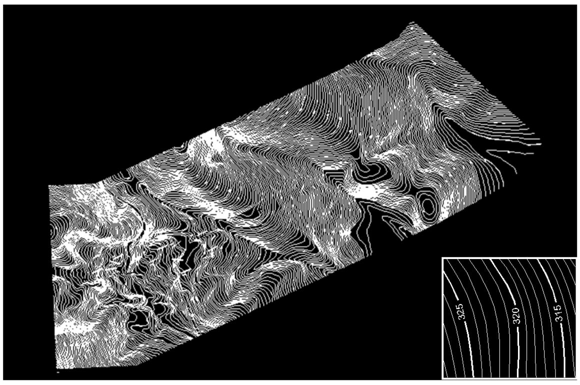


Figure 5: Model of Level Contours

3.3. Coordinates of the Pipeline

The coordinates of the pipeline are of extremely importance for the system feature to detect risks with possible problems, because from these coordinates you can define whether they suffered great effort by the soil shifted relative to its initial position (construction), or there is some kind of curvature or kneading.

A device is passed over the soil instead likely to find the pipeline. The detection and location of coordinates of the pipeline comprise a location horizontally and vertically to the soil level.

Another procedure aggregate with this type of equipment is to define locations in the pipeline where the coating is damaged. This particular feature allows you to find possible sources of deterioration and rupture of the pipeline being in a preventative maintenance and replacing most elusive and costly methods such as excavation.

In the next figure, the line can be identified in red as the coordinates of the pipeline stretch that have been raised level contours.

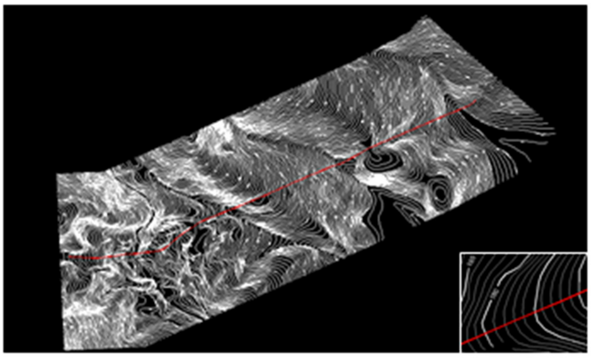


Figure 6: Example of Coordinates of the Pipeline

4. VISUALIZATION AND SIMULATION

In possession of all the data, the system takes care of processing and preparing them for your visualization and decision making on the part of the engineers responsible.

The evaluation of geologic and geotechnical features in the pipeline route is carried out through local inspections and using the three-dimensional visualization model developed from aerial photographs and topography data processing.

As advancing technology allows capturing, storing and processing a large amount of information, the features visualized in flat paper, can be brought to life in 3D forms. The 3D model is built using level contours, Triangular Irregular Network (TIN) and orthophotos.

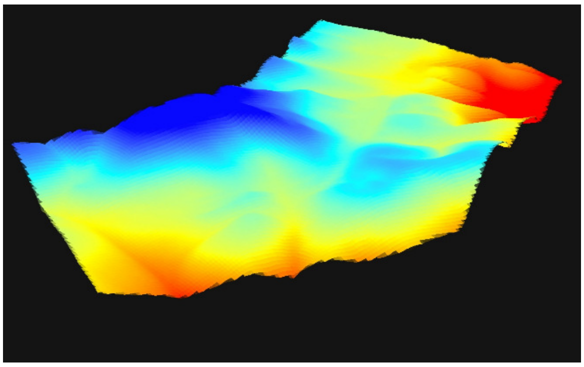


Figure 7: Data Elevation Model (DEM)

Engineers and experts can analyze the geomorphology around the pipeline and it is possible to use two kinds of surface models: Rasters and TIN. Rasters represent a surface as a regular grid of locations with sampled or interpolated values. TINs represent a surface as a set of irregularly located points linked to form a network of triangles with elevation values. Rasters are largely used in USGS (U.S. Geological Survey) in Data Elevation Model (DEM) maps covering the world in arc of 30 seconds. TIN models are not so common and tend to be more expensive to build and process. They are typically used for high precision modeling of smaller areas, such as in geotechnical application

The use of aerial photographs along with TIN model facilitates the visualization instead of using the DEM. The visualization is done in stretches 1 km long and 400 meters wide.

To start this project, criterions data are got by aerial photo. After an image rectification, contour level curves are got in scale of 1:1000, allowing the visualization of curves of 1 meter high.

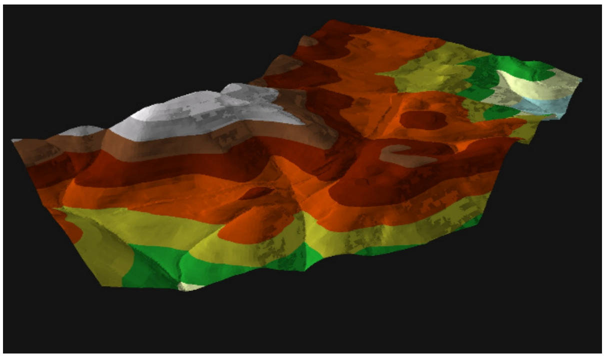


Figure 8: Triangular Irregular Network (TIN)

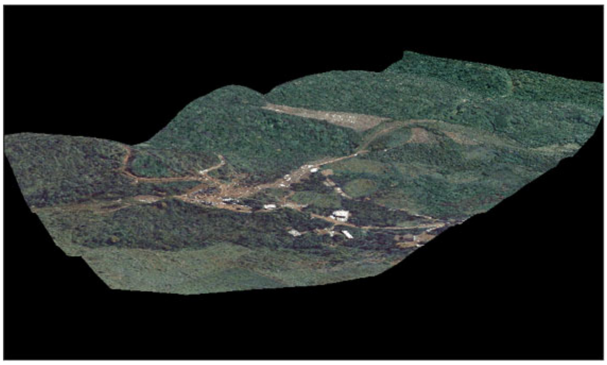


Figure 9: Aerial Photographs Applied the Model TIN

After that, the geomorphologic features can be recognized by specialists and pointed in the image, creating interpretative maps.

All data are preserved in their original state. The transformations are made inside the program. All data are referenced before used in visualization.

The data furnished are aerial photos and cad maps using datum SAD-69. Visualizing a three dimension data gives the observer new perspectives. A 3D view of terrain can provide insights that would be not apparently clear for an observer in the field or looking at a planimetric map.

With this model specialists can observe existing problems without visiting the place, and then perceiving incoming accidents. Therefore the 3D model increases sight inspection.

Besides that, 3D the model is connected to internet, easing the manipulation of the technicians involved in risk analysis. It is possible to show more data by hyperlinks inserted in the model. Field instruments, for instance, can be inserted in the model with original UTM coordinates and the model can access their data.

The next figures demonstrated the utilization of the internet for visualization of features without the need to visit the place of risk.

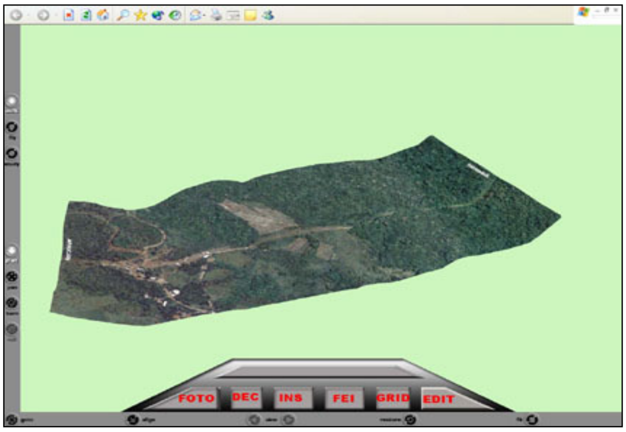


Figure 10: Overall Image of the 3D Model Aerial

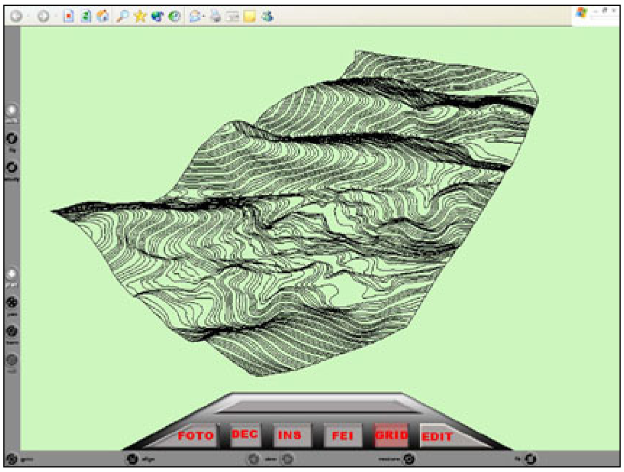


Figure 11: The 3D Visualization of Level Contours

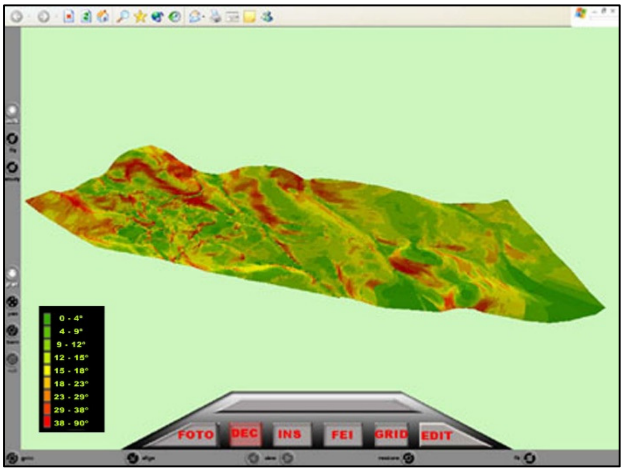


Figure 12: Visualization of 3D Declivity

5. PIPELINE AND SOIL MASS MONITORING

It is very important to monitor areas affected by soil movements. Knowing that soil-pipeline interaction is extremely complex the implementation of an extensive monitoring program including not only the slope but also the pipeline becomes mandatory.

Using a monitoring system it is possible to calibrate and to validate the established soil-pipeline interaction model. The analysis of instrumentation results allows a technical decision about the right moment to intervene on a pipeline. So pipeline operational safety can be guaranteed.

The complexity of the phenomena should always be considered. An extensive program of monitoring is usually recommended. It is necessary to evaluate the causes and consequences of soil movements. Nowadays, in several critical areas, the soil mass is monitored with inclinometers and piezometers. The frequency of readings varies from case to case. In general, they are taken each three months. Sometimes it becomes necessary to increase this frequency during the rain season.

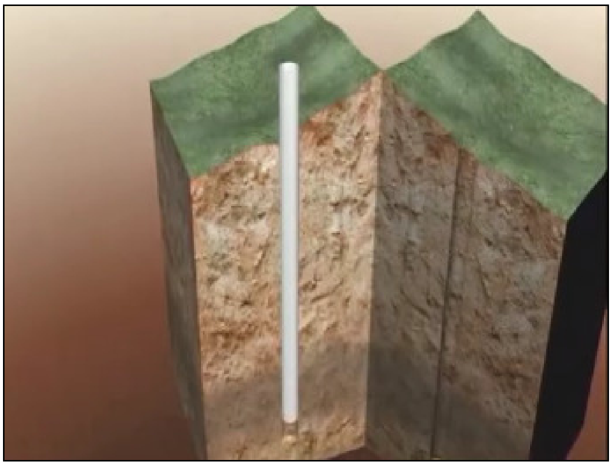


Figure 13: Model of Piezometers

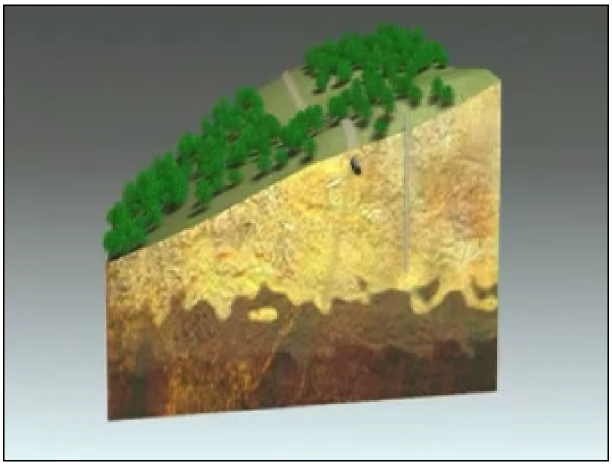


Figure 14: Model of Inclinometer - Initial State

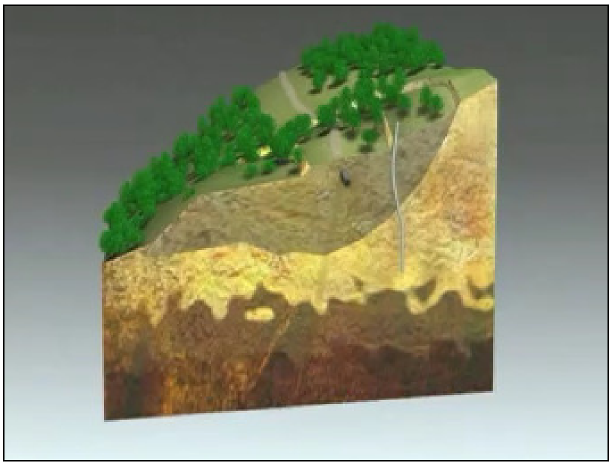


Figure 15: Model of Inclinometer – Final State

Slope monitoring, however, is valuable to understand the process of soil mass movements that did not allow an evaluation of stresses transmitted to the pipeline. On this purpose it was necessary to install strain gauges on the pipeline surface. Another detected problem is that the period between readings, many times can not correspond to the necessity of each point.

In order to improve pipeline's safety, we implemented a new monitoring system including inclinometers and piezometers on the slope and strain gauges on the pipeline with data acquisition in real time. Inclinometers were used to measure the displacements of the soil and an eletric piezometer were used to assess changes in soil pore pressure. Inclinometers were used to measure soil displacements while eletric piezometers was used to evaluate variations of soil pore pressures.

6. THREE-DIMENSIONAL VISUALIZATION AND DECISION-MAKING

Decisions are made in meetings from 3D models. Risk features are traced by experts who then go to the site to ascertain the terms displayed on models. These decisions could take days and huge expenses because they would have to go to where all the engineers who attended the meeting with the 3D model. The next figure presents a detail model demonstrating the risk.

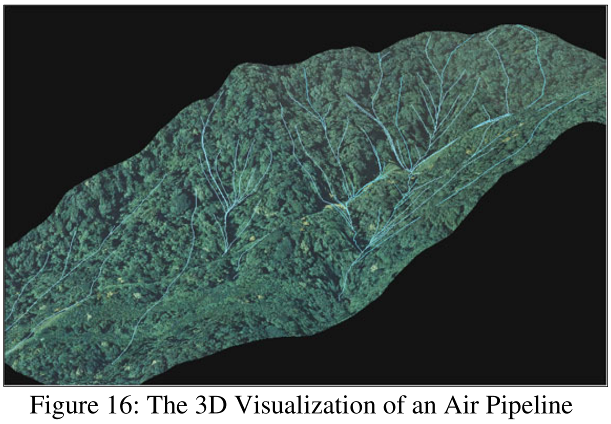


Figure 16: The 3D Visualization of an Air Pipeline

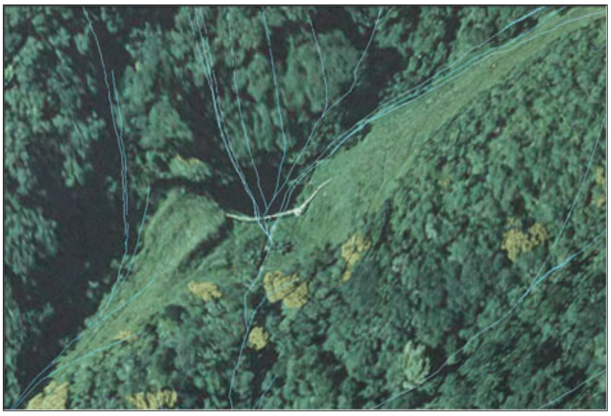


Figure 17: Detail of the Air Pipeline

With the 3D model visualization, engineers were on the scene and confirmed that the air pipeline was running a great risk, because it was a passage of water and debris coming from the mountain.

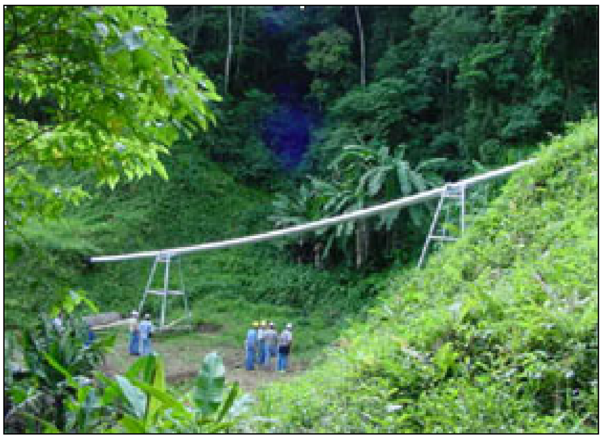


Figure 18: Picture Taken in Place

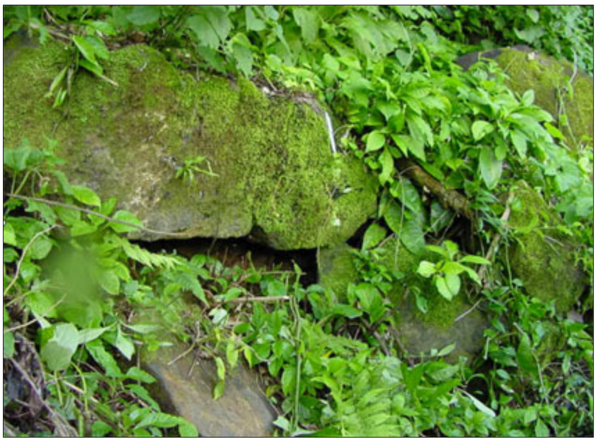


Figure 19: Visualization of Blocks of Loose Rocks

According to the photos by engineers with access to 3D models of the system we arrived at a unanimous view that the pipeline should be buried. In times of many rainfall debris could cause an oil leak at the pipe break.

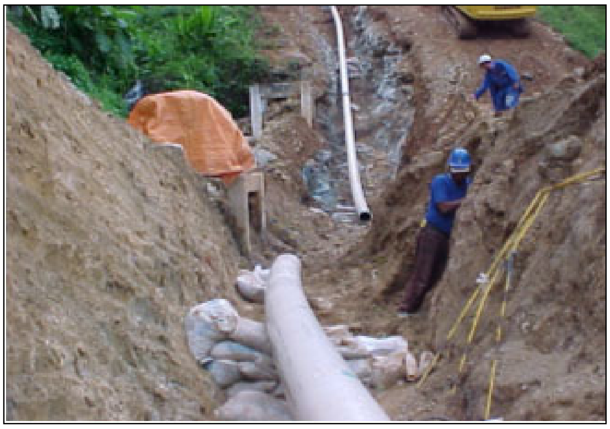


Figure 20: Photo with Pipeline Being Buried

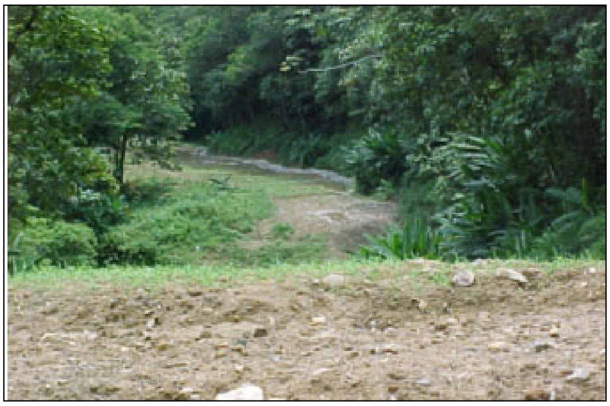


Figure21: Pipeline Buried

7. CONCLUSION

The pipeline installed in unstable areas should be considered a geotechnical work. The interaction of soil with pipeline should permanently be considered in order to ensure the structural integrity of the pipeline. In these areas, new technologies must be developed and implemented to improve safety.

The 3D visualization has a valuable role in the detection of new risk areas and decrease in oil spill accidents, preventing the pollution of nature.

The system enables environmental rehabilitation works and the pipeline system will indicate areas with preventive, corrective or mitigating. It is also used to display the results of specific inspections, routine seasonal and, when necessary, minimize the maximum accidents that impact the environment, whether they are generated by natural or anthropogenic action.

REFERENCES

Andrade, J. B. 1999. *Fotogrametria. Curitiba:* SBEE. 258 p.

Azevedo, E. et al. *Desenvolvimento de Jogos 3D e Aplicações em Realidade Virtual*- Editora: Campus - 2008

Botella, C. G.-*Palacios A, Villa H, Baños RM, Quero S, Alcañiz M, Riva G. Virtual Reality Exposure in the Treatment of Panic Disorder and Agoraphobia*: A Controlled Study. Clin Psychol Psychother 2007

Carey, Rikk e Bell, Gavin 1997- *The Annotated VRML 97 Reference Manual*

Crane, N. 2002 - *Mercator - The Man Who Mapped the Planet* - 320 páginas - Henry Holt & Company - ISBN: 0805066241

Hohl, P. / Mayo, B. - *1990. ArcView GIS Exercise* Book 480 Páginas, 2a. Edição Thomson Learning

INSTITUTO GEOGRÁFICO 1986 (São Paulo, SP). *Projeto Lins Tupã: foto aérea*. São Paulo, 1986. Fx 28, n.15. Escala 1:35.000.

John, R. V. - *VRML clearly explained, AP Professional*, 1998 ISBN 0127100083

Kemp, K. - *Encyclopedia of Geographic Information Science*- Sage Publiation Inc. – 2008.

Plewe, B. 1997 - *GIS Online: Information Retrieval, Mapping, and the Internet* - 311 Páginas Onword Press

Santos, R. C. 2000 – *Determinação e Simulação da Posição da Cabine de Controle em Sondas de Perfuração Através de Algoritmos Genéticos e Realidade Virtual*, Tese de Mestrado, Mestrado em Ciência da Computação/UFF.

SEA, *Surface Temperature Satellite Image Archive 2000*: *Banco de dados mantido pela University of Rhode Island Graduate School of Oceanography.* Disponível em: http://dcz.gso.uri.edu/avhrr-archive/archive.html.

Snyder, J. P, 1993 -Flattening the Earth: 2000 Years of Map Projections- 366 páginas - University of Chicago Press - ISBN: 0226767469

Wolfgram, Douglas E. 1993. *Aventuras em 3D. Traduzido por Marilda Cesar Caselato & Romes Souza Dantas*. Berkeley: Rio de Janeiro.

SIMULATION OF THE DESIGN AREA IN A PRINTING COMPANY

Carlos Quintero Aviles[(a)], Idalia Flores[(b)]

[(a)] Engineering Faculty, UNAM
[(b)] Engineering Faculty, UNAM

[(a)]carlosqa1@comunidad.unam.mx, [(b)]idalia@unam.mx

ABSTRACT

This document presents a study of the activities in the Design area of a company devoted to the printing of promotional products, then the process is simulated to analyze the performance of the area and, with the help of this simulation alternatives were assessed with the aim of determining its ability to respond to customers once they have placed an order. The document begins with an introduction to the activities done in this area, followed by definition of the problem and then we present the methodology followed by the Simulation Model built in Simio and afterwards show the results. Finally we give some conclusions about the activities in the design area.

Keywords: printing design, simulation, printing company (Graphic arts)

1. INTRODUCTION

The Company where this project takes place, prints promotional products, using printing techniques such as silk screen, pad printing and laser engraving, the main products customers request are: pens, keyrings, cups, thermos, among other things.

Table 1: Main products and the printing technique used

No.	Product	Printing technique
1	Pens	Silk screening
2	Metal keyring	Laser engraving
3	Stress Ball	Pad printing
4	USB Memory	Laser engraving
5	Cylinders	Silk screening

The customer/salesperson sends a sketch of the logo that the customer wants printed, through a web system, the DO (design order) that consists of a dummy, the designer is responsibility for checking that the logo has the correct form, in other words the image must be in (vectored) Curves so that will not be distorted during the printing process, it is also necessary to check the logo's position on the product it is to be printed on, as well as its dimensions and the colors of the ink used. Once these activities are completed the dummy is returned to the customer/salesperson for their approval. If approved, an invoice is made out and the DO goes to planning, the dummy is used to make a positive or negative depending on the printing technique, and if the customer requires any modifications the DO is brought back to the design area for the necessary changes to be made and it is sent back to the customer/supplier, until it is accepted.

Figure 1: the main products handled by the company

The Design area has three workers (designers) who work from 8 am to 5 pm, as well as an intern who works just four hours a day from 8 am to 12 pm, the web system can even receive DOs outside the designers' working hours, so the queue of DOs can increase from one day to another because of the DOs received outside working hours.

The designers must know about the printing techniques they use, the different types of inks they have and which ones can be applied to which product, and they must also be able to handle graphic design software such as Adobe Illustrator (AI), all of which they need to do their work to the best of their ability and to advise the customer on the job they require

2. PROBLEM DEFINITION

The following research questions were designed to be used as a guide during the study:

- Is there enough capacity in the Design area to deal with the demand of DO customers?
- What is the average time a customer/salesperson must wait from the moment they send their DO through the web system until the designer returns it to them for authorization?
- What response time can the company establish, as a customer service policy, for returning the DO for authorization once it has been uploaded onto the system?

The scope of the research is confined to the activities of the designer, tracking the DOs from the moment they are received in the system until, once authorized, they are released into planning.

The designers' working hours are from 8 am to 5 pm with an hour for lunch; although outside of this schedule the DOs could be waiting in line owing to the fact that the system is still open and could be generating DOs.

Table 2: Designer's job description

Graphic Designer	
Main Objectives	Activities and/or duties
To communicate a certain message, a message prepared by a person and aimed at a specific context, linked to a selection of elements, colors, shapes, typography selected for the graphic communication of the printed message on a product	To download the dummy template from the system. To review the product's specifications. Vectorization of logotypes. The dummy is filled in as per the design order and is uploaded to the webpage for its revision and/or authorization. Revision of email to see whether the DO has been authorized or needed to be corrected.

The designer may have other activities such as the creation of positives or negatives that shall not be taken into account for this study.

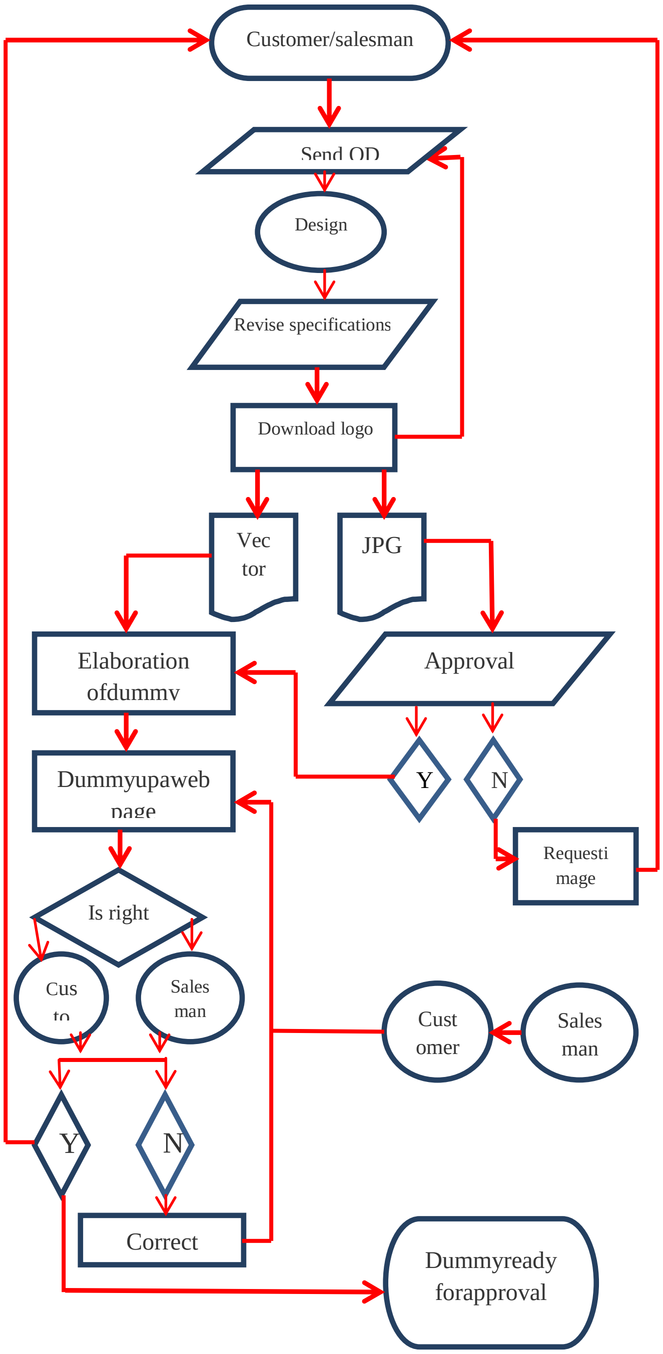


Figure 2: Flow diagram for the design area

The objective is to get the descriptive steps of the performance of the Design area through a simulation study that models the DOs' waiting times in the system, from the moment the designer begins to work on the DO until it is authorized by the customer/salesperson as well as to assess some of the scenarios of interest in order to establish a policy for responding to the customer.

3. METHODOLOGY

We carried out the research in the order followed in the methodology given below (Flores, 2006):

3.1. Problem definition

Here we determine the general objective as well as the specific aspects of the research, the scope and the resources needed.

3.2. System conceptualization

Once we had defined the problem to be researched, we determined the aspects or factors that are most important and most influential in the phenomenon, in order to decide whether or not to include them in the model and in how much detail.

3.3. Data collecting

Once we had conceptualized the system, we determined whether or not the available information was reliable and what other information was needed, in accordance with the requirements of the model.

3.4. Model formulation

Simio software was chosen for the simulation, because of its robustness and ease of use. A model was built that incorporated the previously defined relevant aspects.

3.5. Verification and validation of the model

Some tests were done to find and correct the errors of logic in the model and other tests were implemented to ensure that the results of the model continued to correspond to the real system.

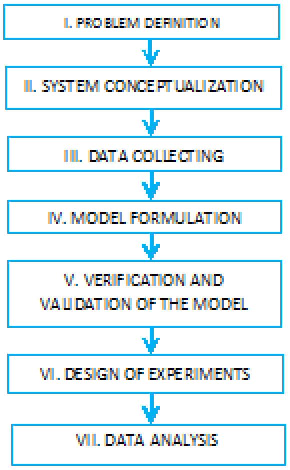

I. PROBLEM DEFINITION
II. SYSTEM CONCEPTUALIZATION
III. DATA COLLECTING
IV. MODEL FORMULATION
V. VERIFICATION AND VALIDATION OF THE MODEL
VI. DESIGN OF EXPERIMENTS
VII. DATA ANALYSIS

Figure 3: Methodology Followed

3.6. Design of experiments

In this section we defined scenarios of interest in accordance with the objective of the research.

3.7. Data analysis

Here the obtained results are compiled and some conclusions given on the present functionality of the system.

4. SIMULATION MODEL

4.1. System conceptualization

The elements relevant to model our study system are: the webpage where the customer and/or the salesperson uploads the DO which must contain the necessary information for the designer to make the dummy, information such as the logo, its dimensions and the substratum where the printing will be done, as well as the tone of the tint that will be used.

The designers take the OD and make the dummy which is a sketch of the final product, this is sent back into the system for the customer's/salesperson's authorization, if it is approved the designer will send it to Planning, otherwise the necessary changes will be made until it is authorized.

In planning the product's manufacturing is programmed this depends on the characteristics of the product and the technique that will be used to do the printing.

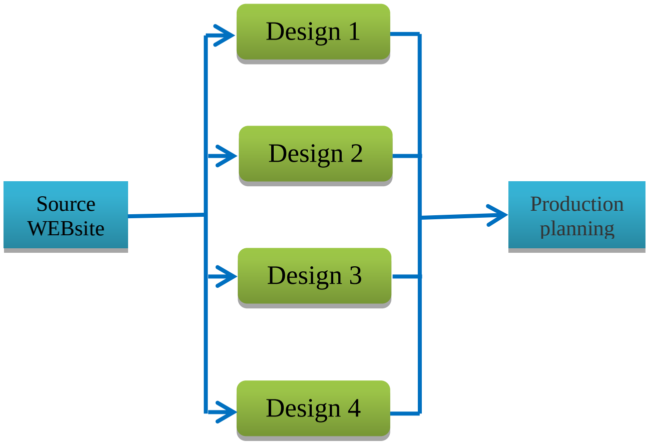


Figure 4: System conceptualization

4.2. Data collecting

The data collection was done in two ways, the first directly from the webpage where the moment a DO is uploaded the date and time are automatically saved, an important factor to mention here is that the page stays available to upload DOs outside the designers' working hours which is from 8am to 5 pm. Which is why in one day the DO queue is made up of the unfinished ones from the day before more than the ones uploaded after 5p.m. and the ones accumulating during the shift. The working days are from Monday to Friday, if a DO is entered into the system on Saturday or Sunday it is considered as though it were entered on Monday at 8:00 am.

Figure 5 shows the DO data collection stage. In 37 consecutive working days 1550 DOs entered into the system, for a λ of 41.9 DOs a day.

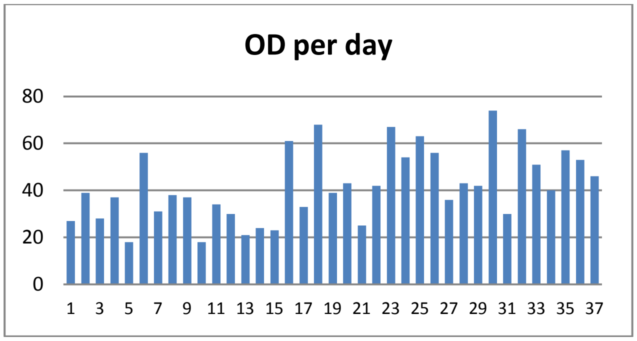


Figure 5: DOs entered per day

The second part of the information was taken directly from within the design area; there are four designers, three of them full-time workers and an intern that only works half a work day. Kenett and Shelenyahu (2000)

4.3. Model formulation

The model will be built as a system with a source that generates the entities (DO), which may be of two types, the ones that arrive with a logo that has already been vectored and the ones that bring it in an image format.

The entities automatically enter the queue, any of the servers (designers), if available, can access the system and download the DO according to FIFO policy, and work on it until it has been authorized to send to planning and thus leave our system under study.

The μ service rate will be analyzed as a whole for all the servers and individually, as their work experience affects their service time.

As we have already mentioned, the software we chose for the simulation is SIMIO. Kelton et al (2012)

The following objects were used:

4.3.1. Entities

As was already mentioned, two type of entities were used. The DO that come with the logo already vectored and the ones who bring it in an image format, the designers need to spend more working time on the latter owing to the fact that that they have to vector the image so that it will not be distorted when printed.

4.3.2. Sources

We used a source that generates the entities (DO), which symbolizes the webpage where the customer or the salesman upload the design orders.

Said source must generate two types of entities, the ones that represent the vectored images and the ones that arrives in image format.

Table 3. Arrival Rate of DO per day

Week	Amount	λ	Classification
1	151	30.2	Medium
2	178	35.6	Medium
3	142	28.4	Low
4	242	48.4	High
5	187	46.75	High

4.3.3. Servers

There are four servers, that correspond to the designers, each of whom has their own work station (desk, phone and computer equipment) where they do their work. When one of them is available they enter the webpage and take a DO under the FIFO discipline. The service time of each designer varies depending on their skill and experience in the job.

4.3.4. Sink

A sink is an object that destroys the entities, in this case we only used a sink to destroy the DOs once the design process had ended, in other words, once the DO has been authorized and sent to the planning area.

4.3.5. Paths and Nodes

The Paths were used to determine the DOs' path within the system and the nodes to represent the points where the DOs are taken by one of the four designers. For example, the paths that come from the DOs source have a node that represents the decision to select a different designer (D1, D2, D3 and D4). The probability of going with one of them, considering a similar service rate for each of them is such that:

$$e_1 + e_2 + e_3 + e_4 = 1 \quad (1)$$

Below are some of the images of the simulation model:

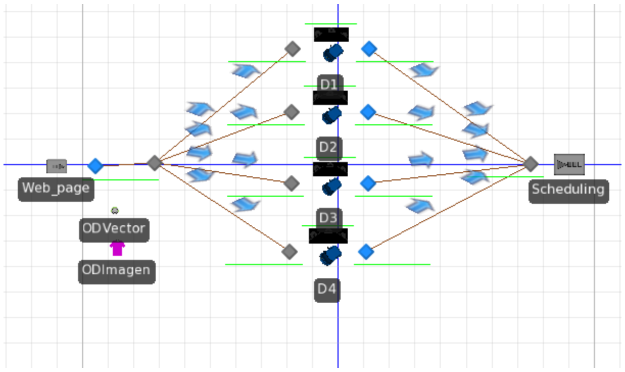


Figure 6: Simulation with SIMIO

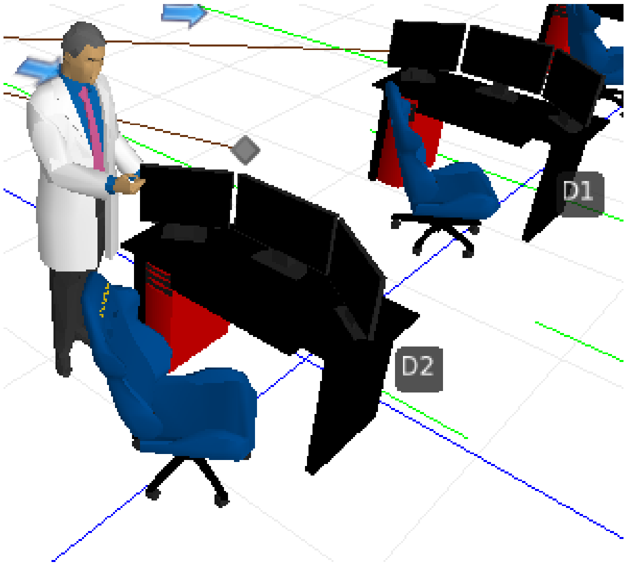


Figure 7: Work area, Designer 2

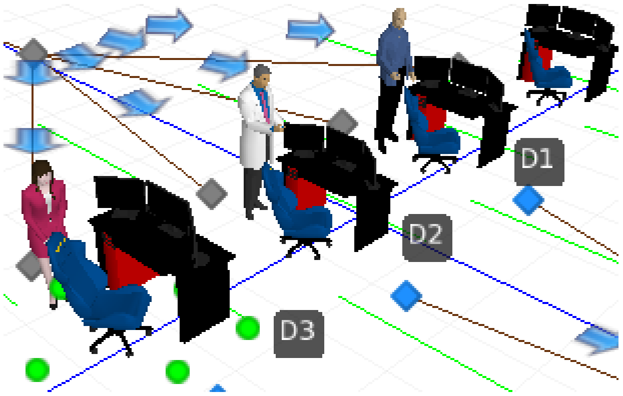


Figure 8: Simulation with 3 designers

4.4. Verification and validation of the model

To validate the model we used the following techniques:

4.4.1. Comparison with the real system

A simulation was done on the number of DOs going into the system, using real data taken from the system after having made the model, then the quantity of these DOs that were dealt with by the designers was compared with the results obtained in the real system and the results were very similar.

4.4.2. Behavior in extreme cases

A simulation is done with a significant input of DOs and a large backlog on the queue, as has occurred on several occasions in the real system, with similar results.

In another test, there was a significant drop in input of DOs in the system, here the results varied in respect of the real system, in which there was an increase of service time on the part of the designers. Law and Kelton (2000)

5. DESIGN OF EXPERIMENTS

5.1. Scenario 1. Normal operating conditions

A simulation is done of the system with the current working conditions, considering an individual service rate for each designer and the input of DOs, which varied over time, as shown in table 3.

The aim is to learn the system's performance measures under normal conditions.

5.2. Scenario 2. With only three designers

In this scenario one of the designers is not working, due to incapacity or absenteeism, which tends to happen fairly frequently in these types of companies. Here an equal service rate is considered for designers at work.

The aim is to determine the impact on the system of working with one less designer, since one of our aims is to find out the company's customer response time.

5.3. Scenario 3. Trained designer 3

In this instance, we consider the case of the designer with the lowest service rate having been trained to do his work better and achieving an average service rate

The purpose is to determine the possibility of lowering waiting time in the DOs queue by increasing the designers' average service rate.

5.4. Scenario 4. Same type of DOs

Here we consider a scenario where the DOs entered into the system are all the same type. Currently the DOs are entered either with a vectored logo or with a logo in an image format. With the former ones the designer takes on average a little more than double the time taken with the ones already vectored, (approximately 12% of the entered DOs come in an Image format)

The aim is to assess the impact on the response time by reducing the percentage of DOs, currently 12%, in image format.

This following table gives a summary of the above scenarios:

Table 4. Table of scenarios.

Scenario	Description
1	Simulation with normal operating conditions
2	One of the designers is absent from work, so only 3 servers are considered
3	The designer with the lowest service rate is trained
4	The same type of DOs (vectored) are considered

6. RESULTS

A simulation was done for each scenario considering a period of eight hours work, ten replicas of each experiment were made. This was repeated for each scenario.

The results are given below:

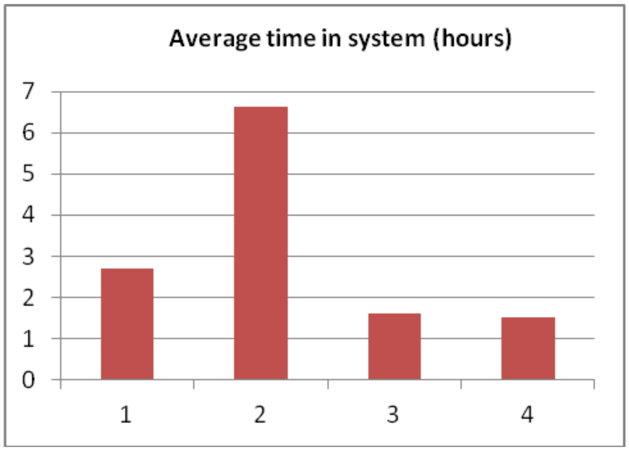


Figure 9: Average Time in System

The average time in the system is 2.71 hours under the present conditions. When a designer is absent, the average time increases by almost seven hours, meaning that there will be DOs pending at the end of the shift (scenario 2). In scenarios 3 and 4 the time is reduced by approximately an hour, meaning that increasing the service time of the least able designer would be almost the same as if the customers were to send all their logos vectored.

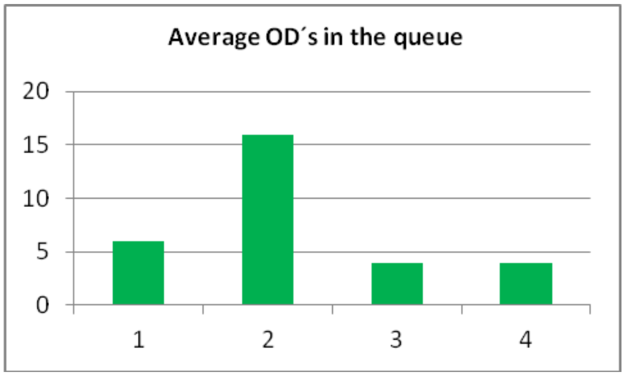


Figure 10: Average DOs in queue

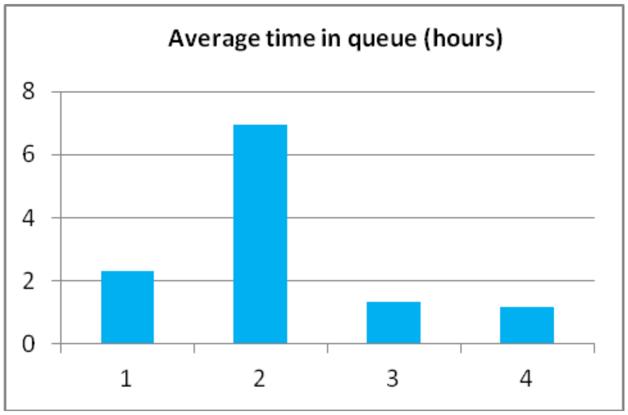


Figure 11: Average time in queue

The average of DOs in the queue increases considerably when a designer is absent, this implies that their time waiting to be done also increases (scenario 2), and the behavior is similar for both scenarios 3 and 4 where, in comparison, we would decrease by approximately the same amount.

7. CONCLUSIONS

This study served to see the impact the designers' work has on the DOs carried out.

With the results obtained we can carry out an analysis to evaluate if the company should decide on a response policy in releasing the DOs to their customers, and, as can be seen in the results, if a policy of no more than three hours response time is stipulated, this would be accomplished as long as all the designers turn up for work, because otherwise the response time increases to approximately seven hours.

If the designers' average service time were to diminish or if they were only received vectored DOs, the response time would be two hours less.

Reformulating the research questions gives us:

- Is there enough capacity in the Design area to deal with the customers' demand for DOs?

Yes, the Design area of this company has enough capacity to meet the customers' demand for DOs.

- What is the average time a customer/salesperson must wait from the moment they send their DO through the web system until the designer returns it to them for authorization?
  Under the present working conditions, it takes almost three hours for the design area to return the DO for authorization.

- What response time can the company set, as a customer service policy, for returning the DO for authorization once it has been uploaded onto the system?
  Maximum seven hours. If one wants to lower this customer response time, some of the following recommendations could be taken implemented:

Train the designers that take the longest to release the DOs, as this would make the average service rate less. The cost of training could be recovered by doing a greater number of DOs in same time.

Asking customers to deliver their logos already vectored rather than in an image format, although this could cause discontent among some customers who might change their supplier and withdraw their orders as a consequence.

Training another worker to do the work of a designer in case any of them is absent for any reason, especially in the case of an incapacity that lasts for several days.

Making the designers aware of the problems they cause when one of them is missing or slow at their work.

And lastly, studying the characteristics of this department’s physical space as while we were collecting the data we noticed some distracters that have an influence the designers’ work.

ACKNOWLEDGMENT

This research was supported by UNAM-PAPIIT grant IN116012. The first author specially acknowledge the support from CONACYT scholarship program.

REFERENCES

Flores de la Mota, Idalia and Elizondo Cortés Mayra, 2006. *Apuntes de simulación*. México, UNAM.

Kenett, S. Ron, Shelenyahu Z., 2000. *Estadística Industrial Moderna.* 2da edición. Thomson editores.

Kelton, W. David, Smith, Jeffrey S., Sturrock, David T. 2012 *Simio and Simulation: Modeling, Analysis, Applications* 2nd. Ed.

Law, Averill M, and Kelton, W. David,2000 *Simulation, Modeling and Analysis*. 3th. ed.

AUTHOR´S BIOGRAPHY

Carlos Quintero Aviles, studied Industrial Engineering in the Instituto Tecnológico de Tlalnepantla (ITTLA, Technological Institute of Tlalnepantla), He Is currently a Masters student in Operational Research in the UNAM. His line of research is simulation and optimization of productive processes.

Idalia Flores de la Mota, studied mathematics at the Sciences Faculty of the UNAM, then a Master and a Ph.D. in Operations Research at the Engineering Faculty. She has taught at various universities and has participated in national and international conferences. Her lines of research are Simulation and Optimization of the Supply Chain.

MODELING AND SIMULATION OF AMPEROMETRIC BIOSENSORS ACTING IN FLOW INJECTION ANALYSIS

Romas Baronas$^{(a)}$, **Darius Baronas**$^{(b)}$

$^{(a)}$Department of Software Engineering, Vilnius University, Didlaukio str. 47, LT-08303, Vilnius, Lithuania
$^{(b)}$Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, LT-08663, Vilnius, Lithuania
Emails: $^{(a)}$romas.baronas@mif.vu.lt, $^{(b)}$darius.baronas@mif.vu.lt

ABSTRACT

This paper deals with amperometric biosensors acting in the flow injection mode when the biosensor contacts with an analyte for a short time. A biosensor-based analytical system is mathematically modeled by reaction-diffusion equations containing a non-linear term related to the Michaelis-Menten kinetics of an enzymatic reaction. The model involves four regions: the enzyme layer where enzymatic reaction as well as the mass transport by diffusion take place, a dialysis membrane and a diffusion limiting region where only the diffusion take place, and a convective region where the analyte concentration is maintained constant. The system of equations was solved numerically by using the finite difference technique. The biosensor operation is analyzed with a special emphasis to the effect of the dialysis membrane on the biosensor response. The biosensor sensitivity is investigated by altering the model parameters influencing the thickness of the dialysis membrane and the catalytic activity of the enzyme. The half maximal effective concentration of the analyte is used as a main characteristic of the sensitivity and the calibration curve of the biosensor.

Keywords: modeling, reaction-diffusion, biosensor, flow injection analysis.

1. INTRODUCTION

Biosensors are analytical devices mainly used for measuring concentrations of analytes (substrates). Main parts composing a biosensor, a biologically active substance, usually an enzyme, and a physicochemical transducer are combined to convert a biochemical reaction result to a measurable quantity (Gutfreund 1995; Turner et al. 1990; Scheller and Schubert 1992). Amperometric biosensors measure changes in the current on the working electrode due to the direct oxidation or reduction of chemical reaction products. The measured current is usually proportional to the concentration of the analyte (substrate). The amperometric biosensors are relatively cheap, sensitive and reliable devices for clinical diagnostics, drug detection, food analysis and environment monitoring (Wollenberger et al. 1997; Gruhl et al. 2011; Viswanathan et al. 2009).

Amperometric biosensors are rather often combined with the flow injection analysis (FIA) for on-line monitoring of raw materials, product quality and the manufacturing process (Ruzicka and Hansen 1988; Mello and Kubota 2002; Nenkova et al. 2010). In the FIA a biosensor contacts with the substrate for short time (seconds to tens of seconds) whereas in the batch analysis the biosensor remains immersed in the substrate solution for a long time (Ruzicka and Hansen 1988). Compared to the batch systems, the FIA systems present the advantages of the reduction in analysis time allowing a high sample throughput, and the possibility to work with small volumes of the substrate (Prieto-Simon et al. 2006; Hernandez et al. 2013). The FIA arrangement also presents a wide response range and high sensitivity (Prieto-Simon et al. 2006).

To improve the efficiency of the development of a novel biosensor as well as to optimize its configuration it is of crucial important to model the biosensor action (Bartlett and Whitaker 1987; Schulmeister 1990; Amatore et al. 2006; Lyons 2009). Biosensors acting in the FIA mode have been already modeled usually at internal diffusion limitations by ignoring the external diffusion (Zhang et al. 2001; Baronas et al. 2002). However, in practical biosensing systems, the mass transport outside the enzyme region is of crucial importance, and it has to be taken into consideration when modeling the biosensor action (Lyons 2009). Recently, the mechanisms controlling the sensitivity of amperometric biosensors acting in FIA mode were numerically modeled taking into consideration the external mass transport (Baronas et al. 2011). The mass transport by diffusion is especially important when dialysis membranes are applied for development of highly stable and sensitive biosensors (Baronas et al. 2010).

The goal of this investigation was to develop a computational model for an effective simulation of the action of an amperometric biosensor containing a dialysis membranes and utilizing FIA as well as to investigate the influence of the physical and kinetic parameters on the biosensor response. The biosensing system was mathematically modeled by reaction-diffusion equations containing a non-linear term related to the Michaelis-Menten kinetics of an enzymatic reaction (Bartlett and Whitaker 1987; Schulmeister 1990). The system of equations was solved numerically by using the finite difference technique (Baronas et al. 2010; Britz 2005). The biosensor operation was analyzed with a special emphasis to the effect of the dialysis membrane on the biosensor response. The biosensor sensitivity was investigated by altering the model parameters influencing the thickness of the dialysis membrane and the catalytic activity of the enzyme. The half maximal effective concentration of the analyte was used as a main characteristic of the sensitivity and the calibration curve of the biosensor (Bisswanger 2008).

2. BIOSENSOR STRUCTURE

The biosensor to be modeled has a layered structure (Simelevicius et al. 2012). Figure 1 shows a principal structure of the biosensor. The biosensor is considered as an electrode with a relatively thin layer of an enzyme (enzyme membrane) entrapped on the surface of the electrode by applying a dialysis membrane. The biosensor model involves four regions: the enzyme layer where the enzyme reaction as well as the mass transport by diffusion take place, a dialysis membrane and a diffusion limiting region where only the mass transport by diffusion take place, and a convective region where the analyte concentration is maintained constant.

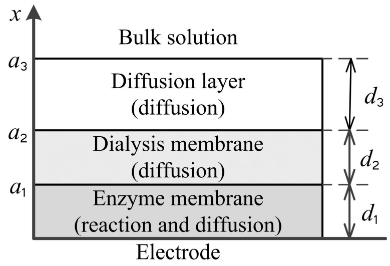


Figure 1: Structural Scheme of the Biosensor

In the enzyme layer we consider the enzyme-catalyzed reaction

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}, \tag{1}$$

where the substrate (S) combines reversibly with an enzyme (E) to form a complex (ES). The complex then dissociates into the product (P) and the enzyme is regenerated (Gutfreund 1995; Scheller and Schubert 1992).

Assuming the quasi steady-state approximation, the concentration of the intermediate complex (ES) does not change and may be neglected when modeling the biochemical behavior of biosensors (Turner et al. 1990; Scheller and Schubert 1992; Segel and Slemrod 1989). In the resulting scheme, the substrate (S) is enzymatically converted to the product (P),

$$\mathrm{S} \xrightarrow{\mathrm{E}} \mathrm{P} \tag{2}$$

It was assumed that $x = 0$ represents the surface of the electrode, a_1, a_2 and a_3 denote the distances from the electrode surface, while d_1, d_2 and d_3 are the thicknesses of the enzyme, the dialysis membrane and the diffusion layers, respectively, $a_i = a_{i-1} + d_i$, $i = 1, 2, 3$, and $a_0 = 0$. The outer diffusion layer ($a_2 < x < a_3$) may be treated as the Nernst diffusion layer (Britz 2005). According to the Nernst approach a layer of thickness $d_3 = a_3 - a_2$ remains unchanged with time. It was assumed that away from it the buffer solution is uniform in concentration.

3. MATHEMATICAL MODEL

Assuming a homogeneous distribution of the enzyme in the enzyme layer of the uniform thickness and symmetrical geometry of the dialysis membrane leads to the mathematical model of the biosensor action defined in a one-dimensional-in-space domain (Schulmeister 1990; Baronas et al. 2010).

3.1. Governing equations

Coupling the enzyme-catalyzed reaction (2) in the enzyme layer with the mass transport by diffusion, described by Fick's law, leads to the following system of the reaction-diffusion equations ($t > 0$):

$$\frac{\partial S_1}{\partial t} = D_{S_1} \frac{\partial^2 S_1}{\partial x^2} - \frac{V_{\max} S_1}{K_M + S_1}, \tag{3a}$$

$$\frac{\partial P_1}{\partial t} = D_{P_1} \frac{\partial^2 P_1}{\partial x^2} + \frac{V_{\max} S_1}{K_M + S_1}, \quad x \in (0, a_1), \tag{3b}$$

where x and t stand for space and time, S_1 and P_1 are the concentrations of the substrate (S) and the product (P) in the enzyme layer, D_{S_1}, D_{P_1} are the constant diffusion coefficients, $V_{\max}$ is the maximal enzymatic rate attainable with that amount of the enzyme, when the enzyme is fully saturated with the substrate, K_M is the Michaelis constant, and $d_1 = a_1$ is the thickness of the enzyme layer (Kulys 1981; Bartlett and Whitaker 1987; Schulmeister 1990). The Michaelis constant K_M is the concentration of the substrate (S) at which the reaction rate is half its maximum value $V_{\max}$. K_M is an approximation of the enzyme affinity for the substrate based on the rate constants within the reactions (1), $K_M = (k_{-1} + k_2)/k_1$.

Outside the enzyme layer, only the mass transport by diffusion of the substrate as well as the product takes place ($t > 0$),

$$\frac{\partial S_i}{\partial t} = D_{S_i} \frac{\partial^2 S_i}{\partial x^2}, \tag{4a}$$

$$\frac{\partial P_i}{\partial t} = D_{P_i} \frac{\partial^2 P_i}{\partial x^2}, \quad x \in (a_{i-1}, a_i),\ i = 2, 3, \tag{4b}$$

where S_i and P_i are the substrate and the product concentrations in the i-th layer, D_{S_i} and D_{P_i} are the diffusion coefficients, and $d_i = a_i - a_{i-1}$ is the thickness of the corresponding layer, $i = 2, 3$.

3.2. Initial conditions

The biosensor operation starts when the substrate appears in the bulk solution. This leads to the following initial conditions ($t = 0$):

$$S_1(x,0) = 0, \quad P_1(x,0) = 0, \quad x \in [0, a_1], \tag{5a}$$

$$S_2(x,0) = 0, \quad P_2(x,0) = 0, \quad x \in [a_1, a_2], \tag{5b}$$

$$S_3(x,0) = \begin{cases} 0, & x \in [a_2, a_3), \\ S_0, & x = a_3, \end{cases} \tag{5c}$$

$$P_3(x,0) = 0, \quad x \in [a_2, a_3], \tag{5d}$$

where S_0 is the substrate concentration in the bulk solution.

3.3. Boundary conditions

During the biosensor operation, the substrate penetrates through the diffusion layer as well as the dialysis membrane and reaches farther boundary of the enzyme layer ($x = a_1$). On the boundary between two adjacent regions having different diffusivities, the matching conditions have to be defined ($t > 0, i = 1, 2$):

$$D_{S_i} \frac{\partial S_i}{\partial x}\bigg|_{x=a_i} = D_{S_{i+1}} \frac{\partial S_{i+1}}{\partial x}\bigg|_{x=a_i}, \tag{6a}$$

$$S_i(a_i, t) = S_{i+1}(a_i, t), \tag{6b}$$

$$D_{P_i} \frac{\partial P_i}{\partial x}\bigg|_{x=a_i} = D_{P_{i+1}} \frac{\partial P_{i+1}}{\partial x}\bigg|_{x=a_i}, \tag{6c}$$

$$P_i(a_i, t) = P_{i+1}(a_i, t). \tag{6d}$$

These conditions mean that fluxes of the substrate and the product through one region are equal to the corresponding fluxes entering the surface of the neighboring region. Concentrations of the substrate and the product in one region versus the neighboring region are assumed to be equal.

Due to the electrode polarization the concentration of the reaction product at the electrode surface is permanently reduced to zero (Schulmeister 1990; Baronas et al. 2010),

$$P_1(0, t) = 0. \tag{7}$$

Due to the substrate electro-inactivity, the substrate concentration flux on the electrode surface equals zero,

$$\frac{\partial S_1}{\partial x}\bigg|_{x=0} = 0. \tag{8}$$

According to the Nernst approach the layer of the thickness d_3 of the outer diffusion layer remains unchanged with time, and away from it the solution is uniform in the concentration (Britz 2005). In the FIA mode of the biosensor operation, the substrate appears in the bulk solution only for a short time period called the injection time (Ruzicka and Hansen 1988). Later, the substrate disappears from the bulk solution,

$$P_3(a_3, t) = 0, \quad t > 0, \tag{9a}$$

$$S_3(a_3, t) = \begin{cases} S_0, & 0 < t \le T_F, \\ 0, & t > T_F, \end{cases} \tag{9b}$$

where T_F is the injection time.

3.4. Biosensor response

The anodic or cathodic current is measured as a result in a physical experiment. The biosensor current is proportional to the gradient of the reaction product concentration at the electrode surface, i.e. on the boundary $x = 0$. When modeling the biosensor action, due to the direct proportionality of the current to the area of the electrode surface, the current is often normalized with that area (Schulmeister 1990; Baronas et al. 2010). The density $I(t)$ of the biosensor current at time t can be obtained explicitly from Faraday's and Fick's laws (Schulmeister 1990),

$$I(t) = n_e F D_{P_1} \frac{\partial P_1}{\partial x}\bigg|_{x=0}, \tag{10}$$

where n_e is a number of electrons involved in a charge transfer, and F is the Faraday constant.

We assume that the system reaches equilibrium when $t \to \infty$. The steady-state current is usually assumed to be the main characteristic of commercial amperometric biosensors acting in the batch mode (Gutfreund 1995; Turner et al. 1990; Scheller and Schubert 1992). In the FIA, due to the zero concentration of the surrounding substrate at $t > T_F$, the steady-state current falls to zero, $I(t) \to 0$, when $t \to \infty$. Because of this, the maximum peak current is the mostly used characteristic in FIA systems,

$$I_{\max} = \max_{t>0} \{I(t)\}, \tag{11}$$

where $I_{\max}$ is the maximal density of the biosensor current.

The corresponding time $T_{\max}$ of the maximal current is used to characterize the response time of the biosensor,

$$T_{\max} = \{t : I(t) = I_{\max}\}. \tag{12}$$

3.5. Characteristics of Biosensor Response

The sensitivity is one of the most important characteristics of the biosensor operation (Gutfreund 1995; Turner et al. 1990; Scheller and Schubert 1992). The sensitivity B_S of the biosensor acting in the FIA mode is defined as the gradient of the maximal current with respect to the concentration S_0 of the substrate in the bulk (Schulmeister 1990; Baronas et al. 2010). Since the biosensor current as well as the substrate concentration vary even in orders of magnitude, a dimensionless expression of the sensitivity is preferable (Baronas et al. 2010). The dimensionless sensitivity $B_S(S_0)$ for the substrate concentration S_0 is given by

$$B_S(S_0) = \frac{\mathrm{d}I_{\max}(S_0)}{\mathrm{d}S_0} \times \frac{S_0}{I_{\max}(S_0)}, \tag{13}$$

where $I_{\max}(S_0)$ is the density of the maximal biosensor current calculated at the substrate concentration S_0.

In the Michaelis-Menten kinetic model, the Michaelis constant K_M as a characteristic of the biosensor calibration curve is numerically equal to the substrate concentration at which half the maximum rate of the enzyme-catalyzed reaction is achieved (Gutfreund 1995; Scheller and Schubert 1992). Under certain conditions, especially under diffusion limitations for the substrate, the half maximal effective concentration C_{50} of the substrate to be determined is often used to characterize the biosensor calibration curve (Bisswanger 2008). In the case of FIA analysis, C_{50} is defined as the concentration of the substrate at which the response of the biosensor reaches half of the maximal response,

$$C_{50} = \left\{ S_0^* : I_{\max}(S_0^*) = 0.5 \lim_{S_0 \to \infty} I_{\max}(S_0) \right\}, \quad (14)$$

where $I_{\max}(S_0)$ is the maximal density of the biosensor current calculated at the substrate concentration S_0.

A greater value of the half maximal effective concentration C_{50} corresponds to a longer linear part of the calibration curve (Bisswanger 2008). At the substrate concentration S_0 corresponding to a linear part of the calibration curve ($S_0 < C_{50}$) the dimensionless biosensor sensitivity $B_S(S_0)$ is approximately equal to unity (Baronas et al. 2010). The concentration C_{50} well characterizes the overall sensitivity of the biosensor.

In the case of biosensors acting in batch mode and exhibiting the Michaelis-Menten kinetics, the concentration C_{50} is usually called the apparent Michaelis-Menten constant (Stikoniene et al. 2010). It has been shown that, under certain conditions, the apparent Michaelis constant highly depends on the biosensor geometry (Ivanauskas et al. 2008). Also, a substantial increase of the apparent Michaelis constant has been shown at restricted diffusion of the substrate through an outer membrane covering an enzyme layer (Stikoniene et al. 2010). This result appears to be of a high practical interest, since it enables to expand the linear dependence of biosensor response on the substrate concentration towards the higher concentrations under the deep diffusion mode of the biosensor operation, whereas the response time increases not very drastic (Stikoniene et al. 2010). This property is especially attractive for biosensors acting in FIA mode because of a relatively short their response time (Cervini and Cavalheiro 2008; Baronas et al. 2002).

3.6. Dimensionless Model

In order to extract the main governing parameters of the mathematical model, thus reducing a number of model parameters in general, a dimensionless model is often derived (Amatore et al. 2006; Schulmeister 1990). The dimensionless model has been derived by replacing the model parameters as defined in Table 1.

Table 1: Dimensional and Dimensionless Model Parameters ($i = 1, 2, 3$)

Dimensional	Dimensionless
x, cm	$\hat{x} = x/d_1$
a_i, cm	$\hat{a}_i = a_i/d_1$
d_i, cm	$\hat{d}_i = d_i/d_1$
t, s	$\hat{t} = tD_{S_1}/d_1^2$
T_F, s	$\hat{T}_F = T_F D_{S_1}/d_1^2$
S_i, M	$\hat{S}_i = S_i/K_M$
P_i, M	$\hat{P}_i = P_i/K_M$
C_{50}, M	$\hat{C}_{50} = C_{50}/K_M$
D_{S_i}, cm^2/s	$\hat{D}_{S_i} = D_{S_i}$ / D_{S_1}
D_{P_i}, cm^2/s	$\hat{D}_{P_i} = D_{P_i}$ / D_{S_1}
I, A/cm^2	$\hat{I} = Id_1/(n_e F D_{P_1} K_M)$

For the enzyme layer, the reaction-diffusion equations (3) can be rewritten as follows ($\hat{t} > 0$):

$$\frac{\partial \hat{S}_1}{\partial \hat{t}} = \frac{\partial^2 \hat{S}_1}{\partial \hat{x}^2} - \alpha^2 \frac{\hat{S}_1}{1 + \hat{S}_1}, \quad (15a)$$

$$\frac{\partial \hat{P}_1}{\partial \hat{t}} = \hat{D}_{P_1} \frac{\partial^2 \hat{P}_1}{\partial \hat{x}^2} + \alpha^2 \frac{\hat{S}_1}{1 + \hat{S}_1}, \quad \hat{x} \in (0, 1), \quad (15b)$$

where α^2 is the diffusion module, also known as Damköhler number (Schulmeister 1990),

$$\alpha^2 = \frac{d_1^2 V_{\max}}{D_{S_1} K_M}. \quad (16)$$

The diffusion module α^2 compares the rate of the enzyme reaction ($V_{\max}/K_M$) with the rate of the mass transport through the enzyme layer (D_{S_1}/d_1^2).

The diffusion equations (4) are transformed as follows ($\hat{t} > 0$):

$$\frac{\partial \hat{S}_i}{\partial \hat{t}} = \hat{D}_{S_i} \frac{\partial^2 \hat{S}_i}{\partial \hat{x}^2}, \quad (17a)$$

$$\frac{\partial \hat{P}_i}{\partial \hat{t}} = \hat{D}_{P_i} \frac{\partial^2 \hat{P}_i}{\partial \hat{x}^2}, \quad \hat{x} \in (\hat{a}_{i-1}, \hat{a}_i), \quad i = 2, 3. \quad (17b)$$

The initial conditions (5) take the following form ($i = 1, 2$):

$$\hat{S}_i(\hat{x}, 0) = 0, \quad \hat{P}_i(\hat{x}, 0) = 0, \quad \hat{x} \in [\hat{a}_{i-1}, \hat{a}_i], \quad (18a)$$

$$\hat{S}_3(\hat{x}, 0) = \begin{cases} 0, & \hat{x} \in [\hat{a}_2, \hat{a}_3), \\ \hat{S}_0, & \hat{x} = \hat{a}_3, \end{cases} \quad (18b)$$

$$\hat{P}_3(x, 0) = 0, \quad \hat{x} \in [\hat{a}_2, \hat{a}_3]. \quad (18c)$$

The matching conditions (6) transform to the following conditions ($\hat{t} > 0, i = 1, 2$):

$$\hat{D}_{S_i}\frac{\partial \hat{S}_i}{\partial \hat{x}}\bigg|_{\hat{x}=\hat{a}_i} = \hat{D}_{S_{i+1}}\frac{\partial \hat{S}_{i+1}}{\partial \hat{x}}\bigg|_{\hat{x}=\hat{a}_i}, \tag{19a}$$

$$\hat{S}_i(\hat{a}_i,\hat{t}) = \hat{S}_{i+1}(\hat{a}_i,\hat{t}), \tag{19b}$$

$$\hat{D}_{P_i}\frac{\partial \hat{P}_i}{\partial \hat{x}}\bigg|_{\hat{x}=\hat{a}_i} = \hat{D}_{P_{i+1}}\frac{\partial \hat{P}_{i+1}}{\partial \hat{x}}\bigg|_{\hat{x}=\hat{a}_i}, \tag{19c}$$

$$\hat{P}_i(\hat{a}_i,\hat{t}) = \hat{P}_{i+1}(\hat{a}_i,\hat{t}). \tag{19d}$$

The boundary conditions (7)-(9) take the following form ($\hat{t} > 0$):

$$\hat{P}_1(0,\hat{t}) = 0, \quad \frac{\partial \hat{S}_1}{\partial \hat{x}}\bigg|_{\hat{x}=0} = 0, \tag{20a}$$

$$\hat{P}_3(\hat{a}_3,\hat{t}) = 0, \tag{20b}$$

$$\hat{S}_3(\hat{a}_3,\hat{t}) = \begin{cases} \hat{S}_0, & \hat{t} \le \hat{T}_F, \\ 0, & \hat{t} > \hat{T}_F. \end{cases} \tag{20c}$$

The dimensionless current (flux) $\hat{I}$ is defined as follows:

$$\hat{I}(\hat{t}) = \frac{\partial \hat{P}_1}{\partial \hat{x}}\bigg|_{\hat{x}=0} = \frac{I(t)d_1}{n_e F D_{P_1} K_M}. \tag{21}$$

Assuming the same diffusion coefficients of the substrate and the product, the initial set of model parameters reduces to the following aggregate dimensionless parameters: $\hat{d}_2$ - the thickness of the dialysis membrane, $\hat{d}_3$ - the diffusion layer thickness, α^2 - the diffusion module, $\hat{T}_F$ - the injection time, $\hat{S}_0$ - the substrate concentration in the bulk during the injection, and $\hat{D}_{S_i} = D_{S_i}/D_{S_1} = D_{P_i}/D_{P_1} = \hat{D}_{P_i}$ - the ratio of the diffusion coefficient in the dialysis membrane (at $i = 2$) or in the diffusion layer (at $i = 3$) to the corresponding diffusion coefficient in the enzyme layer.

The diffusion module α^2 is one of the most important parameters essentially defining internal characteristics of layered amperometric biosensors (Kulys 1981; Bartlett and Whitaker 1987; Schulmeister 1990; Baronas et al. 2010). The biosensor response is known to be under diffusion control when $\alpha^2 \gg 1$. In the very opposite case, when $\alpha^2 \ll 1$, the enzyme kinetics predominates in the response.

4. NUMERICAl SIMULATION

An exact analytical solution is practically possible because of the nonlinearity of the governing equations of the mathematical model (3)-(10) (Schulmeister 1990; Kernevez 1980). Because of this the initial boundary value problem was solved numerically. Solving the problem, an implicit finite difference scheme was built on a uniform discrete grid (Schulmeister 1990; Baronas et al. 2010; Britz 2005; Britz et al. 2009). The computational model was programmed in the C language (Press et al. 1992).

The mathematical model and the numerical solution were validated using a known analytical solution (Schulmeister 1990). Assuming $T_F \to \infty$ and $d_2 \to 0$ or $d_3 \to 0$, the mathematical model (3)-(10) approaches the two compartment model of the amperometric biosensor acting in the batch mode (Schulmeister 1990). The three compartment model approaches the two compartment model also in the unrealistic case where the diffusion coefficients for the dialysis membrane are assumed to be the same as for the diffusion layer, $D_{S_2} = D_{S_3}$ and $D_{P_2} = D_{P_3}$. Additionally assuming $S_0 \ll K_M$, the nonlinear Michaelis-Menten reaction function in (3) simplifies to a linear function $V_{\max}S_1/K_M$. At these assumptions the model (3)-(10) has been solved analytically (Schulmeister 1990). At the steady-state conditions the relative difference between numerical and analytical solutions was less than 1%.

To investigate the effect of the dialysis membrane on the biosensor response, a number of experiments were carried out, while values of some parameters were kept constant (Gough and Leypoldt 1979; van Stroe-Blezen et al. 1993),

$$\begin{aligned} &K_M = 100\mu\text{M}, \quad D_{S_1} = D_{P_1} = 300\,\mu\text{m}^2/\text{s}, \\ &D_{S_2} = D_{P_2} = 0.3D_{S_1}, \quad D_{S_3} = D_{P_3} = 2D_{S_1}, \\ &n_e = 1, \quad d_1 = 200\,\mu\text{m}, \quad d_3 = 20\,\mu\text{m}. \end{aligned} \tag{22}$$

To minimize the effect of the Nernst diffusion layer on the biosensor response, the responses were simulated at a practically minimal thickness ($d_3 = 20\,\mu$m) of the external diffusion layer assuming well stirred buffer solution by a magnetic stirrer (Gough and Leypoldt 1979).

Figure 2 shows the evolution of the density $I(t)$ of the biosensor current simulated at a moderate concentration S_0 of the substrate ($S_0 = K_M$) and different values of the other model parameters: the maximal enzymatic rate $V_{\max}$ (0.75 and 1.5 μM), the injection time T_F (3 and 6 s) and the thickness d_2 of the dialysis membrane (10 and 20 μm). Assuming (22), these two values of the maximal enzymatic rate $V_{\max}$ correspond to the following two values of the dimensionless diffusion module α^2: 1 and 2. Accordingly, d_2 = 10 μm corresponds to the dimensionless relative thickness $\hat{d}_2$ of the dialysis membrane equal to 0.05, while d_2 = 20 μm leads to $\hat{d}_2 = 0.1$.

Figure 2 shows a non-monotonic behavior of the biosensor current. In all the simulated cases the current increases with increasing time t up to the injection time T_F ($t \le T_F$). However, the current also increases some time after the substrate disappearing from the bulk solution ($t \ge T_F$). The time moment $T_{\max}$ of the peak current and the peak current $I_{\max}$ depend on the model parameters: $V_{\max}$, T_F and d_2. In all the simulated cases, the time moment of the peak current was greater than T_F ($T_{\max} > T_F$).

As one can see in figure 2 that at different values of

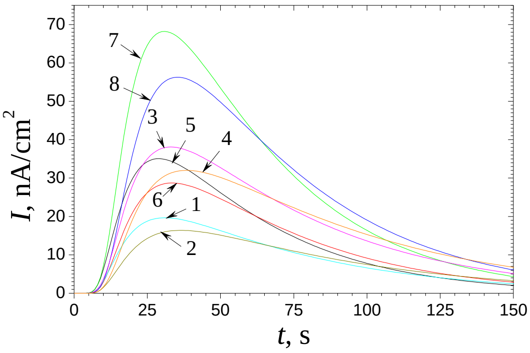


Figure 2: Dynamics of the Biosensor Response; $V_{\max}$: 0.75 (1-4), 1.5 μM (5-8), T_F: 3 (1, 2, 5, 6), 6 s (3, 4, 7, 8); d_2: 10 (1, 3, 5, 7), 20 μm (2, 4, 6, 8)

the model parameters $V_{\max}$ and d_2, the density $I_{\max}$ of the maximal current increases almost two times when the injection time T_F doubles. However, the influence of the doubling the time T_F on the time of the maximal current is rather slight. When comparing curves 1 ($T_F = 3$) and 3 ($T_F = 6\,s$) one can see that the time $T_{\max}$ of the maximal response increases from 31 only to 33 s, while $I_{\max}$ increases from 19.7 event to 38 $\mathrm{nA/cm^2}$ at $V_{\max}$ = 0.75 μM ($\alpha^2 = 2$) and d_2 = 10 μm ($\hat{d}_2 = 0.1$).

Figure 2 also shows that the biosensor response noticeably depends on the thickness d_2 of the dialysis membrane. An increase in d_2 prolongs the time of the maximal current. As one can see in figure 2 that the maximal current decreases when the thickness d_2 of the dialysis membrane increases. FIA biosensing systems have been already investigated by using mathematical models at zero thickness of the dialysis membrane (Baronas et al. 2002, 2011). Figure 2 visually substantiates the importance of the dialysis membrane.

5. RESULTS AND DISCUSSION

Using the numerical simulation, the biosensor action was analysed with a special emphasis to the conditions at which the biosensor sensitivity can be increased and the calibration curve can be prolonged by changing the biosensor geometry (especially the thickness of the dialysis membrane), the injection duration, and the catalytic activity of the enzyme. In order to investigate the influence of the model parameters on the half maximal effective concentration C_{50} of the substrate the simulation was performed at wide ranges of the values of the thickness d_2 of the dialysis membrane, the diffusion module α^2 and the injection time T_F.

The dimensionless half maximal effective concentration $\hat{C}_{50}$ expresses the relative prolongation (in times) of the calibration curve in comparison with the theoretical Michaelis constant K_M. For the biosensor of a concrete configuration, the concentration C_{50} as well as the apparent Michaelis-Menten constant can be rather easily calculated by multiple simulation of the maximal response changing the substrate concentration S_0 (Baronas et al. 2010, 2011).

Figure 3 shows the dependence of the dimensionless half maximal effective concentration $\hat{C}_{50}$ on the thickness d_2 of the dialysis membrane. The the concentration C_{50} was calculated and then normalized with respect to the Michaelis constant K_M at three values of the diffusion module α^2: 0.1 (curves 1 and 2), 1 (3, 4) and 10 (5, 6), and two practically extreme values of the injection time T_F: 1 (1, 3, 5) and 10 s (2, 4, 6). At all these values of α^2 and T_F, the simulations were performed by changing the thickness d_2 from 5 μm ($\hat{d}_2 = 0.025$) to 40 μm ($\hat{d}_2 = 0.2$).

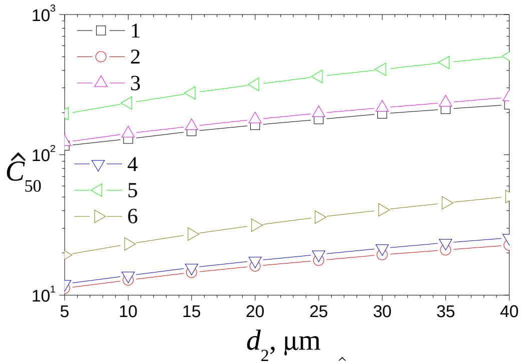


Figure 3: Effective Concentration $\hat{C}_{50}$ vs. Thickness d_2 of the Dialysis Membrane; α^2: 0.1 (1, 2), 1 (3, 4), 10 (5, 6), T_F: 1 (1, 3, 5), 10 s (2, 4, 6)

One can see in figure 3, that the dimensionless half maximal effective concentration $\hat{C}_{50}$ (as well as the corresponding dimensional concentration C_{50}) is a monotonous increasing function of the thickness d_2 of the dialysis membrane. An increase in the thickness d_2 noticeably prologs the linear part of the calibration curve of the biosensor. This can be explained by increasing an addition external diffusion limitation caused by the increasing the thickness of the membrane (Gutfreund 1995; Scheller and Schubert 1992; Ivanauskas et al. 2008; Stikoniene et al. 2010). This figure also shows a significant dependence of C_{50} on the diffusion module α^2 when $\alpha^2 \leq 1$.

To properly investigate the impact of the injection time T_F on the length of the linear part of the calibration curve, the dimensionless half maximal effective concentration $\hat{C}_{50}$ was also calculated by changing T_F from 0.5 up to 10 s. Values of $\hat{C}_{50}$ were calculated at three values of the diffusion module α^2 (0.1, 1 and 10) and two values of the thickness d_2 (10 and 20 μm) of the dialysis membrane. The calculation results are depicted in figure 4.

Figure 4 shows that $\hat{C}_{50}$ approximately exponentially increases with decreasing the injection time T_F. The calibration curve of the biosensor can be prolonged by more than an order of magnitude only by a decrease in the injection time T_F. This impact of T_F on the biosensor sensitivity only slightly depends the thickness d_2 of the dialysis membrane and the diffusion module α^2. A similar effect was also noticed when modeling a

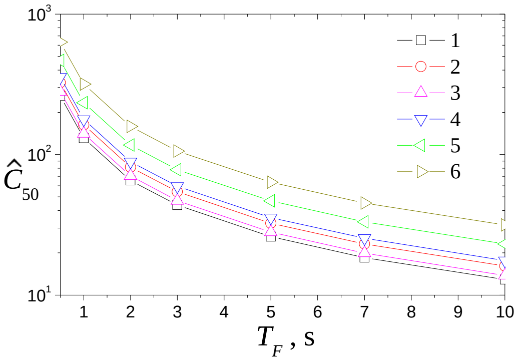


Figure 4: Effective Concentration $\hat{C}_{50}$ vs. Injection Time T_F; α^2: 0.1 (1, 2), 1 (3, 4), 10 (5, 6), d_2: 10 (1, 3, 5), 20 μm(2, 4, 6)

more simple biosensor containing no dialysis membrane (Baronas et al. 2011).

One can also see in figure 4 that the effective concentration $\hat{C}_{50}$ is noticeably higher at greater values of the diffusion module α^2 than at lower ones.

To investigate the impact of the diffusion module α^2 on the effective concentration the biosensor responses were simulated at a wide range of values of α^2. The simulation results are presented in figure 5. The effective concentration $\hat{C}_{50}$ was calculated at two values of the thickness d_2 (10 and 20 μm) of the dialysis membrane and two values of the injection time T_F (1 and 10 s). At concrete values of d_2 and T_F, the calculations were performed by changing the maximal enzymatic rate $V_{\max}$ from 75 nM/s ($\alpha^2 = 0.1$) to 7.5 μM/s ($\alpha^2 = 10$) while keeping the all other parameters constant.

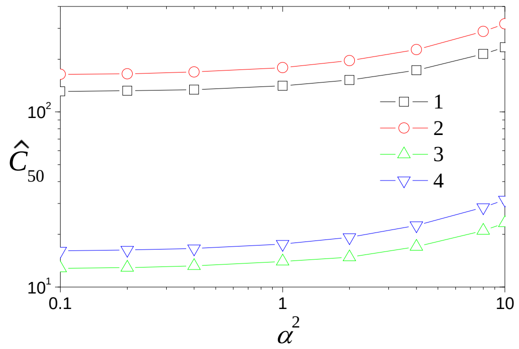


Figure 5: Effective Concentration $\hat{C}_{50}$ vs. Diffusion Module α^2; d_2: 10 (1, 3), 20 μm(2, 4), T_F: 1 (1, 2), 10 s (3, 4)

As one can see in figure 5 that the effective concentration $\hat{C}_{50}$ is a monotonous increasing function of α^2. When the enzyme kinetics predominates in the biosensor response ($\alpha^2 \ll 1$) the concentration $\hat{C}_{50}$ is approximately a constant function. In the opposite case of the biosensor operation when the biosensor response is under diffusion control ($\alpha^2 \gg 1$), the concentration $\hat{C}_{50}$ exponentially increases with an increase in the diffusion module α^2. A similar influence of the diffusion module α^2 to the linear part of the calibration curver was also noticed when modeling the corresponding biosensor with no dialysis membrane (Baronas et al. 2011).

In real applications of biosensors, the diffusion module α^2 can be controlled by changing the maximal enzyme activity $V_{\max}$ as well as the thickness d_1 of the enzyme layer. The maximal enzymatic rate $V_{\max}$ is actually a product of two parameters: the catalytic constant k_2 introduced in (1) and the total concentration of the enzyme (Gutfreund 1995; Scheller and Schubert 1992). Since, in actual applications it is usually impossible to change a value of the constant k_2, the maximal rate $V_{\max}$ as well as the diffusion module α^2 might be changed by changing the enzyme concentration in the enzyme layer.

6. CONCLUSIONS

The mathematical model (3)-(10) of an amperometric biosensor containing a dialysis membrane and utilizing the flow injection analysis can be successfully used to investigate the kinetic peculiarities of the biosensor response. The corresponding dimensionless mathematical model (15)-(21) can be applied to the numerical investigation of the impact of model parameters on the biosensor action and to optimize the biosensor configuration.

By increasing the thickness d_2 of the dialysis membrane, the half maximal effective concentration $\hat{C}_{50}$ can be increased and the linear part of the biosensor calibration curve can be prolonged several fold (see figure 3).

The half maximal effective concentration $\hat{C}_{50}$ approximately exponentially increases with decreasing the injection time T_F. The calibration curve of the biosensor can be prolonged by a few orders of magnitude by decreasing the injection time T_F (figure 4).

The half maximal effective concentration $\hat{C}_{50}$ is a monotonous increasing function of the diffusion module α^2. When the enzyme kinetics distinctly predominates in the response ($\alpha^2 \ll 1$), the $\hat{C}_{50}$ is approximately a constant function of α^2, while at $\alpha^2 \gg 1$ the concentration $\hat{C}_{50}$ exponentially increases with an increase in α^2 (figure 5).

ACKNOWLEDGMENTS

The research by R. Baronas is funded by the European Social Fund under the Global Grant measure, Project No. VP1-3.1-ŠMM-07-K-01-073/MTDS-110000-583. The authors thank participants of the seminar of Vilnius University “BioModa” for valuable discussions.

REFERENCES

Amatore, C., Oleinick, A., Svir, I., da Mota, N., and Thouin, L., 2006. Theoretical modeling and optimization of the detection performance: a new concept for electrochemical detection of proteins in microfluidic channels. *Nonlinear Analysis: Modelling and Control*, 11(4):345–365.

Baronas, D., Ivanauskas, F., and Baronas, R., 2011. Mechanisms controlling the sensitivity of amperometric biosensors in flow injection analysis systems. *Journal of Mathematical Chemistry*, 49(8):1521–1534.

Baronas, R., Ivanauskas, F., and Kulys, J., 2002. Modelling dynamics of amperometric biosensors in batch and flow injection analysis. *Journal of Mathematical Chemistry*, 32(2):225–237.

Baronas, R., Ivanauskas, F., and Kulys, J., 2010. *Mathematical Modeling of Biosensors*. Springer, Dordrecht.

Bartlett, P. N. and Whitaker, R. G., 1987. Electrochemical imobilization of enzymes: Part 1. Theory. *Journal of Electroanalytical Chemistry*, 224(1–2):27–35.

Bisswanger, H., 2008. *Enzyme Kinetics: Principles and Methods*. Wiley-VCH, Weinheim (Germany), 2 edition.

Britz, D., 2005. *Digital Simulation in Electrochemistry*. Springer, Berlin, 3 edition.

Britz, D., Baronas, R., Gaidamauskaitė, E., and Ivanauskas, F., 2009. Further comparisons of finite difference schemes for computational modelling of biosensors. *Nonlinear Analysis: Modelling and Control*, 14(4):419–433.

Cervini, P. and Cavalheiro, É. T. G., 2008. Determination of paracetamol at a graphite-polyurethane composite electrode as an amperometric flow detector. *Journal of the Brazilian Chemical Society*, 19(5):836–841.

Gough, D. and Leypoldt, J., 1979. Membrane-covered, rotated disk electrode. *Analytical Chemistry*, 51:439–444.

Gruhl, F. J., Rapp, B. E., and Länge, K., 2011. Biosensors for diagnostic applications. In *Advances in Biochemical Engineering Biotechnology*, pages 1–34. Springer Berlin Heidelberg.

Gutfreund, H., 1995. *Kinetics for the Life Sciences*. Cambridge University Press, Cambridge.

Hernandez, P., Rodriguez, J. A., Galan, C. A., Castrillejo, Y., and Barrado, E., 2013. Amperometric flow system for blood glucose determination using an immobilized enzyme magnetic reactor. *Biosensors and Bioelectronics*, 41(9–12):244–248.

Ivanauskas, F., Kaunietis, I., Laurinavičius, V., Razumienė, J., and Šimkus, R., 2008. Apparent michaelis constant of the enzyme modified porous electrode. *Journal of Mathematical Chemistry*, 43(4):1516–1526.

Kernevez, J. P., 1980. *Enzyme Mathematics*, volume 10 of *Studies in Mathematics and its Applications*. Elsevier Science, Amsterdam.

Kulys, J., 1981. The development of new analytical systems based on biocatalysts. *Analytical Letters*, 14(6), 377–397.

Lyons, M. E. G., 2009. Transport and kinetics at carbon nanotube - redox enzyme composite modified electrode biosensors. *International Journal of Electrochemical Science*, 4(1):77–103.

Mello, L. and Kubota, L., 2002. Review of the use of biosensors as analytical tools in the food and drink industries. *Food Chemistry*, 77(2):237–256.

Nenkova, R., Atanasova, R., Ivanova, D., and Godjevargova, T., 2010. Flow injection analysis for amperometric detection of glucose with immobilized enzyme reactor. *Biotechnology & Biotechnological Equipment*, 24(3):1986–1992.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK), 2 edition.

Prieto-Simon, B., Campas, M., Andreescu, S., and Marty, J.-L., 2006. Trends in flow-based biosensing systems for pesticide assessment. *Sensors*, 6(10):1161–1186.

Ruzicka, J. and Hansen, E., 1988. *Flow Injection Analysis*. Wiley, New York.

Scheller, F. W. and Schubert, F., 1992. *Biosensors*. Elsevier Science, Amsterdam.

Schulmeister, T., 1990. Mathematical modelling of the dynamic behaviour of amperometric enzyme electrodes. *Selective Electrode Reviews*, 12:203–260.

Segel, L. A. and Slemrod, M., 1989. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review*, 31(3):446–477.

Simelevicius, D., Baronas, R., and Kulys, J., 2012. Modelling of amperometric biosensor used for synergistic substrates determination. *Sensors*, 12(4):4897–4917.

Stikoniene, O., Ivanauskas, F., and Laurinavicius, V., 2010. The influence of external factors on the operational stability of the biosensor response. *Talanta*, 81(4–5):1245–1249.

Turner, A. P. F., Karube, I., and Wilson, G. S., editors, 1990. *Biosensors: Fundamentals and Applications*. Oxford University Press, Oxford.

van Stroe-Blezen, S., Everaerts, F. M., Janssen, L. J. J., and Tacken, R. A., 1993. Diffusion coefficients of oxygen, hydrogen peroxide, and glucose in a hydrogel. *Analytica Chimica Acta*, 273:553–560.

Viswanathan, S., Radecka, H., and Radecki, J., 2009. Electrochemical biosensors for food analysis. *Monatshefte Fur Chemie*, 140(8):891–899.

Wollenberger, U., Lisdat, F., and Scheller, F. W., 1997. *Frontiers in Biosensorics 2, Practical Applications*. Birkhauser Verlag, Basel.

Zhang, S., Zhao, H., and John, R., 2001. Development of a quantitative relationship between inhibition percentage and both incubation time and inhibitor concentration for inhibition biosensors–theoretical and practical considerations. *Biosensors and Bioelectronics*, 16(9–12):1119–1126.

AUTHORS BIOGRAPHY

Romas Baronas was born in 1959 in Kybartai, Lithuania. He enrolled at the Vilnius University, where he studied applied mathematics and received his Ph.D. degree. Now he is a professor and serves as the head of the department of Software Engineering at Vilnius University. His research interests are in database systems, software engineering, and computational modeling of nonlinear phenomena in life sciences.

Darius Baronas is a PhD student at Institute of Mathematics and Informatics of Vilnius University. He graduated that University in 2007. His research interests are focused on computational modeling and optimization of biochemical processes.

MULTI-CONTROLLERS APPROACH APPLIED TO A WRIST OF A ROBOT .

Youcef Zennir[(1)], **Pascal Mouille**[(2)]

[(1)]*Laboratoire d'automatique de Skikda, 26 route El-hadaeik,21000 Skikda, Algérie.*
[(2)]*Polytech Annecy-Chambéry, BP 80439, 74944 Annecy-le-Vieux, France.*

[(1)]youcefzennir@yahoo.com [(2)]Pascal.Mouille@univ-savoie.fr

ABSTRACT

The work presented in this paper focuses on the multi-controller approach of control. In the first time we modeled the process (wrist of a Staubli robot RX 90) and we identified local parametric models around operating points. The originality of our approach lies in the use of an integrator in the process to avoid the use of the operating points in an explicit way in the control law, and the fact that several controllers scanned linear type RST with the delta operator (δ), one hand a working (free switching) to develop the control signal sent to the process. We present the results obtained in the different simulations before opening perspectives for future work.

Keywords: Modelling, Identification, Local Control, Multi-controller control, free commutation, adaptive control

1. INTRODUCTION

Invariant linear model for a physical process can only be an approximation. Indeed, a physical process generally has non-linearities (Slotine 1991) that are not taken into account in the modeling process. For some operating points of the physical process can be determined a local model linear. These linear models can be derived from a priori knowledge of the process or be derived from an identification step. We may then seek to enslave the whole process in operational space using the local information. The objectives of this work are to introduce an integrator in the process and develop a command structure in which control laws together several local synthesized from local models of the process. The purpose of the multi-controller command (Balakrishnan 1997) is to control the output of any process in space operation using controls developed by different local controller use the multi-controller command is to specify:

- The structures of the controllers used.
- The type of switching (Pagès 2000; Duchamp 1998).
- The method of working of the control law.

Different solutions are proposed for the control law such as:

- Use controllers of RST type.
- Use of adaptive controllers.
- Use free or fuzzy commutation (Pagès 2000; Foulloy 1998).
- Use direct or indirect approach to collaboration control law.

In our work we have chosen the solution is to use an indirect approach based on local controllers and switching straightforward.

2. PROCESS MODELING

The process can be represented by the following figure:

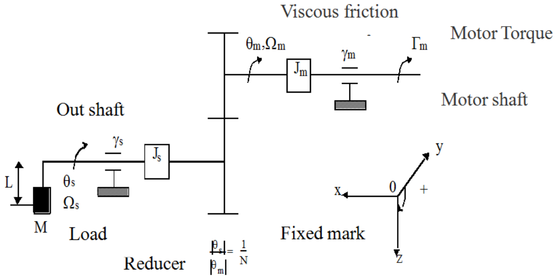


Figure 1: Process model.

It corresponds to a robot wrist (one axis). It is composed of a drive shaft and an output shaft connected by a reducing agent. The output shaft drives a mechanical load. By applying the fundamental law of mechanics (rotation) to the motor shaft and the output shaft we obtain the following equation:

$$\Gamma_m + \frac{M \cdot g \cdot L}{N} \cdot \sin(\theta_s) = J_t \cdot \ddot{\theta}_m + \gamma_t \cdot \dot{\theta}_m \qquad (1)$$

With:

$$J_t = J_m + \frac{J_s^{'}}{N^2} \qquad (2)$$

J_t: Moment of inertia reduced to the motor shaft. J_s: total moment of inertia of the output shaft (output shaft over the mechanical load).

$$\gamma_t = \gamma_m + \frac{\gamma_s}{N^2} \qquad (3)$$

γ_t: total Viscous friction reduced to the motor shaft. The motor torque is given by:

$$\Gamma_m = K_e \cdot u \qquad (4)$$

With : K_e: torque constant, u: control voltage.
The nonlinear model is:

$$X_1 = \theta_m(t);\ X_2 = \dot{\theta}_m(t)\ ;\ \ X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad (5)$$

$$\dot{X} = \begin{bmatrix} 0 & 1 \\ \frac{M \cdot g \cdot L}{N \cdot J_t} & -\frac{\gamma_t}{J_t} \end{bmatrix} \cdot \begin{bmatrix} \sin\left(\frac{X_1}{N}\right) \\ X_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{K_e}{J_t} \end{bmatrix} \cdot u \qquad (6)$$

$$Y = \begin{bmatrix} -\frac{1}{N} & 0 \end{bmatrix} \cdot X \qquad (7)$$

This is a model of a nonlinear system affine. To find the structure of local parametric models, we used the tangent linearization and the linear model is as follows:

$$\delta\dot{X} = \begin{bmatrix} 0 & 1 \\ -\frac{M \cdot g \cdot L}{N^2 \cdot J_t} \cdot \cos\left(\frac{X_{10}}{N}\right) & -\frac{\gamma_t}{J_t} \end{bmatrix} \cdot \delta X + \begin{bmatrix} 0 \\ \frac{K_e}{J_t} \end{bmatrix} \cdot \delta u \qquad (8)$$

$$\delta Y = \begin{bmatrix} -\frac{1}{N} & 0 \end{bmatrix} \cdot \delta X \qquad (9)$$

The transfer function G (p) of the process corresponds the linear model is given by the following formula:

$$G(p) = \frac{-K_p}{p^2 + a_{p1} \cdot p + a_{p2}} \qquad (10)$$

$$\text{With: } K_p = \frac{K_e}{N \cdot J_t};\ a_{p1} = \frac{\gamma_t}{J_t};\ a_{p2} = \frac{M \cdot g \cdot L}{N^2 \cdot J_t} \cdot \mathrm{COS}\left(\frac{X_{10}}{N}\right) \qquad (11)$$

For identify around each operating point considered a linear model of order two, we place the process around the operating point (u_0=0, X_{10}=0) and we excite the process with the following signal:

$$u(t) = 0.2 \cdot [\sin(2\pi t) + \sin(4\pi t) + \sin(8\pi t)] \qquad (12)$$

This command is blocked sampled at the frequency of 1 KHz before being sent to the process. After the identification we obtained the following discrete model:

$$G(z) = \frac{-0.0001109 \cdot z}{z^2 - 1.989 \cdot z + 0.9888} \qquad (13)$$

This model corresponds to discrete continuous model as follows:

$$G(p) = \frac{-0.05566 \cdot p - 111.5}{p^2 + 11.25 \cdot p + 79.14} \qquad (14)$$

We see that the coefficient (-0.05566) is small, more the process model (10) contains no zeros. So we remove this factor and are taken as the continuous model:

$$G(p) = \frac{-111.5}{p^2 + 11.25 \cdot p + 79.14} \qquad (15)$$

We deduce:

$$\begin{cases} K_p = 111.5;\ a_{p1} = 11.25; \\ a_{p2} = C_1 \cdot \mathrm{COS}\left(\frac{X_{10}}{N}\right);\ C_1 = 79.14 \end{cases} \qquad (16)$$

So, we can deduce the two local models corresponding to the operating point's $\theta_{s0}=\pi/3$ and $\theta_{s0}=2\pi/3$ respectively:

$$G(p) = \frac{-111.5}{p^2 + 11.25 \cdot p + 39.57}\ ;\ \ G(p) = \frac{-111.5}{p^2 + 11.25 \cdot p - 39.57} \qquad (17)$$

3. INTRODUCTION OF AN INTEGRATOR

Interest of the integrator does no longer have to introduce the operating points (u0, y0) in an explicit way to compute the control laws. Indeed, it is he who will give the nominal control and guarantee the performance static. The integrator is arranged as follows:

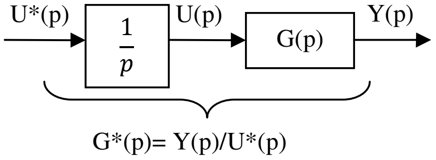


Figure 2: Process with Integrator

From this diagram, it is assumed that the integrator is in the process. So we consider we have a new transfer function G * (p):

$$G^*(p) = \frac{-K_p}{p \cdot (p^2 + a_{p1} \cdot p + a_{p2})} \qquad (19)$$

4. LOCAL CONTROLLERS STRUCTURE

The structure of the local controllers is of type (RST). The command is a command used by the reference model and output feedback (Chebassier 1999; Balakrishnan 1994).We choose the parameters of the reference model for the latter as follows:

$$Y_m(p) = \frac{\gamma^3}{(p+\gamma)^3} \cdot R(p) = G_m(p) \cdot R(p) \qquad (20)$$

With: R (p) is the set of the loop closes. In this case, we can represent the local controller by the following block diagram:

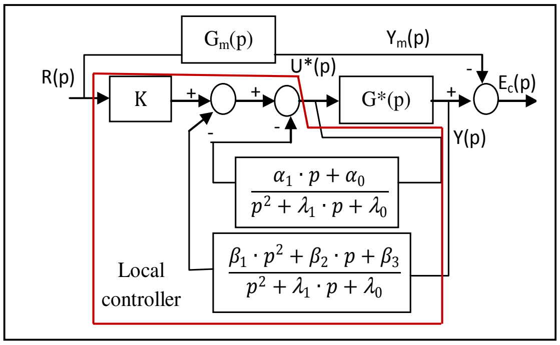

G$_m$(p)
Y$_m$(p)
R(p)
U*(p)
K
+
+
G*(p)
-
+
E$_c$(p)
-
-
Y(p)
$\frac{\alpha_1 \cdot p + \alpha_0}{p^2 + \lambda_1 \cdot p + \lambda_0}$
$\frac{\beta_1 \cdot p^2 + \beta_2 \cdot p + \beta_3}{p^2 + \lambda_1 \cdot p + \lambda_0}$
Local
controller

Figure 3: Structure of local Controller

G * (p): transfer function of the local method with the integrator.

5. THE FIRST SIMULATION

The synthesis of the controllers is continuous. The simulation is done in discrete time operating points are used θ_{s0}=0rad, $\theta_{s0}=\pi/3$rad and $\theta_{s0}=2\pi/3$rad. The parameter values of the reference model λ_0 and λ_1 are:

$$\gamma = 10: \lambda_0 = 900;\ \lambda_1 = 60 \qquad (21)$$

The values λ_0 and λ_1 are chosen so that controllers are stable. The parameters of the controllers around the operating points chosen are:

parameters	K	$\alpha_0(e^{+003})$	α_1
Controller (θ_s=0)	-8.96	1.13	18.75
Controller ($\theta_s=\pi/3$)	-8.96	1.17	18.75
Controller ($\theta_s=2\pi/3$)	-8.96	1.25	18.75
	$B_1(e^{+003})$	B_2	$B_3(e^{+003})$
Controller (θ_s=0)	-8.07	-151.34	-1.51
Controller ($\theta_s=\pi/3$)	-8.07	-157.29	-2.22
Controller ($\theta_s=2\pi/3$)	-8.07	-223.20	-3.72

Table 1: Parameters of the local Controller

Two simulations have been performed for each operating point in order to verify the role of the integrator, the stability of the closed loop and the proper functioning of the controllers around the operating points.

For the validation of the use of the integrator the controller around the operating point θ_{s0}=0rad, the reference signal r (t) is a step of amplitude 0.1rad happens at t=1s. the figure 4.a corresponds to error control and the figure.4.b corresponds to the process and the outputs of reference model.

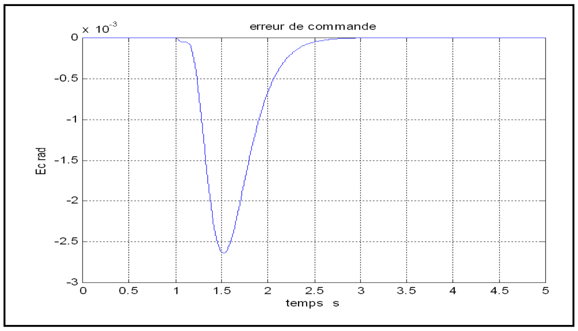


Figure 4.a: Control error e$_c$ (t).

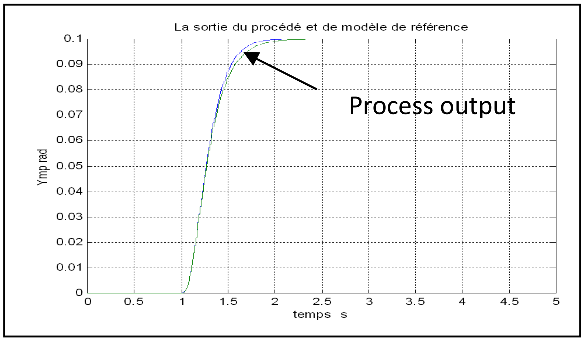


Figure 4.b: outputs of the process and the reference model.

From Figure 4.a, we see that after a transient, the command error tends to zero. From Figure 4.b, we see that the static gain is equal to 1. Thus, we conclude that the integrator guarantees static performance. It was also a good trajectory tracking. We can conclude that the integrator introduced just upstream of the process, plays its role.

For the Operation of the controllers around the operating points the controller around the operating point θ_{s0}=0 rad, the reference signal is equal to:

$$r(t) = 0.1 \cdot \sin(20 \cdot t) \qquad (22)$$

The figure 5.a corresponds to error control and figure 5.b corresponds to the process and the outputs of the reference model.

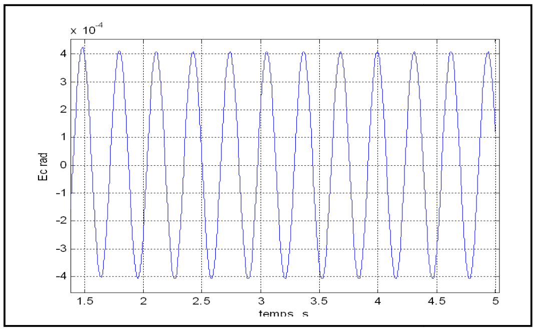


Figure 5.a: control error e_c (t) (rad).

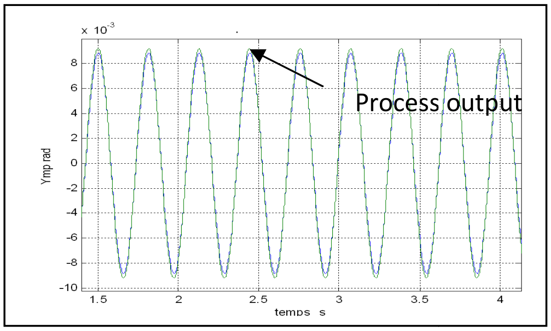


Figure 5.b: outputs of the process and the reference model (rad).

On figure.5.a we see that the control error is very low. We note from the figure.5.b we have a good trajectory tracking of the output of the process compared to the output of the reference model. We can conclude that the local controllers give good performances locally.

6. MULTI-CONTROLLERS STRUCTURE OF CONTROL

The first object of the multi-controllers structure of the control is to control the output of the process in a whole space of variation of parameters under consideration using commands developed by the various local controllers. The diagram is as follows:

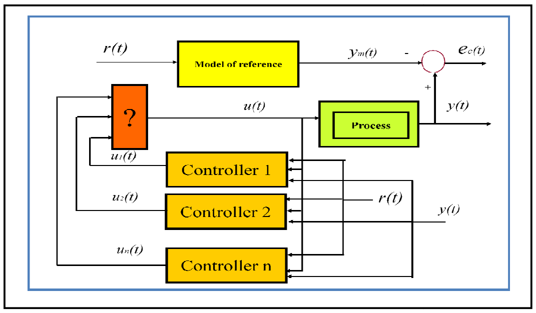


Figure 6: Multi-controller structure of control.

Different solutions are possible to calculate the control to be applied to the process (table 2) and we focus in this study on the use of free switching controllers. That is to say, at each instant a single controller will generate the command to be applied to the process (u (t) = u (t), i = 1 ... n).

Switching or mixing of different orders is overseen by local information about the "distance" between the current state of the process and the different operating points. This information can be measurement (output of the method, for example) or rebuilt (using predictors).

	Frank commutation	**Fuzzy commutation**
Indirect approach based on reconstructed information	**category 1** N linear predictors associated with N linear controllers RST	**category 3** N linear predictors associated with N linear controllers RST
Direct approach commutation based on measured information	**category 2** N linear controllers R.S.T	**category 4** N linear controllers R.S.T

Table 2: Category of multi-controller control.

The work presented in this article relates to first category.

7. INDIRECT APPROACH WITH FRANK COMMUTATION

Control u (t) applied to the process is equal, at every moment, one of the outputs of local controllers. Switching is based on information reconstructed by predictors. These are calculated from the local controllers. The reconstructed information is estimated from the output of the process figure 7. This approach was developed by KS Narendra and J. (Balakrishnan 1997; Toscano 1998; Pagès 2000). Calculating for each predictor an indicator (quadratic criterion) performance defined by the following formula:

$$j_j(t) = \alpha \cdot e_j^2(t) + \beta \cdot \int_0^t e^{(-\lambda\cdot(t-\tau))} \cdot e_j^2(\tau) \cdot d\tau \qquad (23)$$

with : $\alpha \geq 0;\ \beta > 0;\ \lambda > 0$

ej (t) associated with the identification error predictor of index j.
J_j(t): quadratic criterion indicator associated with the performance predictor of index j.

Free switching is based on the following criteria quadratic each time the command is applied to the process equal to the output of the controller associated with the predictor that gives small quadratic criterion. Each local controller is associated with a predictor.

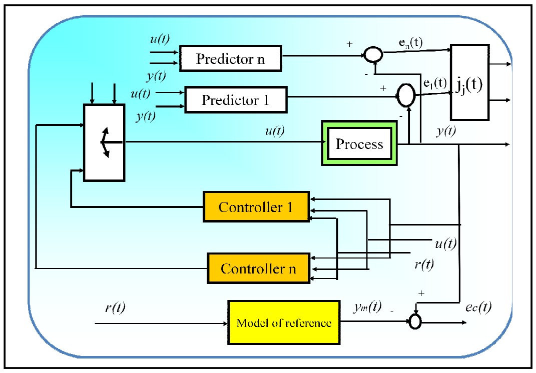


Figure 8: Indirect approach with frank commutation.

We used the predictor proposed by Narendra and Balakrishnan [1] [7], modeled by the following equation:

$$Y(p) = Y_m(p) = G_m(p) \cdot R(p) = G_m(p) \cdot \left[\frac{\varphi(p)}{\lambda(p)} \cdot U^*(p) + \frac{\beta(p)}{\lambda(p)} \cdot Y(p)\right] \qquad (24)$$

Note that the polynomials β(p), φ(p), λ(p) will directly result of the controller.

8. THE SECOND SIMULATION

The synthesis of the predictors is continuous. The simulation is done in discrete time. The operating points are used $\theta_{s0}=0$ rad, $\theta_{s0}=3/\pi$ rad and $\theta_{s0}=2\pi/3$ rad. For predictor around the operating point $\theta_{s0}=0$ rad, the reference signal is equal to:

$$r(t) = 0.1 \cdot \sin(20 \cdot t) \qquad (25)$$

Reference Model: $\gamma = 10$. Quadratic criterion: $\alpha = 1$, $\beta = 4$, $\lambda = 120$. These parameters are selected based on the process dynamics. And we have the following result:

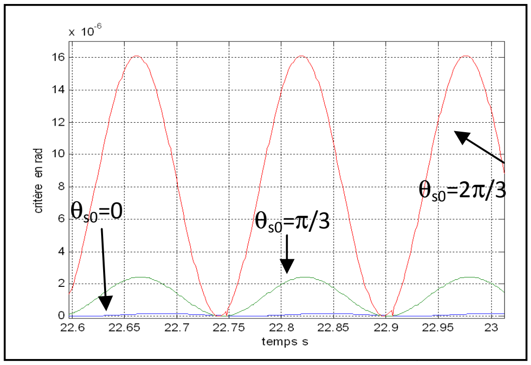


Figure 9: Quadratic criteria.

From the figure 9, we see that the evolution of the quadratic criterion is very low.

In the frank commutation with controllers fixed simulation we use the free switching and the reference signal is:

$$r(t) = \frac{\pi}{3} + \frac{6\pi}{2} \cdot \sin(20 \cdot t) \qquad (26)$$

The simulation is done in discrete time, and the result is shown in the following figures:

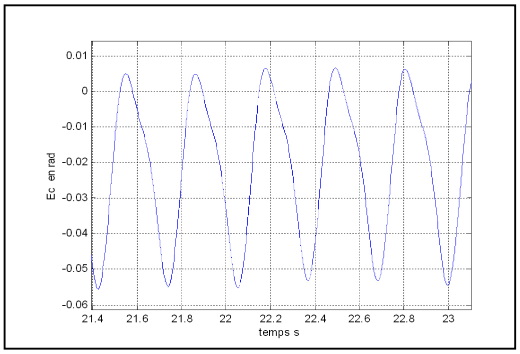


Figure 10: Control error e_c (t) in (rad)

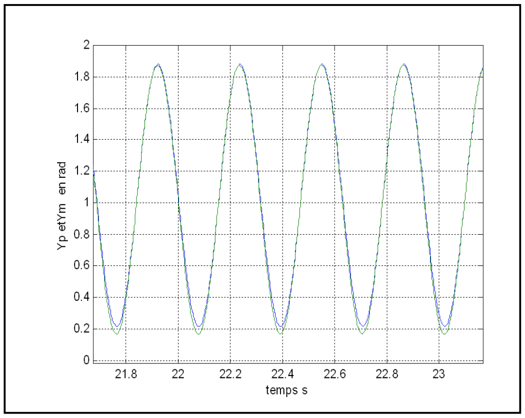


Figure 11: The output of the process and the reference model in (rad).

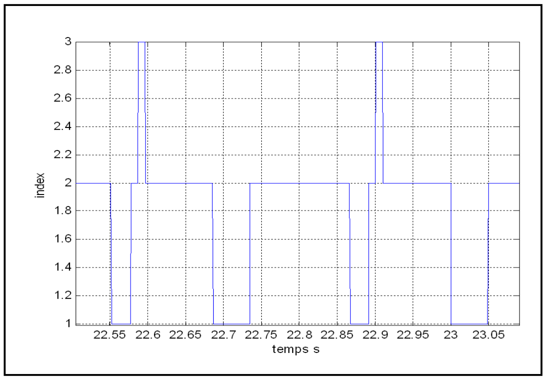


Figure 12: Switching frank local controllers

The figure 10 and figure 11 show the evolution of the error control signal and the output of the process and reference model. The figure 12 shows the commutation signal. We see from the results that the evolution of the criteria is very low. Predictors and the local controllers ensure proper operation of the process around the operating points.

9. RE-INITIALIZED PREDICTOR ADAPTIVE

The re-initialized predictor adaptive with the same structure of the predictor. The parameters of this re-initialized predictor adaptive are re-initialized by the values of the parameters of a fixed predictor who gives the minimum error of the identification (Karimi 1998). The adaptive predictor has as a role to adapt the parameters. The objective to use the re-initialized predictor adaptive is to ensure the good performance of the process between the points of operation.

Considering the fixed predictors give good performances only around the operation points. We used the algorithm of least squares standardized with a factor of lapse of memory to adapt the parameters of the predictors [5]. While using like a reference signal:

$$\mathrm{r(t)} = \frac{\pi}{3} + \frac{6\pi}{2} \cdot \sin(20 \cdot \mathrm{t}) \qquad (27)$$

After simulation wee obtains the following figures:

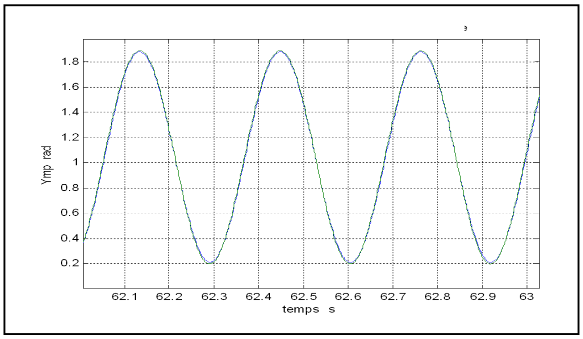


Figure 13: The output of the process and the reference model in (rad).

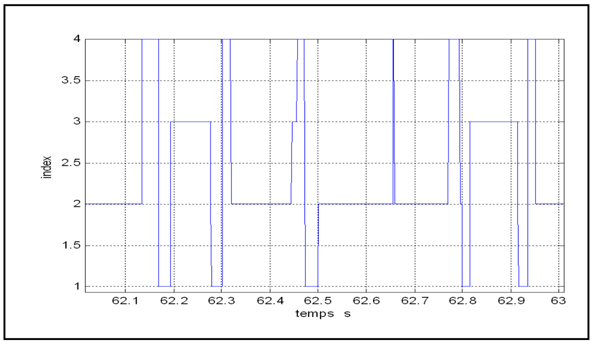


Figure 14: Switching frank local controllers

10. CONCLUSION

In this work, we presented the modeling of nonlinear process. Then, we calculated linear models about operating points considered. Then, we identified the parameters of local linear models. After that, we did a study on the introduction of an integrator.

The results obtained allow concluding that the local controllers give good results around the operating points. But the results are local.

Therefore, we must seek a collaborative approach these local control laws to obtain good results in all operating space. To this end, we presented the different types of control structures multi-controllers with indirect approaches, and the principle of frank switching.

Then we presented the structures of the predictors used. According to the simulation results, we conclude that the predictors fixed premises give good results. It can be seen that the use of the free switching gives acceptable results in the entire space of the system.

It is also concluded that the results obtained by the introduction of a re-initialized predictor adaptive are good compared to the results obtained by frank commutation without the adaptive one.

In the continuation of our work we will study at the first time fuzzy commutation with the indirect approach, then the same category with other type of controller as example numerical PID fractional

REFERENCES

Balakrishnan J. and Narendra S. 1994. *Improving Transient Response of Adaptive Control Systems Using Multiple Models and Switching*. IEEE Trans. on Automatic Control, Vol. 39, n°9, Septembre 1994, pp. 1861-1866.

Balakrishnan J. and Narendra S. 1997. *Adaptive Control using Multiple Models*. IEEE Trans. on Automatic Control, Vol. 42, n°2, Février 1997, pp. 171-187.

Chebassier J.1999. *Méthadologies pour la conception d'un système de commande par calculateur*. Thèse Laboratoire d'Automatique de grenoble (INPG) , 1999.

Duchamp J-M. 1998. *Commutation Floue de lois de Commande applique à la Robotique*. Rapport de DEA , LAMII /CESALP.

Slotine JJE. 1991. *Applied Nonlinear Control*. Prentice-Hall International, ISBN: 0-13-040049.

Karimi A. and Landau I-D. 1998. *Robust Adaptive Control of a Flexible Transmission System Using Multiple Models*. Laboratoire d'Automatique de Grenoble (CNRS-INPG-UJF). Design-CSD, 1998.

Pagès O., Mouille P. and Caron B. 2000. *Two approaches of the multi-model control. Real time Implementation for a wrist of a Robot*. Mechatronics 2000, 1st IFAC Conference on mechatronic systems, Darmstadt, Allemagne, Septembre, 2000.

Pagès O., Mouille P. and Caron B. 2000. *Multi-Model Control by Applying a Symbolic Fuzzy Switcher*. Control Systems Design-CSD 2000, IFAC Conference, Bratislava, République Slovaque, Juin, 2000.

Toscano R., Martin-Calle D. and Passerieu P. 1997. *Adaptation paramétrique floue d'une commande au premier ordre en fonction du point d'équilibre courant.* Laboratoire d'Automatique E.N.I St ETIENNE LFA'97 -Lyon-décembre, pp.3-10.

Foulloy L. and Ramdani M. 1998. *Logique Floue Exercices corrigés et exemples d'applications*. cEpaduEs-EDITIONS, juillet, 1998.

A STUDY ON PERFORMANCE EVALUATION OF THE “DOGS OF THE DOW” INVESTMENT STRATEGY FOR THE THAI STOCK MARKET

Kittipong Tissayakorn[(a)], **Yu Song**[(b)], **Fumio Akagi**[(c)]

[(a)]Graduate student, Graduate School of Engineering, Fukuoka Institute of Technology, Fukuoka, 811-0295, Japan
[(b)]Professor Department of System Management, Fukuoka Institute of Technology, 811-0295, Japan
[(c)]Professor Department of System Management, Fukuoka Institute of Technology, 811-0295, Japan

[(a)]Killua.benz@gmail.com, [(b)]Song@fit.ac.jp, [(c)]Akagi@fit.ac.jp

ABSTRACT

In stock markets, investors constantly seek ways to make profits or outperform benchmarks. However, this goal is not easy to achieve even for professional fund managers. In this study, we purpose applying the “Dogs of the Dow” investment strategy to the Thai market. With this strategy, we buy the ten highest yielding SET50 (Stock Exchange of Thailand 50) stocks and rebalance the portfolio annually. We conduct a simulation for data from 1995 to 2012. The simulation results show that, on average, the “Dogs of the Dow” strategy outperforms the stock market indices. Even after risk adjustment, the “Dogs of the Dow” strategy is still superior the benchmark.

Keywords: the “Dogs of the Dow” strategy, portfolio selection, sharp ratio, the Stock Exchange of Thailand

1. INTRODUCTION

Investors are constantly seeking ways to outperform benchmarks in stock markets. However, it is quite difficult even for professional investors.

In recent years, the “Dogs of the Dow” investment strategy, also known as the Dow 10 strategy, has become widely recognized for its ease of maneuverability and high performance. It is a portfolio selecting strategy devoted to picking the highest dividend stocks from the Dow Jones Industrial Average (DJIA) stocks. The strategy was first proposed by J. Slatter (1988). It involves investing equal amounts in the 10 highest yielding stocks of the DJIA stocks and rebalancing the portfolio every year. The ten stocks are called “dogs”, which means “losers”, because high yields implies that the stocks are not approved by the market. Slatter examined the performance of the strategy for several years and found that it outperformed the DJIA index by 7.6% on an annual basis. Similar results were reported in investment books like (Knowles and Pretty 1992) and (O’Higgins and Downes 1991). These books highlighted the “Dogs of the Dow” strategy and prompted its increasing popularity among both institutional and individual investors.

In this paper, we examine the performance of the “Dogs of the Dow” strategy in a different market setting and during different time periods. In particular, our purpose is to analyze the performance of the “Dogs of the Dow” strategy in the Thai stock market to examine its validity. We implement simulations for data from 1995 to 2011 and compared the performance of the “Dogs of the Dow” strategy with two popular market indices, the SET (Stock Exchange of Thailand) Index and the SET50.

The results of simulations show that the “Dogs of the Dow” strategy outperforms the market indices, though the superiority is not statistically significant, and portfolios with fewer than ten stocks have even better performance than the original ten-stock portfolio.

The remainder of this paper is organized as follows: In the next section, we review the related literature on the “Dogs of the Dow” investment strategy. In Section 3 we briefly introduce the simulation. Section 4 compares performance of the “Dogs of the Dow” with the SET Index. Then, we adjust the risk for the “Dogs of the Dow” investment strategy and make the comparison again (Section 5). Portfolios with other numbers of dogs are described in Section 6. Finally, Section 7 concludes the paper with remarks on the future.

2. LITERATURE REVIEW

2.1. Study on American Market

The “Dogs of the Dow” investment strategy was originally proposed by John Slater (1988). With this strategy, an investor selects the 10 highest dividend yielding stocks from the DJIA stocks at the end of each calendar year and invests equal amounts to each stock. After 1 year, the portfolio is rebalanced and updated with equally weighted investments in the new highest yielding stocks. It was reported that from 1972 to 1987, the average annual return of such a portfolio outperformed the DJIA by 7.6 percentage points. For longer time horizons, O’Higgins and Downes and Knowles and Petty published books to introduce further information regarding the “Dogs of the Dow” strategy in American market. In (O’Higgins and Downes 1991), the authors reported that the average annual return of the “Dogs of the Dow” is 6.2 points higher than the

DJIA during the period 1973 to 1991. Reference (Knowles and Petty 1992) shows that while the “Dogs of the Dow” portfolio had an average annual return of 14.2%, the DJIA only had an average annual return of 10.4% from 1957 to 1990. They also examined an alternative version of the strategy in which the “Dogs of the Dow” portfolio consists of the five highest dividend yielding stocks. The reported average annual return for this five dog stock strategy is 15.4%.

The idea was that the dividend yield was often an inverse indicator of popularity, and that buying Dow stocks when they were temporarily out of favor was a shrewd way to beat the market. The theoretical basis for the strategy could be traced to the theory of corporate dividend policy. Corporations strive to main stable dividend payouts in order to avoid sending undesirable signals to the markets about the company’s future business prospect.

The first academic study on the “Dogs of the Dow” strategy in American stock market was performed by McQueen, Shields and Thorley (1997), whose study of the phenomenon produced mixed results. They found that the “Dow 10” strategy outperformed the “Dow 30” strategy over a period of 50 years from 1946 to 1995 by approximately 3 percentage points. Breaking the sample into five 10-year periods, the authors found that the strategy was successful in each period, however, they argued that the strategy would lose effectiveness after adjustment for risk (in term of company - specific risk from inadequate diversification), transaction costs, and tax treatment. After they incorporated these factors, the Dow 10 strategy’s premium over the Dow 30 shrank to 0.95 percentage points.

2.2. Studies on Other Market

The “Dogs of the Dow” investment strategy has been examined in many other stock markets.

Reference Visscher and Filbeck (1997) examined the “Dogs of the Dow” strategy in the British stock market. The authors simulated UK data from 1985 to 1994 and applied the “Dogs of the Dow” strategy to stocks included in the FTSE100 (Financial Times Stock Exchange 100) index. The “Dogs of the Dow” portfolio was documented to outperform the market index on a risk adjusted basis in only 4 years out of the 10, thereby indicating that the “Dogs of the Dow” strategy was not particularly effective in the UK.

The effectiveness of the “Dogs of the Dow” strategy in the Canadian stock market was focused on the Toronto35 index from 1987-1997 and reported an average annual excess return of 6.6% for the “Dogs of the Dow” portfolio. More importantly, the study showed that the “Dogs of the Dow” strategy produced significantly higher risk adjusted return than the Toronto35 and TSE300 (Toronto Stock Exchange 300) indices and the reported excess returns were also high enough to compensate for the higher taxes and transaction costs (Visscher and Filbeck 2003).

Andre and Silva studied its performance in Latin American stock markets from 1994 to 1999. They found that the “Dogs of the Dow” slightly outperformed the market indices in Argentina, Chile, Colombia, Mexico, Peru and Venezuela, while the strategy seemed to underperform relative to market index in Brazil. Moreover, they conclude that the result lacks statistical significance, probably because of the short test period (Andre and Silva 2001).

Furthermore, Brzeszczynski and Gajdka (2008) focused on the Polish stock market from 1997 to 2007. The study showed that there were important implications for investors’ regarding their investment horizon choices. Portfolios were proven to be a profitable investment during the entire sample period even though their returns varied considerably in shorter periods. Thus, the new empirical evidence from Poland, confirms the findings from some other markets that investors should view this type of a trading strategy as a long term, rather than a short term investment.

In Japan, Song and Hagio (2007) proposed to apply the “Dogs of the Dow” strategy to the Tokyo Stock Price Index 30 (TOPIX30) and NIKKEI 225. They showed that for data from 2002 to 2006, the “Dogs of the Dow” strategy is only slightly superior to the benchmark when applied to the TOPIX30, while the performance is much better when applied to the NIKKEI 225. Therefore they concluded that the strategy should be applied to the NIKKEI 225 in the Japanese market. For a longer period (1981 – 2010), Qiu, Song and Hasama investigated the strategy, and showed that it outperformed the NIKKEI 225 and the result is statistically significant.

In Finland, Rinne and Vahamaa (2011) summarized the performance of the “Dogs of the Dow” investment strategy in Aktiebolaget Optionsmäklarna / Helsinki Stock Exchange (OMX25) index from 1998 to 2008. They indicated that the strategy can be successfully replicated in different types of markets and in different market conditions. Their result reported an annual abnormal return of 4.5% and the outperformance of the strategy appeared particularly pronounced during a stock market downtown.

Qiu, Yan and Song (2012) focused on the Hong Kong stock market from 2001 to 2011. Based on the result of the simulation, they found that the “Dogs of the Dow” strategy outperformed the Hang Seng Index. However, the result was not statistically significant. They also found that the portfolios with fewer than 10 dogs outperformed the benchmark. Thus, they concluded that the “Dogs of the Dow” strategy was effective in in the Hong Kong stock market.

3. APPLICATION TO THE THAI STOCK MARKET

3.1. Market Indices in the Thai Stock Market

In this paper, we propose applying the “Dogs of the Dow” investment strategy to the Stock Exchange of Thailand (SET), which is the only stock exchange in the country. As of 31 December 2012, the SET had about 600 listed securities.

The most popular market index for the Thai stock market is the SET Index, which is calculated from the prices of all common stocks (including unit trusts of property funds) on the main board of the SET, except for stocks that have been suspended for more than one year. The index is a market capitalization-weighted price index, which compares the current market value of all listed common shares with its value on the base date of April 30, 1975, when the SET Index was established and set at 100 points. In addition to the SET Index, the SET also provides other indices to investors which include the Market for Alternative Investment index (mai index), industry group and sectorial indices, SET50 index and SET100 index.

Here we apply the "Dogs of the Dow" strategy to the SET50 index, which is "calculated from the stock prices of the top 50 listed companies on the SET in terms of large market capitalization, high liquidity and compliance with requirements regarding the distribution of shares to minor shareholders"(The Stock Exchange of Thailand 2013). It is also a capitalization-weighted index, and was calculated from August 1995 with a base value of 1000 points. The component stocks in the SET50 Index are reviewed every six months in order to adjust for any changes that may have occurred in the stock market, such as new listings or public offerings.

3.2. Simulation

We implemented the simulation of the "Dogs of the Dow" strategy in the following steps.

Step 1. Collect data on all of the 50 companies of the SET50 on 30 September, and then select the 10 highest dividend yielding stocks, invest in the 10 stocks with equal amounts on 1st October.

Step 2. Hold these stocks for 1 year, and then sell them out on 30th September of the following year. After updating the list of the SET 50, invest in the new top 10 stocks equally.

Step 3. Repeat the above process every year.

In this study, we conducted a simulation for the data for the years 1995–2011. We searched for the price of the stocks from the Internet and obtained the dividend data from the SET. October 1st was chosen as the investment date because it is the beginning of the fiscal year for most Thai companies.

4. COMPARISON OF PERFORMANCE

In this section, we compare the performance o of the "Dogs of the Dow" strategy with the SET Index from various points of view.

4.1. Difference between Annual Returns

Figure 1 plots the difference in annual return between the "Dogs of the Dow" strategy and the SET Index portfolios for each of the 17 years. A positive difference indicates that the "Dogs of the Dow" strategy outperformed the SET Index portfolio.

From Figure 1, we can see that the "Dogs of the Dow" strategy portfolio outperformed 1 the SET Index 11 times from 1995 to 2011. In particular, in 1999, the "Dogs of the Dow" strategy was 86.15 percentage points greater than the SET Index. In 2003, the "Dogs of the Dow" strategy was 85.87 percentage points greater than the SET Index. However, the performance of the "Dogs of the Dow" strategy was also poor in several years. The worst performance was in 2008, when the difference between the two strategies was -60.19 percentage points.

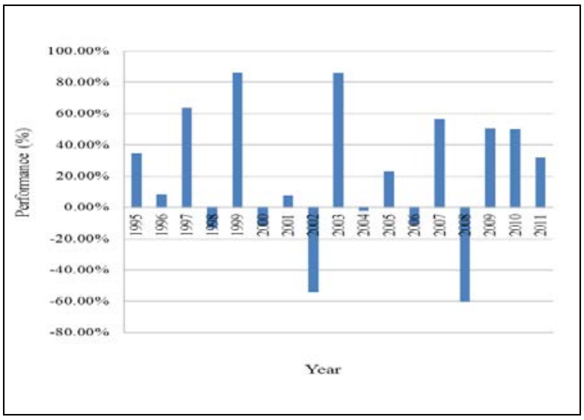


Figure 1: Annual Difference in Return between the "Dogs of the Dow" strategy and the SET Index

4.2. Average Return

Table 1 shows the average return and deviation of the "Dogs of the Dow" strategy and the SET Index. We can see that the "Dogs of the Dow" strategy had an average return of 23.68% and a standard deviation of 35.45%, while the SET Index portfolio had a lower mean return and deviation of 3.32% and 33.03%, respectively. Table 1 also shows data on the difference between the two portfolios. The "Dogs of the Dow" strategy had an average 20.36 percentage points higher return, and the difference of the standard division was 2.42 percentage points.

From the result, we can see that the "Dogs of the Dow" strategy outperformed the SET Index during the 17 years on average. To check the statistical significance of the result, we conducted a T-test at a 5% significance level, and the result was $p = 0.10258084 > 0.05$. Therefore the difference of 20.36 percentage points is not statistically significant.

Table 1: Annual Return Summary Statistics (1995 to 2011)

Portfolio	Average annual return	Standard deviation
"Dogs of the Dow"	23.68%	35.45%
SET Index	3.32%	33.03%
Difference	20.36%	2.42%

4.3. Accumulated Performance

Figure 2 shows the accumulated performance of the "Dogs of the Dow" strategy, the SET Index and the SET50. From 1995 through 2011, the "Dogs of the Dow" strategy always had a higher accumulated return than both the SET Index and the SET50. In contrast, the

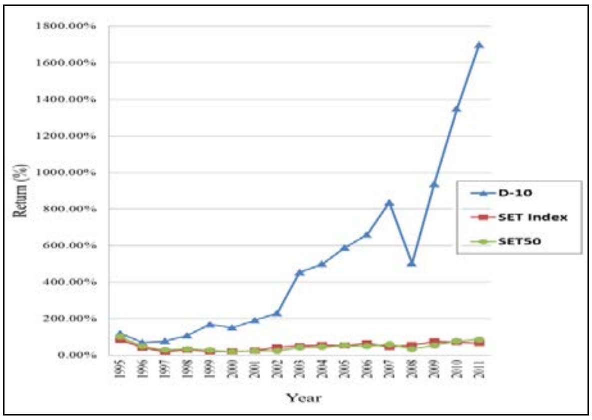


Figure 2: Accumulated Performances

SET Index and the SET lines in the lower part of the graph show a steady trend.

In 2011, the "Dogs of the Dow" strategy had an accumulated value of 1698.60%, which is about 17 times the value in 1995. In contrast, the value of the SET Index in 2011 was only 94.28%, which is at about the same level with the data of 1995, and the accumulated value of the SET50 Index was quite similar to that of the SET Index.

Therefore, we conclude that in the long term, the "Dogs of the Dow" strategy is very powerful for making profit and beating the benchmarks.

4.4. Subperiod Analysis

Table 2 reports a 5-year subperiod analysis regarding the mean return and nominal difference between the "Dogs of the Dow" strategy and the SET Index. From table 2 we can conclude that the "Dogs of the Dow" strategy outperformed the SET Index in all of the 5 year subperiods. From 1995-1999, the "Dogs of the Dow" strategy had a very large difference from that of the SET Index, the difference being 35.99 percent. During the periods 2000-2004, 2005-2009, and 1995-2011, the differences were 5.15, 11.73, and 20.36 percentage points, respectively. As a result, the "Dogs of the Dow" strategy is useful in making profit in the medium term. However, the standard deviation of the "Dogs of the Dow" strategy was greater than the SET Index.

5. RISK ADJUSTMENT

Table 2 shows that the "Dogs of the Dow" strategy has a higher mean return than the SET Index. It also shows that in most periods, the "Dogs of the Dow" strategy had higher standard deviations than the SET Index. With only 10 stocks in the portfolio, there were some unsystematic risks that led to the higher standard deviations. Therefore, we need to adjust the risk of the "Dogs of the Dow" strategy to judge the performances of different strategies more precisely.

The Sharpe Ratio tells us whether a portfolio's returns are due to smart investment decisions or the result of excess risk. This measurement is very useful because although one portfolio or fund can reap higher returns than its peers, it is only a good investment if those higher returns do not come with too much additional risk. The greater a portfolio's Sharpe ratio, the better its risk-adjusted performance has been. A negative Sharpe ratio indicates that a risk-less asset would perform better than the security being analyzed (Sharp 1966). By assuming that the investor allocates part of his portfolio to some riskless assets, the Sharpe Ratio eliminates the risk premium from the portfolio, thus enabling the comparison of two different risk degree portfolios.

In this paper, we use Thai government bonds as the risk-free asset. Then the adjustment for the entire 17 years period is the same to invest 93 percent (33.03% / 35.45% = 93.17%) of the wealth in the "Dogs of the Dow" strategy and the remaining 93 percent (1 - 93%) in government bonds. With this 93% investment in the national debt, we can adjust the higher risk of the "Dogs of the Dow" strategy to have nearly the same standard deviation as that of the SET Index. After that, using the government bonds mean annual return of 2.68%, the return of the "Dogs of the Dow" strategy can be transformed to 22.25% (i.e., (23.68% - 2.68%) (33.03% / 35.45%) + 2.68%). Apparently, the Dogs of the Dow strategy outperformed the SET Index even after the adjustment, although the difference between average return now shrinks to 18.93 percentage points (Table 3).

The second column in Table 3 is the risk-adjusted average returns of the "Dogs of the Dow" strategy, and the fourth column is the difference between the return of the risk-adjusted "Dogs of the Dow" strategy and those of the SET Index. Before the adjustment, the "Dogs of the Dow" strategy performed better than SET Index for all three subperiods, and the results remained

Table 2: Subperiod Analysis

5 years Subperiod	Mean Return		The Standard Deviations		Nominal Difference
	"Dogs of the Dow"	SET Index	"Dogs of the Dow"	SET Index	
1995-1999	17.13%	-18.86%	33.72%	38.81%	35.99%
2000-2004	28.70%	23.54%	36.70%	26.26%	5.15%
2005-2009	20.71%	8.97%	40.35%	23.39%	11.73%
1995-2011	23.68%	3.32%	35.45%	33.03%	20.36%

Table 3: The Difference between the Risk-Adjusted "Dogs of the Dow" and SET Index

5 years Subperiod	Return of Risk Adjusted "Dogs of the Dow"	Return of SET Index	Risk - Adjusted Difference
1995–1999	16.31%	-18.86%	38.17%
2000–2004	21.29%	23.54%	-2.25%
2005–2009	13.13%	8.97%	4.16%
1995–2011	22.25%	3.32%	18.93%

the same even after the adjustment. In particular, in the 2000 - 2004 subperiod, the returns of SET Index becomes greater than the return of the risk adjusted "Dogs of the Dow". Consequently, we can see that the "Dogs of the Dow" strategy remains superior to the SET Index even after eliminating risk factors.

6. PORTFOLIOS WITH OTHER NUMBERS OF DOGS

In this section, we test other portfolios with fewer stocks and compared the performance of each portfolio using the "Dogs of the Dow" strategy with the SET Index and SET50. We named the portfolio with the top N stocks as Dow N strategy. The simulation of this strategy was conducted in a manner similar to that of the "Dogs of the Dow" strategy but by using the top N stocks instead of the top-10 stocks.

From Figure 3, we can conclude that all portfolios of the "Dogs of the Dow" investment strategy outperformed the SET which is the original "Dogs of the Dow" strategy, had the ninth highest average return, with a value of 23.68% during the 1995–2011 periods. The Dow 3 portfolio had the highest average annual return. The Dow 10 strategy, SET Index and SET50, however, were as low as 3.32% and 4.93%, respectively.

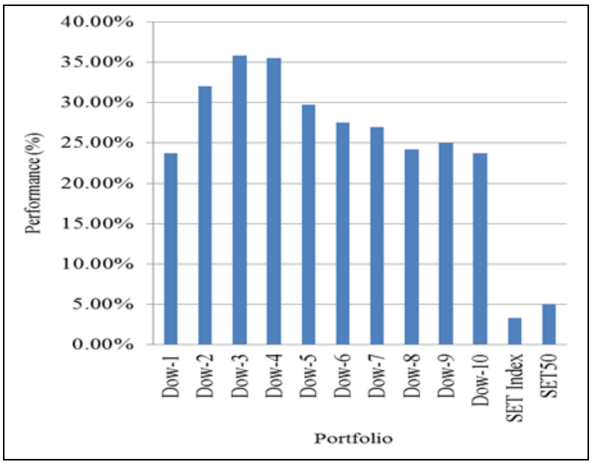


Figure 3: Average Annual Returns of other Numbers of Dogs, SET Index and SET 50

7. CONCLUSION

In this study, we proposed applying the "Dogs of the Dow" investment strategy for the Thai stock market and compared the performance of the "Dogs of the Dow" strategy with benchmarks over several years. On average, the strategy outperformed the SET Index and the SET50 Index in the Thai stock market. However, the result is not statistically significant. We found that the portfolios with fewer than ten "dogs" also outperformed the SET Index. Therefore, we can conclude that in the long term, the "Dogs of the Dow" strategy is effective in Thai stock market to make profits and outperform the benchmarks.

ACKNOWLEDGMENTS

We would like to express our deepest appreciation to those people who helped in even the smallest way in making the completion of this paper possible. In particular, we would like to express our sincere gratitude and appreciation to Ms. Nattaporn Yuenyongpisit for providing the basic data regarding the SET 50 stocks.

REFERENCES

Andre L.C., Silva D., 2001. Empirical test of the Dogs of the Dow strategy in Latin American stock markets, *International Review of Financial Analysis* vol. 10(2): pp. 187-199.

Brzeszczynski J., Gajdka J., 2008. Performance of High Dividend Yield Investment Strategy on the Polish Stock Market 1997-2007, *Investment Management and Financial Innovations* vol. 5(2): pp. 86-92.

Knowles, and Petty D. H., 1992. *The Dividend Investor, A safe and sure way to beat the market with high-yield dividend stocks*. Chicago: Probus Publishing.

McQueen G., Shields K. and Thorley S., 1997. Does the "Dow-10 Investment Strategy" Beat the Dow Statistically and Economically? *Financial Analysts Journal* vol. 53(4): pp. 66-72.

O'Higgins, and Downes J., 1991. *Beating the Dow*, New York: Harper Perennial.

Rinne E., Vahamaa S., 2011. The 'Dogs of the Dow' Strategy Revisited: Finnish Evidence, *The European Journal of Finance* vol. 17(5-6): pp. 451-469.

Sharpe W.F., 1966. Mutual Fund Performance, *The Journal of Business* vol. 39(1): pp. 119-138.

Slatter J., 1988. “Study of Industrial Averages Finds Stocks with High Dividends Are Big Winners,” *Wall Street Journal (Eastern edition)*, August 11.

Song Y., Hagio K., 2007. A Study on Portfolio Selection Strategies for Stock Investment: pp. 29-36. (*In Japanese*).

The Stock Exchange of Thailand website. Available from :http://www.set.or.th/en/products/index/setindex_p3.html [accessed 25 May 2013]

Visscher S., Filbeck G., 1997. Dividend Yield Strategies in the British Stock Market, *The European Journal of Finance* vol. 3(4): pp. 227-289.

Visscher S., Filbeck G., 2003. Dividend-Yield Strategies in the Canadian Stock Market, *Financial Analysts Journal* vol. 59(1): pp. 99-106.

Qiu M., Song Y. and Hasama M., 2011. Applying the Dow 10 Investment Strategy to Japanese Stock Market, *Asian Conference of Management Science & Application,* No. 157.21-23 December 2011, Sanya, Hainan, China.

Qiu M., Yan H. and Song Y., 2012. Empirical Analyses of the Dogs of the Dow strategy: Hong Kong Evidence, *European Journal of Management* vol. 12(3): pp. 183 – 187.

AUTHORS BIOGRAPHY

Kittipong Tissayakorn was born in Sukhothai, Thailand. He earned his bachelor degree at King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. Currently, he is a master’s degree student at the Graduate School of Engineering, Fukuoka Institute of Technology, Fukuoka, Japan.

Yu Song was born in China. He earned his Ph.D. at Tohoku University, Japan. He is a professor at Fukuoka Institute of Technology, Fukuoka, Japan.

Fumio Akagi was born in Japan. He earned his Ph.D. at Osaka City University, Japan in 1985. Currently he is a professor at Fukuoka Institute of Technology, Fukuoka, Japan.

MODELING, SIMULATION AND ANALYSIS APPLIED TO A NAPHTHA STABILIZER TOWER

Felipe Sanchez Nagata$^{(a)}$, Wilson Hideki Hirota$^{(a)}$, Roberto Nasser Jr.$^{(a)}$, Rosimeire Aparecida Jerônimo$^{(b)}$

$^{(a)}$Departamento de Ciências Exatas e da Terra - Universidade Federal de São Paulo - UNIFESP. Rua Prof. Artur Riedel, 27, Jd. Eldorado, Diadema, SP, Brasil, CEP 09972-270.
$^{(b)}$Universidade Federal de Itajubá – UNIFEI. Rua Irmã Ivone Drumond, 200, Distrito Industrial II, Itabira, MG, Brasil, CEP 35903-087.

$^{(a)}$roberto.nasser@unifesp.br, $^{(b)}$ rosijeronimo@unifei.edu.br

ABSTRACT

The use of simulation and analysis in the unit operation has the objective of checking the operation of an equipment under specified conditions and perform a possible operation optimization using many different tools. One of the objectives is to get the specification data of naphtha stabilizer tower and resorting Aspen HYSYS® commercial software. Applying this software it was possible to find the convergence of this process. The specified temperature of outlet stream in the heat exchanger was changed to have a possibility of getting the equipment convergence. Thus, the profile of the product streams was obtained for this operation, checking contamination of light chain by the presence of NBP 11, NBP 26, and NBP 40. Changing the temperature profile of the tower, could observed the decrease of contamination, which is already a desired result.

Keywords: Simulation, analysis, operation optimization, naphtha stabilizer tower.

1. INTRODUCTION

The Petroleum is used since ancient times for different purposes. The society started its application for simple purposes of medical use and building construction, and over time, their purposes has been expanded, mainly from the nineteenth century, with the advent of Petroleum wells.

Today, compounds obtained by purification have application in several areas, and he is best known for use as fuel, such as natural gas, gasoline and diesel.

Although these compounds have high recognition, naphtha is the most important. It has a composition similar to gasoline, but its energy use is not feasible, but due to their wide application, to obtain fuels, as well as compounds for the application in several process industries, has high value.

Because of this importance it is necessary a high control in their production, by performing the separation of the lightest compounds in its mixture through a distillation column known as naphtha stabilizer tower.

This control intended to keep the quality to ensure the maximum production, but this requires high costs, which makes modeling and simulation tools quite attractive.

Garcia (2009) defined a modeling being *the mathematic abstraction of a real process*. The statement of Chapra and Canale (2008) complements this theory, indicating that *modeling is a formulation that presents the essential features of a physical system or process in mathematic language*.

The major difficulties of these tools are the large number of environmental factors influences, as well as naphtha's infinite composite components number. These situations can be overcome with the use of commercial software, which are designed to simulate or simplify these problems, especially with the components creation based on the common characteristics of the substances in the mixture to be studied.

The operating systems evolution and the source codes simplification allow to obtain simpler and more accurate software, which made them very attractive in the industrial environment.

Despite to this facilitation, analysis of these simulations is still needed and is one of the more complex steps. This allows to define the success or failure of the experiment and requires extensive technical knowledge of the subject, and sometimes extreme attention to the smallest details.

The detailed analysis allows the identification of points in the process to be optimized, ensuring energy economy, raw material and process for low maintenance costs and equipment.

Publications related of modeling, simulation and analysis, as opposed up to the three decades ago, seek to optimize processes that exist today. Despite of this goal, few studies allow for a thorough analysis of the results obtained by the lack of information regarding the experiment.

The reason behind this information's absence is the treatment given to the process simulation as a secondary step, insignificant when compared to the study of a controller.

Today, mostly articles have as main objective obtaining or applying different controllers. As examples are the publications of Almeida Neto, Odloak and

Rodrigues (1999) and Ventin (2010), which seek to replace the employed controller for source codes with better response time and more robust results.

Few studies of simulation and analysis in chemical processes have been published in the last decade. Due to the expansion of this tool, the use of simulation in other areas is allowed, and a greater attention is generated to optimization studies of administrative systems and other scientific fields. The publication examples are the publication of Silva (2002), which deals the simulation for accelerated analysis of the air traffic, and Bleicher et al (2002), who seek a better learning method of the sound waves operation through mathematical and computational study, listing the frequencies of musical scales and different beats.

For industrial plants equipments simulation, has been seen more papers involving controller innovations, as the cited works and Marquini et al (2007), which demonstrate a simulation of a distillation system in a ethanol production. There are also works who seek study chemical treatments, escpecially recovery methods, as has Sadighi et al (2009), that seeks recovery of naphtha.

Maitelli et al (2006) presented at the 2006 *Rio Oil & Gas* Conference a naphtha stabilizer tower simulation and the application of a control that would provide greater profitability than that used in the Potiguar Clara Camarão Refinery, located in the Guamaré city, Brazil.

This Paper has the objective to propose a methodology for modeling and simulating a unit operation responsible for naphtha stabilization, obtaining parameters for use in future controller studies and analyzing possible procedure optimizations.

2. METODOLOGY

Due to the complexity of the mixture, it was decided to use commercial software. We adopted the software were whose operation is more acquainted, Aspen HYSYS®.

The simulation in these tools requests the knowledge of most appropriate thermodynamic model to the physical and chemical characteristics of the involved compound or mixture.

To define the best model, we used the model suggested by the Publication of Almeida et al (1999) and Ventin (2010), with the model of Peng-Robinson, and the procedure proposed by Carlson (1996), verifying a better fit for Grayson-Streed model, since it fits better to the physicochemical properties of the mixture.

The input current composition is based on the Paper presented by Ventin. This is listed in Table 1 as attached.

Table 1. Composition and physical properties of the compounds present in the Naphtha Stabilizer Tower input flow.

Components		NBP (ºC)	Molecular weight	% Volume liquid
Hydrogen	H_2	-252,60	2,02	0,0095
Nitrogen	N_2	-195,80	28,01	0,1149
Carbon Monoxide	CO	-191,45	28,01	0,0149
Methane	CH_4	-161,52	16,04	0,0088
Ethylene	C_2H_4	-103,75	28,05	0,1558
Ethane	C_2H_6	-88,60	30,07	0,1868
Propane	C_3H_8	-42,10	44,10	2,7790
Iso-Butane	C_4H_{10}	-11,73	58,12	1,9431
1-Butene	C_4H_8	-6,25	56,11	0,0409
n-Butane	C_4H_{10}	-0,50	58,12	5,4717
Iso-Pentane	C_5H_{12}	22,88	72,15	0,0450
NBP 11		11,03	60,61	1,6629
NBP 26		25,96	65,48	2,4385
NBP 40		40,37	72,37	4,2889
NBP 54		54,02	78,10	6,8701
NBP 67		67,32	83,86	7,1400
NBP 82		82,36	90,50	6,8055
NBP 97		96,58	97,45	7,9180
NBP 111		110,59	104,80	8,7766
NBP 125		124,80	112,38	8,2854
NBP 139		139,10	120,23	8,0347
NBP 153		153,25	128,47	8,0594
NBP 168		167,57	137,37	7,4376
NBP 181		181,09	145,28	4,3618
NBP 196		195,65	154,33	2,9562
NBP 210		209,98	163,78	2,2413
NBP 225		224,80	174,05	1,9526

Source: Ventin (2010)

The state variables were specified according to the article published by Almeida, as shown below:

1. Heat exchanged feed stream data (*Feed*):
   - Flow: 1445 m³/d;
   - Temperature: 40 ºC;
   - Pressure: 8 kgf/cm2;
   - Feeding in the heat exchanger shell;

2. Naphtha Stabilization Tower input stream data (*FeedHot*):
   - Vapour Fraction: 0.06;
   - Temperature: 136 ºC;

3. Heat exchanger parameters (TC-01):
   - Differential pressure in the tube (ΔP): 0,5 kgf/cm²;
   - Heat exchanger specification standardized by Tubular Exchanger Manufactures Association (TEMA): A-F-L, where "A" indicates removable lid and channel, "F" tells shell longitudinal deflector with two

steps, and "L" shows the bundle tube with fixed stationary head;

4. Naphtha stabilizer tower parameters (D-01):
   - Column with 30 actual stages, with increasing count from the top.;
   - Feed in stream FeedHot on stage 17;
   - Partial condenser;
   - Standard HYSYS® Reboiler;
   - Condenser pressure equivalent to 7 kgf/cm^2;
   - Pressure in the reboiler equal to 7.8 kgf/cm^2;

5. Temperature Profile Settings, on the *Parameters/Profiles* tab:
   - Top temperature: 54ºC
   - Bottom temperature: 163 ºC;

6. Murphree efficiency definition on the *Parameters/Efficiences* tab: 0.75 in all stages;

7. Heat exchange in the condenser: 1.0 MMcal/h.

The choice of a partial condenser is justified by the formation of a gas output stream (FG - Fuel Gas) and a liquid output stream (LPG - Liquefied Petroleum Gas).

The selection of the standard reboiler HYSYS® was done due to the process simplification, the lack of information regarding this simulation.

The process is defined in accordance with the software flowchart in Figure 1 as below.

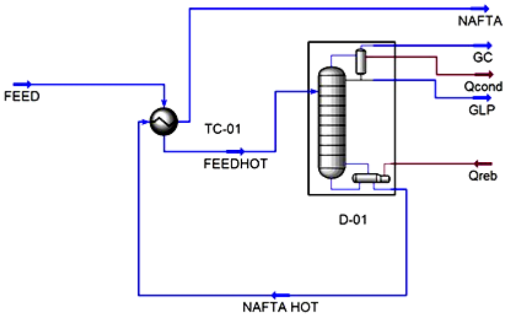


Figure 1: Flowchart of naphta stabilize unit operation.

In order to eliminate the freedom degrees, the design variables specifications are made on the tab *Design/ Specs*:

8. Distillate flow (constant): 108 m^3/d;
9. Temperature in stage 5 (state variable):
   - Minimum: 60 ºC;
   - Maximum: 92.5 ºC;
   - Fixed value obtained: 76.25 ºC.

To obtain a restriction results have been specified limits for the output variables:

10. Temperature in the LPG stream (output): 20 ºC;
11. Reboiler Heating:
    - Minimum: 1.2 MMcal/h;
    - Maximum: 3.5 MMcal/h;

The variable specified in item 10 allows to define the behavior of the LPG stream at the output in the top of the stabilizer tower, while the parameters established in item 11 allows an adjustment of the input stream of the same tower, in addition to ensure the production of naphtha with the desired characteristics, guaranteeing the absence or minimization of contamination by the light products.

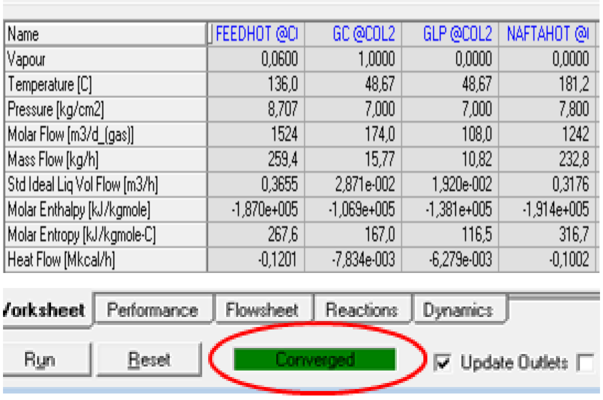
The variables manipulated are the temperature in the input stream in the stabilization tower and the temperature in stage 5 of the distillation tower. The selection of this stage should be generally defined as more sensitive stage for the temperature perturbations in this equipment.

The variables which affect these parameters are the temperature profile throughout the column and the temperature of the naphtha input stream in the heat exchanger, which will influence the FeedHot stream temperature. This enables to consider them as disturbance variables.

3. RESULTS

One of the commercial software problems is a partial analysis of the problem, identifying only the convergence mathematics. This causes certain illusion of the experiment's success.

Providing all data input, the convergence can be achieved, as shown in Figure 2.



Name	FEEDHOT @Cl	GC @COL2	GLP @COL2	NAFTAHOT @l
Vapour	0,0600	1,0000	0,0000	0,0000
Temperature [C]	136,0	48,67	48,67	181,2
Pressure [kg/cm2]	8,707	7,000	7,000	7,800
Molar Flow [m3/d_(gas)]	1524	174,0	108,0	1242
Mass Flow [kg/h]	259,4	15,77	10,82	232,8
Std Ideal Liq Vol Flow [m3/h]	0,3655	2,871e-002	1,920e-002	0,3176
Molar Enthalpy [kJ/kgmole]	-1,870e+005	-1,069e+005	-1,381e+005	-1,914e+005
Molar Entropy [kJ/kgmole-C]	267,6	167,0	116,5	316,7
Heat Flow [Mkcal/h]	-0,1201	-7,834e-003	-6,279e-003	-0,1002

Figure 2: Example of convergence in a simulation, demonstrated in a results table.

With the same consideration, previously made to verify convergence, it is also possible to check the convergence of the physical states that are the output streams, which can be obtained by analyzing the total mass balance and the energy balance. The vapor fraction was found for the fuel gas was equivalent to 1 while for the other streams was equal to 0, which indicates a certain agreement with the convergence and good phase separation.

The phase separation can also be observed through the net molar flow chart versus stage position in the distillation column, as shown in Figure 3.

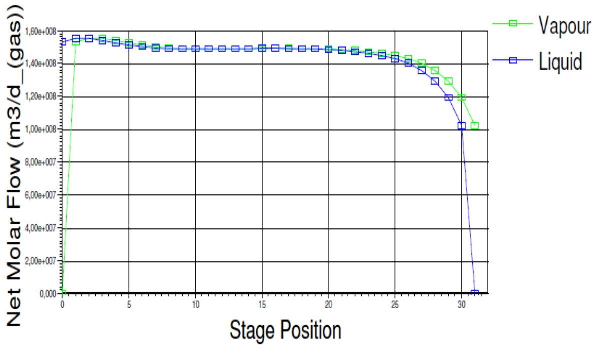


Figure 3: Net Molar Flow vs Stage Position

The higher molar flow observed at the highest position of the stabilizer tower is due to the greater volatility of lighter compounds. This allows having a greater molar flow. The heavier compounds have less volatility, therefore, a lower molar flow, as shown in the lower position of the column.

In despite of this result, the analysis should be done also considering the mass conservation per component, which will be discussed later.

By observing the temperature distribution of the equipment according to Figure 4, there is the expected behavior with sigmoidal curve, indicating good distribution of the trays, which ensures a good simulation of the equipment, but not the process simulation.

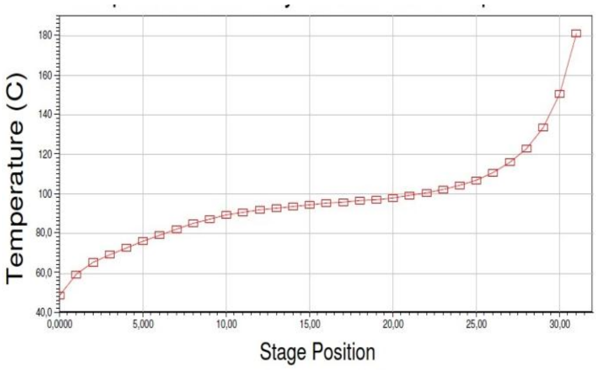


Figure 4: Temperature vs Stage position

Analyzing the composition obtained at the output, there is certain contamination in the overhead stream in the stabilizer tower with the presence of heavy compounds NBP11, and NBP26 NBP40, as shown in the Table 2.

Table 2: Stream out compositions (in molar fraction).

Components	Feed Hot	GC	GLP	Nafta Hot
Hydrogen	0,0040	0,0039	0,0000	0,0000
Nitrogen	0,0045	0,0390	0,0006	0,0000
Carbon Monoxide	0,0006	0,0050	0,0001	0,0000
Methane	0,0002	0,0019	0,0001	0,0000
Ethylene	0,0029	0,0239	0,0024	0,0000
Ethane	0,0030	0,0238	0,0040	0,0000
Propane	0,0434	0,2928	0,1415	0,0000
Iso-Butane	0,0256	0,1365	0,1410	0,0000
1-Butene	0,0006	0,0030	0,0035	0,0000
n-Butane	0,0747	0,3551	0,4824	0,0000
Iso-Pentane	0,0005	0,0001	0,0002	0,0006
NBP 11	0,0202	0,0861	0,1467	0,0000
NBP 26	0,0299	0,0288	0,0775	0,0259
NBP 40	0,0508	0,0000	0,0001	0,0623
NBP 54	0,0790	0,0000	0,0000	0,0969
NBP 67	0,0793	0,0000	0,0000	0,0973
NBP 82	0,0726	0,0000	0,0000	0,0890
NBP 97	0,0800	0,0000	0,0000	0,0982
NBP 111	0,0840	0,0000	0,0000	0,1030
NBP 125	0,0751	0,0000	0,0000	0,0922
NBP 139	0,0691	0,0000	0,0000	0,0848
NBP 153	0,0658	0,0000	0,0000	0,0807
NBP 168	0,0574	0,0000	0,0000	0,0705
NBP 181	0,0322	0,0000	0,0000	0,0395
NBP 196	0,0207	0,0000	0,0000	0,0255
NBP 210	0,0150	0,0000	0,0000	0,0184
NBP 225	0,0124	0,0000	0,0000	0,0152

Verifying the composition of the streams and the obtained temperature in the simulation, it is observed that the reason this contamination is due to the overhead stream temperature, which is higher than the boiling temperature of these three components.

This same contamination can be observed in the graphics obtained in this simulation, shown in Figures 5 to 7, with the high presence of these compounds in the top and intermediate stages.

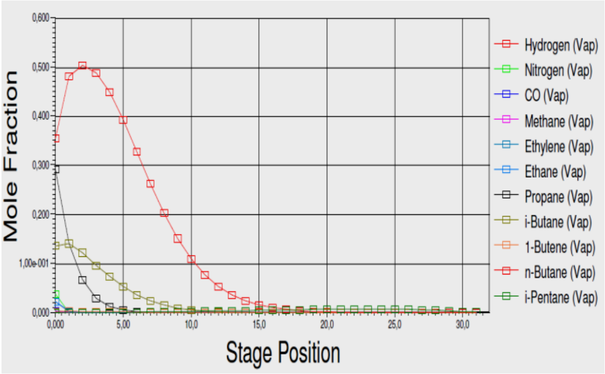


Figure 5: Light key (vapour) composition vs stage position

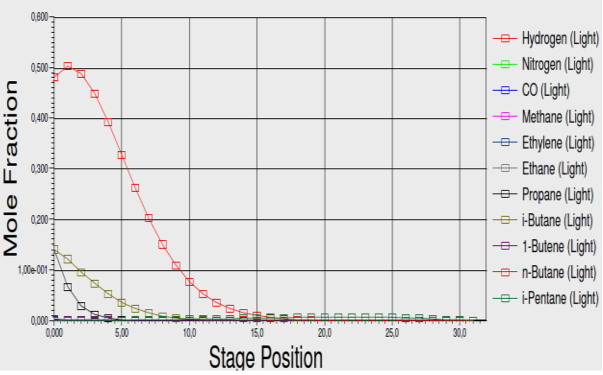


Figure 6: Light key (Liquid) composition vs stage position

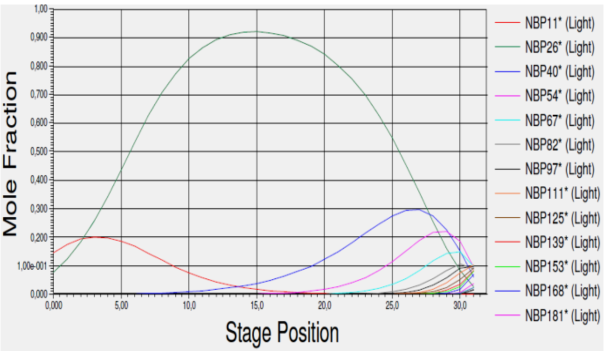
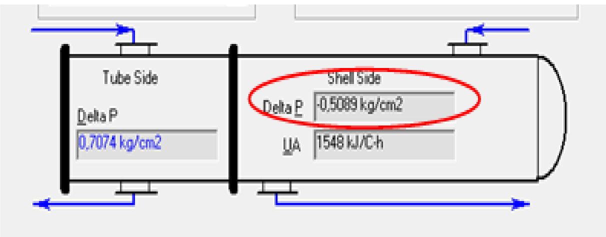


Figure 7: Heavy key composition vs stage position

To avoid this contamination, it is necessary to control the temperature of this stream below 11.03 °C, boiling point of NBP11, but it is more suitable to use a control to keep this temperature close to 2 °C as indicated by Almeida (1999).

This obstacle can have a second question which allows a better fit of the process. It was found that, in despite of the convergence, the occurrence of negative pressure (-0.51 kgf/cm²) in the shell side of the heat exchanger, which indicates a process malfunction.



Tube Side

Delta P 0,7074 kg/cm2

Shell Side

Delta P -0,5089 kg/cm2

UA 1548 kJ/C-h

Figure 7: Presence of a negative pressure difference in a heat exchanger.

The simplest method to fix this problem is to decrease the temperature of the output stream present in the portion of the heat exchanger with negative differential pressure, since the temperature is an independent variable.

Changing the vapor fraction in this stream will affect the process of undesired manner, besides being dependent variable mentioned above.

Adopting the temperature to a value below the specified (130 °C) allows an increase in this pressure difference, allowing the flow of the feed stream in favor of the process feed (0,076 kgf/cm²) without a perceptible change in the output composition.

The temperature profile of the naphtha stabilizer tower has changed to observe the effect on the composition. The temperature was changed in stage 5 to 67.5 ° C by checking the change in composition as shown in Table 3, below.

There was a considerable contamination reduction of the compounds NBP26 and NBP40, with this measure.

Table 3: Stream out composition after temperature profile change.

Componen ts	Feed Hot	GC	GLP	Nafta Hot
Hydrogen	0,0040	0,0039	0,0000	0,0000
Nitrogen	0,0045	0,0390	0,0006	0,0000
Carbon Monoxide	0,0006	0,0050	0,0001	0,0000
Methane	0,0002	0,0019	0,0001	0,0000
Ethylene	0,0029	0,0239	0,0024	0,0000
Ethane	0,0030	0,0238	0,0040	0,0000
Propane	0,0434	0,2929	0,1415	0,0000
Iso-Butane	0,0256	0,0030	0,1410	0,0000
1-Butene	0,0006	0,1365	0,0035	0,0000
n-Butane	0,0747	0,3551	0,4824	0,0000
Iso-Pentane	0,0005	0,0001	0,0002	0,0006
NBP 11	0,0202	0,0861	0,1467	0,0000
NBP 26	0,0299	0,0288	0,0775	0,0259
NBP 40	0,0508	0,0000	0,0001	0,0623
NBP 54	0,0790	0,0000	0,0000	0,0969

NBP 67	0,0793	0,0000	0,0000	0,0973
NBP 82	0,0726	0,0000	0,0000	0,0890
NBP 97	0,0800	0,0000	0,0000	0,0982
NBP 111	0,0840	0,0000	0,0000	0,1030
NBP 125	0,0751	0,0000	0,0000	0,0922
NBP 139	0,0691	0,0000	0,0000	0,0848
NBP 153	0,0658	0,0000	0,0000	0,0807
NBP 168	0,0574	0,0000	0,0000	0,0705
NBP 181	0,0322	0,0000	0,0000	0,0395
NBP 196	0,0207	0,0000	0,0000	0,0255
NBP 210	0,0150	0,0000	0,0000	0,0184
NBP 225	0,0124	0,0000	0,0000	0,0152

The thermodynamic package was changed for an appropriate package to the mixture (Grayson-Streed), obtaining a wide process improvement, with the elimination of the contamination by heavy component NBP40 and a minimal presence of NBP26, as shown in Table 4.

Table 4: Stream out composition after thermodynamic model change.

Componen ts	Feed Hot	GC	GLP	Nafta Hot
Hydrogen	0,0040	0,0040	0,0000	0,0000
Nitrogen	0,0045	0,0394	0,0008	0,0000
Carbon Monoxide	0,0006	0,0051	0,0001	0,0000
Methane	0,0002	0,0019	0,0001	0,0000
Ethylene	0,0029	0,0237	0,0031	0,0000
Ethane	0,0030	0,0241	0,0042	0,0000
Propane	0,0434	0,2963	0,1417	0,0000
Iso-Butane	0,0256	0,0030	0,0036	0,0000
1-Butene	0,0006	0,1371	0,1427	0,0000
n-Butane	0,0747	0,3558	0,4883	0,0000
Iso-Pentane	0,0005	0,0001	0,0002	0,0006
NBP 11	0,0202	0,0858	0,1468	0,0000
NBP 26	0,0299	0,0238	0,0663	0,0276
NBP 40	0,0508	0,0000	0,0000	0,0622
NBP 54	0,0790	0,0000	0,0000	0,0968
NBP 67	0,0793	0,0000	0,0000	0,0972
NBP 82	0,0726	0,0000	0,0000	0,0889
NBP 97	0,0800	0,0000	0,0000	0,0980
NBP 111	0,0840	0,0000	0,0000	0,1029
NBP 125	0,0751	0,0000	0,0000	0,0920
NBP 139	0,0691	0,0000	0,0000	0,0847
NBP 153	0,0658	0,0000	0,0000	0,0805
NBP 168	0,0574	0,0000	0,0000	0,0703
NBP 181	0,0322	0,0000	0,0000	0,0394
NBP 196	0,0207	0,0000	0,0000	0,0254
NBP 210	0,0150	0,0000	0,0000	0,0183
NBP 225	0,0124	0,0000	0,0000	0,0152

4. CONCLUSION

The use of the methodology proposed in this work has enabled the convergence of both equipments in the studied unit operation.

Furthermore, makes possible to get parameters required for the experiment repeatability, as well as using it in other works and the procedure improvement with the adoption of milder temperatures and thermodynamic package more appropriate, achieving the proposed objective.

The small presence of contaminants was not completely solved, but the path to be used for obtaining better refined data has already been established.

5. REFERENCES

Chapra, S. C.; Canale, R., 2006. *Métodos Numéricos para Engenharia*. New York: McGraw-Hill, 5ª. ed. 832 p.

Garcia, C., 2009. *Modelagem e Simulação de Processos Industriais e de Sistemas Eletromecâmicos*. São Paulo: EDUSP, 2ª. ed. 688 p.

Ventin, F. F., 2010. *Controle Robusto de uma Torre Estabilizadora de Nafta*. 126 f. Thesis (Master's Degree), Instituto Alberto Luiz Coimbra de Pós_Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro. Rio de Janeiro. Available from: objdig.ufrj.br/60/teses/coppe_m/FabyanaFreireVentin.pdf [acessed 15 September 2012]

Silva, A. M. G., 2001. *Sistema de simulação acelerado para análise de fluxo de tráfego aéreo*. 252 f. Thesis (Master's Degree), Instituto Nacional de Pesquisas Espaciais, Ministério da Ciência e Tecnologia, São José dos Campos. Available from: http://150.163.34.246/col/sid.inpe.br/iris@1913/2005/08.01.12.31/doc/publicacao.pdf [acessed 14 July 2013]

Bleicher, L. et al., 2002. Análise e Simulação das Ondas Sonoras Assistidas por Computador. *Revista Brasileira de Ensino de Física*, v. 24, n. 2, p. 129-13. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S180611172002000200008 [accessed 13 July 2013]

Marquini, M. F. et al., 2007. Simulação e análise de um sistema industrial de colunas de destilação de etanol. *Acta Scientiarum Technology*, v. 29, n. 1, p 23-28. Available from:

http://eduemojs.uem.br/ojs/index.php/ActaSciTechnol/article/view/81/55 [acessed 13 July 2013]

Maitelli, A. L. et al, 2006. Simulação de uma debutanizadora real utilizando um software comercial. In: Rio Oil & Gas Expo and Conference 2006, 11, 11-14 set. 2006, Rio de Janeiro. *Rio Oil & Gas Expo and Conference 2006 Annals*, Rio de Janeiro: Instituto Brasileiro de Óleo e Gás. v. 1. Available from: http://www.dca.ufrn.br/~maitelli/FTP/artigos/RioOilGas_CONPETRO.pdf [acessed 15 March 2013]

Carlson, E. C., 1996. Don’t Gamble with Physical Properties for Simulations. *Chemical Engeneering Progress*, v.10, October, p. 35-46.

AN AGENT-BASED ELECTRONIC MARKET SIMULATOR ENHANCED WITH ONTOLOGY MATCHING SERVICES AND EMERGENT SOCIAL NETWORKS

Virgínia Nascimento(a), Maria João Viamonte(a), Alda Canito(a), Nuno Silva(a)

(a) GECAD – Knowledge Engineering and Decision Support Research Center,
Institute Of Engineering – Polytechnic of Porto (ISEP/IPP)
Porto, Portugal
(a){vilrn, mjv, alrfc, nps}@isep.ipp.pt

ABSTRACT

AEMOS is a simulator which aims to support the development of agent-based electronic markets capable of dealing with the natural semantic heterogeneity existent in this kind of environment. AEMOS simulates a marketplace which provides ontology matching services, enhanced with the exploitation of emergent social networks, enabling an efficient and transparent communication between agents, even when they use different ontologies. The system recommends possible alignments between the agent's ontologies, and lets them negotiate and decide which alignment should be used to translate the exchanged messages. In this paper we propose a new ontology alignment negotiation process, which promotes the reutilization and combination of already existent alignments, as well as the involvement of the business agents in the alignment composition process. With this new model, we aim to achieve a higher adequacy of the used alignments, as well as a more accurate and trustful evaluation of the alignments.

Keywords: agent mediated e-commerce, agent-based simulation, ontology alignment negotiation, emergent social networks

1. INTRODUCTION

E-commerce is a widely used technology which presents several advantages when compared to the traditional commerce (Du et al., 2005). Among these advantages is the availability and accessibility of information. However, the amount of available information also becomes a problem, being difficult for a human user to compare all possible deals in order to achieve the best one.

The rapid growth of e-commerce has increased the demand for automated processes to support both customers and suppliers in buying and selling products (Huang et al., 2010). In this context, the use of software agents as mediators in e-commerce has been receiving an increasing attention (Zhang et al., 2011). However, in e-commerce, the involved entities may possess different conceptualizations about their needs and capabilities, giving rise to a semantic heterogeneity problem that is seen as a corner stone for agents' interoperability (Nascimento et al., 2013b).

In order to provide a solution for this problem we developed the AEMOS system (Nascimento et al., 2013a, Nascimento et al., 2013b, Viamonte et al., 2012, Viamonte et al., 2011). AEMOS is an agent mediated e-commerce (AMEC) simulator which simulates a marketplace that provides ontology services in order to facilitate the interoperability between agents that have different conceptualizations, i.e., use different ontologies. The system follows an ontology-based information integration approach, exploiting the ontology matching paradigm (Euzenat and Shvaiko, 2007), selecting and suggesting possible alignments between the agents' ontologies and letting them choose which ones to use to translate the subsequent exchanged messages.

Conversely to other similar approaches for AMEC (Malucelli et al., 2006), AEMOS is not restricted to the use of a determined ontology matching technique, nor does it include such a complex and time consuming process as the discovery of correspondences between ontologies (i.e. ontology matching process) within the business negotiation process itself. In our system, this process is performed by specialized matching agents in parallel to the market activities as new ontologies are registered. Moreover, considering that agents may use publicly shared ontologies, our approach also allows collecting already existent ontology alignments from web repositories, promoting their reutilization.

Nevertheless, this approach raises the possibility of multiple alignments between a pair of ontologies. Each alignment might be more or less adequate depending on the context of the negotiation and therefore affect its efficiency and result. To overcome this issue, we developed a simulator where we can explore relationships that emerge as the agents interact with each other, applying social network analysis (SNA) techniques (Wasserman and Faust, 1994), in order to improve ontology alignment recommendations as well as supporting agents in their decisions.

Despite being successful in providing an efficient and transparent negotiation between agents, even when they use different ontologies for the same domain, we consider that this process can be further improved by ensuring a higher quality/adequacy of the used alignments. This can be achieved by combining the results of different specialized matching agents, i.e. combining parts of different ontology alignments. For that, we propose a new model for the ontology services where agents can negotiate the composition of the ontology alignment to be used to translate the subsequent exchanged messages, instead of selecting an existing one. This new model promotes not only the reutilization of existent ontology alignments, but also their combination, allowing achieving more adequate or complete alignments. Involving the business agents in the ontology alignment composition process allows excluding irrelevant correspondences, achieving a more adequate alignment as well as a more accurate and trustful evaluation.

In this paper we present a detailed description of the new ontology services model proposal. We start by presenting a brief overview of AEMOS (Section 2). Then we detail the new ontology alignment negotiation process (Section 3) and present the required adaption of the social network based support model (Section 4). Then we present a brief comparison with our previous model (Section5). Finally we draw some conclusions and suggest follow-up research efforts (Section 6).

2. AEMOS SYSTEM OVERVIEW

AEMOS (Agent-based Electronic Market with Ontology Services) is an innovative project (PTDC/EIA-EIA/104752/2008) supported by the Portuguese Agency for Scientific Research (FCT). In this system, agents representing consumers and suppliers negotiate with each other autonomously in order to satisfy the business goals of the entity each represents. The agents customize their behaviour adaptively by learning each user's preference model and business strategies (Viamonte et al., 2007).

The system simulates a marketplace which provides ontology matching services in order to enable communication between agents that use different ontologies. In order to overcome issues related to how the used ontology alignment may influence the business negotiation efficiency, the system includes a component based in emergent social networks (SN), capable of improving the ontology alignments recommendations and supporting the agents' decisions about which alignment to choose.

In this section we present only the key aspects in understanding the functioning of the system. This description is based in (Nascimento et al., 2013a), which presents a recent overview of the AEMOS project. A more detailed description of the ontology services model and SN-based support component can also be found in (Nascimento et al., 2013b).

2.1. Multi-Agent Model

The multi-agent model includes several types of agents divided into two main groups namely, business agents and supporting agents.

The business agents are those representing real world entities with business goals to satisfy. The main types of business agents are: Buyers (B) – representing consumers; and Sellers (S) – representing suppliers.

The supporting agents are those supporting the communication and negotiation between business agents, being responsible for the market's correct functioning. The most relevant supporting agents in the interaction protocol are: Market Facilitator (MF) – an intermediary to the negotiation process, responsible for the establishing communication between potential business partners and ensure they are able to understand each other; Ontology Matching intermediary (OM-i) – agent responsible for the ontology matching services; and Social Network intermediary (SN-i) – agent responsible for the SN-based support.

2.2. Interaction Protocol

To participate in the market, the business agents must register first. During the registration they provide information about the ontologies they use and share (parts of) the profile of the entity they represent. This information is stored by MF and SN-i agents. Once registered, the agents are allowed to negotiate. For that, B agents start announcing their buying products and wait for S agents to formulate proposals. Figure 1 illustrates the interactions between the main actors during a business negotiation.

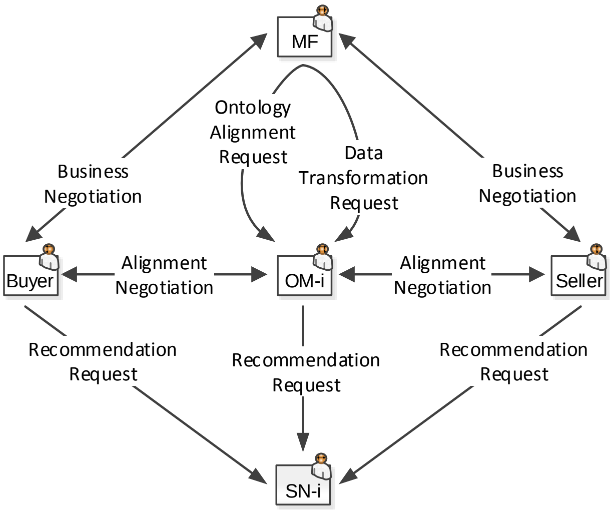


Figure 1: Main Interactions between Agents during a Business Negotiation (Nascimento et al., 2013a)

When the negotiation starts, the responsible MF selects the S agents that might be able to satisfy the B agent's request. For that it follows an ontology-based approach, selecting: (i) the S agents that use the same ontology as the B; and, (ii) supported by an OM-i, the ones that use ontologies that can be aligned with it. Therefore, the business negotiations may occur in two different scenarios: (i) a scenario where both agents use the same ontology; and (ii) a scenario where the agents use different ontologies.

In the first scenario the MF acts as a proxy between B and S, simply receiving and forwarding messages. While in the second, it is necessary to find an agreement about the alignment between the respective ontologies. For that the MF requests an OM-i to mediate an ontology alignment negotiation between B and S. If an agreement is achieved, the subsequent exchanged messages are sent to the OM-i, which translates their content according to the agreed alignment, ensuring that the message receiver will be able to understand it.

During the business negotiation the involved agents, B and S, exchange proposals and counterproposals, following a protocol based on the FIPA's "Iterated Contract Net Interaction Protocol Specification" (FIPA, 2002). The negotiation terminates when an agreement is achieved or when the agents have no more proposals to formulate. When a business agent satisfies all its business goals, or its deadlines are reached, it must terminate its activity, notifying the market and declaring the achieved results.

2.3. Ontology Alignment Negotiation

The ontology alignment negotiation initiates when a MF sends a request to an OM-i identifying (i) both business agents, (ii) the respective ontologies and (iii) providing information about the B agent's request.

The OM-i selects, from its repository, all the possible alignments between the indicated ontologies. Then, it performs sorting and filtering actions, following its internal criteria and/or requesting a SN-i to rank the alignments, obtaining a list of possible alignments and their respective score. Both B and S, analyze the recommended alignments taking into account their preferences, replying to the OM-i with the list of the alignments which they consider acceptable.

The OM-i analyzes both replies and checks if there is an agreement, i.e., if some alignment was selected by both agents. If there is no agreement, depending on the system configuration, the negotiation may terminate, or proceed, with the OM-i refining its list of recommended alignments and asking agents to reconsider their options and criteria. Otherwise, if there is an agreement, the OM-i notifies both agents and the MF about the agreement and proceeds with the transformation of the B agent's request. From that moment on, all the subsequent exchanged messages between the agents are forward to the OM-i for transformation.

2.4. Ontology Matching Services

When two agents that use different ontologies wish to exchange messages, a set of intermediary steps are necessary, namely: (i) discovering the correspondences between both ontologies – ontology matching process; (ii) represent the discovered correspondences so they can be applied in data transformation – ontology alignment document specification; and (iii) transform the content of the message according to the ontology alignment – ontology's instances transformation process.

In order to improve performance, in AEMOS, the ontology matching process is performed by specialized matching agents, in parallel to the market activity. When a new ontology is registered (e.g. during a business agent's registration), the specialized matching agents are notified. These then try to find correspondences between this new ontology and the already existing ones, using different techniques. The discovered alignments are reported to the OM-i which stores them in a repository. These alignments are then recommended during the ontology alignment negotiation process.

2.5. Social Network based Support

During the market activity, the SN-i collects information about its participants and their interactions. The SN-i then builds and maintains a relationship graph, applying SNA techniques (Wasserman and Faust, 1994) in order to capture proximity relations between agents, and adequacy relations from alignments to agents, which emerge during the agents' activities in the market. By combining this information, the SN-i is able to evaluate the adequacy of the alignments to each business negotiation.

3. ONTOLOGY ALIGNMENT NEGOTIATION PROCESS PROPOSAL

The results achieved following the approach described in the previous section, shown that by selecting more adequate alignments the agents normally achieve a higher business satisfaction, as the negotiation efficiency is improved (Nascimento et al., 2013b).

In this paper we propose a new ontology alignment negotiation process in order to increase the adequacy of the used alignments. We support the idea that, since the alignments are discovered using different techniques, each may include correspondences that are not included in the others. Therefore, more adequate/complete alignments may be achieved by combining parts of the already existing ones.

In this new ontology alignment negotiation model, we propose to reduce the granularity in the negotiation in order to achieve more adequate alignments. The agents negotiate each correspondence between ontology entities (i.e. classes, properties) separately from the original ontology alignment document.

Figure 2 (below) illustrates the main tasks which compose the new ontology alignment negotiation process, each one performed by a determined agent or group of agents. As illustrated, in order to achieve an ontology alignment agreement between two business agents, a set of steps is followed. In the first step the OM-i selects and proposes a set of possible correspondences to both B and S agents. The business agents then analyze the proposed correspondences deciding for each one if it should be included in the alignment or if it should be rejected (Step 2). When the OM-i receives the responses of both agents (Step 3) it checks if there is an agreement or if there are conflicting correspondences (Step 4). In the latter case, the OM-i decides if it is worth to continue negotiating, i.e. checks if an agreement seems probable (Step 5). If so, the OM-i formulates a request to both agents indicating the mutual agreed correspondences and the ones in conflict (Step 6).

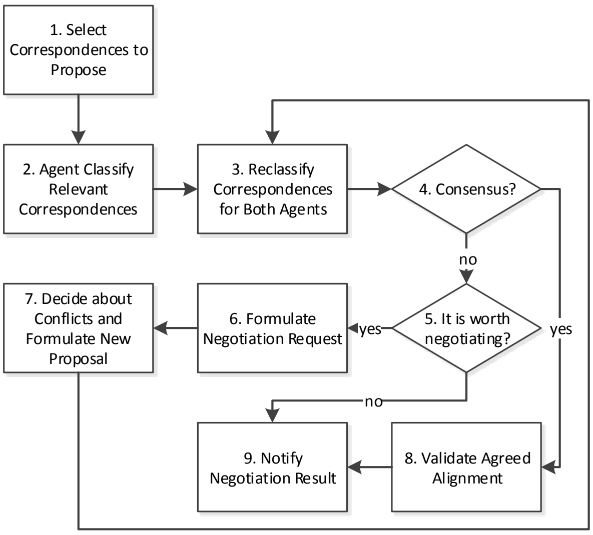

1. Select Correspondences to Propose
2. Agent Classify Relevant Correspondences
3. Reclassify Correspondences for Both Agents
4. Consensus?
no
7. Decide about Conflicts and Formulate New Proposal
6. Formulate Negotiation Request
yes
5. It is worth negotiating?
yes
no
9. Notify Negotiation Result
8. Validate Agreed Alignment

Figure 2: Ontology Alignment Negotiation Process

The agents analyze the negotiation request evaluating the conflicting correspondences and deciding if they can concede on their restrictions (Step 7). The responses are returned to the OM-i and the process (from Step 3) is repeated. If an agreement is achieved, the OM-i checks the final alignment for inconsistencies (Step 8). Finally, when an alignment is found, or the negotiation is terminated, the OM-i notifies both B and S, as well as the MF, about the negotiation result (Step 9).

The following subsections present further details about each step in this process.

3.1. Selecting Correspondences to Propose (Step 1)

The OM-i starts by selecting all possible correspondences between the indicated ontologies. For each ontology entity, the agent verifies if the amount of possible correspondences is considered elevated (i.e. it is above a defined threshold). If it is, the OM-i should reduce the amount of correspondences to propose. For that, the agent may simply consider the confidence value attributed by the matcher responsible for the correspondence's discovery. Or, in alternative, it may request an SN-i to indicate its confidence on each correspondence's correctness/adequacy to the business negotiation in question.

OM-i then sends to both B and S, the set of possible correspondences, including, for each one, the SN-i agent's evaluation (if it was performed) along with other additional information.

3.2. Analyzing Correspondences by Business Agents (Step 2)

Each business agent analyses the received set of correspondences taking into account its own preferences. The agent starts by selecting only the correspondences related to ontology entities considered relevant (i.e. the ones that are used by the agent to describe business goals/restrictions). For each selected correspondence the agent evaluates its confidence on the correspondence's adequacy taking into account the information provided by the OM-i or, alternatively, requesting an SN-i to support this evaluation. The agent then classifies each correspondence as: (i) mandatory, must be included in the final alignment, (ii) acceptable, might be included if the other agent agrees, and (iii) rejected, should not be included in the final alignment. In order to perform this classification the agent considers two types of threshold, namely: (i) mandatory threshold, above which the correspondence is classified as mandatory; and (ii) rejection threshold, below which the correspondences are rejected, the remaining are classified as acceptable. The agents respond to the OM-i indicating the correspondences' classification.

3.3. Analyzing the Business Agents' Responses (Steps 3 and 4)

The OM-i checks the agents' responses classifying the correspondences as: (i) mutually accepted, if it is mandatory for both agents, mandatory for one and acceptable for the other, or acceptable for both; (ii) mutually rejected, if it is rejected by both agents or rejected by one and not mandatory for the other; and conflicting, if it is mandatory for one and rejected by the other. Following this classification the OM-i verifies if there is a consensus, i.e. if there are no conflicting correspondences.

3.4. Deciding if the Negotiation Should Continue (Step 5)

If there are conflicting correspondences, the agent verifies if it is worth continuing the negotiation, i.e., it verifies if an agreement seems probable. For that we adopted a simplified approach, where the agent verifies the level of agreement between the agents (i.e. checks if the number of conflicting correspondences is reduced in relation to the number of agreements) and then analysis the interest of the agents in continuing negotiating. The result of this evaluation is a value from the range [0,1] which the agent compares with its defined threshold for negotiation.

3.5. Formulating a Negotiation Request (Step 6)

If the OM-i decides to proceed with the negotiation it will formulate a new request to the agents indicating the agreed correspondences and the conflicting ones. In order to provide an additional incentive to the agents, the OM-i might request an SN-i to evaluate the adequacy of the alignments which would result if the agents concede on their restrictions, and include this information in the request. For example, the OM-i could include in the request for each agent the adequacy of the alignment which results from (i) including the correspondences that it rejected originally and that are mandatory for the other agent, and (ii) not including correspondences that it classified as mandatory and that were rejected by the other agent.

3.6. Deciding about Conflicts (Step 7)

Each agent then decides upon the conflicting correspondences. Here, three situations may arise: (i) the correspondence was rejected by the agent and was classified as mandatory by the other; (ii) the correspondence was not evaluated by the agent (was not considered relevant) and was classified as mandatory by the other; and (iii) the correspondence was classified as mandatory by the agent and rejected by the other;

In the first two scenarios the agent may simply evaluate the impact that including the correspondences might have in the final alignment's adequacy. For that it may request support of a SN-i agent. In the last situation the agent has to decide if the correspondence can be conceded. This decision will depend on different factors, other than the impact in the alignment's adequacy. An important factor considered is the usage of the correspondent ontology entity in the definition of restrictions. Other important factor is the proximity of the deadline the agent has to accomplish its business goals. These factors are also important in deciding the level of interest of the agent in continuing negotiating when conflicts remain unresolved.

The agents respond to the OM-i indicating the correspondences which they find acceptable along with the correspondences that remain in conflict and their level of interest in continuing negotiating.

3.7. Validating the Final Alignment and Notifying the Negotiation Result (Steps 8 and 9)

The OM-i checks the final alignment for inconsistencies (e.g. redundant correspondences) resolving the ones it finds.

Finally, when an agreed alignment is found, or when the OM-i decides to terminate the negotiation, the result is reported to both B and S as well as the MF which initiated the process.

4. SOCIAL NETWORK BASED SUPPORT COMPONENT

During the alignment negotiation process, the agents may resort to SN-i agents in order to receive additional support for their decisions (cf. sections 3.1, 3.2, 3.5 and 3.6). In our previous models, SN-i agents would evaluate the adequacy of alignments essentially taking into account their previous usage, without giving much relevance to its content (although it was considered to evaluate its coverage).

In our current model, all the previous features of SN-i agents are maintained. However, these agents will now be capable of evaluating correspondences separately from the original ontology alignment documents, as well as evaluating new ontology alignments taking into account the included correspondences. Therefore, two new types of evaluation are added to SN-i agents' model, namely: (i) evaluating correctness/adequacy of a correspondence to an agent or business negotiation; (ii) evaluating the correctness/adequacy of a new ontology alignment to an agent or business negotiation;

When requested, the SN-i evaluates its confidence in the adequacy of a correspondence or alignment to an agent or a business negotiation (i.e. to a pair of agents). For that, it follows similar principles to the ones considered in our previous model for the alignment's adequacy to business negotiation evaluation (Nascimento et al., 2013b).

In each evaluation the agent considers a series of factors. For instance, to determine the adequacy of a correspondence the SN-i evaluates: (i) the confidence in the correspondence's correctness; (ii) the correspondence's adequacy to the agent (or pair of agents); and (iii) the correspondence's adequacy to the related agents (i.e. agents with high proximity relations to the agent). To determine a new alignment's adequacy, among other factors, the SN-i evaluates its coverage of the agents' relevant ontology entities.

In the following subsections we describe how each of the considered factors is evaluated. Then, in the final two subsections we describe how the SN-i combines these factors in order to determine its confidence values.

4.1. Correspondences Correctness

In this evaluation the agent considers information provided by its source, as well as its previous usage in business negotiations. More specifically, the agent considers: the confidence value attributed by the matcher (*cv*); the confidence/trust in the matcher (*cm*); the success rate in business negotiations where the correspondence was included (*src*); and the satisfaction in deals where the correspondence was included (*sdc*). The confidence in the correspondence's correctness (*cc*) is given by:

$$cc = \frac{w_1.cm.cv + w_2.src + w_3.sdc}{w_1 + w_2 + w_3} \quad (1)$$

where $w_{1\text{-}3}$ are the weights attributed to each factor, which are defined in the SN-i agent's configuration.

The agent considers the matcher's confidence in the correctness of the correspondence (*cv*). However, since different matchers may determine their confidence in different manners, the agent should consider the confidence in the matcher itself (*cm*). This confidence is normally defined in the agent's configuration, and may evolve during the agent's activity as correspondences from the matchers are used and evaluated.

In order to determine the correspondence's success rate in business negotiations (*src*), the SN-i analyses the outcomes of negotiations where the correspondence was included in the used alignments (similarly to how the alignment's success rate is determined in our previous model's description). The correspondence's success rate is given by:

$$src = \begin{cases} \frac{sn - fn}{tn} : tn > 0 \\ 0\ otherwise \end{cases} \quad (2)$$

where tn is the total number of negotiations where the correspondence was included, sn is the number of successful ones and *fn* is the number of failed ones.

The satisfaction in deal is a value in the range [0,1] provided by the B at the end of a successful negotiation, indicating its level of satisfaction with the achieved deal. Normally, it is determined by analyzing the similarity between the purchased product and the desired one (Nascimento et al., 2013b). The correspondence's satisfaction in deals is obtained by simply averaging the satisfaction value of each deal where the correspondence was included in the used alignment (also similar to the evaluation of the alignment's satisfaction).

4.2. Correspondence's Adequacy to an Agent

The evaluation of a correspondence's adequacy to an agent (*cata*) is given by:

$$cata(a,c) = \frac{\sum(w_i.f_i(a,c))}{\sum w_i} \quad (3)$$

where *c* is the evaluated correspondence, *a* is the agent, f_i is each evaluation factor and w_i is the weight associated to each factor. The considered factors: the success rate of the agent in negotiations where the correspondence was included (*srac*); the satisfaction in deals of the agent where the correspondence was included (*sac*); and the relevance the agent attributes to the ontology entity related to the correspondence (*re*).

The first two factors are determined in a similar way to the ones described in the previous subsection, only now the agent will evaluate only the negotiations where the agent participated.

The business agents attribute a relevance value (range [0,1]) to each of the used ontology entities, considering its usage frequency, as well as their use in specifying restrictions (Nascimento et al., 2013b). This information is provided by the agent during its registration, and can be used to evaluate the last factor.

4.3. Correspondence's Adequacy to the Closest Agents

The adequacy of the correspondence (c) to the agents closest to an agent (a) is given by:

$$catra(a,c) = \frac{\sum atar(a,c_i).cata(c_i,c)}{\sum atar(a,c_i)} \quad (4)$$

where c_i are the closest agents to a, i.e. those that have a high proximity relation with a, and $atar(a,c_i)$ gives the value of the proximity relation between agents c_i and a. Note that c_i can be related to a directly (there is a direct connection from a to c_i) or indirectly (there is a multi-steps path from a to c_i). In the latter case the value of the relation from a to c_i is obtained by the accumulated product of each relation value in the path.

4.4. Alignment's Coverage

The alignment's coverage evaluation differs depending on if it is related to a specific agent or to a business negotiation. In the first case the agent evaluates the alignment's coverage in relation to the agent's used ontology entities, taking into account their respective relevance. While in the second case, the SN-i will evaluate the coverage in relation to the ontology entities used in the initial request. In this case the ontology entities are considered to have the same relevance. The alignment coverage is given by:

$$cov = \frac{\sum r_i.ce_i - \sum r_j.ne_j}{\sum r_i + \sum r_j} \quad (5)$$

where ce_i is an ontology entity that is both relevant to the agent (or used in the request) and contemplated in the alignment, ne_j is an ontology entity that is relevant to the agent but not covered by the alignment, r_i and r_j are the relevance values assigned to the respective ontology entities.

4.5. Confidence in Correspondence's Adequacy

The SN-i agent's confidence in the correspondence's adequacy is given by:

$$cca = \frac{w_1.cc + w_2.cata + w_3.catra}{w_1 + w_2 + w_3} \quad (6)$$

where $w_{1\text{-}3}$ are the weights assigned to each evaluation factor, and *cc*, *cata*, *catra* are the considered factors previously described (cf. sections 4.1, 4.2, and 4.3 respectively). Note that, when the evaluation is performed considering a pair of agents (rather than a specific one), the evaluation of *cata* and *catra* will be obtained by averaging the results of this evaluation for each agent.

4.6. Confidence in New Alignment's Adequacy

The confidence of the SN-i agent in a new alignment's adequacy is given by:

$$caa = \frac{w_1.cov + w_2.acca}{w_1 + w_2} \quad (7)$$

where w_1 and w_2 are the weights assigned to each evaluation factor, *cov* is the evaluation of the alignment's coverage (cf. section 4.4) and *acca* is the average confidence in the adequacy of each included correspondence to the agent (or pair of agents) (cf. previous subsection).

5. COMPARISON WITH PREVIOUS MODEL

Consider an e-commerce scenario such as the one detailed in (Nascimento et al., 2013b), where a B uses the MP3P ontology and a S uses the CEO ontology. Figure 3 depicts the considered correspondences between these ontologies, some of which are incorrect (correspondences C9, C10 and C11).

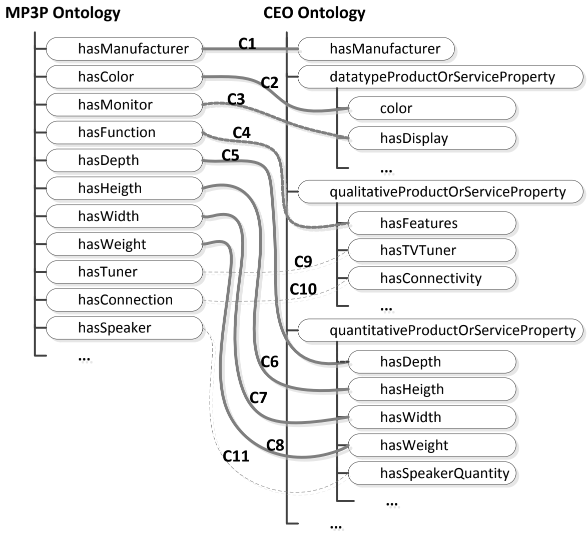


Figure 3: Considered Correspondences between MP3P and CEO Ontologies

Normally, in our previous model, we would consider two alignment possibilities: (i) one containing more correct correspondences, but with lower coverage (e.g. a_1={C1-C3,C5-C8}); and (ii) another with less correct correspondences, but with a higher coverage (e.g. a_2={C1,C4,C5-C11}). The alignments would then be evaluated by the SN-i as they were used, promoting the usage of the more adequate alignments.

However, in some cases, an inadequate alignment might include some correct correspondences which are not included elsewhere (e.g. a_2 includes C4 which is not included in a_1). In these cases, the process would benefit from evaluating each correspondence separately from the original alignment document.

Following the proposed approach, the OM-i would propose all possible correspondences (from C1 to C11) to both B and S agents. Since agents normally possess different classification thresholds, different combinations of correspondences may result from the ontology alignment negotiation process, being then used (tested) in business negotiations.

On the other hand, the agents normally classify the proposed correspondences resorting to an SN-i. The SN-i evaluates each correspondence taking into account, among other aspects, its previous usage in the marketplace, especially by the B and S agents and agents with high proximity relations to these (cf. section 4.5). As the correspondences are included in different alignments and used in business negotiations, their adequacy evaluation is refined, allowing the discovery of more complete/adequate alignments.

6. CONCLUSIONS AND FUTURE WORK

In this paper we propose a new model for the AEMOS' ontology services, which promotes the reuse and combination of already existent ontology alignments, with the goal of improving the adequacy of the alignments used in business negotiations. In this new model, rather than simply selecting/negotiating the alignment that should be used, the agent will also negotiate its composition, allowing excluding irrelevant correspondences, and achieving more adequate alignments as well as more accurate and trustful evaluations.

At this stage, we were interested in improving the alignment negotiation model taking advantage of the already developed components and mechanisms. Following the results achieved with this approach, the process should be significantly improved by adopting more sophisticated models and negotiation protocols.

ACKNOWLEDGMENTS

This work is supported by FEDER Funds through the "Programa Operacional Factores de Competitividade - COMPETE" program and by National Funds through FCT "Fundação para a Ciência e Tecnologia" under the projects: FCOMP-01-0124-FEDER-PEst-OE/EEI/UI0760/2011; PTDC/EIA-EIA/104752/2008.

REFERENCES

Du, T.C., Li, E.Y. and Chou, D., 2005. Dynamic vehicle routing for online B2C delivery. *Omega,* 33**,** 33-45.

Euzenat, J. and Shvaiko, P., 2007. *Ontology matching*. Secaucus, NJ: Springer-Verlag New York, Inc.

FIPA., 2002. *FIPA Iterated Contract Net Interaction Protocol Specification*. FIPA. Available from: http://www.fipa.org/specs/fipa00030/ [Accessed 28-01-2013].

Huang, W., Jin, J., Wang, N. and Wang, F., 2010. Technology and Application of Intelligent Agent in Electronic Commerce. *Proceedings of the 2010 International Conference on Measuring Technology and Mechatromics Automation (ICMTMA '10),* pp. 730-733. March 13-14, Changsha (China).

Malucelli, A., Palzer, D. and Oliveira, E., 2006. Ontology-based Services to help solving the heterogeneity problem in e-commerce negotiations. *Electroic Commerce Ressearch and Appliplations,* 5**,** 29-43.

Nascimento, V., Viamonte, M.J., Canito, A. and Silva, N., 2013a. Agent-Based Electronic Commerce with Ontology Services and Social Network Based Support. *Proceedings of the 15th International Conference on Enterprise Information Systems (ICEIS'13).* July 3-7, Angers (France).

Nascimento, V., Viamonte, M.J., Canito, A. and Silva, N., 2013b. Enhancing Agent Mediated Electronic Markets with Ontology Matching Services and Social Network Support. *Journal of Research and Practice in Information Technology, Special Collection on Ontologies and semantics in communication systems and networks*.

Viamonte, M.J., Nascimento, V., Silva, N. and Maio, P., 2012. AEMOS: An Agent-Based Electronic

Market Simulator With Ontology-Services And Social Network Support. *Proceedings of the 24th European Modeling & Simulation Symposium (Simulation in Industry) (EMSS '12)*, pp. 144-149. September 19-21,Vienna (Austria).

Viamonte, M.J., Ramos, C., Rodrigues, F. and Cardoso, J., 2007. ISEM: A Multi-Agent System That Simulates Competitive Electronic MarKetPlaces. *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications: Special Issue on Decision Support*, 15**,** 191-199.

Viamonte, M.J., Silva, N. and Maio, P., 2011. Agent-Based Simulation of Electronic Marketplaces With Ontology-Services. *Proceedings of the 23rd European Modeling & Simulation Symposium (Simulation in Industry) (EMSS '11)*, pp. 496-501. September 12-14, Rome (Italy).

Wasserman, S. and Faust, K., 1994. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences).* Cambrige: Cambrige University Press.

Zhang, L., Song, H., Chen, X. and Hong, L., 2011. A simultaneous multi-issue negotiation through autonomous agents. *European Journal of Operational Research*, 210**,** 95-105.

AUTHORS BIOGRAPHY

Virgínia Nascimento is a fellowship researcher in the Knowledge Engineering and Decision Support Research Center (GECAD) of the School of Engineering at the Polytechnic of Porto. Her research areas include multi-agent simulation, agent mediated electronic commerce, decision support systems and learning techniques. Nascimento has an MsC degree in computer science engineering from the School of Engineering at the Polytechnic Institute of Porto. Contact her at vilrn@isep.ipp.pt

Maria João Viamonte is a professor and researcher in the Knowledge Engineering and Decision Support Research Center (GECAD) of the School of Engineering at the Polytechnic of Porto. Her research areas include multi-agent systems, simulation, agent mediated electronic commerce and decision support systems. Viamonte has a PhD in electrical engineering from the University of Trás-os-Montes and Alto Douro. Contact her at mjv@isep.ipp.pt

Alda Canito is a graduate student in the Knowledge Engineering and Decision Support Research Center (GECAD) of the School of Engineering at the Polytechnic of Porto. Her research areas include Information Integration, Knowledge Engineering and the Semantic Web. Contact her at alrfc@isep.ipp.pt

Nuno Silva is a professor and researcher in the Knowledge Engineering and Decision Support Research Center (GECAD) of the School of Engineering at the Polytechnic of Porto. His research areas include Information Integration, Knowledge Engineering and the Semantic Web. Silva has a PhD in electrical engineering from the University of Trás-os-Montes and Alto Douro, Portugal. Contact him at nps@isep.ipp.pt

CREATING SORTED LIST FOR SIMULINK MODELS WITH GRAPH TRANSFORMATION

Péter Fehér [(a)], Tamás Mészáros [(a)], Pieter J. Mosterman [(b)] and László Lengyel [(a)]

[(a)] Department of Automation and Applied Informatics
Budapest University of Technology and Economics
Budapest, Hungary
[(b)] Research and Development
MathWorks
Natick, MA, USA

[(a)]{feher.peter, mesztam, lengyel}@aut.bme.hu, [(b)]pieter.mosterman@mathworks.com

ABSTRACT

Embedded systems are usually modeled to simulate their behavior and facilitate design space exploration. Nowadays, this modeling is often implemented in the Simulink® environment, which offers strong support for modeling complex systems. As design progresses, models are increasingly elaborated by gradually adding implementation detail. An important elaboration is the execution order of the elements in a model. This execution order is based on a sorted list of all semantically relevant model elements. Thus, to create an implementation or to execute a model, Simulink generates the dependency list of the model entities, which is referred to as the Sorted List. The work presented in this paper raises the level of abstraction of the model elaboration by modeling the Sorted List generation in order to unlock the potential for reuse, platform independence, etc. The transformation is implemented by applying graph transformation methods. Moreover, an analysis of the transformation is also provided.

Keywords: embedded systems, model-based design, model transformation, Simulink

1. INTRODUCTION

Nowadays a growing amount of software is modeled during the development phase. This is especially true with respect to the embedded systems. By modeling these systems, they can be examined based on, for example, their functionality, performance or robustness.

Selecting the most appropriate framework for modeling systems requires a remarkable amount of attention. Recently, domain-specific modeling has become a popular approach to describe complex systems. It is a powerful, but still understandable technique, the main strength of which lies in the application of domain-specific languages. Since domain-specific languages are specialized for a certain application domain, their application is more efficient than that of general purpose languages (Fowler 2010, Kelly and Tolvanen 2008).

MATLAB® (Matlab 2012) and Simulink® (Simulink 2012a) have undoubtedly become some of the leading tools for model-based system design and synthesis in the past years (Mosterman, Prabhu, and Erkkinen 2004, Mosterman and Vangheluwe 2002, Nicolescu and Mosterman 2010). As many other modeling tools, Simulink also offers the possibility to simulate the modeled system enabling thus to examine the behavior of the system before realizing it.

In order to simulate the system under design, Simulink must perform numerous preprocessing steps on the model. An important step of this preprocessing phase is inferring the execution order of the entities used in the model. This execution order is referred to as the *Execution List*. To establish the Execution List, Simulink must determine the relationships between the blocks. This is the aim of generating the so called *Sorted List* that constitutes a dependency list (Fehér et al. 2012, Simulink 2012b). That is, the Sorted List contains the elements of the modeled system in a specific order based on the control and data dependencies that determine how the different blocks can follow each other in the overall execution.

At present, generating the Sorted List is implemented in the Simulink code base. Though efficient, this makes difficult or even prevents unlocking value for which a higher level of abstraction is more appropriate (e.g., reasoning about the implementation and modularization of operations). Therefore, by implementing the Sorted List generation procedure at a higher level of abstraction, the advantageous properties of domain-specific modeling can be utilized. This is a fundamental premise of Computer Automated Multiparadigm Modeling; to use the most appropriate formalism for representing a problem at the most appropriate level of abstraction (Mosterman and Vangheluwe 2004, Mosterman, Sztipanovits, and Engell 2004).

This paper focuses on a novel solution to generate the Sorted List for Simulink models. This approach is based on a model transformation created in the Visual Modeling and Transformation System (VMTS) framework. VMTS has been prepared to communicate with the Simulink environment; therefore the model transformations designed in VMTS can be directly applied to the various Simulink models. Using model transformation to solve the Sorted List generation issue helps to raise the abstraction level from the API programming to the level of software modeling. The solution possesses all advantageous characteristics of model transformations, such as transparency, reusability and platform independence.

The rest of the paper is organized as follows. Section 2 presents the algorithm used for generating the Sorted List. Next, Section 3 introduces the implemented transformation. In Section 4, the properties of the transformation are examined. A simple example for the transformation is presented in Section 5. Finally, concluding remarks are elaborated.

2. CREATING A SORTED LIST

As it was previously described, Simulink creates the Sorted List based on the dependency of the elements. Previous work has captured in detail the dependencies that Simulink accounts for (Han and Mosterman 2010). A block *b* depends on the block *a* if the *direct feedthrough* (DF) property (Simulink 2012c) on its inport block obtaining the signal from *a* is set to *true*. In this case block *a* must appear before block *b* in the Sorted List. Else, there is no dependency, that is, the output of the block *b* with the input port can be computed without knowing the value on the input port. In this manner, there generally are many lists that satisfy the dependencies and it does not make any difference in semantics which list is selected.

At this phase of the processing, Simulink has already flattened the virtual subsystems of the model, therefore only nonvirtual subsystems left. Nonvirtual subsystems are, for example, the *Enabled Subsystem, Triggered Subsystem, Atomic Subsystem, Function-call Subsystem*, etc. These nonvirtual subsystems have their own Sorted List with the same principle discussed above.

The nonvirtual subsystems are treated as opaque blocks (in terms of execution) at the hierarchical level where they are used and so the input ports of a nonvirtual subsystem also have the DF attribute. The setting of this attribute is inferred from the content of the nonvirtual subsystem. Generally, for each input port of the nonvirtual subsystem the DF is set to be same as the DF attribute of the first block that the input connects to internally (to the nonvirtual subsystem). Note that a more sophisticated analysis may be applied to resolve some DF issues, as presented in other work (Mosterman and Ciolfi 2004, Denckla and Mosterman 2004).

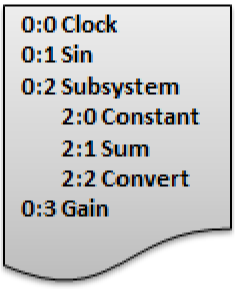

0:0 Clock
0:1 Sin
0:2 Subsystem
  2:0 Constant
  2:1 Sum
  2:2 Convert
0:3 Gain

Figure 1: An example Sorted List

The Sorted Lists have hierarchical layering, as Figure 1 depicts. The 0 : *x* says that *x* is the position of the block in the Sorted List for the 0 hierarchical layer (with 0 being the top). Similarly, 2 : *y* says that *y* is the position of the block in the Sorted List for the 2 hierarchical layer, which is chosen the same as the position of the block with the hierarchical layering in its parent's Sorted List. For example, 2 : 0 *Constant* says that the *Constant* block is at the first position for the 2 hierarchical layer, which relates to the *Subsystem* element.

In this section the algorithms for creating a Sorted List are presented as well as the complexity of the entire process.

2.1. The Algorithms

The main part of the Sorted List generation algorithm is shown in Algorithm 1. The SL algorithm contains a simple *Repeat-Until* block with only three algorithms. These three algorithms are responsible for processing the blocks.

Algorithm 1 The algorithm of the transformation SL

```
procedure SL()
repeat
    PROCESSFIRSTBLOCKS(null)
until not PROCESSSUBSYSTEM(null) and not PROCESSNORMALBLOCKS(null)
return
```

The three algorithms differ from each other in the type of block that they processing. First, the PROCESSFIRSTBLOCKS algorithm processes only blocks without any incoming edge. Since without incoming edges a block cannot depend on any other block, the processing of these blocks does not need the examination of the DF properties. The PROCESSFIRSTBLOCKS algorithm is presented in Algorithm 2.

Algorithm 2 The algorithm of processing blocks without incoming edges

```
procedure PROCESSFIRSTBLOCKS(Subsystem $sub$)
while PROCESSBLOCK($sub, true$) do
    DELETEBLOCKSANDEDGES()
return
```

The PROCESSFIRSTBLOCKS algorithm obtains a parameter, which sets the actual subsystem the blocks are being processed within. In case the current hierarchical level is the root of the model, then the parameter is *null*. The ProcessFirstBlocks algorithm contains a *while* block. The condition of this loop is the return value of the ProcessBlock algorithm, which is responsible for the processing itself. In case there was at least one block without incoming edges, the ProcessBlock returns *true* and the DeleteBlocksAndEdges algorithm is called. Since the DeleteBlocksAndEdges deletes the processed blocks and their edges, the next time the ProcessBlock algorithm is called, there may be new blocks without incoming edges. If there is no such a block, the algorithm terminates.

The ProcessBlock algorithm is shown in Algorithm 3. It has two parameters: the *sub* sets the actual subsystem, while the *onlyFirsts* parameter determines whether only the blocks without incoming edges should be processed. Since each nonvirtual subsystem is basically an opaque block for execution purposes, it has its own Sorted List. Therefore, if this parameter is set to a subsystem, then only those blocks should be processed that are contained by this subsystem. This containment can be checked by the examination of the *Parent* property of the block to be processed. The algorithm only processes "simple" blocks, it is referred to blocks that are not composite elements. Moreover, the *Inport* and *Outport* blocks of the subsystems should not be processed, therefore, they are not part of the *SimpleBlock* set.

Algorithm 3 The algorithm of block processing

```
procedure ProcessBlock(Subsystem $sub$, Boolean $onlyFirsts$)
Boolean $processedAny \leftarrow$ false
Block $parent \leftarrow null$
if $sub$ not $null$ then
    $parent \leftarrow sub$
for all $b|b \in SimpleBlocks \wedge b.Parent = parent \wedge b.Tag \neq "Processed"$ do
    if $(onlyFirsts \wedge b.Sources = \emptyset) \vee (\neg onlyFirsts \wedge \nexists e|e \in b.Source \wedge e.DF = true)$ then
        print CalculateIndent() + $b.Name$
        $b.Tag = "Processed"$
        if not $processedAny$ then
            $processedAny \leftarrow$ true
return $processedAny$
```

If the *onlyFirsts* parameter is set to *true*, the algorithm processes only the blocks without any incoming edge, that is, the *Sources* property is empty. Otherwise, if the *onlyFirsts* parameter is *false*, the algorithm must check if the DF properties on the port of the incoming edges are set to *false*.

The actual processing of a block is straightforward. The CalculateIndent method determines the actual indent and position of the block. These values depend on the hierarchical level and the number of previously processed blocks. The name of the block is appended to the calculated value. After the list is maintained the algorithm depicts the block as "Processed". If the algorithm processed at least one block, then it returns *true*.

As it was mentioned before, the DeleteBlocksAndEdges algorithm deletes the processed blocks and its edges. It is depicted in Algorithm 4.

Algorithm 4 The algorithm of deleting processed elements

```
procedure DeleteBlocksAndEdges()
for all $b|b \in Blocks \wedge b.Tag = "Processed"$ do
    for all $e \in b.InEdges$ do
        $e.Delete()$
    for all $e \in b.OutEdges$ do
        $e.Delete()$
    $b.Delete()$
return
```

After there are no "simple" blocks left in the model without incoming edges, the SL algorithm moves on to process possible subsystems. This is achieved by the ProcessSubsystem algorithm shown in Algorithm 5. Since it is possible in Simulink that a subsystem contains another subsystem, the algorithm obtains the current subsystem as a parameter. If it is set to *null* then the algorithm looks for a subsystem on the root level. The return value of the CheckForSubsystem algorithm determines if there is any processable subsystem on the given hierarchical level. If there is any, then it is stored in the *newSub* variable, and processed in a *repeat-until* loop.

Algorithm 5 The algorithm of processing a Subsystem

```
procedure ProcessSubsystem(Subsystem $sub$)
Subsystem $actualSub \leftarrow sub$
Subsystem $newSub \leftarrow null$
Boolean $processedAny \leftarrow$ false
if CheckForSubsystem($actualSub, newSub$) then
    repeat
        $processedAny \leftarrow$ true
        DeleteBlocksAndEdges()
        ProcessInOutPorts($newSub$)
        ProcessFirsts($newSub$)
    until not ProcessSubsystem($newSub$) and not ProcessNormals($newSub$)
    $newSub.Tag = "Processed"$
return $processedAny$
```

This loop is the same as the one in the SL algorithm but it has three additional commands. First, this algorithm sets the *processedAny* variable to *true*, which will be the return value. Next, it calls the DeleteBlocksAndEdges to delete the already processed elements if there are any. After this, the ProcessInOutPorts algorithm deletes the *Inport* and *Outport* blocks and the edges connecting to them. This is necessary, since the Sorted Lists do not contain these blocks. After these steps, the ProcessSubsystem algorithm processes the blocks in the same manner as the SL algorithm. As it can be seen in the Algorithm 5, the ProcessSubsystem algorithm can be called recursively, in case the subsystem contains another subsystem. The algorithms are called with the *newSub* parameter as the current hierarchical level.

In order to determine if there is any processable Subsystem on the current hierarchical level, the CheckForSubsystem algorithm is used. The algorithm is shown in Algorithm 6. A Subsystem is processable, if it is contained by the appropriate parent element and does not have any input port with the DF property set to *true*. In case the algorithm finds such a Subsystem, then it assigns the Subsystem to the *newSub* variable and returns *true*. Otherwise, the return value is *false*.

Algorithm 6 The algorithm of checking processable Subsystem

```
procedure CHECKFORSUBSYSTEM(Subsystem $actSub$, Subsystem $newSub$)
if $\exists s | s \in Subsystems \wedge s.Parent = actSub \wedge \nexists e | e \in s.Sources \wedge e.DF = true$ then
    print CALCULATEINDENT() + $s.Name$
    $newSub \leftarrow s$
    return true
return false
```

As it was mentioned before, the *Inport* and *Outport* blocks of the Subsystem are not part of the Subsystem. Therefore, their outgoing edges should be deleted without processing the blocks, as it is depicted in Algorithm 7. Note, that a nonvirtual Subsystem is only processable if none of its *In* ports has its DF property set to *true*. In case of virtual Subsystems this restriction does not hold, but the virtual Subsystems must be flattened before the processing of the Sorted List begins. This makes it possible to delete all edges connected to the *Inport* blocks.

Algorithm 7 The algorithm of processing edges related to *In*- and *Outport* blocks

```
procedure PROCESSINOUTPORTS(Subsystem $sub$)
for all $b | b \in InBlocks \wedge b.Parent = sub$ do
    for all $e \in b.OutEdges$ do
        $e.Delete()$
for all $b | b \in OutBlocks \wedge b.Parent = sub$ do
    for all $e \in b.InEdges$ do
        $e.Delete()$
return
```

The PROCESSNORMALS algorithm (Algorithm 8) is called from both *repeat-until* blocks. This algorithm is responsible for processing those blocks that have an arbitrary number of incoming edges but none of its input ports has a DF property with a value of *true*. It is similar to the PROCESSFIRSTBLOCKS algorithm but calls the PROCESSBLOCK algorithm with different parameter.

Algorithm 8 The algorithm of processing blocks with incoming edges

```
procedure PROCESSNORMALS(Subsystem $sub$)
if PROCESSBLOCK($sub$, $false$) then
    DELETEBLOCKSANDEDGES()
    return true
return false
```

Note that both *until* blocks (in SL and PROCESSSUBSYSTEM) have the same condition. That is, the algorithm stays in the loop if either the PROCESSSUBSYSTEM or the PROCESSNORMALS algorithm returns *true*. In other words this means that either of these two algorithms processes at least one element. In this case, after the deletion of the related edges, there may be new, processable blocks. Otherwise, if there are no Subsystems and no “simple” blocks to process, then either all elements have been processed or there is an algebraic loop in the current hierarchical level. The SL algorithm terminates if this condition is met in the root level, that is, no elements are left in the model or there is an algebraic loop.

This section has presented the SL algorithm, which was designed to create a Sorted List from the input model. Next, the complexity of the algorithm will be determined.

2.2. Complexity Analysis

In order to use an algorithm in production applications its complexity must be established. In this section the algorithms are examined based on their execution time. Therefore, the attributes that determine their computational complexity must be determined.

To determine the complexity of the SL algorithm, the following algorithms must be examined (the rest of the algorithms call these ones):

- PROCESSBLOCK
- DELETEBLOCKSANDEDGES
- CHECKFORSUBSYSTEM
- PROCESSINOUTPORTS

The PROCESSBLOCK algorithm is called from the PROCESSFIRSTBLOCKS and the PROCESSNORMALS algorithms. This is the one and only algorithm that processes “simple” blocks. On the one hand, since the Sorted List must contain all of these blocks, the body of the PROCESSBLOCK algorithm is executed at least n_{sb} times, where n_{sb} means the number of “simple” block on the current hierarchical level. On the other hand, since each processed block is deleted by the DELETEBLOCKSANDEDGES algorithm, the PROCESSBLOCK algorithm is executed at most n_{sb} times. Assume, that the complexity of the CALCULATEINDENT method is O(*1*). In this manner, the complexity of the PROCESSBLOCK algorithm is O(n_{sb}).

The DELETEBLOCKSANDEDGES algorithm deletes all processed elements and their edges. Based on the previous reasoning, all “simple” blocks are processed, therefore the *for* loop in the algorithm is executed at least *sb* times. However, the algorithm is also called after a Subsystem is processed. Let *n* denote the sum of the “simple” blocks and the Subsystem blocks. In this case the aforementioned *for* loop is run exactly *n* times. At each iteration the algorithm deletes the related edges as well. Assume, that the complexity of a deletion is O(1). In this manner, the complexity of the algorithm is O($n*(e_i+e_o)/2$), where e_i represents the average number of incoming edges to a processed element and e_o represents the average number of outgoing edges from a processed element. Let e_s denote all edges connected to either a “simple” block or a Subsystem. In this manner, the complexity of the DELETEBLOCKSANDEDGES algorithm is O($n+e_s$), every element and edge is deleted exactly once.

The CHECKFORSUBSYSTEM algorithm simply examines whether there is a processable Subsystem. If there is, then the algorithm assigns it to the *newSub* variable, and adds it to the Sorted List with the help of the CALCULATEINDENT method. Therefore, the complexity of the algorithm is $O(n_s)$, where n_s means the number of Subsystems in the model.

Finally, the PROCESSINOUTPORTS algorithm deletes all edges connected to the *Inport* or *Outport* blocks. In this case the complexity of the algorithm is $O(e_{io})$, where e_{io} represents the edges connected to the *Inport* or *Outport* blocks.

To summarize, the complexity of the SL algorithm is $O(n_{sb}+n+e_s+n_s+e_{io})$, which is equal to $O(2*n+e_s+e_{io})$. Let e denote all edges in the model, that is, the sum of e_s and e_{io}. In this manner, the complexity of the algorithm is $O(2*n+e)$, that is, each block ("simple" or Subsystem) is processed exactly once and all edges with the processed blocks are deleted.

3. IMPLEMENTING THE ALGORITHM

The previous section presented an algorithm that creates a Sorted List from a Simulink model. This section provides a novel approach to realize this algorithm: the algorithm is implemented via graph transformation.

With this approach, the solution utilizes the strong mathematical background of the graph rewriting-based model transformation. Moreover, the result possesses all the advantageous properties of the model transformation, for example, it is reusable, transparent and platform independent.

3.1. The Modeling Environment

To create transformations of a Simulink model, the Visual Modeling and Transformation System (VMTS) (Angyal, et al. 2009, VMTS 2012) was used. The VMTS is a general purpose metamodeling environment supporting an arbitrary number of metamodel levels. Models in VMTS are represented as directed, attributed graphs. The edges of the graphs are also attributed. VMTS is also a transformation system. It utilizes a graph rewriting-based model transformation approach or a template-based text generation. Whereas templates are used mainly to produce textual output from model definitions in an efficient way, graph transformation can describe transformations in a visual and formal way.

In VMTS the Left-Hand Side (LHS) and the Right-Hand Side (RHS) of the transformation are depicted together. In this manner, the process of the transformation is more expressive. VMTS applies different colors to distinguish the LHS from the RHS in the presentation layer. Imperative constraints can also be applied.

In VMTS a control flow determines the order of the transformation rules. Each controls flow has exactly one start state and one or more end states. The applicable rules are defined in the rule containers. This means that exactly one rule belongs to each rule container. The application number of the rule can be defined here as well. By default, the VMTS attempts to locate just one match for the LHS of the transformation rule. However, if the IsExhaustive attribute of the rule container is set to true, then the rule will be applied repeatedly as long as its LHS pattern exists within the model.

The edges are used to determine the sequence of the rule containers. The control flow follows an edge, which corresponds to the result of the rule application. In VMTS, the edge to be followed in case of a successful rule application is depicted with a solid gray flow edge, in case of a failed rule application with a dashed gray flow edge. Solid black flow edges represent the edges that can be followed in both cases.

3.2. The Transformation

As it was mentioned, a control flow determines the application order of the rules. The control flow of the TRANS_SL transformation is shown in Figure 2.

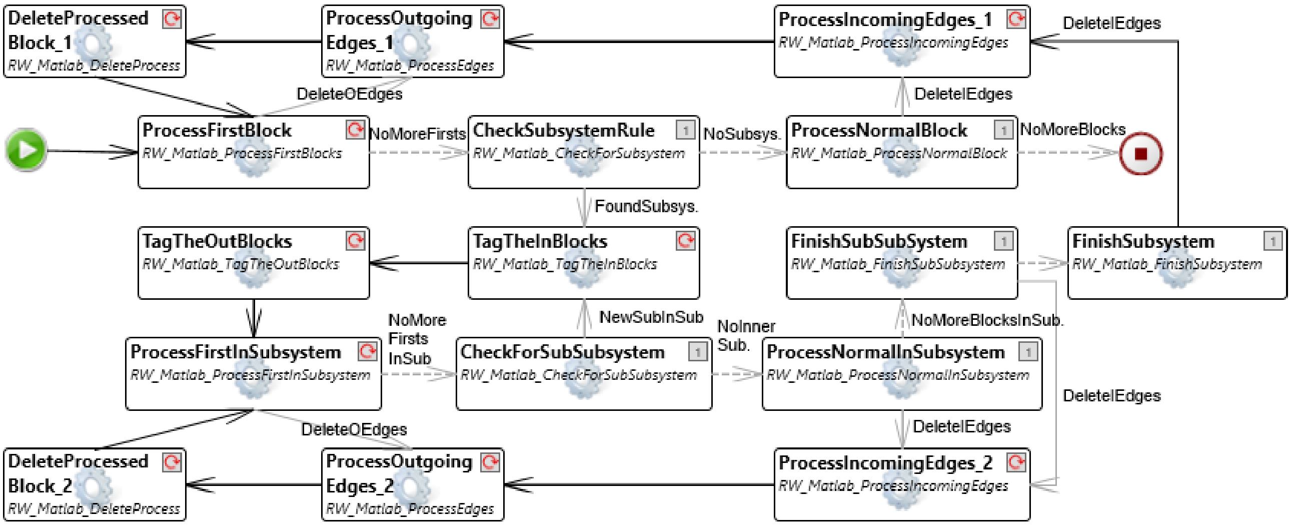


Figure 2: The control flow of the TRANS_SL transformation

At first, the transformation attempts to apply the RW_Matlab_ProcessFirstBlocks rule. This transformation rule matches for “simple” blocks, which have no incoming edges, therefore they do not depend on any other block. If a match is found, then the imperative code part of the rule writes out the name of the block with the necessary indentation, hierarchical level, and positioning. In this manner, the CalculateIndent method of the ProcessBlock algorithm is implemented via imperative code.

The rule is applied exhaustively, that is, as long the transformation finds an unprocessed “simple” block without incoming edges, and each processed block is tagged as “Processed”. In case there was at least one successful match, the transformation moves to ProcessOutgoingEdges_1 rule container, which applies the RW_Matlab_ProcessEdges rule (depicted in Figure 3). This rule attempts to find matches of the outgoing edges from the already processed elements and deletes them.

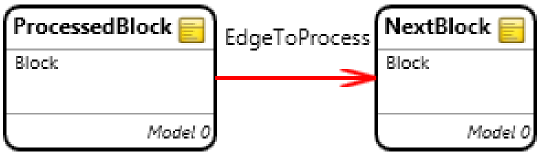


Figure 3: The transformation rule RW_Matlab_-ProcessEdges

After there are no edges left to delete, the RW_Matlab_DeleteProcessed transformation rule (shown in Figure 4), which is contained by the DeleteProcessedBlock_1 rule container, deletes the appropriate elements.

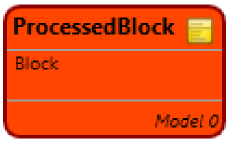


Figure 4: The transformation rule RW_Matlab_-DeleteProcessed

These three transformation rules are applied exhaustively, this way imitating the behavior of a *foreach* or *while* loop. The RW_Matlab_ProcessFirstBlocks implements the ProcessBlock algorithm with the parameter list of (*null, true*); and the RW_Matlab_ProcessEdges with the RW_Matlab_DeleteProcessed correspond to the DeleteBlocksAndEdges algorithm. Although the transformation does not delete any incoming edges related to the processed elements unlike the DeleteBlocksAndEdges algorithm, this is not necessary here, since these blocks do not have any incoming edges. Moreover, since the transformation moves to the RW_Matlab_ProcessFirstBlocks after deleting the already processed elements, the *while* loop of the ProcessFirstBlocks algorithm is implemented as well.

When the application of the RW_Matlab_ProcessFirstBlocks transformation rule was unsuccessful, the transformation moves along the dashed grey line of the control flow, which leads to the RW_Matlab_CheckForSubsystem rule. As it is shown in Figure 5, the rule attempts to find a match for a Subsystem, an Inport block and an ordinary Block. In Simulink, the DF property is an attribute of the In port of the blocks. However, in VMTS, this property is moved, and is a characteristic of the edges. In this manner, a Subsystem is processable, if the DF property of the *FirstEdge* edge (the edge starting from the Inport block) is set to *false*. In case such a Subsystem is found, it is written out with the help of the imperative code, and the transformation starts processing its elements.

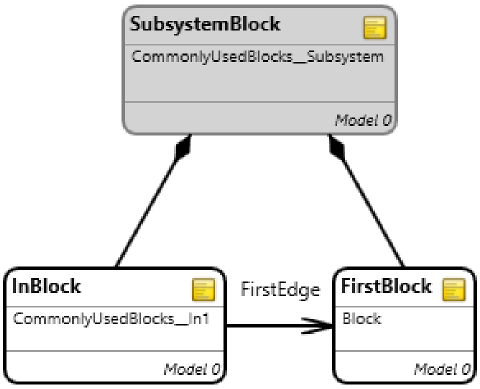


Figure 5: The transformation rule RW_Matlab_-CheckForSubsystem

This processing starts with deleting the edges connected to the Inport and Outport blocks. These deletions are accomplished in the RW_Matlab_TagTheInBlocks (depicted in Figure 6) and RW_Matlab_TagTheOutBlocks, respectively. The first rule may create blocks without any incoming edge, which means they are independent from any other block. The latter is only necessary to avoid any dangling edges after the transformation terminates.

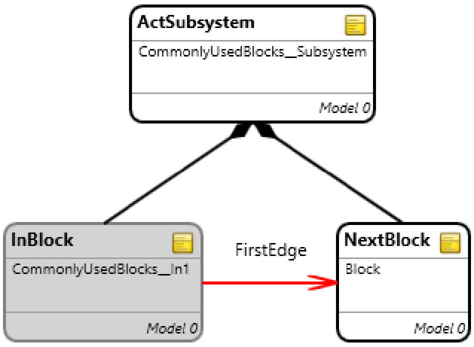


Figure 6: The transformation rule RW_Matlab_-TagTheInBlocks

The next three transformations are similar to the first three. The RW_MATLAB_PROCESSFIRSTINSUBSYSTEM rule (contained by the PROCESSFIRSTINSUBSYSTEM rule container), which is shown in Figure 7, processes the blocks without incoming edges. These blocks must be contained by the found Subsystem element, which is the only difference between this rule and the RW_MATLAB_PROCESSFIRSTBLOCKS. In case the RW_MATLAB_PROCESSFIRSTSINSUBSYSTEM rule is applied at least once, the transformation executes the same rules to delete the edges and blocks as in the beginning of the process. The application of a new rule container is necessary though, since the transformation now must move to the RW_MATLAB_PROCESSFIRSTSINSUBSYSTEM instead of the RW_MATLAB_PROCESSFIRSTBLOCKS.

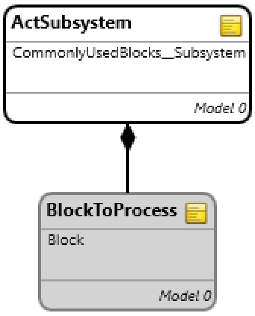


Figure 7: The transformation rule RW_MATLAB_PROCESSFIRSTSINSUBSYSTEM

When there is no other processable block without incoming edges, the transformation attempts to find a match for the RW_MATLAB_CHECKFORSUBSUBSYSTEM (presented in Figure 8) transformation rule. This rule basically attempts to find a Subsystem element inside the current one. Therefore, the applied rule is very similar to the already described one, the only difference is the presence of a parent Subsystem, which needs to be the current one.

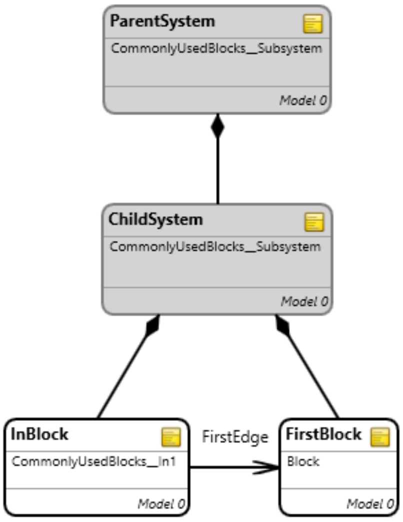


Figure 8: The transformation rule RW_MATLAB_CHECKFORSUBSUBSYSTEM

In case the transformation found a processable inner Subsystem, the transformation then starts to process with the already presented RW_MATLAB_TAGTHEINBLOCKS rule. This is essentially the implementation of the recursive CHECKFORSUBSYSTEM algorithms in a model transformation environment.

However, if there is no processable inner Subsystem, the transformation attempts to process the remaining blocks with the help of the RW_MATLAB_PROCESSNORMALINSUBSYSTEM rule. This rule is exactly the same as the RW_MATLAB_PROCESSFIRSTSINSUBSYSTEM but it allows the matched blocks to have incoming edges, if their DF properties are set to *false*. In case the transformation processed any block here, then it moves to the RW_MATLAB_PROCESSINCOMINGEDGES rule, which deletes the incoming edges of the processed elements. Next, the transformation returns to the RW_MATLAB_PROCESSEDGES rule. This structure corresponds to the *repeat-until* block of the CHECKFORSUBSYSTEM algorithm.

In case there are no blocks to process in the actual Subsystem, the transformation applies the RW_MATLAB_FINISHSUBSUBSYSTEM (shown in Figure 9) and RW_MATLAB_FINISHSUBSYSTEM rules. These are the exit points of the recursion and tag the found Subsystem as “Processed”. If the first rule can be matched, then it means that there was an inner Subsystem and the current Subsystem must be set back to its parent. After this, the edges connected to this processed Subsystem and then the Subsystem itself is deleted. The transformation continues with the processing of the parent Subsystem. In case the second rule is matched, then it means the transformation returns to the root level again. The applied transformation rules are the same, but the transformation returns to the RW_MATLAB_PROCESSFIRSTBLOCKS instead of its equivalent rule in the Subsystem level.

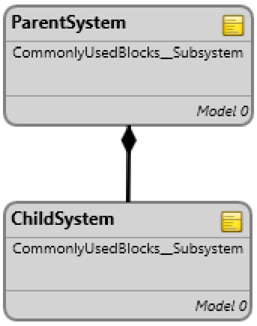


Figure 9: The transformation rule RW_MATLAB_FINISHSUBSUBSYSTEM

The transformation terminates when there is no processable element in the root level. This means that the RW_MATLAB_PROCESSNORMALBLOCK transformation rule does not find any match.

In this section the VMTS framework was briefly introduced. Moreover, a model transformation, which implements the algorithms defined in Section 2.1, was also presented. During the presentation, the mechanisms of defining *foreach*, *while* loops, *repeat-until* blocks and recursion were introduced. The transformation is also a good example of how well the declarative and imperative approaches can complement and extend each other.

4. THE ANALYSIS OF THE TRANSFORMATION

As complexity analysis is essential for applying a new algorithm, the analysis of a model transformation is necessary before the transformation is included in a robust engine, which is the subject of this section. First, the functionality of the transformation is examined and then its further attributes, such as termination and correctness, are verified.

Definition 1. *The **simple elements** of a Simulink model are all elements, that are neither a composite element (e.g. Subsystems) nor a mandatory element of a composite element (e.g. Inport and Outport block of a Subsystem).*

Before the transformation is examined in detail, note, that the transformation rules can be divided into two parts. The first part consists of the following:

- RW_MATLAB_PROCESSFIRSTBLOCKS
- RW_MATLAB_PROCESSEDGES
- RW_MATLAB_DELETEPROCESSED
- RW_MATLAB_PROCESSINCOMINGEDGES
- RW_MATLAB_CHECKFORSUBSYSTEM
- RW_MATLAB_FINISHSUBSYSTEM
- RW_MATLAB_PROCESSNORMALBLOCK

These transformation rules process the simple elements and find Subsystems on the root level. The second group contains the following:

- RW_MATLAB_PROCESSFIRSTSINSUBSYSTEM
- RW_MATLAB_PROCESSEDGES
- RW_MATLAB_DELETEPROCESSED
- RW_MATLAB_PROCESSINCOMINGEDGES
- RW_MATLAB_CHECKFORSUBSUBSYSTEM
- RW_MATLAB_FINISHSUBSUBSYSTEM
- RW_MATLAB_PROCESSNORMALIN-SUBSYSTEM

The rules contained by the second group behave exactly the same way as the ones in the first, but they processes the elements in deeper levels. The transformation rules match the same pattern with the exception of matching a parent Subsystem as well. So the reason behind the existence of the rules in the second group is the need of parent matching.

By the examination of the transformation we assume there is no algebraic loop in the model.

Proposition 1. *After the transformation TRANS_SL, all simple elements and Subsystems of the Simulink model are processed, therefore they are contained by the Sorted List. These elements are processed exactly once.*

Proof: The RW_MATLAB_PROCESSFIRSTBLOCKS transformation rule processes all simple elements without incoming edges. Throughout the process, its imperative code implements the CALCULATEINDENT method, which inserts the block into the Sorted List in the appropriate format. Each processed element is marked as “Processed”. Both the RW_MATLAB_PROCESSEDGES and RW_MATLAB_-DELETEPROCESSED rules match these marked blocks and delete their outgoing edges, moreover, the rules delete the blocks themselves. This may results in other blocks without incoming edges, therefore the RW_MATLAB_PROCESSFIRSTBLOCKS may be applicable again.

When these rules cannot be applied anymore, it means, that a nonvirtual Subsystem is in the way (which require special treatment) or there is a directed cycle in the model. If a Subsystem is found, the RW_MATLAB_CHECKFORSUBSYSTEM rule puts the element into the list and the transformation starts processing the elements of the Subsystem. This is achieved by the rules corresponding to the ones on the root level, therefore they are not discussed in more detail. After a Subsystem is processed, the transformation marks it as “Processed” and moves on to delete their edges, and finally the marked Subsystem as well. The processing of the model continues with the RW_MATLAB_PROCESSFIRSTBLOCKS again.

In case there is a directed cycle, the RW_MATLAB_PROCESSNORMALBLOCK looks for blocks which have none of their incoming edges marked with a DF property set to *true*. If the application of the rules was successful, then the rule inserts the found elements into the Sorted List and marks them as “Processed”. In this case, the transformation moves on to the rules responsible for deleting the related edges. In case, however, the application of the rule is unsuccessful, then the transformation terminates. This means that either there are no elements left in the model, or there are no elements left without having at least one incoming edges with the DF property set to *true*. The first case means that the transformation successfully processed all simple elements of the model. However, the latter case means that there is an algebraic loop in the Simulink model. Since it was assumed, there is no algebraic loop in the source model, the transformation cannot come to this result. In this manner, the transformation terminates if and only if there is no element left to process. ■

Proposition 2. *The transformation TRANS_SL processes the elements of the Simulink model in an appropriate order, that is, a block **a** is always processed later, than a block **b**, on which **a** is depend.*

Proof: There are three rules on the root level, which insert elements into the Sorted List:

- RW_MATLAB_PROCESSFIRSTBLOCKS, which processes only simple blocks without any incoming edges. A block without incoming edges does not depend on any other blocks, therefore it can be placed into the Sorted List. If a block *a* had an incoming edge *e* with the DF property set to *true*, but this edge *e* has been deleted after processing its source block *b*, then it means, *a* is now processable, since its dependency has been processed already. In this manner, this rule cannot insert any block into the Sorted List, that has any unprocessed dependency.
- RW_MATLAB_CHECKFORSUBSYSTEM, which processes Subsystem elements without incoming edges, or Subsystem elements with incoming edges with DF property set to *false*. These conditions ensure that the Subsystem is processable at the moment, it has no unprocessed dependency.
- RW_MATLAB_PROCESSNORMALBLOCK, which has the same conditions:

1. The processed block has no incoming edges,
2. The processed block has incoming edges, but none of them has its DF property set to *true*.

In this manner, these rules never insert any block into the Sorted List, that has any unprocessed dependency. These rules have their pairs in the Subsystem level, which behave the same in this regard. This means, the transformation TRANS_SL processes the elements of the Simulink model in an appropriate order. ■

Proposition 3. *The Sorted List created by the transformation TRANS_SL is a valid Sorted List for the input Simulink model.*

Proof: Proposition 1 states that every simple element is processed by the transformation, and Proposition 2 state that the elements are processed in the right order. This means that the Sorted List created by the transformation TRANS_SL is a valid Sorted List of the model, since it contains all relevant elements in an appropriate order. ■

Proposition 4. *The transformation TRANS_SL always terminates.*

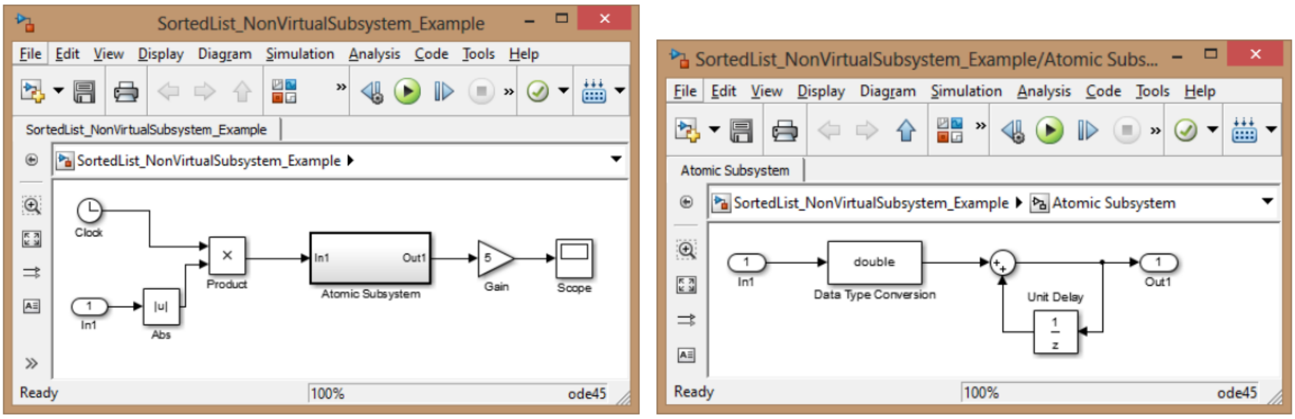
Proof: In order to prove the transformation always terminates, the following two statements have to be proved:

1. Each transformation rule is applied only a bounded number of times,
2. The transformation does not contain any infinite loop.

In VMTS the application mode of a transformation rule is set to either “Once” or “IsExhaustive”. In case the rule is applied “Once” then after the transformation attempts to apply the rule, it moves on to the next rule. The result of the application defines only the direction of the movement. Therefore, these rules are only applied a bounded number of times.

However, this is not the case when the application mode is set to “IsExhaustive”. In this case the transformation attempts to apply the LHS of the rule as long as there is a corresponding pattern in the host graph. In this manner, it has to be checked whether the rules applied in this way terminates. These rules are the following:

- RW_MATLAB_PROCESSFIRSTBLOCKS, where the rule attempts to match unprocessed simple elements. Since it marks the elements as “Processed” after each application, the rule is applied at most *n* times, where *n* means the number of simple elements in the Simulink model. The Simulink models contain only a bounded number of blocks, therefore the number of application of the rule is always bounded.
- The RW_MATLAB_PROCESSEDGES rule attempts to match processed simple elements with at least one outgoing edge, and deletes the matched edge. A block must have a bounded number of edges, therefore the rule cannot be applied indefinitely.
- The RW_MATLAB_DELETEPROCESSED rule is applied to process simple elements. Since the rule deletes the matched block, and a Simulink model has a bounded number of blocks, the rule is applied only bounded number of times.
- The RW_MATLAB_PROCESSINCOMINGEDGES rule attempts to match processed elements with at least one incoming edge. Since the rule is the same as the RW_MATLAB_PROCESSEDGES, but with incoming edges, the reasoning is analogous.
- The RW_MATLAB_TAGTHEINBLOCKS, where the rule attempts to match Inport blocks and deletes their outgoing edges. A Subsystem must have a bounded number of Inport blocks and an Inport block must have a bounded number of outgoing edges, the rule is applied a bounded number of times.
- The RW_MATLAB_TAGTHEOUTBLOCKS is similar to the RW_MATLAB_TAGTHEINBLOCKS, but it attempts to match Outport blocks with incoming edges. The same reasoning can be applied here as well, that is, the Subsystem must have a bounded number of Outport blocks and an Outport block must have a bounded number of incoming edges. Therefore, the rule is applied a bounded number of times.
- The RW_MATLAB_PROCESSFIRSTSINSUBSYSTEM rule is the pair of the RW_MATLAB_PROCESSFIRSTBLOCKS rule at a deeper level. Since it looks for the same pattern with the extension of a parent element, the same reasoning can be applied here as well.


SortedList_NonVirtualSubsystem_Example
File Edit View Display Diagram Simulation Analysis Code Tools Help
Clock
Product
In1
Abs
Atomic Subsystem
Out1
Gain
Scope
Ready
100%
ode45
SortedList_NonVirtualSubsystem_Example/Atomic Subs...
Atomic Subsystem
double
Data Type Conversion
Unit Delay
1/z
Out1

(a) The root level of the model

(b) The Atomic Subsystem

Figure 10: The example Simulink model

Based on the previous, none of the rules in the TRANS_SL can be applied an indefinite number of times.

Now, it has to be checked, that the transformation does not contain infinite loops. It can be stated, that none of the transformation rules create any new element, that is no new edges or blocks are created throughout the process. However, every processed element (and its edges) is deleted by the appropriate rule. Since the Simulink model contains a bounded number of elements, the processing rules, and the related rules deleting elements, can be applied only a bounded number of types.

Furthermore, in Simulink models the number of Subsystems that can be used is limited and hierarchies of Subsystems cannot be created in a recursive manner. Moreover, it is not possible to create a hierarchy, where Subsystem S_A contains Subsystem S_B and S_B also contains S_A. With these restrictions it is ensured that the application of the RW_MATLAB_CHECKFORSUBSYSTEM transformation rule cannot lead to an infinite loop because each found Subsystem will be processed and deleted.

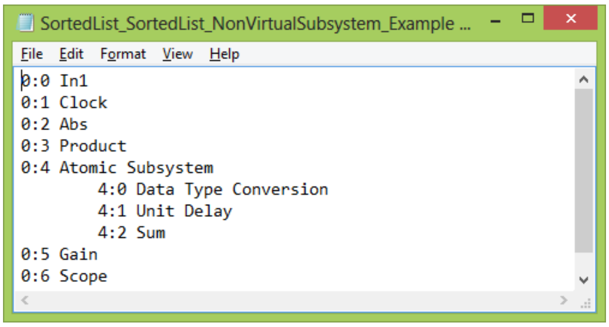
In this manner, since none of the transformation rules can be applied indefinitely and the transformation does not contain an infinite loop, it is proven that the transformation always terminates. ■

5. EXPERIMENTAL RESULTS

After introducing and analyzing the transformation, this section presents a simple example to demonstrate its functionality.

Figure 10 shows an example Simulink model. The top level of the model is depicted in Figure 10a, and the elements contained by the nonvirtual Subsystem are shown in Figure 10b. It is a simple example, which contains only one nonvirtual Subsystem, and a couple of simple elements. The transformation was examined on more complex models as well, and produced the expected results.

After the transformation for Figure 10b finished its execution, it resulted in the Sorted List depicted in Figure 11.


SortedList_SortedList_NonVirtualSubsystem_Example ...
File Edit Format View Help

```
0:0 In1
0:1 Clock
0:2 Abs
0:3 Product
0:4 Atomic Subsystem
          4:0 Data Type Conversion
          4:1 Unit Delay
          4:2 Sum
0:5 Gain
0:6 Scope
```

Figure 11: The resulting Sorted List

6. CONCLUSIONS

Nowadays Simulink is a popular tool for modeling embedded system. Simulink, in order to precisely model the functionality of the modeled system, can automatically elaborate a source model. Such elaboration is a form of model transformation process that is currently implemented in software as part of the Simulink code base.

Part of the elaboration is creating a Sorted List, which represents the dependency between the elements in the source model. In this paper, an algorithm is presented in detail, which is suitable for creating such a list. This algorithm is examined in terms of complexity.

Moreover, a detailed model transformation-based solution is also presented for creating Sorted Lists. This approach enables taking advantage of the benefits of model transformation such as reusability and platform independence. In this manner, the abstraction level of the model transformation problem can be raised. Besides the transformation details, the analysis of the transformation is also discussed and a simple example is given to presents its applicability.

Future work intends to study whether with the help of this transformation, the execution list can also be implemented via model transformation. In this manner, the abstraction level could be raised even further and more benefits unlocked.

ACKNOWLEDGEMENTS

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013) organized by VIKING Zrt. Balatonfüred. This work was partially supported by the Hungarian Government, managed by the National Development Agency, and financed by the Research and Technology Innovation Fund (grant no.: KMR_12-1-2012-0441).

REFERENCES

Angyal, L., Asztalos, M., Lengyel, L., Levendovszky, T., Madari, I., Mezei, G., Mészáros, T., Siroki, L., and Vajk, T., 2009, Towards a fast, efficient and customizable domain-specific modeling framework, *Software Engineering*.

Denckla, B. and Mosterman, P. J., 2004, An intermediate representation and its application to the analysis of block diagram execution, *Proceedings of the 2004 Summer Computer Simulation Conference (SCSC'04)*, pp. 167–172, July 2004.

Fehér, P., Mosterman, P. J., Mészáros, T., and Lengyel, L., 2012, Processing Simulink models with graph rewriting-based model transformation, *Model Driven Engineering Languages and Systems (MODELS '12) – Tutorials*.

Fowler, M., 2010, *Domain Specific Languages*, Addison-Wesley.

Han, Z. and Mosterman, P. J., 2010, Detecting data store access conflict in Simulink by solving boolean satisfiability problems, *Proceedings of the 2010 American Control Conference (ACC'10)*, pp. 5702–5707, June 2010.

Kelly, S. and Tolvanen, J.-P., 2008, *Domain-Specific Modeling: Enabling Full Code Generation*, Wiley.

Matlab® 2012b, http://www.mathworks.com/, 2012.

Mosterman, P. J. and Ciolfi, J. E., 2004, Using interleaved execution to resolve cyclic dependencies in time-based block diagrams, *Proceedings of 43rd IEEE Conference on Decision and Control (CDC'04)*, pp. 4057–4062, December 2004.

Mosterman, P. J., Prabhu, S., and Erkkinen, T., 2004, An industrial embedded control system design process, *Proceedings of The Inaugural CDEN Design Conference (CDEN'04)*, pp. 02B6–1–02B6–11, 2004.

Mosterman, P. J. and Vangheluwe, H., 2002, Computer automated multi-paradigm modeling, *ACM Transactions on Modeling and Computer Simulation*, vol. 12, no. 4, pp. 249–255.

Mosterman, P. J. and Vangheluwe, 2004, H., Computer automated multi-paradigm modeling: An introduction, *SIMULATION: Transactions of The Society for Modeling and Simulation International*, vol. 80, no. 9, pp. 433–450.

Mosterman, P. J., Sztipanovits, J., and Engell, S., 2004, Computer automated multi-paradigm modeling in control systems technology, *IEEE Transactions on Control Systems Technology*, vol. 12, no. 2, pp. 223–234.

Nicolescu, G. and Mosterman, P. J., 2010, Model-Based Design for Embedded Systems, *Computational Analysis, Synthesis, and Design of Dynamic Models Series*. CRC Press.

Simulink® 2012b, http://www.mathworks.com/simulink/, 2012.

Simulink® 2012b user's manual, http://www.mathworks.com/help/simulink/index.html, 2012.

Simulink® 2012b - direct feedthrough, http://www.mathworks.com/help/simulink/sfg/s-functionconcepts.html, 2012.

VMTS website, http://vmts.aut.bme.hu/, 2012.

EXPERIMENTAL AND NUMERICAL ANALYSIS OF THE KNOCKING PHENOMENON IN A GDI ENGINE

M. Costa[(a)], U. Sorge[(b)], B. M. Vaglieco[(c)], P. Sementa[(d)],F. Catapano[(e)]

[(a), (b), (c), (d), (e)] CNR – Istituto Motori, Viale Marconi, 8 – 80125, Naples, ITALY

[(a)]m.costa@im.cnr.it, [(b)]u.sorge@im.cnr.it, [(c)]b.m.vaglieco@im.cnr.it, [(c)]p.sementa@im.cnr.it, [(e)]f.catapano@im.cnr.it

ABSTRACT

The in-cylinder auto-ignition process leading to knocking in a GDI (gasoline direct injection) engine equipped with a high-pressure injector is numerically simulated. The turbocharged engine, 4-stroke, 4-cylinder, is also experimentally characterized at the test bench under stoichiometric conditions by varying the time of spark advance so to promote the occurrence of the knocking phenomenon. This last is numerically investigated by resorting to the Shell model, that allows reproducing within the combustion chamber the chemical activity of the mixture at low temperature in the *end-gas zone*. The simultaneous use of an Extended Coherent Flamelet Model (ECFM) for the combustion initiated by the spark plug allows calculating the pressure cycle, as well as the amount of CO produced under each considered operating condition. The link between the knocking occurrence and the CO amount at the exhaust is analyzed, together with the role of an intermediate species of the Shell model in localizing both spatially and temporally the occurrence of an undesired auto-ignition.

Keywords: multidimensional engine model, GDI, knocking.

1. INTRODUCTION

The tendency to knocking in spark ignition engines is favorably affected by the direct injection of gasoline within the combustion chamber, due to the decrease of the charge temperature consequent the evaporation process (Alkidas, 2007). Despite many investigations, the knocking phenomenon, however, is still a crucial topic in the design and development also of GDI engines, with remaining uncertainties and unsolved questions.

The knocking phenomenon limits the performance and the efficiency of an engine, preventing to exceed certain values of the compression ratio and spark advance. Knocking is known to appear as a characteristic metallic noise involving power loss, vibration, and, under particularly severe conditions, damage of mechanical parts.

Modeling and previewing the knocking occurrence is a really complicated matter, since it implies the need to take into account tents of reactions including hundreds of chemical species to reproduce the detailed underlying chemical mechanisms in the mixture not yet reached by the flame front initiated by the spark plug.

The auto-ignition of the so-called in-cylinder *end-gas zone*, in fact, results from a set of pre-flame or low temperature reactions, which lead to the start of combustion without an external source, but through the formation of not stable products of partial oxidation (peroxides, aldehydes, hydroperoxides, etc..) and thermal energy release. When the energy of chemical exothermic reactions exceeds the amount of heat transferred by the reagent system to the external environment, the combustion occurs spontaneously. As a result, the mixture temperature increases, rapidly accelerating the subsequent oxidation reactions.

The complex kind of the pre-flame reactions, of chain type between highly reactive species, produced in a manner superior to the consumed ones by different propagation reactions, may be controlled through the introduction of small amount of additives in the base fuel, which hinder or enhance the formation of radicals acting as chain propagators (Leppard, 1991; Li *et al.*, 1994).

Aim of this paper is to provide a better understanding of how the knocking is initiated in a GDI spark ignition engine. To this purpose, a synergic experimental and numerical study is performed: high temporal resolution pressure measurements are realized on a high performance turbocharged engine, together with multidimensional simulation of the in-cylinder energy conversion process carried out through a properly developed 3D model. In particular, different spark advances are tested to reach different levels of knocking intensity. The Fast Fourier Transforms (FFT) of the measured cycles allows identifying the frequency range of knocking. The lacks of the experimental data consequent the local nature of the phenomenon under study are overwhelmed through the in-cylinder numerical simulation, that relies on the simultaneous use of a *flamelet* model for the combustion initiated by the spark and a reduced kinetic scheme reproducing the auto-ignition of the *end-gas* mixture. The kinetic scheme, namely the Shell model, assumes the low-temperature activity as occurring through 8 steps involving 3 main fictitious species (indeed group of chemical species), that allows reproducing the major chemical events, as branching, propagation or linear termination.

The Shell model has been widely used in engine applications related to both Diesel and spark ignition engines (Sazhina *et al.*, 1999; Costa *et al.*, 2005). Although this model was developed more than thirty years ago, it is still used for computational fluid dynamics (CFD) applications, due to its simplicity combined with a generalized description of the kinetic mechanism of ignition, which has been proved as adequate to predict the main phenomena of interest in a variety of situations. Other kinetic mechanisms developed subsequently to the Shell, on the other hand, have a much higher number of reactions and species involved, which makes for their use being of little interest in complex computational domains (Griffiths *et al.*, 1994; Sahetchian *et al.*, 1995).

2. EXPERIMENTAL APPARATUS

The experimental apparatus of the present work includes the following modules: the spark ignition engine, an electrical dynamometer, the fuel injection line, the data acquisition and control units, the emission measurement system. The electrical dynamometer allows operating the engine under both motoring and firing conditions, hence detecting the in-cylinder pressure data and exploring the engine behaviour under stationary and simple dynamic conditions.

A spark ignition GDI, inline 4-cylinder, 4-stroke, displacement of 1750 cm3, turbocharged, high performance engine is considered. It has a wall guided injection system with a 6-hole nozzle located between the intake valves and oriented at 70° with respect to the cylinder axis. The engine is equipped with a variation valve timing (VVT) system in order to optimize the intake and exhaust valves lift under each specific regime of operation. The engine is not equipped with after-treatment devices. Further details are reported in Table 1.

Table 1: Characteristics of the engine under study.

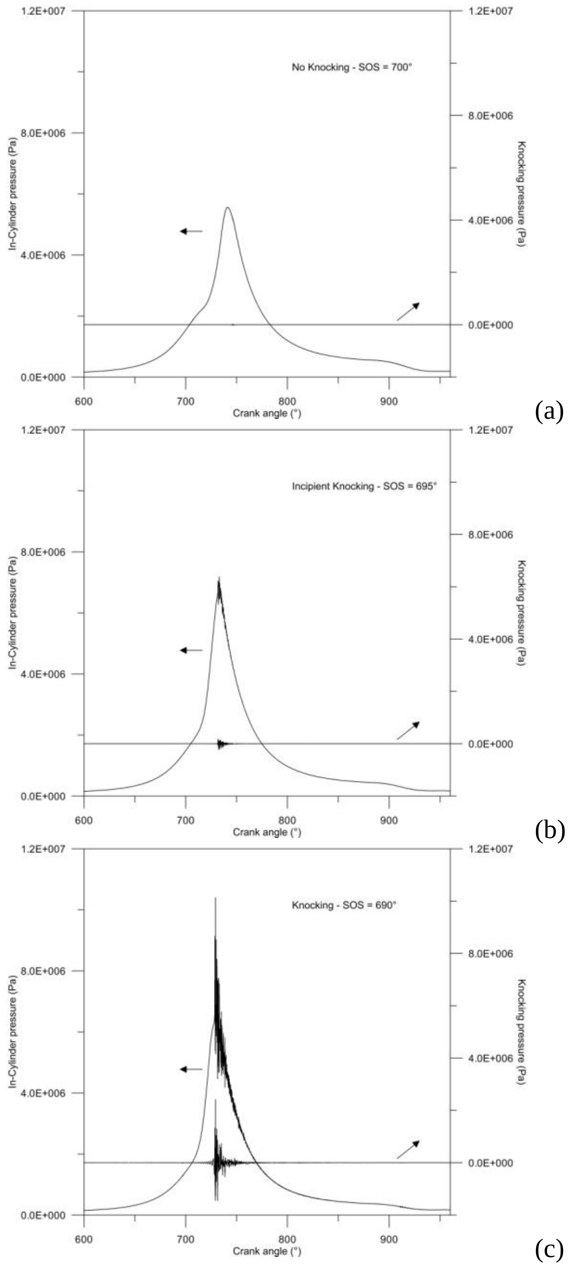
Engine characteristics	
Unitary displacement [cm^3]	435.5
Bore [mm]	83
Stroke [mm]	80.5
Turbine	Exh. gas turbocharger
Max. boost pressure [bar]	2.5
Valve timing	Int. and Exh. VVT
Compression ratio	9.5:1
Max. power [kW]	147.1 @ 5000 rpm
Max. torque [Nm]	320.4 @ 1400 rpm

An optical shaft encoder is used to transmit the crank shaft position to the electronic control unit for the electronic control. The information is in digital pulses, the encoder has two outputs, the first is the Top Dead Center (TDC) index signal with a resolution of 1 pulse/revolution, and the second is the crank angle degree marker (CDM) 1pulse/0.2degree. The engine is 4-stroke and the encoder gives as output two TDC signals per engine cycle. In order to determine the right crank shaft position, one pulse is suppressed via the dedicated software.

A quartz pressure transducer is installed into the spark plug in order to measure the in-cylinder pressure with a sensitivity of 19 pC/bar and a natural frequency of 130 kHz. Thanks to its characteristics, a good resolution at high engine speed is obtained. The in-cylinder pressure, the rate of heat release and the related parameters are evaluated on an individual cycle basis and/or averaged on 400 cycles (Zhao and Ladommatos, 2001, Heywood, 1988).

All the tests presented in the paper are carried out at the engine speed of 1500 rpm and high load. The absolute intake air pressure depends on the variable geometry turbocharger and it remains constant around 1300 mbar. The intake air temperature is kept at 323 K. The start of spark (SOS) is initially fixed at 700° (20° before TDC, BTDC), which corresponds to the maximum brake torque (MBT) condition achievable in the absence of knocking. In order to induce this last, the spark advance is progressively increased by steps of 5° crank angle. Commercial gasoline with 92 octane number is used. The fuel injection always occurs directly in the combustion chamber at the pressure of 10 MPa. The Start Of Injection (SOI) is fixed at 520° (200° BTDC). The direct fuel injection strategy is able to affect the combustion process and the pollutants formation (Sementa *et al.*, 2012). The duration of injection is fixed as equal to 4700 μs, that allows realizing a stoichiometric equivalence ratio, as measured by a lambda sensor installed in the engine exhaust. All the measurements are performed under steady state conditions. Each engine point is stabilized for around 10 seconds before starting the acquisition of the optical data and of the pressure signal.

Figure 1 shows the in-cylinder pressure in three different situations. Figure 1a shows the normal combustion cycle, having the spark advance at 20° BTDC. Figure 1b and 1c represents pressure cycles collected at two different SOSs, each greater than 5° with respect to the previous one. In particular, Figure 1a reports the average over 300 consecutive cycles in the MBT condition, Figure 1b and 1c report the instantaneous pressure curves corresponding to the 150th cycle of 300 consecutive ones. In Figure 1b and 1c the pressure traces show the typical ripples of knock. Their intensity increases as the spark advance is increased. During the test, the temperature in the combustion chamber increases because of the thermal evolution typical of engines under knocking condition.



(a)

(b)

(c)

Figure 1: In-cylinder pressure in the no knocking, incipient knocking and knocking case and 5-30 kHz band pass filter.

To evaluate the knock signal, the 5 - 30 kHz band-pass filtering of the pressure signals is performed as also shown in Figure 1 (Draper, 1938; Checkel and Dale, 1989; Brunt *et al.*, 1998). The knock intensity is obtained by the peak-to-peak amplitude of the knock signal. The combustion cycles, therefore, are classified with respect to their knock intensity in the normal combustion, borderline knocking, (standard) knocking and heavy knocking cycle. If the knock intensity is lower than the 5% of the motored pressure at TDC, the engine works in normal combustion conditions. Until the 15%, a borderline knocking is considered. The standard knocking is associated to the range 15-20%. For knock pressures higher than the 20% of the motored pressure, the heavy knocking occurred (Mittal *et al*, 2007). According to this classification, the knock signals for the selected cycle reported in Figure 1c represents the heavy knocking. For this case, the knocking onset occurs after the combustion initiation by the spark plug and it continues for part of the expansion stroke.

For the selected cycles, the FFT shown in Figure 2 indicates that primarily two ranges of frequencies are excited, the range of the calculation only include the combustion phase from 20° BTDC at 50° ATDC. The lower frequency is around 5 kHz and it corresponds to the first circumferential frequency of the combustion chamber (Figure 3). Three different peak around these frequencies are well defined. A second range of frequencies is observed between 7 and 21 kHz without a single well-defined peak. This range of frequencies corresponds to the first axial mode due to the motion of the piston and to the second and higher circumferential modes. At increasing knock, an increase in amplitude for the lower and higher frequency is detected.

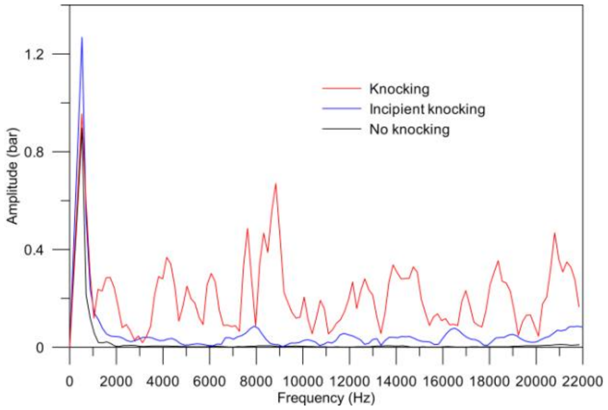
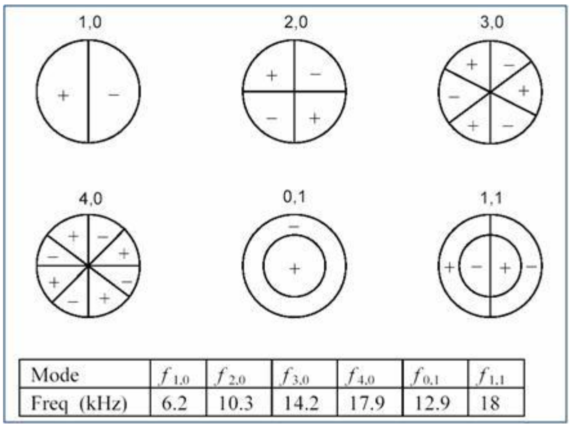


Figure 2: Pressure cycles Fast Fourier Transforms (FFT).



Mode	$f_{1,0}$	$f_{2,0}$	$f_{3,0}$	$f_{4,0}$	$f_{0,1}$	$f_{1,1}$
Freq (kHz)	6.2	10.3	14.2	17.9	12.9	18

Figure 3: Frequency modes of cylinder vibration.

3. 3D ENGINE MODEL

The CFD 3D model for the engine under study is developed within the AVL Fire^{TM} environment. The discretization of the computational domain corresponding to the cylinder and the intake and exhaust ducts is made through the pre-processing module Fame Engine Plus (FEP), with part of the domain discretized "manually" to increase the mesh regularity and assure stability of computations. An idea of the mesh size is given by Table 2.

Table 2: Typical size of the employed meshes.

Grid size			
Range	Starting	Dead Center	Ending
Exhaust Stroke	413868	421788	287349
Intake Stroke	315080	336947	342839
Closed Valves	197951	159808	197951

The whole 4-stroke engine cycle is simulated. Boundary conditions for the 3D model are obtained from the test bench analysis. Mixture formation occurs through a six-hole injector manufactured by Magneti-Marelli, 0.140 mm of hole diameter and solenoid actuation, mounted between the two intake valves.

The delivered spray was preliminary experimentally characterized both at the mass flow rate test bench and in an optically accessible vessel. In a previous work, the collected data were used to develop a proper 3D model of the spray dynamics by Costa *et al.* (2012). The model considers the droplets break-up according to the Huh-Gosman model (Huh and Gosman, 1991), the effects on the droplets dynamics of the turbulent dispersion through the sub-model by O'Rouke (O'Rouke and Bracco 1980), the coalescence through the sub-model by Nordin (2001), the evaporation through the sub-model by Dukowicz (1979). In developing the 3D engine model here discussed, the spray model is modified to also account for the gasoline droplets impingement on the piston or cylinder walls. This is considered according to the model by Mundo-Sommerfeld (Mundo *et al.*, 1994).

Combustion is simulated through the Extended Coherent Flamelet Model (ECFM) (Colin, Benkenida, and Angelberger 2003), NO formation follows the Zeldovich's mechanism (Zeldovich *et* al., 1947). The ECFM model is properly tuned to well catch the in-cylinder pressure curve by acting on the initial flame surface density and the flame stretch factor. For the sake of brevity, further details of the validation procedure of the 3D model are here not reported. The interested reader may refer to the paper by Allocca *et al.* (2012). The comparison between the numerically computed in-cylinder pressure and the experimentally measured one in the normal combustion case of Figure 1a is shown in Figure 4. A good agreement is visible in the intake, compression, combustion and expansion phases. In particular, the start of combustion is well reproduced.

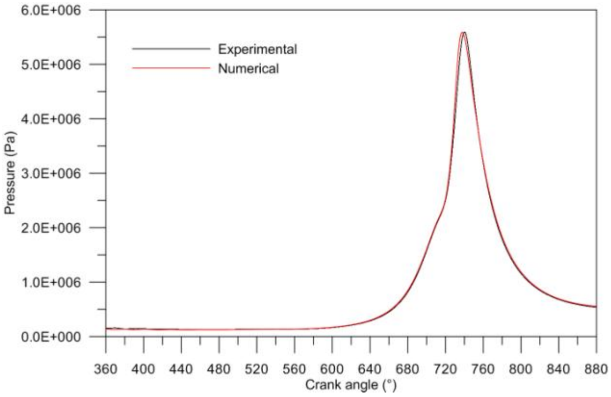


Figura 4: Experimentally measured and numerically computed in-cylinder pressure cycles.

The simulation of the auto-ignition process of an air-fuel mixture, as said within the introduction paragraph, can be carried out at different orders of approximation. A model that has proved successful in predicting both spatially and temporally the occurrence of the auto-ignition, and, at the same time, that does not require excessive computational time, is the so-called Shell model, developed by Halstead *et al.*, (1977). This model allows describing the process by schematization of the incubation phase of combustion through a reduced number of reactions involving not individual species, but groups of chemical compounds of similar behaviour.

Introducing the hydrocarbon RH, namely the fuel of composition C_xH_y, the Shell model is constituted by the following chemical reactions:

- primary initialization kinetic costant

  $RH + O_2 \rightarrow 2R^*$ K_q

- pre-flame propagation

  $R^* \rightarrow R^* + P$ K_p

  $R^* \rightarrow R^* + B$ f_1K_p

  $R^* \rightarrow R^* + Q$ f_4K_p

  $R^* + Q \rightarrow R^* + B$ f_2K_p

- branching

  $B \rightarrow 2R^*$ K_b

- endings linear and quadratic

  $R^* \rightarrow NR$ f_3K_p

  $2R^* \rightarrow NR$ K_t

where the letter P indicates the reaction products (CO_2, H_2O), while B and Q, respectively, represent branching agents and generic intermediate species. With the term NR are indicated not reacting compounds created at the end of the pre-flame reactions.

Into detail, the model contemplates the start of combustion, with the breaking of the chains of carbon-hydrogen fuel and the formation of radicals R*, and its development through the formation of oxygenated products. As already mentioned, the species that have a similar role in the kinetics of pre-flame are treated uniquely, as if they were a single entity. The advantages of using a reduced kinetic scheme, compared to a detailed scheme, consist just in the identification of groups of radicals or radicals which lead to branching of the chains of reaction or simple propagation of the linear type, and in the possibility to follow the variation in time of the concentration of these radicals.

The chemical pre-flame kinetics leading to the auto-ignition of the mixture not yet reached by the main flame front, within the present work, is reproduced after a proper tuning of the constants regulating the specific fuel reaction speed in the Shell model.

Some results of the 3D engine model are shown in Figures 5, 6 and 7 for the three conditions of Figure 1.

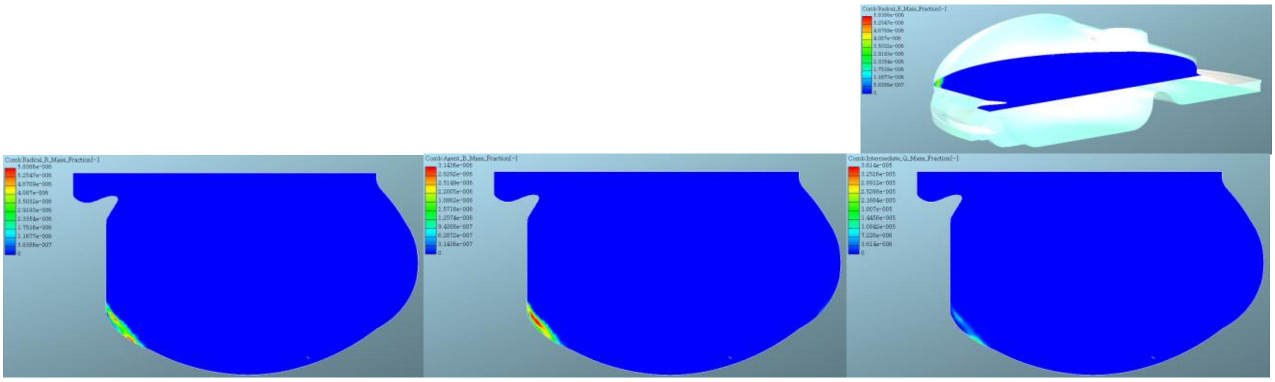


Figure 5: R* (left), B (middle) and Q (right) species on the plane orthogonal to the cylinder axis passing for the location of maximum Q for the no knocking case.

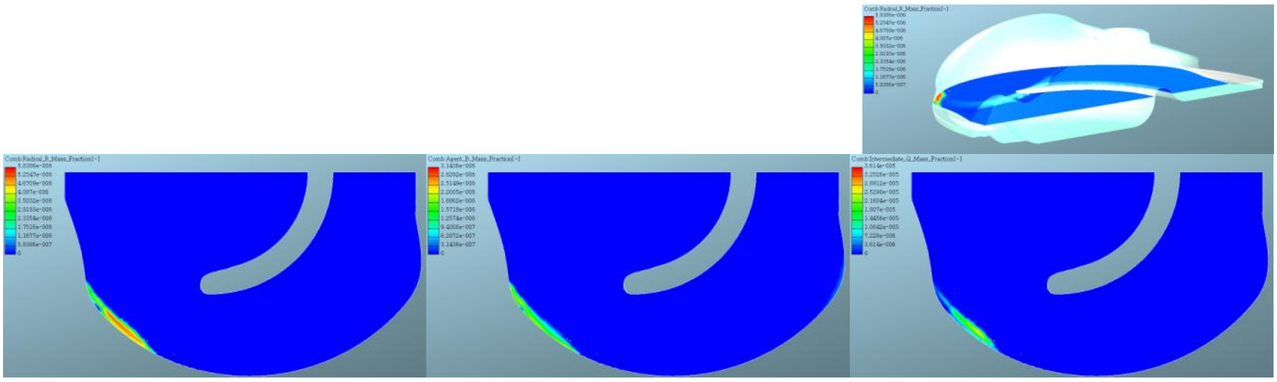


Figure 6: R* (left), B (middle) and Q (right) species on the plane orthogonal to the cylinder axis passing for the location of maximum Q for the borderline knocking cycle.

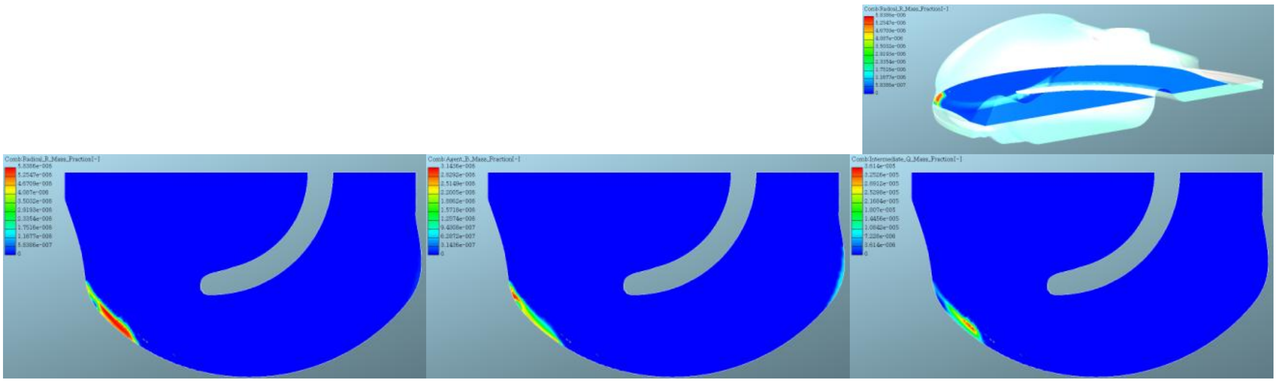


Figure 7: R* (left), B (middle) and Q (right) species on the plane orthogonal to the cylinder axis passing for the location of maximum Q for the heavy knocking cycle.

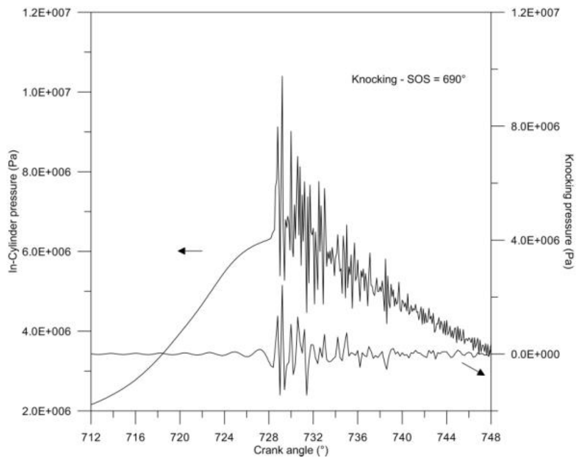


Figure 8. Particular of Figure 1C. Pressure cycle in the heavy knocking cycle.

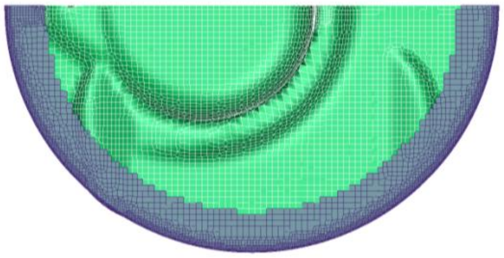


Figure 9: Piston surface and cells selection chosen to quantify the amount of formed Q species.

The R*, Q and B distributions in the combustion chamber are shown on planes orthogonal to the cylinder axis corresponding to the ones over which the species Q has its maximum value for each engine operation. This assures that the location where the most intense chemical activity is present is investigated for each engine operating condition defined by the specific value of SOS.

The crank angle considered for the case with SOS at 690°, is 729° (9 ° ATDC), that is in good agreement with the experimentally measured crank angle of knocking occurrence, as shown in Figure 8.

The chemical reactivity in the *end-gas zone* is higher in the knocking case, with a particularly high formation of the species Q, that may be taken as an indicator of the probability of knocking occurrence. From Figure 7, one may note that the most dangerous point for knocking onset within the combustion chamber is in the proximity of the wall on the intake valves and injector side, in agreement with the physical consideration that a decrease of the mixture temperature occurs in the opposite zone, where the spray is directed and mainly concentrated.

To better identify the variation of the Q species within the *end-gas zone*, a selection *ad hoc* is defined in the computational domain. In other words, the value of Q is calculated only in a volume made of the grid cells comprised within the annulus shown in Figure 9, having a thickness of 6 mm starting from the cylinder wall. As in Figures 5, 6 and 7, the intake valves and the injector are located on the left side, while the exhaust valves are on the right side.

In Figure 10 the mass fraction of the formed Q species in the afore mentioned cells selection is reported as a function of crank angle for the three different SOSs of Figure 1. The Q formation increases with increasing the spark advance, with the highest value occurring for SOS equal to 690°, thus confirming that the species Q is as a good knock indicator, identifying quite precisely the instant of crank angle where knocking occurs, as well as its location in space. The maximum of the Q relevant to the knocking case is just at 729°, namely at the knocking onset of Figure 8 (and 1.c). Figure 11 represents the maximum local value of the species Q mass fraction as a function of SOS, together with the angles measuring the interval of flame initiation and the flame propagation. These may be quantified by the interval of crank angle comprised between SOS and $\theta_{10\%}$ and the interval between $\theta_{10\%}$ and $\theta_{90\%}$, respectively. $\theta_{10\%}$ is the crank angle where the 10% of the mixture is burnt, while $\theta_{90\%}$ is the crank angle where the 90% of the mixture is burnt.

It is clear that by increasing the spark advance both the flame initiation and the flame propagation get slower. The greatest intervals needed for flame initiation and development at the highest spark advance, hence the unfavourable conditions of temperature and pressure at spark timing, give the mixture enough time to auto-ignite in the end-gas zone.

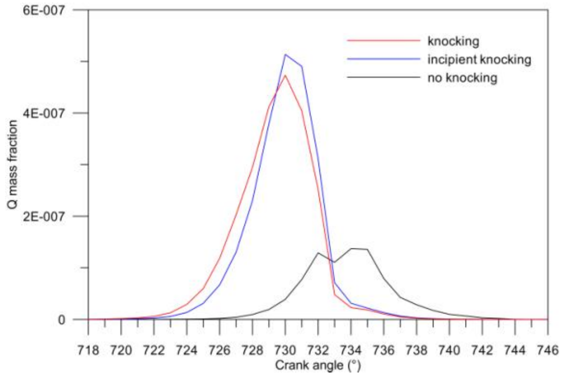


Figure 10: Trend of the species Q in the combustion chamber.

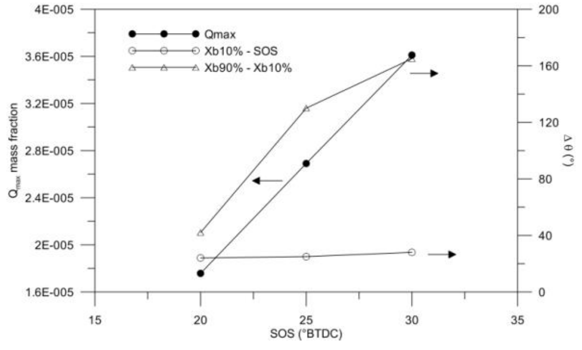


Figure 11: Maximum local value of species Q, flame initiation angle and flame development angle a function of SOS.

A good indicator of the mixture reactivity is also identified in the CO concentration at the exhaust valve opening, according to the paper by Li et al. (2007). As shown in Figure 12, the knocking case of Figure 1a exhibits a higher CO concentration at the exhaust valve opening with respect to cases of Figure 1b and 1c, having a 5° and 10° lower spark advance. This circumstance is also confirmed by Figure 13, where the numerically evaluated CO as a function of SOS are represented.

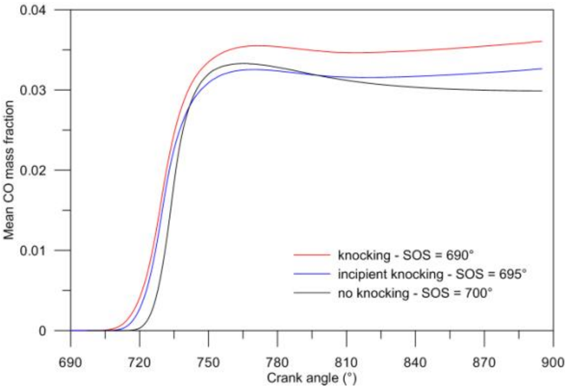


Figure 12: CO mass fraction computed by the 3D engine model in the combustion chamber.

4. CONCLUSION

A high performance GDI inline 4-cylinder, 4-stroke engine is experimentally tested under stoichiometric charge conditions to highlight the occurrence of knocking. The experimental measurements show that the increase of the spark advance increases the knocking intensity.

The knocking combustion is revealed through the FFT of the in-cylinder pressure that allows identifying the characteristic frequencies of the phenomenon.

Experimental data are employed to tune a properly developed 3D engine model, where the main combustion process is simulated through the ECFM model and the auto-ignition of the *end-gas zone* through the Shell model.

The computations show a good agreement with experiments as regards the knocking onset and its temporal location. The spatial position being the most probable for knocking is also highlighted. The chemical reactivity in the zone not yet reached by the flame front increases as the spark advance is increased, also as a consequence of the greatest time needed for flame initiation consequent the lower in-chamber value of temperature and pressure at spark timing.

The role of the species CO at exhaust valves opening, is shown as an indicator of the knocking occurrence. Indeed higher values are obtained in the case experimentally recognized as of heavy knocking.

REFERENCES

Alkidas, A.C., 2007. Combustion Advancements in Gasoline Engines, *Energy Conversion and Management*, 48, 2751-2761.

Allocca L., Costa M., Montanaro A., Sementa P., Sorge U., Vaglieco B.M., 2012. Characterization of the Mixture Formation Process in a GDI Engine Operating in Stratified Mode, *ICLASS 2012, 12th Triennial International Conference on Liquid Atomization and Spray Systems*, Heidelberg, (Germany), ISBN 978-88-903712-1-9.

Brunt, M.F., Pond, C.R., Biundo, J., 1998. Gasoline Engine Knock Analysis using Cylinder Pressure Data, *SAE Paper n. 980896*.

Checkel, M.D. and Dale, J.D., 1989. Pressure Trace Knock Measurements in a Current S.I. Production Engine. *SAE Paper n. 890243*.

Colin O., Benkenida A., Angelberger C., 2003. 3D Modeling of Mixing, Ignition and Combustion Phenomena in Highly Stratified Gasoline Engines, Oil & Gas Science and Technology – Rev. IFP Energies Nouvelles, 58, 47-62.

Costa M., Sorge U., Allocca L., 2012. CFD optimization for GDI spray model tuning and enhancement of engine performance, *Advances in Engineering Software*, 49, 43-53.

Costa M., Vaglieco B.M., Corcione F.E., 2005. Radical species participating the cool-flame regime of diesel combustion: a comparative numerical and experimental study, *Experiments in Fluids*, 39, 512-524.

Draper, C.S., 1938. Pressure Waves Accompanying Detonation in Internal Combustion Engine. *J. Aeronautical Sci*. Vol. 5.

Dukowicz J.K., 1979. Quasi-steady droplet change in the presence of convection, *informal report Los Alamos Scientific Laboratory*, Los Alamos Report LA7997-MS.

Griffiths, J.F., Hughes, K.J., Schreiber, M., Poppe, C., Dryer, F.L., 1994, A unified approach to the reduced kinetic modeling of alkane combustion, *Combustion and Flame*, 99 (3-4), 533-540.

Halstead, M.P., Kirsch, L.J., Quinn, C.P., 1977. The auto-ignition of hydrocarbon fuel at high temperatures and pressures – fitting of a mathematical model *Combustion and Flame*, 30, 45-60.

Heywood, J.B., 1988. *Internal Combustion Engine Fundamentals*, New York: McGraw-Hill.

Huh K.Y., Gosman A.D., 1991. A phenomenological model of diesel spray atomisation, *International Conference on Multiphase Flows*, Tsukuba, Japan.

Leppard, W.R., 1991. The autoignition chemistries of octane-enhancing ethers and cyclic ethers: A motored engine study, *SAE Paper 912313*.

Li, T., Nishida, K., Zhang, Y., Hiroyasu, H., 2007. Effect of split injection on stratified charge formation of direct injection spark ignition engines, *International Journal of Engine Research*, 8, 205-219.

Li, H., Prabhu, S., Miller, D., Cernansky, N., 1994. Autoignition Chemistry Studies on Primary Reference Fuels in a Motored Engine, *SAE Technical Paper 942062*.

Merola S.S, Sementa P., Tornatore C., Vaglieco B.M., 2009. Knocking Diagnostics in the Combustion Chamber of Boosted PFI SI Optical Engine. *International Journal of Vehicle Design*, Inderscience Publishers ed., 49, 70-90.

Mittal, V., Revier, B.M., Heywood, J.B., 2007. Phenomena that Determine Knock Onset in Spark-Ignition Engines, *SAE Paper n. 2007-01-0007*.

Mundo C., Sommerfeld M., Tropea M.C., 1994., Experimental Studies of the Deposition and Splashing of Small Liquid Droplets Impinging on a Flat Surface, *ICLASS-94 Rouen*, France.

Nordin W.H., 2001. *Complex Modeling of Diesel Spray Combustion*, Thesis (PhD), Chalmers University of Technology.

O'Rourke P.J., Bracco F.V., 1980. Modeling of Drop Interactions in Thick Sprays and a Comparison with Experiments, *Institution of Mechanical Engineers (IMECHE)*, London.

Sahetchian, K., Champoussin, J.C., Brun, M., Levy, N., Blin-Simiand, N., Aligrot, C., Jorand, F., Guerassi, N., 1995. Experimental study and modeling of dodecane ignition in a diesel engine, *Combustion and Flame*, 103 (3), 207-220.

Sazhina, E.M., Sazhin, S.S., Heikal, M.R., Marooney, C.J., 1999. The Shell Autoignition Model: Applications to Gasoline and Diesel Fuels, *Fuel*, 78, 389-401.

Sementa P., Vaglieco B.M., Catapano F., 2012. Thermodynamic and optical characterizations of a high performance GDI engine operating in homogeneous and stratified charge mixture conditions fueled with gasoline and bio-ethanol. *Fuel*, 96, 204-219.

Zeldovich Y.B., Sadovnikov P.Y., Frank-Kamenetskii D.A., 1947. Oxidation of Nitrogen in Combustion, *Translation by M. Shelef, Academy of Sciences of USSR, Institute of Chemical Physics*, Moscow-Leningrad.

Zhao, H., Ladommatos, N., 2001. *Engine Combustion Instrumentation and Diagnostics*, SAE Int., Inc.

A SIMULATION CASE STUDY OF EMAIL MANAGEMENT WITHIN A LARGE I.T. COMPANY IN GREECE

Agapi Vouvoudi[(a)], Makrina Viola Kosti[(b)], Lefteris Angelis[(c)]

[(a) (b) (c)]Department of Informatics, Aristotle University of Thessaloniki, Greece

[(a)]vouvoudiagapi@gmail.com, [(b)]mkosti@csd.auth.gr, [(c)]lef@csd.auth.gr

ABSTRACT

The use of Internet technologies in business processes has significantly altered the way people communicate and interact with each other, introducing new manners of communications, such as e-mails. Thus e-mail processing has become a considerable part of the everyday workload in companies. The time spent in replying to customers' requests and the overall response times of various e-mail categories are critical operational indices that contribute to efficient management decision making and work planning. In this paper we present a case study of a Greek company which provides IT and financial products and services. The study is based on a discrete event simulation model that represents the system of professional e-mail flow and processing within the company. The model is used to investigate different scenarios with experimentation and statistical analysis of the output.

Keywords: e-mail, simulation, mailbox, corporation

1. INTRODUCTION

During the last decades the corporation operational environment is continuously changing due to technological progress in many fields, especially in that of Information Technology and Internet. The use of Internet in business processes has significantly altered the way people communicate and interact with each other, introducing new manners of communications, such as e-mails. A study by Rogen International (Thomas et al., 2006) reported that from 1995 to 2001 e-mail communication grew by 600% and also that executives spent approximately 2 hours each working day checking and sending e-mails.

Using e-mails in the business context and the benefits of such use have been widely discussed and documented in the literature. These benefits originate from the relatively low cost of e-mails and a number of other characteristics such as the fact of being asynchronous (Thomas et al., 2006), i.e. the communicating parts do not have to be simultaneously present in order to interact. Furthermore, e-mails are text-based, can be forwarded or shared, depending on the recipient's needs, can be easily stored and traced and they are considered quite efficient according to various studies in the literature (Tyler & Tang, 2003; Dabbish & Kraut, 2006; Clark, 1996; Monk, 2003; Renaud, Ramsay & Hair, 2006). For the aforementioned reasons, electronic mail has proven to be an incredibly appealing and eventually a necessary means of communication not only in the corporate discipline but also in the interpersonal every day relations.

Since managers and business staffs are depending on e-mails to a great extent, researchers are trying to recognize and study problems associated to the aspects of this type of communication (Weber, 2004; Whittaker et al. 2005). One out of several challenges, managers have to cope with, is the handling of a high volume of mails that arrive on daily bases at their firms. Because of the increased volume of mails, working staff spends more time on e-mails than they used to, as is shown by several surveys. Particularly, the American Management Association in 2004 conducted a survey from 840 Organizations and reported that 47% of the Organization workers spent around 2 hours per working day.

Despite the positive aspects and the facility e-mail communication brings, it also leads to significant work time shortages and information overload (Denning 1982; Markus 1994; Berghel 1997; Jackson et al. 2003). Generally the mail response rates are quite low, however in Jackson et al. (2003) is shown that the accumulative time loss is considerable when we refer to very large organizations with very high e-mail flows.

Furthermore, many studies in literature have designated simulation as one of the most effective tools for management decision making **(**Hovey and Wagner, 1958; Green et al. 1977). Extended work has also taken place related to the practical application of simulation in industry and in the field of business and management (Watson, 1978; Christy and Watson, 1983; Lee et al. 1981; Millichamp, 1984; Hillier and Liebman, 1986).

This paper presents a study of a Greek company with a large number of departments and working employees. The collaboration allowed us to collect information about the amount of e-mails they deal with at a daily basis, the way they manage their mailbox and the importance of using e-mails for the operation of the organization. The aim was to develop a discrete event simulation model in order to study the real system of e-mail communication and to enlighten management operations regarding the impact of e-mails on employees’ engagement.

In Section 2 we analyze the problems rising by the entrance of e-mails in a company’s working flows. Section 3 presents the methodology used in order to develop our simulation model. Section 4 presents the results of our simulation with the corresponding validation and verification of the model. Section 6 describes some results after applying regression analysis to data derived from different scenarios we conducted. Finally we conclude with some reflections end discussion of our results.

2. PROBLEM DEFINITION

Discrete Event Simulation has been widely applied in business processes. Some of the research areas include decision making in military applications, health systems, economic studies, social analysis etc. A respectable amount of work also exists in simulation in sociology, which is extended in areas like iterated game theory, neural networks, multilevel simulation, simulation of social networks, policy oriented tax-benefit micro simulation etc. (Halpin, 1999). However, disproportionately with the importance of the role of e-mail in business operations, limited work has been performed with respect to mailbox simulation and management.

Specifically, Gupta et al. (2004) performed simulation based studies and concluded that if other tasks are more important and e-mail communication is secondary, e-mail messages should be checked 4 times a day with each processing period not exceeding 45 minutes. A significant contribution was made by Greve et al. (2007) with the eSIM model, who provided a decision support tool for testing the effectiveness of alternative e-mail processing strategies, given individual knowledge workers' e-mail environments. Another study working in the same direction is that of Narasimha (2007), which focuses on e-mail behavior of knowledge workers.

These studies may benefit in the area of productivity reduction and the more efficient e-mail management by the employees; it could be very helpful on the other hand though to study the system at a much higher level, that is that of the company’s manager or CEO. Thereby, we have issues regarding the knowledge of managers about the real time the employees dedicate to e-mail processing and the combination of occupation time slots in order to achieve better e-mail processing.

Another point is that in different working environments and operational areas the findings might be surprisingly different, always according e-mail load management. In this direction the study of a specific case could benefit the case itself and also help us reflect further over the problems of information flow inside a company.

In the context of conducting a master thesis, we contacted a large company, with its headquarters in Thessaloniki, Greece, in order to be able to discuss with the administration and employees about topics concerning the amount of e-mail they deal with every day and the way they manage it. We also wanted to extract information in relation to the impression they had about the time their employees spend handling e-mails.

The organization began its course of action in 1992, providing educational seminars regarding tax and labor issues. The company entered the computer market in 1997, having today three main operational directions: software production, network implementation and business training. Main axes of activities of the company are:

- Application of its long-time knowledge in order to accomplish high technology IT products.
- High quality services, implementation, operation support and training for systems and computer programs.
- IT consulting, organization and implementation of financial, management and other corporation studies.
- Establishment of educational seminars and lectures for the training of business staff on financial and taxation issues.

Several issues had to be taken into account; concerning the volume of e-mails the company coped with on a regular basis and which were straightly connected with the functionality of the company. These topics deserve thorough investigation within an operational research framework because they have direct impact on the available working time of the organization employees and their working performance and productivity.

The present simulation study is an attempt to describe the system systematically, as a model, in order to determine and measure various functional characteristics such as the time needed for e-mail processing and closure and the time needed from the employees to process the e-mails regarding their field of expertise. Moreover, the simulation gives the opportunity to study hypothetical functioning scenarios and also to determine and analyze the crucial factors that affect mailbox management

3. METHODOLOGY

For the purposes of our simulation we chose SIMUL8® (http://www.simul8.com). This software package is used by the academic users mainstream (Hlupic, 2000) for its editing capabilities, the animated processes and the visualization of the results (Johansson, 2002). Moreover, where statistical analysis was needed and performed, we used the statistical package PASW® (http://www-01.ibm.com/software/analytics/spss/).

After presenting the characteristics of the organization and defining the problem, the study was conducted in phases, presented in the subsequent sections, regarding: the system description, the designing of the simulation model, the collection and processing of the input data, the implementation of the simulation model and finally the analysis and interpretation of results.

In favor of simulation process correctness we conducted verification and validation of both model and results.

3.1. System Description

As already pointed out, the main objective of this simulation process is to identify the level of engagement of employees on e-mails and how that engagement can affect the organization. The paths an e-mail follows and the time consumed for its processing will be analyzed using internal information, retrieved by the company under study. By the term "company e-mails" we refer to the communication channels within the company, among customers and partners. The e-mails processed by the organization can be divided into two main categories: inbox and sent messages. The received e-mails include:

- customers' questions, i.e. queries on specific technical programming issues
- customers' requests on further functions for their software packages or the capability to extend them
- customers' suggestions to include new functions for future versions of logistic packages
- people's application forms for seminars participations.

As for outgoings mails, they include:

- company information directed to customers or vendors regarding the company seminars
- organization's promotion messages about its new software releases.

When an e-mail arrives in the company's inbox, it is classified in three major folders which are referred to as "Technical", "Business" and "Financial". Quite often employees promote an e-mail to other relative departments. They always save data about older, already processed messages as they desire to maintain the profile of each client. Finally, the failed-to-deliver e-mails are saved in a folder named "Error reports".

For the needs of the study we dealt with "Technical", "Business", "Finance", "Feedback", "New Editions", "Seminars" and "Error reports" company e-mail folders (Figure 3). E-mails for which their processing has been completed, are saved in specific folders, named "Closed", i.e. "Closed technical", "Closed business", "Closed finance" etc.

The information provided by the organization regarding an e-mail path is described as follows. The e-mail enters the inbox and waits in a queue to be categorized by the secretary. Afterwards, it is processed by the Directors' Department or is forwarded to other departments in order to reply to clients' requests. Finally, it is moved to one of the so-called "Closed" folders.

On the other hand, the procedure for outgoing e-mails unfolds as follows. The R&D department writes and sends e-mails about new software releases, while Business Trainers prepare and send e-mails about educational seminars. Most of the times, the company's staff tries to reply to customers as soon as possible (usually within a workday). If this is not possible, they prefer to communicate their reply by phone for faster service supply.

3.2. Model Building

For the reasons analyzed above, the system was divided into two separate independent subsystems. The first one refers to incoming messages while the second one to the Organization's outgoing mails. Each one was modeled with the Activity Cycle Diagrams (ACD), which represent the main entities and their mutual interactions. ACDs constitute a technique of modeling the interactions of system objects and are particularly useful for systems with a queuing structure, using only two symbols to describe the life cycle of the system's objects or entities (Paul, 1993). A green square illustrates the "active state" where collaboration of different classes of entities takes place and the red circle illustrates the "dead state" where an entity is waiting. Figure 1 illustrates the described schema:

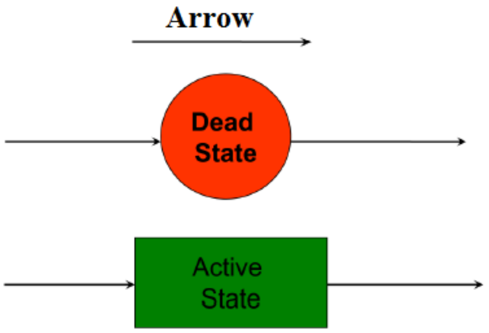


Figure 1: Symbols used in ACDs.

The procedures of arriving and sending e-mail systems are shown in the corresponding ACDs in Figure 2 where a number of abbreviations are used (see Table 1).

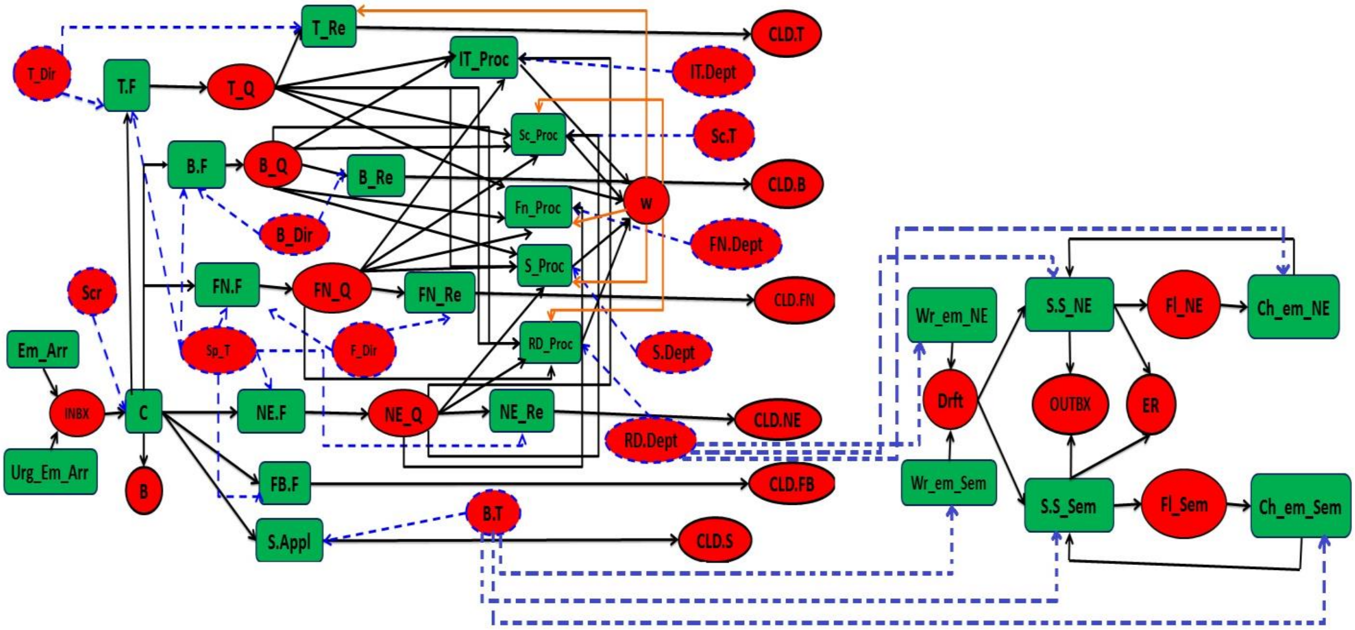


Figure 2: The activity cycle diagram of e-mail paths in the company

Table 1: System abbreviations

Em_Arr	**Urg_Em_Arr**	**C**	**T.F**	**B.F**	**FN.F**	**NE.F**	**FB.F**	**S.Appl**
E-mail Arrivals	Urgency Emails Arrivals	Categorization	Technical Files	Business Files	Finance Files	New Editions Files	Feedback Files	Seminars Applications
Sc_Proc	**Fn_Proc**	**S_Proc**	**RD_Proc**	**INBX**	**B**	**T_Q**	**B_Q**	**FN_Q**
Science Process	Financial Process	Sales Process	R&D Process	INBOX	Bin	Technical Queue	Business Queue	Finance Queue
CLD.FN	**CLD.NE**	**CLD.FB**	**CLD.S**	**Scr**	**T_Dir**	**B_Dir**	**F_Dir**	**B.T**
CLOSED Finance	CLOSED New Editions	CLOSED Feedback	CLOSED Seminars	Secretary	Technical Director	Business Director	Finance Director	Business Trainer
S.Dept	**RD.Dept**	**Wr_em_NE**	**Wr_em_Sem**	**Drft**	**OUTBX**	**S.S_NE**	**S.S_Sem**	**Fl_NE**
Sales Department	R&D Department	Write emails New Editions	Write emails Seminars	Draft e-mails	OUTBOX	Select-Send New Editions	Select-Send Seminars	Failed New Editions
T_Re	**B_Re**	**FN_Re**	**NE_Re**	**IT_Proc**	**NE_Q**	**w**	**CLD.T**	**CLD.B**
Technical Reply	Business Reply	Finance Reply	New Editons Reply	IT Process	New Editions Queue	waiting	CLOSED Technical	CLOSED Business
Sp_T	**IT.Dept**	**Sc.T**	**FN.Dept**	**Fl_Sem**	**ER**	**Ch_em_NE**	**Ch_em_Sem**	
Support Team	IT Department	Science Team	Financial Department	Failed Seminars	Error Reports	Check emails New Editions	Check e-mails Seminars	

As we can see from the systems' ACD, there are two entrances in the incoming e-mail subsystem: "E-mail arrivals" and "Urgent e-mail arrivals". Subsequently, the e-mails are categorized by the secretary and then processed by the department directors (technical, business and financial), by the support team and by the business trainers. If an e-mail needs further input or feedback, it is forwarded to the residing department (IT, Sales, R&D, Financial or Science Team). Moreover, in this simulation model the e-mail moves to the corresponding "closed" folder.

Regarding the outgoing mail subsystem, we also have two system entrances. One entrance new releases and the other for upcoming seminars. The resources that are utilized by this system are the R&D Department employees and Business Trainers, who write the corresponding e-mails, select the receivers and finally send the e-mail In the case of successful delivering, the aforementioned e-mails are moved to the "outbox". Otherwise, if failure occurs, the responsible department undertakes the re-check of the e-mail header elements (i.e. receiver's e-mail address etc.) and resends the faulty e-mail. If the problem persists (i.e. e-mail address no longer exists or the incoming mail server of the receiver is not working), the e-mail ends up to the "Error reports" folder. This helps the company to be able to check faulty address entries and update them.

Each one of the resources follows specific shifts (specific hours in workdays and weekends) according to their work-loads, that can manage e-mails. Something that is always important to be mentioned is that there always is at least one employee to service a client's request in each category and we assume that urgency e-mails are first in queues while the others follow FIFO (first in first out) distribution. In storage bins certain capacity or minimum wait-time has not been defined.

3.3. Input Data

In order to collect input data regarding the arrivals and the processing of e-mails, one of the authors had continuous collaboration with the company and arranged several meetings and interviews with employees, according to the needs of the study. Our conversations focused on the quantity, quality and the type of e-mail data, trying to represent efficiently the e-mail flow. During these meetings, the operations of the company were explained and also the basic paths that an e-mail followed between clients and company staff, or among organization departments. Furthermore, a meeting with the company's CEO concerned the company's policies on the management of the e-mails, the replying procedure, the folders and sub folder used in their mailbox and their use.

Overall, we were able to obtain data from e-mails for a period of four months. As already mentioned, the accumulated data concerned especially the "technical", "financial", "business", "feedback", "New Releases" and "Seminars" folders of the company mailbox. In total, we counted 6,136 e-mails, while only 5,560 of them were processed and used for this study.

Each e-mail consists of three parts:

- Conversation ID: This is the id number that each e-mail receives once it is sent, either from a client or from an employee. That id remains the same as long as it is forwarded or replied.
- Time stamp: This declares the date and the accurate time that the e-mail entered the system, was replied or was forwarded to the organization's departments.
- Location: This is the folder where each e-mail registers, according to its subject-topic.

Some other points that should be taken into account are:

- The ordinary working timetable of the company is from 8:00 to 16:00 every day, 5 days per week (however sometimes it is overcome because of excessive work necessities).
- The e-mail address of the company is able to accept mails 24 hours per day.
- The company's most important target is to reduce the volume of e-mails. That is why they occasionally prefer to reply to senders via telephone.
- Urgent e-mails are replied in much less time than the common ones.

3.4. Simulation Model

The simulation model was formulated by the activity cycle diagram in Figure 2 and implemented with SIMUL8®.

In order to use the collected data for our simulation model, we analyzed and classified them respectively. The percentages of the concluded e-mail categories are shown in Figure 3.

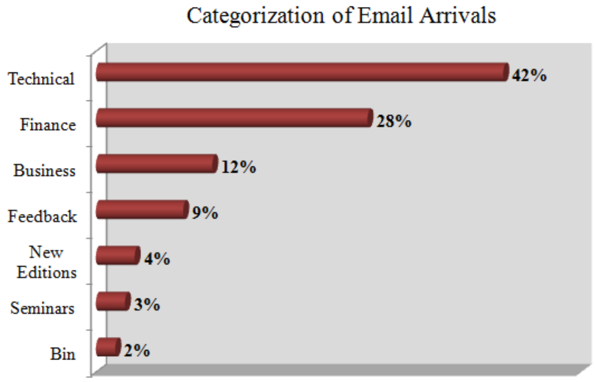


Figure 3: Main types of e-mails in the Organization mailbox (in %)

Once analyzed, the data were divided into four time intervals. Based on these intervals, the following time dependent distributions were defined:

- 8:01-12:00 (morning period with increased arrival rate)
- 12:01-16:00 (afternoon period with relatively lower arrival rate)
- 16:01-20:00 (evening period with significantly arrival rate)
- 20:01-8:00 (night with very slow arrival rate)

For the data classified in these intervals according to their time stamp, we estimated their arrival distributions. Furthermore, we extensively studied the time it takes an e-mail to get answered.

4. RESULTS

The time-unit used in the system is the minute; the simulation model runs 24 hours per day and 7 days per week. 3 working days were defined as warm-up period. Moreover, the time period simulated is 4 months.

The simulation model ran 100 times in order to have average values for the model entities and consequently information about e-mail manipulation from the company. We were interested in observing the occupation of the company's working units in the direction of e-mail processing on one hand and the time needed for e-mails to be considered as "closed". These two aspects could show if managers have a realistic image of employee's utilizations from the e-mail processing point of view and additionally would give us a clear picture of the quality of service of the company, expressed in e-mail processing time.

According to the data provided by the company and used as input, the simulation gave information regarding the utilization of the directors and the support team as presented in Table 2. We can see that the utilization ranged from 63% to 71%,.

Table 2: Department Directors and Support Team Utilization

Model Resource	Utilization %
Technical Director	69.23
Business Director	62.98
Finance Director	66.06
Support Team	70.65

The results indicate that the utilization of the department directors and the support team although are not low, they are not aligned with the general impression that their available e-mail processing time is fully exploited and in some occasions it is insufficient. The utilization of each department staff is shown in Table 3 and the results were surprisingly low.

Table 3: Department employees Utilization

Model Resource	Utilization %
IT Department	11.52
Science Team	3.84
Financial Department	11.53
Sales Department	5.33
Business Trainer	14.63
R&D Department	6.77

According to the incoming mail flux information and according to the processing time for each e-mail in average, the department employees' utilization does not concord with the manager's perception. This outcome reveals how difficult it is for managers to have a complete and clean view of the operation times needed in the organization context. As for the e-mails to be declared as closed, the average time in the system did not exceeded one hour. We have to point out that the system resources are available to process e-mails as suggested from the company itself and the model simulation concerns strictly e-mail manipulation and does not take into account other jobs employees might have in hands. It would be much more realistic to have a broader view of the company's activities and the according information in order to be able to infer more reliable results. In this case the e-mails would have been taken into account as interrupts in the system.

4.1. Verification and validation of the model

The quality of a simulation model depends on the content, the process and the outcome of the simulation itself (Robinson, 2001). This can be achieved through validation and verification techniques for each modeling step and its results.

The verification and validation needed for every phase were conducted with several tests after each step. To make the evaluation easier, features of Simul8 were utilized, such as animation, step by step monitoring and debugging. Also, during the model development's phase, sub model testing was conducted with extreme condition tests. Finally, at the phases of simulation model and results, statistical techniques, consistency tests and extreme condition tests were applied, which were supported by the simulation software.

Additionally, the parameters affecting the performance of the model were examined. Sensitivity analysis was applied, that is the process of testing the significance of the data input parameters and their impact on the system. We set up two tests to evaluate the impact of the arrival rates to the number utilization rates and the average time in simulation of the categorized e-mails. The results are shown in Tables 4 and 5. In Test 1 we increased e-mail arrival rates while in Test 2 we decreased them. In Table 4 the "+" shows if the change in the arrival rates affects the time in system of the e-mails. The opposite is shown with the "-" symbol.

Table 4: Sensitivity analysis results measuring the time in simulation

	Impact on time in system	
Categorized e-mail	**Test 1**	**Test 2**
Closed technical	+	+
Closed business	+	+
Closed finance	+	+
Closed N_Releases	+	+
Closed seminars	+	+
Closed feedback	+	+

Table 5: Sensitivity analysis results measuring employee's utilization

	Standard results	**Impact on employee utilization (%)**	
Resource		**Test 1**	**Test 2**
Secretary	74.28	83.69	37.70
Technical Director	69.23	76.96	40.85
Business Director	62.98	70.94	33.26
Finance Director	66.06	73.80	37.23
Support Team	70.65	79.80	34.40
IT Department	11.52	12.29	8.49
Science Team	3.84	4.09	2.85
Financial Department	11.53	12.31	8.47
Sales Department	5.33	5.63	3.96
Business Trainer	14.63	15.75	11.99
RD Department	6.77	7.13	5.26

In both cases, Test 1 and 2, the time in system is significantly affected by the change in the arrival rates.

Validation has to do with the assessment of behavioral accuracy of the model (Balci, 2003). It is the technique that allows us to demonstrate that the simplified built model, behaves with satisfactory accuracy, consistent to the study objectives. Therefore the results of the simulation were compared with observations from the realistic environment of the company and were discussed with the staff and the manager. Also, the software provided the statistical analysis, i.e. confidence intervals needed for the validation of the model, so no further analysis was conducted.

5. SCENARIOS AND REGRESSION ANALYSIS

We conducted an experimental design in order to test hypothetical scenarios and their affect to some of the system's aspects. The organization interest and concern mainly regarded the average time an e-mail stayed in their inbox before being replied. We found interesting though, to investigate how the queues are affected when changes take place in the available e-mail processing working hours.

To achieve the aforementioned task, we divided each working day in 8 equal time slots, lasting one hour each. Afterwards, depending on the working hours of each resource for the e-mail processing activity, we designed and conducted 40 scenarios. The structure of the scenarios is indicatively shown in Table 6, only for one of the scenarios. Hence all scenarios are similar matrices which can also interpreted as working timesheets. The existence of "1" in a cell shows that the corresponding resource works at the specific time slot.

Table 6: Structure of conducted scenarios

Time slots	Secretary	Technical Dir.	Business Dir.	Finance Dir.	Support Team	IT Dep.	Science Team	Financial Dep.	Sales Dep.	Business Trainer	R&D Dep.
1. 8:00-9:00	1	1	1	1							
2. 9:00-10:00		1	1	1	1						
3. 10:00-11:00											
4. 11:00-12:00	1										
5. 12:00-13:00	1	1	1	1	1						
6. 13:00-14:00					1						
7. 14:00-15:00		1	1	1		1	1	1	1	1	1
8. 15:00-16:00											

Thereby, we were able to monitor the aspects of the system mentioned in the first paragraph of this section, namely the exits of the e-mails and their average time in the system, in relation to the conducted scenarios. In detail, we investigated dependency tendencies of: (a) the average remaining times in the system of the technical, business and finance e-mail categories and (b) the average queue sizes of the same categories, in relation to the working timesheets performing regression analysis. Our aim was to model the aforementioned relationships rather than to perform prediction. It is important to mention at this point that our variables were not normally distributed. So, in order to be able to perform a parametric method we transformed the variables using Blom's rank transformation (Blom, 1958). For our regression analysis, we used PASW® (SPSS) with the stepwise predictor selection method.

According to the regression results regarding the average remaining time in the system of technical e-mails, we had a significant model with p<<0.005, explaining the 44% of the variability of the dependent variable (R^2=0.44). All the independent variables selected were statistically significant. In this case we observed that the average remaining time is negatively affected (it increases) when the support team and the business director work between 13:00 and 14:00 o'clock. For the average remaining time in the system of the business e-mails we had a statistically significant model explaining 59% of the variability of the dependent variable. In more detail, we observed that the average time decreases when the finance director works between 8:00-9:00 o'clock ($p<<0.001$), the support team between 10:00-11:00 o'clock ($p=0.001$) and the business director between 12:00-13:00 o'clock. On the other hand the occupation of the IT department between 8:00-9:00 o'clock seems to delay the remaining of business e-mails in the system.

Finally, regarding the category of finance e-mails and their average remaining time in the system we found a statistically significant model explaining 49% of the total variability of the dependent variable; that is the average remaining time, with three statistically significant independent factors. In this case we observed that the occupation of the finance director between 13:00-14:00 o'clock predisposes the increase of the average time of these e-mails in the system, while the occupation of the technical director and support team between 12:00:13:00 and 10:00-11:00 correspondingly seams to decrease the average remaining time in the system of finance e-mails.

As already mentioned we found also interesting to investigate the queue sizes of the technical, business and finance e-mail categories in relation to the conducted scenarios. In this direction ($R^2=0.824$, $p<0.001$) we observed an increase of the technical e-mails queue when the support team, finance director and finance department work between 11:00-12:00, 8:00-9:00 and 10:00-11:00 o'clock correspondingly. On the other hand the occupation of the finance director between 12:00-13:00 o'clock and of the IT department 8:00-9:00 o'clock seems to ease the queue load. Additionally, for the business queue average size we found a statistically significant model explaining 89% of the variability of the dependent variable. This model revealed that the business average queue size tends to increase when the finance director works between 11:00-12:00 o' clock and the secretary between 9:00-10:00 o'clock.

Finally, regarding the average size of the finance queue we came to the following results. The model ($p<0.001$) explained 84% of the total variability of the dependent variable with the subsequent remarks: the average size of the queue decreases when the finance director is occupied between 11:00-12:00 o'clock and 14:00-15:00 o' clock; the technical director is occupied between 13:00-14:00 o'clock, the support team between 9:00-10:00 o'clock and the secretary between 11:00-12:00 o'clock.

In general one can build numerous models with different dependent variables each time in order to obtain insight of how the structure of a timesheet affects the functional characteristics of an e-mail manipulation system.

6. DISCUSSION

Simulation provides low cost methods that allow us to gain detailed understanding of the processes of e-mail handling. Additionally, it allows the experimentation with scenarios that can benefit cost and time saving.

This simulation model which was built on a specific e-mail handling system of an IT company is an accurate depiction of the processes, based on necessary simplifications, which do not reduce the validity of the model and the exported information. Of course similar models can be constructed for any other company according to its needs and functionalities.

The experience gained by this research outlined a number of advantages of simulation in the e-mail management procedure. The IT organization was more than willing to assist this research which gave them the opportunity to perceive the nature of their complex system and study in organized manner its operational processes.

Overall we came to two main conclusions. The simulation model, which was based on historical data and information provided by the company, showed that the actual occupation of resources was divergent from the manager's impression. Secondly, with the assistance of linear regression analysis, we were able to explore several combinations of resources working hours (timesheets) and their impact in the queue sizes and average e-mail remaining times in system of the categories that mostly were in the interest of the company's manager.

These results are potentially very useful and should be carefully and methodically studied by the company manager taking into account other parameters which were not included in the present simulation model. These parameters can be financial, operational, even related to personal relationships. The comprehensive view of a system through the results of a model combined with the personal experience of a manager can lead to accurate and effective decision making of how to handle e-mails in a company.

Of course these models have limitations due to restrictions in the available information. However additional improvements could be achieved by creating and using a larger historical database in order to harvest more accurate data for the simulation.

Conclusively, this study can be seen as a first step to develop more sophisticated methodologies that can provide decisions concerning human resource time planning, in the direction of having efficient e-mail processing management.

REFERENCES

Balci, O. (2003, December). Verification, validation, and certification of modeling and simulation applications: verification, validation, and certification of modeling and simulation applications. In *Proceedings of the 35th conference on Winter simulation: driving* *innovation* (pp. 150-158). Winter Simulation Conference.

Berghel, H. (1997). Email—the good, the bad, and the ugly. *Communications of the ACM*, *40*(4), 11-15.

Blom, G. (1958). *Statistical estimates and transformed beta-variables* (Doctoral dissertation, Stockholm).

Christy, D. P., & Watson, H. J. (1983). The application of simulation: a survey of industry practice. *Interfaces*, *13*(5), 47-52.

Clark, H. H. (1996). *Using language* (Vol. 4). Cambridge: Cambridge University Pre

Dabbish, L. A., & Kraut, R. E. (2006, November). Email overload at work: an analysis of factors associated with email strain. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 431-440). ACM.

Denning, P. J. (1982). ACM president's letter: electronic junk. *Communications of the ACM*, *25*(3), 163-165.

Green, T. B., Newsom, W. B., & Jones, S. R. (1977). A survey of the application of quantitative techniques to production/operations management in large corporations. *Academy of Management Journal*, *20*(4), 669-676.

Greve, R. A., Sharda, R., Kamath, M., & Gupta, A. (2007). The Email Strategy Investigation Model (eSIM): A DSS for Analysis of Email Processing Strategies. In *Decision Support for Global Enterprises* (pp. 113-138). Springer US.

Gupta, A., Sharda, R., Greve, R., Kamath, M., & Chinnaswamy, M. (2004). *How often should we check our email? Balancing interruptions and quick response times*. Working paper, Department of MSIS, Oklahoma State University, Stillwater.

Halpin, B. (1999). Simulation in sociology. *American behavioral scientist*, *42*(10), 1488-1508.

Hillier, F. S. (1990). *Intro To Operations Research 8E (Iae)*. Tata McGraw-Hill Education.

Hlupic, V. L. A. T. K. A. (2000). Simulation software: a survey of academic and industrial users. *International Journal of Simulation*, *1*(1-2), 1-12.

Hovey, R. W., & Wagner, H. M. (1958). Letters to the Editor—A Sample Survey of Industrial Operations-Research Activities. *Operations Research*, *6*(6), 876-881.

Jackson, T. W., Dawson, R., & Wilson, D. (2003). Understanding email interaction increases organizational productivity. *Communications of the ACM*, *46*(8), 80-84.

Johansson, B., Johnsson, J., & Eriksson, U. (2002). An Evaluation of Discrete Event Simulation Software for Dynamic Rough-Cut Analysis. In *CIRP-International Seminar on Manufacturing Systems* (Vol. 35, No. 1, pp. 348-355).

Lee, S. M., Moore, L. J., & Taylor III, B. W. (1981). Management Science. Wm. C.

Markus, M. L. (1994). Finding a happy medium: Explaining the negative effects of electronic communication on social life at work. *ACM Transactions on Information Systems (TOIS)*, *12*(2), 119-149.

Millichamp, J. M., (1984). Simulation models are a flexible, efficient aid productivity improvement efforts. *Ind. Eng.*, *28*(8), 78-85.

Monk, A. (2003). Common ground in electronically mediated communication: Clark's theory of language use. *HCI models, theories, and frameworks: Toward a multidisciplinary science*, 265-289.

Narasimha, C. Y., Kamath, M., & Sharda, R. (2007, September). A Semi Markov Decision Process Approach to E-mail Management In A Knowledge Work Environment. In *Automation Science and Engineering, 2007. CASE 2007. IEEE International Conference on* (pp. 1051-1056). IEEE.

Paul, R. J. (1993, December). Activity cycle diagrams and the three-phase method. In *Proceedings of the 25th conference on Winter simulation* (pp. 123-131). ACM.

Renaud, K., Ramsay, J., & Hair, M. (2006). " You've got e-mail!"... shall I deal with it now? Electronic mail from the recipient's perspective. *International Journal of Human-Computer Interaction*, *21*(3), 313-332.

Robinson, S. (2002). General concepts of quality for discrete-event simulation. *European Journal of Operational Research*, *138*(1), 103-117.

Thomas, G. F., King, C. L., Baroni, B., Cook, L., Keitelman, M., Miller, S., & Wardle, A. (2006). Reconceptualizing e-mail overload. *Journal of Business and Technical Communication*, *20*(3), 252-287.

Tyler, J. R., & Tang, J. C. (2003, January). When can I expect an email response? A study of rhythms in email usage. In *ECSCW 2003* (pp. 239-258). Springer Netherlands.

Watson, H. J. (1978). An empirical investigation of the use of simulation. *Simulation & Games*.

Weber, R. (2004). *The grim reaper: The curse of e-mail. MIS Quarterly*, 28(3), 1–12.

Whittaker, S., Bellotti, V., & Moody, P. (2005). Introduction to this special issue on revisiting and reinventing e-mail. *Human-Computer Interaction*, *20*(1), 1-9.

MODELLING AND SIMUILATION OF THERMAL INDUCED STRESS IN 3D NANO PMOS

Abderrazzak El Boukili

Al Akhawayn University in Ifrane, Morocco

Email: a.elboukili@aui.ma

ABSTRACT

We are presenting a new physically based numerical model in 3D to calculate the intrinsic stress due to thermal mismatch in Silicon Germanium for 3D nanometer PMOSFETs after deposition. This intrinsic stress is used to calculate the extrinsic stress distribution in the channel which has the advantage of enhancing performances of devices and circuits. Numerical results of channel stress based on this novel model will be presented and discussed for Intel 45 nanometers PMOSFETs. Results obtained with this model are in good agreement with those found in literature using other models.

Keywords: workstation modeling and simulation in 3D, physically based model, thermal induced stress, Intel nano PMOSFETs

1. INTRODUCTION

The exponential growth predicted by Moore's Law stayed valid for the last four decades. But, it cannot continue the trend forever. The industry already enters the nanometer regime, where the transistor gate length drops down to 45 nm or below and the gate oxide thickness to 1 nm or below. In this nanometer regime, physical limitations such as off-state leakage current and power density pose a potential threat to enhance performance by simple geometrical scaling. Right now, the industry needs a new scaling vector. Front-end process induced extrinsic stress has thereby emerged as the new scaling vector for the 90 nm node technology and below. The extrinsic stress has the advantage of improving the performances of PMOSFETs and NMOSFETs transistors by the enhancing mobility. This mobility enhancement fundamentally results from alteration of electronic band structure of silicon due to extrinsic stress.

The extrinsic stress is the stress that exits in the whole transistor (or circuit). It is produced by an intrinsic stress that exists in different materials (or films) that make up the transistor. The intrinsic stress is either introduced intentionally or unintentionally or both. The unintentional intrinsic stress is induced by the processing steps as: implantation, etching, deposition of thin films, oxidation, or diffusion. The intentional intrinsic stress is introduced intentionally by the manufacturers to increase performance.

Actually, most of nano semiconductor device manufacturers as Intel, IBM and TSMC are intentionally using the intrinsic stress to produce uniaxial extrinsic stress in the Silicon channel. And, it is now admitted that the channel stress enhances carrier mobilities for both nano PMOS and NMOS transistors (by 30% or more (Krivokapic 2003).

In this paper, we are presenting a new physically based numerical model to calculate the intrinsic stress due to thermal mismatch in Silicon Germanium (SiGe) films after deposition. This intrinsic stress is caused by the different thermal expansion coefficients of the thin films of SiGe and the Silicon (Si) substrate. In fact, the temperature is high during deposition of SiGe films on the top of Si substrate. And, the temperature will cool down to room temperature after deposition. Then, an intrinsic stress will develop in the SiGe thin films and in the Si substrate during the cooling down to room temperature.

This paper is organized as follows. Section 2 will outline different sources of intrinsic stress. Section 3 will present the new physically based model in 3D to calculate the intrinsic stress in SiGe due to thermal mismatch between SiGe and Silicon. Section 4 will present the 3D numerical results of the channel extrinsic stress that is calculated using the intrinsic stress calculated using the proposed model. This section will also analyze qualitatively and quantitatively the numerical results and provide some comparisons with the results found in the literature. Section 5 outlines the concluding thoughts and future work.

2. MA DIFFERENT SOURCES OF INTRINSIC STRESS IN SIGE

The deposition step is the main processing step in determining the intrinsic stress in SiGe films. The deposition takes place at elevated temperatures. When the temperature is decreased, the volumes of the grains of SiGe film shrink and the stresses in the material increase. The stress gradient and the average stress in the SiGe film depend mainly on the Silicon-Germanium ratio, the substrate temperature and orientation, and the deposition technique which is usually LPCVD (low pressure chemical vapor deposition) or PECVD (plasma enhanced chemical vapor deposition). The intrinsic stress existing in thin films has generally the following main sources.

2.1. Intrinsic stress due to lattice mismatch

During deposition, thin films are either stretched or compressed to fit the substrate on which they are deposited. After deposition, the film wants to be smaller if it was stretched earlier, thus creating tensile intrinsic stress. And similarly, it creates a compressive intrinsic stress if it was compressed during deposition. The intrinsic stress generated due to this phenomenon can be quantified by Stoney's equation by relating the stress to the substrate curvature.

2.2 Intrinsic stress due to doping

Boron doping in p-channel source/drain regions introduces a local tensile strain in the substrate due to its size mismatch with Silicon. Boron (B) atom is smaller in size than Silicon atom and when it occupies a substitutional lattice site, a local lattice contraction occurs because the bond length for Si-B is shorter than for Si-Si (Randell 2005, Horn 1955).

It was reported in (Horn 1955) that a single boron atom exerts 0.0141 Angstrom lattice contraction per atomic percentage of boron in Silicon at room temperature. The stress induced in the channel due to boron doping was insignificant for long-channel devices. But, for nanoscale CMOS transistors where the channel lengths are in the nanometer realm, this stress plays a significant role in determining the carrier mobility enhancement.

This tensile stress can be deleterious to the compressive stress intentionally induced by embedded Si(1-x)Ge(x) in source and drain and can result in carrier mobility much lower than expected. Also, the boron solubility in Silicon Germanium increases much beyond its limit in Silicon. So the doping stress generation problem proves to be even more significant in advanced CMOS devices where Germanium concentration is expected to be close to 30%.

Methods to counter and suppress the doping induced stress are very important issues and are still under ongoing research.

2.3 Intrinsic stress due to thermal mismatch

Thermal mismatch stress occurs when two materials with different coefficients of thermal expansion are heated and expand or contract at different rates. During thermal processing, thin film materials like SiGe, Polysilicon, Silicon Dioxide, or Silicon Nitride expand and contract at different rates compared to the Silicon substrate according to their thermal expansion coefficients. This creates an intrinsic strain and stress in the film and also in the substrate. The thermal expansion coefficient is defined as the rate of change of strain with temperature.

In this paper, we are focusing on the 3D modeling of the intrinsic stress due to thermal mismatch. We are proposing a new second order numerical model. It will be presented and analyzed in the section 3.

3. MODELING OF THERMAL INDUCED INTRINSIC STRESS IN SIGE

Thermal mismatch intrinsic stress occurs when two materials with different coefficients of thermal expansion are heated and expand or contract at different rates. During thermal processing, thin film materials like SiGe, polysilicon, SiO$_2$, or silicon nitride expand or contract at different rates compared to the Silicon substrate according to their thermal expansion coefficients. The intrinsic strain tensor, ε_0, is defined by:

$$\varepsilon_0 = (\varepsilon_0^{xx}, \varepsilon_0^{yy}, \varepsilon_0^{zz}, \varepsilon_0^{xy}, \varepsilon_0^{yz}, \varepsilon_0^{xz}), \tag{1}$$

where $\varepsilon_0^{xx}, \varepsilon_0^{yy}, \varepsilon_0^{zz}$ are the intrinsic normal strain components, and $\varepsilon_0^{xy}, \varepsilon_0^{yz}, \varepsilon_0^{xz}$ are the intrinsic shear strain components. We assume that the intrinsic shear strain components are all zero. We also assume that in SiGe:

$$\varepsilon_0^{xx} = \varepsilon_0^{yy} = \varepsilon_0^{zz} = \varepsilon_{SiGe}^{0}(T)\,. \tag{2}$$

The thermal expansion coefficient of SiGe, $\alpha_{SiGe}(T)$, is defined as the rate of change of the intrinsic strain component, $\varepsilon^0_{SiGe}(T)$, in SiGe with respect to temperature T. Its unit is micro strain/Kelvin (με/K) and it is given by:

$$\alpha_{SiGe}(T) = \frac{d\,\varepsilon^0_{SiGe}(T)}{dT} \tag{3}$$

Then, the intrinsic strain component will be given by:

$$\varepsilon^{0}{}_{SiGe}(T) = \int_{T0}^{T} \alpha_{SiGe}(t)dt \tag{4}$$

where T0 is the ambient temperature and T is the processing temperature.
For example, T could be the temperature during deposition of SiGe thin film. In this paper, we are choosing $\alpha_{SiGe}(T)$ to be linear with respect to temperature, then our model for intrinsic strain component $\varepsilon^{0}{}_{SiGe}(T)$ given by the equation (4) will be a second order model with respect to temperature T. The models we have found in the literature (Nirav 2005, Freund 2003) are just first order models since they are taking $\alpha_{SiGe}(T)$ as a constant.

On the other hand, to include the effects of the thermal expansion coefficient of the Si substrate, $\alpha_{Si}(T)$, on the intrinsic strain in SiGe thin film, we add the term $\Delta\alpha(T)\Delta T$ to the model given by the equation (4). Then, the proposed model for the intrinsic strain component $\varepsilon^{0}{}_{SiGe}(T)$ is given by:

$$\varepsilon^{0}{}_{SiGe}(T) = \int_{T0}^{T} \alpha_{SiGe}(t)dt + \Delta\alpha(T)\Delta T$$

where

$$\Delta\alpha(T) = \alpha_{SiGe}(T) - \alpha_{Si}(T), \tag{6}$$

$$\Delta T = T - T0\,. \tag{7}$$

The component form of the intrinsic stress, σ_0, is defined by:

$$\sigma_0 = (\sigma_0^{xx}, \sigma_0^{yy}, \sigma_0^{zz}, \sigma_0^{xy}, \sigma_0^{yz}, \sigma_0^{xz}),$$

(,,)

where $\sigma_0^{xx}, \sigma_0^{yy}, \sigma_0^{zz}$ are the intrinsic normal stress components and $\sigma_0^{xy}, \sigma_0^{yz}, \sigma_0^{xz}$ are the intrinsic shear stress components that are also assumed to be zero. We also assume that in SiGe:

$$\sigma_0^{xx} = \sigma_0^{yy} = \sigma_0^{zz} = \sigma^{0}_{SiGe}(T)\,. \tag{8}$$

On the other hand, we are using a physically based model to define the intrinsic stress tensor σ_0 in the SiGe thin film, since, we are using the Hookean's elastic law to express the relation between strain and stress as follows:

$$\sigma_0 = \begin{bmatrix} \sigma_0^{xx} \\ \sigma_0^{yy} \\ \sigma_0^{zz} \\ \sigma_0^{xy} \\ \sigma_0^{yz} \\ \sigma_0^{xz} \end{bmatrix} = D \,. \begin{bmatrix} \varepsilon_0^{xx} \\ \varepsilon_0^{yy} \\ \varepsilon_0^{zz} \\ \varepsilon_0^{xy} \\ \varepsilon_0^{yz} \\ \varepsilon_0^{xz} \end{bmatrix}, \tag{9}$$

In three dimensions, the stiffness tensor D, used in our model, for isotropic elastic materials as SiGe is given by:

$$D = \begin{bmatrix} c11 & c12 & c12 & 0 & 0 & 0 \\ c12 & c11 & c12 & 0 & 0 & 0 \\ c12 & c12 & c11 & 0 & 0 & 0 \\ 0 & 0 & 0 & c44 & 0 & 0 \\ 0 & 0 & 0 & 0 & c44 & 0 \\ 0 & 0 & 0 & 0 & 0 & c44 \end{bmatrix} \tag{10}$$

(5)

where the elastic constants c11,c12, and c44 for each material are given by:

$$\begin{aligned} c11 &= \frac{E\,.(1-\gamma)}{(1+\gamma)(1-2\,.\gamma)} \\ c12 &= \frac{E\,.\gamma}{(1+\gamma)(1-2\,.\gamma)} \\ c44 &= \frac{E}{1+\gamma}. \end{aligned} \tag{11}$$

The term E represents the Young modulus and the term γ represents the Poisson's ratio. The terms E and γ depend strongly on the material. They depend on the interface's orientation of the substrate that are (001), (110), or (111). For SiGe, they also depend on the Germanium mole fraction.
The model we have found in the literature (Nirav 2005, Freund 2003) for the intrinsic stress component $\sigma^{0}{}_{f}(T)$ is only a 2D model and is given by:

$$\sigma_f^0(T) = \left(\frac{E}{1-\gamma}\right)(\alpha_f - \alpha_s)\Delta T\ , \qquad (12)$$

where the subscript ‘f’ represents the film, and ‘s’ represents the substrate. The film could be SiGe, Silicon dioxide, Silicon nitride, or another film. We should note that this model is only linear with respect to temperature.

In our simulation program, the intrinsic stress tensor σ_0 is used as a source term to calculate, in the whole 3D nano MOSFET structure, the extrinsic stress tensor $\sigma = (\sigma^{xx}, \sigma^{yy}, \sigma^{zz}, \sigma^{xy}, \sigma^{yz}, \sigma^{zx})$. We note that σ^{xx}, σ^{yy}, and σ^{zz} represent the extrinsic stress along the channel, vertical to the channel, and across the channel. This channel stress is used to enhance the mobility of holes in 3D nano PMOSFETs based on Intel technology (Ghani et al., 2003). We assume that Silicon and Silicon Germanium are elastic materials. And, to calculate the stress tensor σ, we use the elastic stress model based on Newton's second law of motion, and the following Hooke’s law relating stress to strain:

$$\sigma = D\varepsilon + \sigma_0\,. \qquad (13)$$

Here σ_0 is the intrinsic stress given by the proposed 3D model in the equation (9). A detailed description of this elastic model is given in (El Boukili 2010).

4. 3D NUMERICAL RESULTS AND ANALYSIS

The proposed second order model of intrinsic stress is used to simulate numerically the 3D extrinsic stress in the channel of an Intel 45 nm gate length PMOSFET shown in Figure 1. For the following numerical results, we used (001) for the substrate orientation and 17% as the Germanium mole fraction. In the future, we will do more investigations using different models of temperature.

The results in Figures 2 and 3 show the 3D distribution of x stress components along channel for 300°K and 1000°K respectively. Figure 4 shows 3D distribution of z stress component across channel at 1000°K. This Figure shows also that the stress component σ^{zz} across the channel is also significant. This is an important finding of this paper. A similar stress distribution has been reported in (Victor et al. 2004). The values of the calculated 3D extrinsic stress are also qualitatively and quantitatively in good agreement with those calculated in (Victor et al. 2004). Figure 5 shows the contour lines of the x stress component at 300°K.

All these results did show that the processing temperatures have a great effects on intrinsic and extrinsic stress profiles. These temperature effects will also affect the performances of the MOSFETs devices. On the other hand, these numerical results confirm that our implementation of thermal induced intrinsic and extrinsic stress models in 3D provide valid and correct results. We also believe that these results are of great interest to the semiconductor community including industrials and academia.

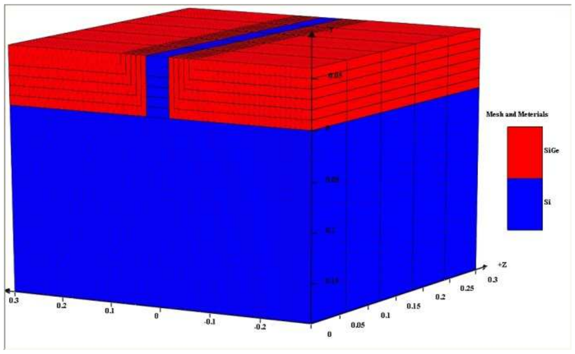


Figure 1: Materials and Mesh of The Simulated Structure

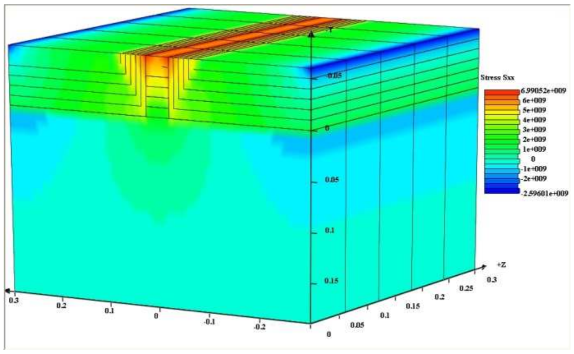


Figure 2: 3D Distribution of x Stress Component Along Channel at 300°K

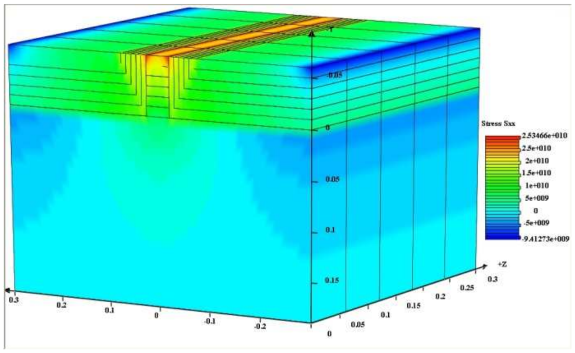


Figure 3: 3D Distribution of x Stress Component Along Channel at 1000°K

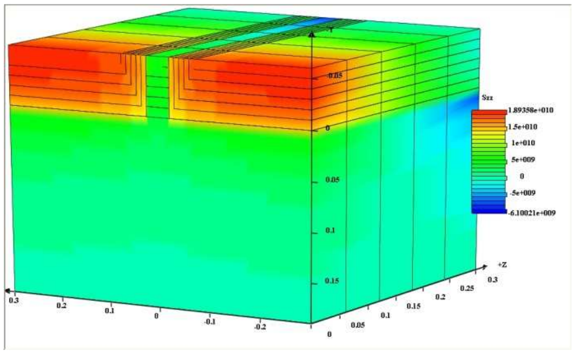


Figure 4: 3D Distribution of z Stress Component Across Channel at 1000°K

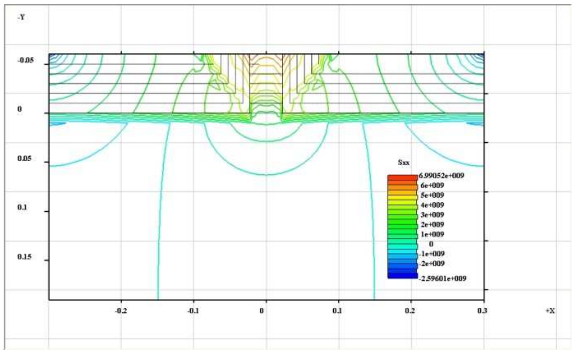


Figure 5: Contour Lines of x Stress Component at 300°K

5. CONCLUSIONS

In this paper, we have developed a new physically based second order model to calculate the intrinsic stress that is due to thermal mismatch between SiGe and Si substrate after deposition of SiGe pockets in source and drain of a strained nano PMOSFETs. This model has been implemented and used successfully to simulate the extrinsic stress in the channel of an Intel 45 nm gate length PMOSFET shown in Figure 1. The important finding of this paper is that all the stress components σ^{xx} and σ^{zz} along the channel, and across the channel respectively are significant. On the other hand, this paper did show that the distribution of the z stress component is really non-uniform in the channel. The quantitative and qualitative behavior of the numerical results is in good agreement with those found in literature (Victor et al. 2004) for similar 3D structure.

REFERENCES

Brash, B. et al., 2004. Mobility enhancement in compressively strained SiGe surface channel PMOS transistors with HFO2/TIN gate stack. *Electrochemical society proceedings,* Vol. 7, pp. 12-30, San Antonio, (California, USA).

El Boukili, A., 2010. 3D Stress Simulations of Nano Transistors. *Progress in Industrial Mathematics at* 85-91.*ECMI,85-91.*

El Boukili, A., 2013. New analytical model and simulation of intrinsic stress in silicon germanium for 3D nano PMOSFET. *International Journal of Control Theory and Computer Modeling,* 85-91.

Fischetti, M., Laux, S., 1996. Band Structures Deformation Potentials, and Carrier Mobility in Strained Si, Ge, and SiG Alloys. *J. Appl. Phys.,* Vol. 80, 2234-2240.

Freund, L.B. and Suresh, S., 2003. *Thin Film: Materials:Stress Defect Formation and Surface Evolution.* Cambridge, United Kindom, Cambridge Unievrsity Press.

Ghani, T. et al., 2003. A 90nm High Volume Manufacturing Logic Technology Featuring Novel 45nm Gate Length Strained Silicon CMOS Transistors. *Proceedings of IEDM Technical Digest,pp.978-980* Washington, DC, USA.

Horn, F., 1955. Densitometric and Electrical Investigation of Boron in Silicon. *Physical Review*, Vol. 97, 1521-1525.

Hollauer, C., 2007. *Modeling of thermal oxidation and stress effects*. Thesis (PhD), Technical University of Wien.

Krivokapic, Z. et al., 2003. Locally strained ultra-thin channel 25nm Narrow FDSOI Devices with Metal Gate and Mesa Isolation. *Proceedings of IEDM, IEEE International,* pp. 445-448, Washington, DC, USA.

Nirav, S., 2005. *Stress modeling of nanoscale MOSFETs*. Thesis (PhD), University of Florida.

Randell, H., 2005. *Applications Of Stress From Boron Doping And Other Challenges in Silicon Technology*. Thesis (Master), University of Florida.

Rieger, M Vogl, P., 1993. Electronic-band parameters in strained Si(1-x)Ge(x) alloys on Si(1-y)Ge(y) substrate. *Phy. Rev. B,* Vol. 48, No 19, 14276-14287.

Rim, E. et al., 2000. Fabrication and Analysis of Deep Submicron Strained-Si N-MOSFET's. *IEEE*

Transactions on Electron Devices, Vol.47, No 7, 1406-1415.

Takagi, S. et al., 2003. Channel Structure Design, Fabrication and Carrier Transport Properties of Strained-Si/SiGe-On-Insulator Strained-SOI) MOSFETs. *IEDM Technical Digest*, pp. 57-60, 10 December, Washington, DC, USA.

Van de Walle, C., Martin, R., 1986. Lattice constants of unstrained bulk Si(1-x)Ge(x). *Phy. Rev. B.*, Vol. 34, 5621-5630.

Victor, M. et al., 2004. *Analyzing strained-silicon options for stress-engineering transistors. July Edition of Solid State Technology Magazine.*

AUTHOR'S BIOGRAPHY

Abderrazzak El Boukili received both the PhD degree in Applied Mathematics in 1995, and the MSc degree in Numerical Analysis, Scientific Computing and Nonlinear Analysis in 1991 at Pierre et Marie Curie University in Paris-France. He received the BSc degree in Applied Mathematics and Computer Science at Picardie University in Amiens-France. In 1996 he had an industrial Post-Doctoral position at Thomson-LCR company in Orsay-France where he worked as software engineer on Drift-Diffusion model to simulate hetero junction bipolar transistors for radar applications. In 1997, he had European Post-Doctoral position at University of Pavia-Italy where he worked as research engineer on software development for simulation and modeling of quantum effects in hetero junction bipolar transistors for mobile phones and high frequency applications. In 2000, he was Assistant Professor and Research Engineer at the University of Ottawa-Canada. Through 2001-2002 he was working at Silvaco Software Inc. in Santa Clara, California-USA as Senior Software Developer on mathematical modeling and simulations of vertical cavity surface emitting lasers. Between 2002-2008, he was working at Crosslight Software Inc. in Vancouver-Canada as Senior Software Developer on 3D Process simulation and Modeling. Since Fall 2008, he is working as Assistant Professor of Applied Mathematics at Al Akhawayn University in Ifrane-Morocco. His main research interests are in industrial TCAD software development for simulations and modeling of opto-electronic devices and processes.
http://www.aui.ma/personal/~A.Elboukili.

PROCESS-INTERACTION MODELING AND SIMULATION: A JAVA-BASED APPROACH

Brahim Belattar [(a)], Abdelhabib Bourouis [(b)]

[(a)] Department of computer Science, University Colonel El Hadj Lakhdar, Batna 05000, Algeria
[(b)] Department of computer Science, University Larbi Ben M'Hidi, Oum El Bouaghi 04000, Algeria

[(a)]brahim.belattar@univ-batna.dz, [(b)]a.bourouis@univ-oeb.dz

ABSTRACT

A large research effort has been devoted to enrich mainstream languages as C, C++, Java, Python with simulation capabilities. The most common choice is to provide the additional simulation functionality through a software library. Independently of the architectural level at which they are provided (application, library, language), the simulation capabilities embody a world view for their users. In this paper we present the architecture and major components of an object-oriented simulation library written in Java. The process-interaction worldview adopted by the library is discussed. A practical example is given in order to ascertain important features of the library. Further motivations are discussed and suggestions for improving our work are given.

Keywords: Discrete-Event Simulation, Object-Oriented Simulation, Process-Interaction Worldview, Java-based modeling and simulation

1. INTRODUCTION

Today, Object Oriented Modeling (OOM) is largely recognized as an excellent approach that deals with large and complex systems through abstraction, modularity, encapsulation, layering and reuse. A conceptual model is obtained by decomposing a real system in a set of objects in interaction. Each object represents a real world entity that encapsulates state and behavior. A class is a template for creating objects that share common related characteristics. Object Oriented Simulation (OOS) benefits from all the powerful features of the OOM especially model conceptualization which is one of the early steps in a simulation study.

The formalism used by a simulation language to conceptualize a domain or system is called its "worldview". Three worldviews are commonly used to model the dynamics of discrete-event systems: Event-Scheduling, Process-Interaction and Activity Scanning. The process-interaction worldview is often convenient for describing the queuing nature of higher-level stochastic systems. From an external point of view, the principal component of simulation software is the simulation language (SL) which allows description of simulation models and their dynamic behavior (Korichi and Belattar 2008).

A large research effort has been devoted to enrich mainstream languages as C, C++, Java, Python with simulation capabilities. The most common choice is to provide the additional simulation functionality through a software library. Independently of the architectural level at which they are provided (application, library, language), the simulation capabilities embody a world view for their users. The world view is essentially the set of concepts that constitute the basic elements available to the modeler to compose and to specify the simulation. The diverse world views are functionally equivalent, but differ in expressive power and in terms of computational efficiency. Native support for multithreaded execution is a fundamental aspect to the implementation of a natural process-oriented modeling worldview. This can be achieved using special programming languages that offer at least a SIMULA's coroutine like mechanism, thus programming languages offering multithreading like Java are suitable.

JAPROSIM is an object-oriented simulation library, free and open source that adopts the popular process-interaction worldview. It is written in Java and was deliberately kept simple, easy to use and extensible. The library is divided into packages to organize the collection of classes into important functional areas. It is easy to build discrete event simulation models using JAPROSIM, either for experimented programmers in Java or for simulation experts with elementary programming knowledge. JAPROSIM can also serve as a basis for the development of dedicated object-oriented simulation environments.

The rest of the paper is organized as follows: In section 2, we present an overview of related work. In section 3 major components of the simulation library and its architecture are detailed. In section 4 we describe the process-interaction worldview adopted by JAPROSIM. An example is given in section 5 in order to ascertain important features of JAPROSIM. Section 6 summaries the paper and provides suggestions for future improvements of our work.

2. RELATED WORK

The idea of building process-oriented simulations using a general purpose object-oriented programming language is not original and several tools were developed in this way. For example, both of CSIM++ (Schwetman 1995) and YANSL (Joines and Roberts 1996) are based on C++, while PsimJ (Garrido 2001), JSIM (Miller et al. 1998) are based on Java. Discrete Event Simulation tools written in Java, like PsimJ and SSJ (L'ecuyer et al. 2002) are well designed and freeware libraries but not open source. Silk (Kilgore 2000) is also well designed but is a commercial tool. There is also a large collection of free open source libraries, we may consider for instance:

- JavaSim (Little 1999) is a set of Java packages for building discrete event process-based simulation, similar to that in Simula and C++SIM.
- JSIM (Miller et al. 1998) is a Java-based simulation and animation environment supporting Web-Based Simulation.
- Simjava (Howell and McNab 1998) is a process based discrete event simulation package for Java, similar to Jade's Sim++, with animation facilities.
- jDisco (Helsgaun 2000) is a Java package for the simulation of systems that contains both continuous and discrete-event processes.
- DESMO-J (Page and Wolfgang 2005) is a framework which supports both event and process worldviews.
- SimKit (Buss 2002) is a component framework for discrete event simulation, influenced by MODSIM II and based on the event graph modelling.

Many simulators aim at replicating the functionality and design of Simula in Java. For example, SSJ is designed for performance, flexibility and extensibility. It offers its users the possibility to choose between many alternatives for most of the internal algorithms and data structures of the simulator. SimJava and JSim are among the first implementations of the thread-based class of simulators. These early efforts pay particular attention to web-based simulation and to the Java Applet deployment model (Cuomo et al. 2012).

JAPROSIM is not a java version of any existing simulation language as Simjava or JavaSim. There are, however, unique aspects in JAPROSIM that lead to fundamental distinctions between our work and others. For example, JAPROSIM embeds a hidden mechanism for automatic collection of statistics. This approach enables a clean separation between implementing the dynamics of the model and gathering data, so traditional performance measurements are automatically computed. The model can thus be created without any concern over which statistics are to be estimated, and the model classes themselves will not contain any code involved with statistics. This leads in more code source clarity. Nevertheless, users could, if needed, implement specific statistics collection using different classes offered by the JAPROSIM statistics package. This feature makes the key difference between JAPROSIM and the other discrete event simulation libraries written in Java. Exception is made for SimKit which already offers this possibility, but which uses a different modeling approach based on event graphs.

3. THE JAPROSIM LIBRARY

The JAPROSIM library is part of an ongoing project that aims at providing an advanced visual interactive simulation and modeling environment for DES (Bourouis and Belattar 2008). The library is currently divided into six main packages:

- kernel: a set of classes dealing with active entities, scheduler, queues and resources.
- random: contains classes for uniform random stream generation.
- distributions: contains a rich set of classes for useful probability distributions.
- statistics: contains classes representing intelligent statistical variables.
- gui: a set of graphical user interface classes to use for project parameterization, trace and simulation results presentation.
- Utilities: a set of useful classes for express model development.

We will focus on the simulation kernel, random, and statistics packages.

3.1. The Kernel Package

The kernel package is at the heart of JAPROSIM. A UML class diagram of the kernel is given below.

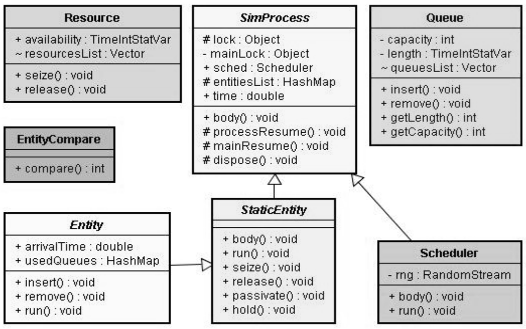


Figure 1: The Kernel class diagram

As we can see, the kernel package is made up of classes dealing with active entities, scheduler, queues and resources. The coroutine like mechanism is implemented trough SimProcess, Scheduler, StaticEntity and Entity classes. A coroutine program is a collection of coroutines which run in quasi-parallel with one another. Each coroutine is an object with its own execution state, so that it may be suspended and resumed. Our aim in the design of JAPROSIM was putting a great emphasis into following the semantic of SIMULA but the design itself is not close to it. The advantage of this approach is that design is simpler without explicit coroutine class support and the semantics of facilities that are well-known and thoroughly tested through many years use of SIMULA are completely supported. Native support for multithreaded execution is a fundamental aspect to the implementation of a natural process-oriented modeling capability in Java. Every active entity's life cycle is executed in a single separate thread.

3.2. The Random and Statistics Packages

Random number generators (RNGs) are the basic tools of stochastic modeling. The random package provides the RandomStream interface which represents a base reference for creating Random Number Generators. Each RNG must rewrite the RandU01() method which normally returns a uniformly distributed number (a Java double) in the interval [0, 1]. JAPROSIM provides a set of well known good RNGs see [[L'ecuyer (1998)]]13] and [14[L'ecuyer and Panneton (2005)], as Park-Miller, McLaren-Marsaglia and RandMrg in which the backbone generator is the combined multiple recursive generator (CMRG) proposed in [15[L'ecuyer (1999)]. The setSeed(long[] seed) method is used to specify seeds instead of default values. The user can define its own RNG by implementing the RandomStream interface. To be used with JAPROSIM, an instance of the user-defined RNG must be assigned to the Scheduler's static public attribute rng. A prosperous set of discrete and continuous Random Variate Generators (RVGs) is offered by the distribution sub-package. This set covers typically most practical distributions to be used in discrete event simulation. However, the user could supply it with additional RVGs.

The statistics package provides two useful classes. DoubleStatVar class dealing with time-independent statistical variables (having double values) as response time and waiting time in a queue. It implements the mechanisms for keeping track of observational-based statistics and must be updated every time its value change using the update() method. TimeIntStatVar class is used for time-dependent statistics (with integer values) such as a queue length or number of customers in a system. Typically, the user instantiates the desired class, then puts and updates it in the appropriate code locations. The placement of statistical variables and their update is a source of several pitfalls. For this reason we have enhanced automatic placement and update of those variables for the most known and useful performance measures.

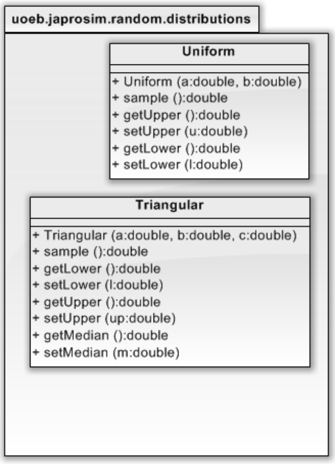


Figure 2: The distribution sub-package

4. PROCESS-INTERACTION WORLDVIEW IN JAPROSIM

The origins of the process-interaction worldview can be traced to the authors of SIMULA. It provides a way to represent a system's behavior from the active entities point of view. A system is modeled as a set of active entities in interaction. Interaction is a consequence of competition and/or cooperation for the acquisition of critical resources. A process-oriented model is a description of the sequence of processing steps these entities experience as they flow through the system. Each active entity's life cycle consists of a sequence of events, activities and delays. A routine implementing an active entity requires special mechanisms for interrupting, suspending and resuming its execution at a later simulated time under the control of an internal event scheduler. This can be achieved using special programming languages that offer at least a SIMULA's coroutine like mechanism, thus programming languages offering multithreading like Java are suitable.

In JAPROSIM, active entities are transient entities moving through the system (dynamic entities). An entity's life cycle is a sequence of active and passive phases. On one hand, an active phase is characterized by the execution of the relevant process. Normally this corresponds to the events during which system state changes without progression of simulation time. On the other hand, passive phases are characterized by activities and delays. So the relevant process is suspended while simulation time advances. Events are the criterion of scheduling which explain the use of a future event list (FEL). After a process is suspended, the scheduler resumes and decides of which is the next process to reactivate according to the system state and the FEL. The scheduler is a special process that coordinates the execution of a simulation model. Processes are executed in pseudo-parallel and only one (which has the imminent simulation time) is running at any instance of real time. Simulation processes may execute concurrently at any instance of simulation time. Hence the scheduler executes in alternation with other simulation processes.

In JAPROSIM, this shared behavior is modeled through the SimProcess abstract class which extends the Java Thread class. The method processResume(Entity e) is called by the scheduler to reactivate a simulation process and mainResume() is called by a simulation process to reactivate the scheduler. Each simulation process has its own lock object. Locks are used in combination with wait() and notify() to synchronize implementation threads instead of the Java deprecated methods suspend() and resume(). A thread which calls any of the previous methods will block on its own lock after notifying the appropriate one.

Schedule(Entity e) is a synchronized method offered by the SimProcess class which could be called by the scheduler or by a newly created simulation process for an appropriate insertion into the FEL. At the end of its life cycle, a simulation process calls automatically the dispose() method to reactivate the scheduler without blocking itself. So the corresponding thread could be terminated. This leads to free occupied memory and improve simulation performance. Otherwise this may cause a Java runtime error as we experienced with an academic version of the commercial package Silk.

Specific behavior of a simulation process is normally described using the dedicated abstract method body(). It must be rewritten to be an ordered sequence of method invocations terminated by an implicit automatic call to dispose(). The behavior of the scheduler is also described using this method. Since SimProcess is abstract, it is intended to be extended. A new class is created to model simulation processes. The Entity class provides the basis for defining classes that obey to the process-oriented simulation worldview. This class is declared to be abstract, so instances of Entity cannot be created directly. Instead, modelers define their own classes that extend Entity and describe the dynamic behavior of the corresponding system components in terms of the process-oriented methods inherited in particular from those classes.

Each class derived from Entity runs in its own thread of execution, a capability inherited from SimProcess. The Entity class provides the implementation of the run() method which in turn invokes body(). The user is required to supply the body() method. Four remarkable methods are offered: insert(), remove(), seize(), hold() and release(). They could be used to model familiar queuing scenarios. The passivate() method is used to wait until a specific system state is reached (ex: waiting for a resource to be free). Since the thread will be suspended and inserted into the passive list (PL) after a call to passivate(), this call is typically used within a while() loop. Each time the scheduler takes control; it starts reactivating suspended threads in the PL first, then dealing with the FEL. So such a reactivated thread would have the opportunity to return back to the PL, if there is no expected evolution in the system state.

The abstract class StaticEntity is used to model the behavior of active entities that have not the ability to move. Typical examples of those entities are “intelligent resources”. StaticEntity derives directly from SimProcess. Since The Entity class is used to model dynamic entities, it derives from StaticEntity and defines two new methods insert() and remove(). The other methods: seize(), hold(), release() and passivate() discussed previously are defined in the StaticEntity and hence inherited by Entity.

The scheduler proceeds in two phases. First, it reactivates each thread in the PL. So the reactivated thread checks for expected changes in the system state and may return back to the PL as it may continue executing the rest of its operations. Secondly, the scheduler picks the imminent simulation process from the FEL and reactivates the corresponding thread. These two phases are repeated as long as the simulation experiment termination condition isn’t verified. The Scheduler class has an attribute rng which is an instance of a random number generator and could be customized by the user. The EntityCompare class implements the Java Comparator interface and is used to implement priority queuing mechanism.

The Resource class represents a passive entity characterized by a capacity. Generally, a simulation process seizes some units of a resource to accomplish a service and releases them later. The hold() method of the StaticEntity class is used to specify the service duration. The Queue class models a space for waiting which may be limited. It provides an ordered list where entities (or other user-defined types) can reside. Typically, an entity is inserted into a queue by having it activate the insert(Queue q) method of the Entity class. There is no implicit conditional status delay logic associated with queues, which means that the entity's thread of execution is not suspended pending some system status evolution.

Modeling conditional status delays is the realm of the while() and passivate() constructs. As a consequence, an entity can reside simultaneously in any number of queues. This feature can be particularly convenient in collecting certain types of system statistics related to waiting times or queue lengths. Another important distinction is that the removal of an entity from a queue could be independent of the ordering of the queue at the time of removal. Users are required to explicitly identify the entity to be removed at the time of removal. Typically this is accomplished by having the corresponding entity activate the remove(Queue q) method of the Entity class. While entities are generally inserted and removed from queues using the insert(Queue q) and remove(Queue q) methods of the Entity class, the same tasks can be accomplished using the insert(Entity e) and remove(Entity e) methods defined in the Queue class.

5. A MODELING EXAMPLE USING JAPROSIM

5.1. Example Description

The modeling example illustrates a simplified simulation model of a TVs inspection and adjustment process as described in (Pegden et al. 1990).

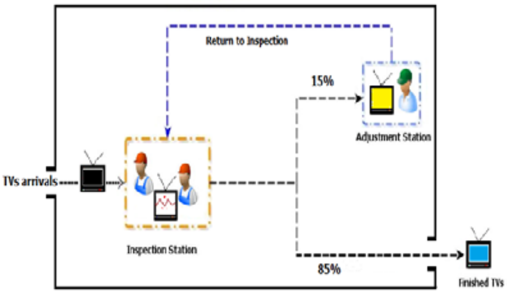


Figure 3: The TVs Inspection example

In this model, an arriving TV is first inspected at an inspection station. If a TV is found to be functioning improperly, it is routed to an adjustment station. After adjustment, the TV is sent back to the inspection station where it is again inspected. TVs passing inspection, whether to the first time or after one or more routings through the adjustment station, are sent to a packing area. A probabilistic branching is used when a TV passes the inspection station. It specifies that 15% of the TVs inspected are sent to the adjustment station and 85% are sent to the packing area. The inter-arrival time between TVs to the system, the inspection delay and the adjustment delay are all modeled as uniform variates. (See the source code in Figure. 5).

5.2. The JAPROSIM Simulation Model

In JAPROSIM we can model each active entity in a separate class derived from the Entity class. A class diagram of the JAPROSIM simulation model for this example is shown below:

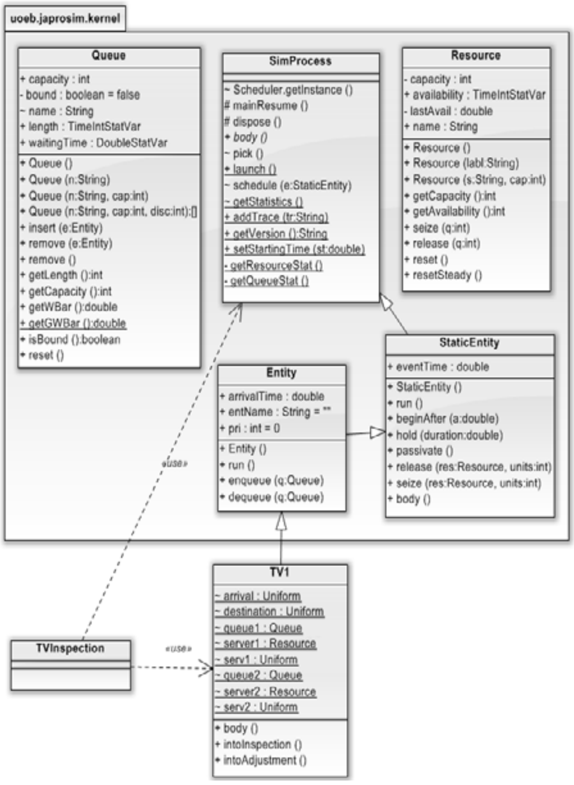
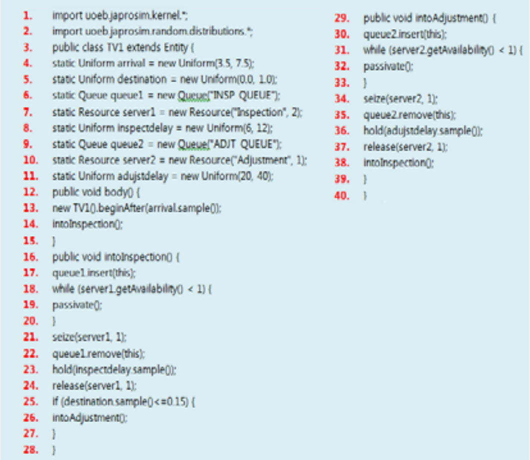


Figure 4: A class diagram of the simulation model

From Figure 4, it appears that the JAPROSIM simulation model of the example uses two classes: TVInspection and TV1. The source code of each class is given below.



```
1.  import uoeb.japrosim.kernel.*;
2.  import uoeb.japrosim.random.distributions.*;
3.  public class TV1 extends Entity {
4.  static Uniform arrival = new Uniform(3.5, 7.5);
5.  static Uniform destination = new Uniform(0.0, 1.0);
6.  static Queue queue1 = new Queue("INSP QUEUE");
7.  static Resource server1 = new Resource("Inspection", 2);
8.  static Uniform inspectdelay = new Uniform(6, 12);
9.  static Queue queue2 = new Queue("ADJT QUEUE");
10. static Resource server2 = new Resource("Adjustment", 1);
11. static Uniform adujstdelay = new Uniform(20, 40);
12. public void body() {
13. new TV1().beginAfter(arrival.sample());
14. intoInspection();
15. }
16. public void intoInspection() {
17. queue1.insert(this);
18. while (server1.getAvailability() < 1) {
19. passivate();
20. }
21. seize(server1, 1);
22. queue1.remove(this);
23. hold(inspectdelay.sample());
24. release(server1, 1);
25. if (destination.sample()<=0.15) {
26. intoAdjustment();
27. }
28. }
29. public void intoAdjustment() {
30. queue2.insert(this);
31. while (server2.getAvailability() < 1) {
32. passivate();
33. }
34. seize(server2, 1);
35. queue2.remove(this);
36. hold(adujstdelay.sample());
37. release(server2, 1);
38. intoInspection();
39. }
40. }
```

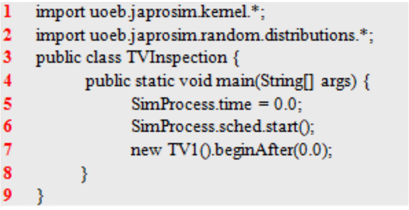
Figure 5: Source code of The TV1 class

We can easily distinguish four parts in the source code of The TV1 class. The first part (from line 4 to line 11) serves to set the parameters of the model. We can see that the inspection delay, the adjustment delay and the inter-arrival time are defined as uniform variates with specific arguments. We have also to define the inspector and adjustor resources and their associated queues. The variable destination is defined as a uniform variate and is used when deciding if a TV just inspected is to be routed to the adjustment station or to exit the system.

The second part (from line 12 to line 15) serves to route the active entity to the inspection station and to create next TVs arrivals with respect to the inter-arrival time between TVs. The third part (from line 16 to line 28) represents the classical scheme of resource allocation. A TV arriving at the inspection station is inserted in the associated queue. When a resource unit is free, it is allocated to a waiting TV with respect to the queue priority. An inspection delay associated to this TV is sampled, and the TV will hold the resource unit seized until the associated delay is elapsed. The resource unit is then released and can be allocated to other waiting TVs. Line 28 serves to decide if the TV just inspected is to be routed to the adjustment station or to exit the system.

The fourth part (from line 29 to line 39) models the adjustor resource allocation scheme. A TV arriving at the adjustment station is inserted in the associated queue. When the adjustor resource is free, it is allocated to a waiting TV with respect to the queue priority. An adjustment delay associated to this TV is sampled, and the TV will hold the adjustor resource seized until the associated delay is elapsed. The adjustor resource is then released and the TV is sent back to the inspection station.

To run a JAPROSIM simulation model, we need another class which constitutes a starting point for any Java program. This class contains the main() method for standalone programs or the init() method for browser-based applets. It is where simulation model would be initialized, and the scheduler started. In our example, this class is called TVInspection. The source code is as follows:

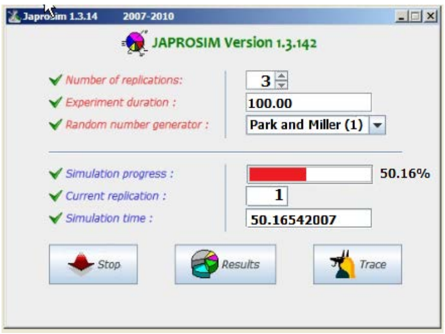


```
import uoeb.japrosim.kernel.*;
import uoeb.japrosim.random.distributions.*;
public class TVInspection {
        public static void main(String[] args) {
                SimProcess.time = 0.0;
                SimProcess.sched.start();
                new TV1().beginAfter(0.0);
        }
}
```

Figure 6: Source code of the TVInspection class

5.3. Running the Simulation Model

When running the simulation model, the JAPROSIM window is first displayed. It consists of an experimentation frame where simulation parameters are to be set. Parameters like the number of replications, the simulation duration, the RNG used must be specified here by the user. A button Run/Stop allows user to start simulation, stop and resume it at any time during execution. Two other buttons are used for presentation of simulation results and trace execution.

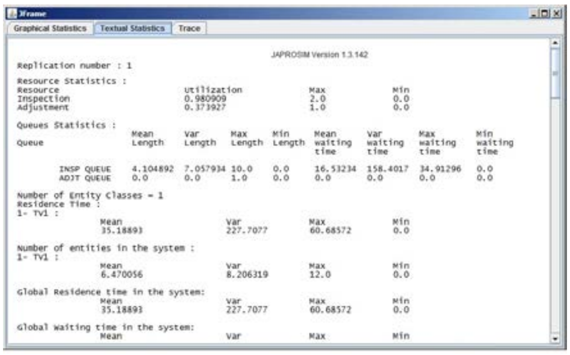


✓ Number of replications: 3
✓ Experiment duration : 100.00
✓ Random number generator : Park and Miller (1)

✓ Simulation progress : 50.16%
✓ Current replication : 1
✓ Simulation time : 50.16542007

Figure 7: JAPROSIM Experimentation Frame

At the end of each simulation run, the simulation results can be viewed in a textual form or in a graphical one.



```
                                              JAPROSIM Version 1.3.142
Replication number : 1

Resource Statistics :
Resource                 Utilization            Max           Min
Inspection               0.980909               2.0           0.0
Adjustment               0.373927               1.0           0.0

Queues Statistics :
                      Mean      Var       Max      Min      Mean      Var       Max       Min
Queue                 Length    Length    Length   Length   waiting   waiting   waiting   waiting
                                                            time      time      time      time

        INSP QUEUE    4.104892  7.057934  10.0     0.0      16.53234  158.4017  34.91296  0.0
        ADJT QUEUE    0.0       0.0       1.0      0.0      0.0       0.0       0.0       0.0

Number of Entity Classes = 1
Residence Time :
1- TV1 :
                Mean                  Var             Max           Min
                35.18893              227.7077        60.68572      0.0

Number of entities in the system :
1- TV1 :
                Mean                  Var             Max           Min
                6.470056              8.206319        12.0          0.0

Global Residence time in the system:
                Mean                  Var             Max           Min
                35.18893              227.7077        60.68572      0.0

Global waiting time in the system:
                Mean                  Var             Max           Min
```

Figure 8: Textual Simulation Results

As we can see, the textual simulation results are expressed as statistical quantities which resume resources and queues utilization during a run. On the other hand, the graphical form uses plots, bar charts or pie charts. For example, Figure 9 shows the utilization of the two resources used in the simulation model during each replication.

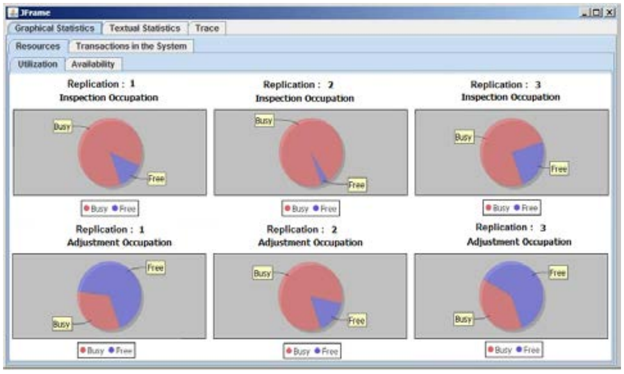


Figure 9: Graphical Simulation Results

5.4. Summary of JAPROSIM Important Features

From the example presented we can draw many advantages of the object-orientation of JAPROSIM and the process-interaction worldview adopted. The relationship between the simulation model and the real system is more obvious and therefore easier to teach and to understand. The java source code of the simulation model is easy to understand and users can learn far more than if they have to experiment with sophisticated commercial simulation packages in which important details of the simulation implementation are hidden and thus never understood.

Furthermore, we can observe in the source code of the classes used in the JAPROSIM simulation models, that no class of the statistics package is explicitly used. In addition, no Java constructs are clearly used to do so. This is the key feature of JAPROSIM that all well known and useful performance measures are implicitly and automatically handled. The user doesn't worry about how many, or what kind of statistical variables to use, nor where to place and update them. This mechanism is embedded in the library.

The SimProcess class declares a protected static entitiesList which is a Java HashMap to collect the residence time of each simulation entity class (a Java class that extends the JAPROSIM Entity class). The key for the HashMap is the class name and values are DoubleStatVar. In the Entity constructor, each time a new entity class is created, the above HashMap is updated. In the run() method of the Entity Class and after the call to the body() method, the residence time is updated using the simulation time and the arrivalTime attributes.

Each Queue object possesses a statistical variable to hold waiting time in it. This variable is updated trough insert()/remove() methods. The number of entities in a queue is handled by a length time-dependent statistical variable. The resource availability is also a time-dependant variable. It is used to compute resource utilization. The Queue class has a static Java Vector to register all queues used in the simulation model. In the same way, the Resource class also has an analogous list to keep track of all used resources. Those lists have a package visibility; hence they could be accessed by all the simulation processes. They are updated each time a new resource or queue instance is created.

6. CONCLUSION

Our aim in the design of JAPROSIM was putting a great emphasis into following the semantic of SIMULA but the design itself is not close to it. The advantage of this approach is that design is simpler without explicit coroutine class support and the semantics of facilities that are well-known and thoroughly tested through many years use of SIMULA are completely supported. Advanced process-oriented modeling features supported by JAPROSIM include: capacity-constrained resources, conditional waiting and special process relationships. The later is supported through the utilities package which offers pre-specified entities with specific behavior. For example, the SimpleServiceStation entity is used to model intelligent servers which are able to take decisions like "batch servers". The SymetricServiceStation entity models a service station with identical servers while AsymetricServiceStation models a service station with multiple heterogeneous servers.

Furthermore, JAPROSIM embeds a hidden mechanism for automatic collection of statistics. This approach enables a clean separation between implementing the dynamics of the model and gathering data, so traditional performance measures are automatically computed. The model can thus be created without any concern over which statistics are to be estimated, and the model classes themselves will not contain any code involved with statistics. This leads in more code source clarity.

JAPROSIM is distributed as an Open Source project (http://sourceforge.net/projects/japrosim/). The source code is available freely along with some documentation. Future improvements will focus on increasing the JAPROSIM performances, integrating a graphical model building facility, providing animations of simulation models and using xml standards for web-based simulation.

REFERENCES

Bourouis. A, Belattar. B, 2008. *JAPROSIM: A Java Framework for Discrete Event Simulation, in Journal of Object Technology*, vol. 7, no. 1, January-February 2008, pp. 103-119, Available from:<http//www.jot.fm/> [accessed June 16, 2013].

Buss, A., 2002. Component Based Simulation Modeling with SimKit. *Proceedings of the 2002 Winter Simulation Conference*, pp. 243-249, 2002, Piscataway, New Jersey.

Cuomo, A., Rak, M., Villano, U., 2012. Process-oriented Discrete-event Simulation in Java with Continuations - Quantitative Performance Evaluation, *Proceedings of the 2nd International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, pp. 87-96, 2012, Rome, Italy.

Garrido, J.M., 2001. *Object-oriented Discrete Event Simulation with Java: a practical introduction*. New York, Kluwer Academic/Plenum Publishers.

Helsgaun, K., 2004. *Discrete Event Simulation in Java*. DATALOGISK SKRIFTER (writings on computer science), Roskilde University, Denmark.

Howell, F., McNab, R., 1998. simjava: a discrete event simulation package for Java with applications in computer systems modelling. *First International Conference on Web-based Modelling and Simulation*, San Diego CA.

Joines, J.A.; Roberts, S.D. 1996. Design of object oriented simulations in C++. *Proceedings of the 1996 Winter Simulation Conference*, pp. 65-72, 1996, Piscataway, New Jersey.

Kilgore, R.A., 2000. Silk, Java and Object-Oriented simulation. *Proceedings of the 2000 Winter Simulation Conference*, pp. 246-252, 2000, Piscataway, New Jersey.

Korichi Ahmed, Belattar Brahim, 2008. Towards a Web Based Simulation Groupware: Experiment with BSCW, *WSEAS transactions on Business and Economics*, Issue 1, Volume 5, pp. 9-15, January 2008.

L'Ecuyer, P., 1998. Uniform Random Number Generator. *Proceedings of the 1998 Winter Simulation Conference*, pp. 97-104, 1998, Piscataway, New Jersey.

L'Ecuyer, P., 1999. Good parameters and implementations for combined multiple recursive random number generators", Operations Research, vol. 47(1), 159–164.

L'Ecuyer, P.,, Melian, L., Vaucher, J., 2002. SSJ: A framework for stochastic simulation in Java. *Proceedings of the 2002 Winter Simulation Conference*, pp. 234–242, 2002, Piscataway, New Jersey.

L'Ecuyer, P., Panneton, F., 2005. Fast Random Number Generators Based on Linear Recurrences Modulo 2: Overview and Comparison. *Proceedings of the 2005 Winter Simulation Conference*, pp. 110-119, 2005, Piscataway, New Jersey.

Little, M. C., 1999. *The JavaSim User's Manual*. Department of Computing Science, University of Newcastle upon Tyne.

Miller, J.A., Ge, Y., Tao, J., 1998. Component Based Simulation Environments: JSIM as a Case Study Using Java Beans. *Proceedings of the 1998 Winter Simulation Conference*, pp. 373-381, 1998, Piscataway, New Jersey.

Page, B., Wolfgang, K., 2005. *The Java Simulation Handbook - Simulating Discrete Event Systems with UML and Java*. Aachen, Shaker Verlag.

Pegden, C.D., Shannon, R.E. and Sadowski, R.P., 1990. *Introduction to Simulation Using SIMAN*. New York, McGraw-Hill Inc.

Schwetman, H. 1995. Object-Oriented simulation modeling with C++/CSIM17. *Proceedings of the 1995 Winter Simulation Conference*, pp. 529-533, 1995, Piscataway, New Jersey.

AUTHORS BIOGRAPHY

B. Belattar is a professor at the University of Batna since 1992. He has also taught at the University of Constantine from 1982 to 1985. He received his BS degree in Computer science from the University of Constantine in 1981 and his MS and PhD degrees from the University Claude Bernard of Lyon (French) respectively in 1986 and 1991. His research interests include simulation, databases, semantic web and AI.

A. Bourouis is a lecturer at the University of Oum el Bouaghi since 2003. He received his BS degree in Computer science from the University of Constantine in 1999 and his MS and PhD degrees from the University of Batna respectively in 2003 and 2009. His research interests include Artificial intelligence, performance evaluation, parallel and distributed simulation.

TOWARDS LARGE SCALE ROAD TRAFFIC SIMULATION EXPERIMENT

Marek Małowidzki, Tomasz Dalecki, Przemysław Bereziński, Michał Mazur

Military Communication Institute
Zegrze, Poland

{m.malowidzki,t.dalecki,p.berezinski,m.mazur}@wil.waw.pl

ABSTRACT

The Insigma project goals include traffic optimization and control. It is assumed that advanced functions including traffic monitoring and prediction, route planning, and, finally, traffic optimization and control, will be able to utilize the available road infrastructure in an optimal way in order to minimize traffic jams and related social and environmental losses. The efficiency of these mechanisms will be evaluated through large-scale simulation experiments. However, before such experiments become feasible, an appropriate simulation environment must be prepared. In the paper, we discuss the application of SUMO as a simulation environment and its integration with Insigma's traffic control layer.

Keywords: road traffic, routing, simulation, traffic simulator, SUMO, Open Street Map (OSM).

1. INTRODUCTION

The Insigma project is aiming at the development of an intelligent information system for global monitoring, detection and identification of threats. The system collects data from various kinds of sensors, cameras, and users, and processes the data to identify threats and notify appropriate public services. One of Insigma's tasks is road traffic optimization and control, which includes traffic lights, information boards, and route planning.

Insigma's goals with respect to traffic control are ambitious. On one hand, they contain a number of features related to public security (support for emergency services and special vehicles, collecting and reporting events, etc.); on the other hand, one of goals is traffic optimization and control. It is assumed that all control mechanisms will be based either on dynamic traffic data, collected and updated in real time, or on traffic forecast.

The first obvious problem is the (feasible and right) approach to evaluation of designed and implemented mechanisms. It seems that large-scale experiments are only possible in a simulated environment. Such an environment should include a road traffic simulator integrated with a control layer, responsible for traffic management. In the paper, we discuss our work on integration the SUMO simulator with the routing service we have already developed (Małowidzki et al. 2012, 2013a, 2013b).

The paper is organized as follows: First, we discuss the traffic subsystem in Insigma and the routing service's architecture and functions. Then, we overview SUMO and its key ideas related to performing simulations. Next, we comment on how we have integrated SUMO with the routing service, include an example, and describe our experience. We propose simulation scenarios. Finally, we overview related work and end the paper with summary.

2. THE INSIGMA'S TRAFFIC SUBSYSTEM

The Insigma's ultimate (and somewhat ambitious) vision is presented in Figure 1. Insigma's goals with respect to road traffic include traffic control, traffic prediction, route planning, and related functions. Despite the fact that ongoing work includes real-world equipment (advanced cameras located at crossroads, collecting detailed data about observed traffic license plate recognition software (Janowski et al. 2012), etc.), the only way to verify control algorithms is simulation. (Drivers would not be pleased finding themselves to be beta testers of our ideas.) Thus, we integrate the SUMO road traffic simulator with the control layer in order to prepare a complete simulation environment for traffic measurements and control.

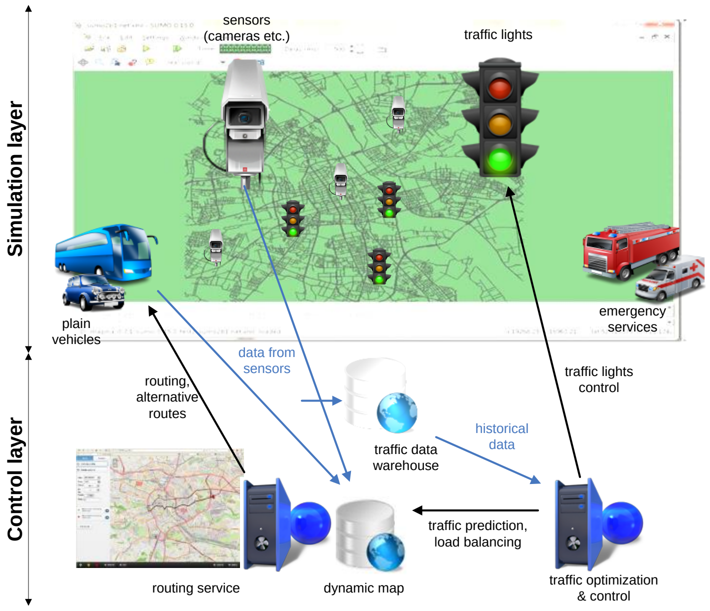


Figure 1. Insigma's traffic subsystem architecture: control signals (arrows pointing up or left) and data flows (arrows pointing down or right)

The control layer consists of a number of components. Some of them have already been implemented (the routing service, the static and dynamic maps), some are under development (traffic data warehouse), some remain to be designed and developed (load balancing, traffic optimization and control). At present, the control layer is represented by the routing service (and the maps it utilizes), which is the main integration target; this is the reason we focus on the routing service in our paper.

There are a number of such services commercially available and successful on the market (with Google Maps as a premier example), however, specific functions that the service was to support as well as planned integration with higher-level traffic control algorithms in practice required implementing a new one from scratch.

In the following section, we discuss the routing service's architecture and key ideas.

3. THE ROUTING SERVICE

The routing service's internal architecture is shown in Figure 2. The architecture has been discussed in detail in our previous work (Małowidzki at al. 2012) but, for the completeness of the discussion, we briefly summarize it here.

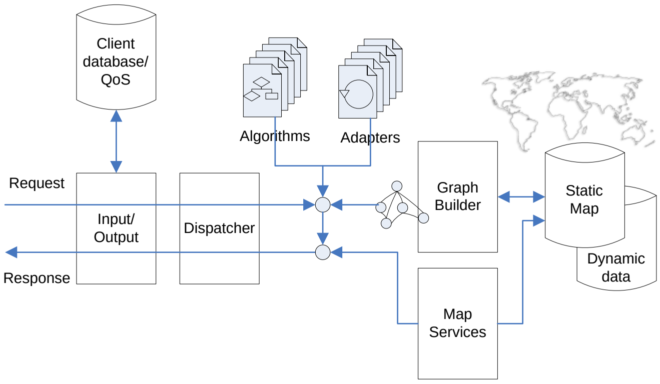


Figure 2. The route server's internal architecture

The main elements (components) are as follows:

- Input/output, a component responsible for handling messages for clients, providing optional QoS and security functions;
- Database, containing the static map and dynamic data (traffic statistics: drive and turn times, speeds, etc.). At present, the static map contains the Open Street Map (OSM) data, although the database format has been significantly modified for our purposes (refer to Małowidzki et al. (2013b) for details).
- Graph Builder, responsible for transforming map data into a graph (a set of nodes and edges) that may be used for route computations;
- Adapter(s), computing graph weights. Usually, each route type (Fast, Short, Optimal, etc.) requires a separate adapter. Their implementation may be trivial (e.g., for a Short route) or fairly complex, as it is in case of a privileged route adapter, described in Małowidzki et al. 2013a. Note that we assume that *weights are functions of time* (the start time of the drive at a given edge), which allows to take into account dynamic (current and predicted) traffic data.
- Algorithm(s), performing route optimizations. Algorithms are separated from road data by adapters; they only see the graph with edge weights computed by adapters. We have tested a number of algorithms, including Ant Colony Optimization (Bedi et al. 2007) and an adapted version of SAMCRA (Góralski et al. 2011), but found that Dijkstra-based algorithms (an optimized version using a priority queue or the A* algorithm (Hart et al. 1968)), perform best.
- Finally, the Dispatcher, managing the above-mentioned elements, and providing additional functions (e.g., alternative routes) and debugging capabilities.

The software is implemented in the Microsoft .NET 4 framework environment. The internal architecture is organized around a number of interfaces. Most crucial elements (.NET classes implementing well-known interfaces) that affect the service behavior (the graph builder, adapters, algorithms) are dynamically loaded according to the server's configuration.

We have also developed a client, implemented in JavaScript/OpenLayers environment. The client's capabilities allow to make use of most of the routing service's functions.

4. SUMO

SUMO (Simulation of Urban MObility) (Behrisch at al. 2011, Krajzewicz et al. 2012) is a microscopic traffic simulator that models the movement of vehicles in space-continuous map and uses a discrete time (with one-second resolution). Each vehicle is modeled separately and is described by a departure time (the time it starts its drive) and a route described by a set or roads.

SUMO has been implemented in the Institute of Transportation Systems at the German Aerospace Center. The version used in our simulation is 0.16.0 (released in December 2012).

The core of simulation software is written in C++. There are additional software libraries that allow to affect the simulation with Python and Java code.

During the selection of the simulator, we also considered MATSim but we found SUMO to be more user-friendly and contain most functions we would require.

SUMO's key features are as follows:

- Open source code;
- Maturity of the project; new versions appearing regularly;
- Sufficient documentation and examples;
- Usage of XML, which provides configuration flexibility. We successfully developed a map converter for SUMO (see section 6).
- An included converter for importing OSM maps;
- A convenient API (called TraCI; available in C++, Python and Java) that allows to control the simulation;
- Last but not least, a good GUI for simulation visualization.

Regarding SUMO drawbacks, it does not support privileged vehicles (that would not have to obey the rules of the road), which is important in Insigma (support for emergency services is one of explicit goals of the project, see Małowidzki et al. (2013a)). This poses a problem with simulating this type of vehicles.

5. PERFORMING SIMULATIONS IN SUMO

This section describes key simulation issues in SUMO. In the next section, we comment on how our "control loop" affects the way simulations are performed.

Map preparation. SUMO's approach to modeling map data is quite convenient. Three separate XML files describe, respectively, nodes, edges and connections. There are additional tools for map data processing. The main tool, *netconvert*, enables to coalesce these files into a SUMO map. OSM data import is easy.

Routing vehicles. SUMO provides a dedicated tool, *activitygen*, for modeling traffic demands, but we decided to implement our own tool. (The main reason was that we wanted to have a better control of the demands. Additionally, we found activitygen's configuration to be complex and insufficiently documented.) The tool supports three traffic classes: driving to work, business traffic, and transit traffic. We assume our map covers a city with residential districts and some center/production area. Our tool, given a population of the city, uses heuristics to generate flows for each traffic class. The flows are defined by a start and an end point for each vehicle. Then, we are either able to use SUMO's routing functions or compute the routes ourselves (section 6).

Moving vehicles. During each simulation step, which equals 1 second, and for every simulated vehicle, SUMO's engine checks whether the vehicle can move ahead and then selects an appropriate speed value. The value may be set to:

- The maximum allowed for a road;
- The previous value increased by an acceleration factor;
- The previous value decreased by a braking factor, if a vehicle is approaching a crossroad or an obstacle (a slower vehicle ahead).

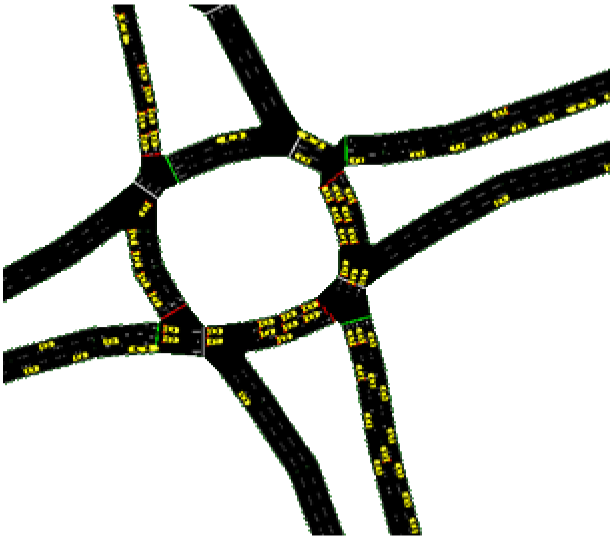


Figure 3. A screenshot that presents a crossroad simulated in SUMO (Dmowski Roundabout in Warsaw)

Then, a vehicle is moved according to the speed value.

Road sensors. SUMO provides abstract detectors that can be used to collect data in any place we would like to observe:

- Aerial induction loops provide average speed of vehicles passing them;
- Crossroads may be monitored as black boxes providing average times needed to pass a crossroad in a particular relation (direction).

Traffic control. Traffic control may be performed mainly through traffic lights. Two main control types are considered: Configuring time slices for a green light (in a given direction) and synchronizing subsequent crossroads along selected major roads to assure a "green wave," that is, uninterrupted traffic through a number of crossroads.

6. SUMO AND ROUTING SERVICE INTEGRATION

This section describes the integration of SUMO and our routing service.

Map preparation. Our database model is based on OSM but contains important extensions. Thus, we needed a dedicated tool able to convert our graphs into SUMO maps. Through an internal Graph Handler interface, it is possible to intercept the graph after it has been constructed by the Graph Builder. The intercepted graph is an input to a converter component, which generates the three required XML files (section 5), which are finally converted to a SUMO map by netconvert.

The conversion retains crucial road parameters such as speed limits, lane counts, etc. Database identifiers are preserved as well, as they are later needed to identify roads in computed routes. Traffic control at crossroads is based on either traffic lights, priorities (specified on the basis of road signs) or the right hand rule.

Routing vehicles using the routing service. The traffic demands are prepared as usually, using our tool described above. Having for each vehicle the start and the destination points, we do not rely on SUMO but instead ask the service to compute routes. Note that the routes are based on dynamic data, which are continuously updated (read from SUMO sensors and delivered to the dynamic map, see below). The interface is implemented in Python.

Requests to the routing service may be synchronous or asynchronous. The synchronous mode may cause some time synchronization issues – see the discussion below – and is supported for testing purposes. In a more realistic approach, with asynchronous requests, a vehicle issues a request and continues its drive along the previous route; as soon as new route is available, it is passed to the vehicle. In case the new route cannot be applied (e.g., a vehicle has just passed a place where, according to the new route, it should have taken a turn), it is discarded and the vehicle continues its drive while a new routing request is scheduled.

The data flow is presented in Figure 4. Data from SUMO sensors feed the dynamic map, and are consumed by the routing service when calculating routes. The simulation/control loop is thus closed.

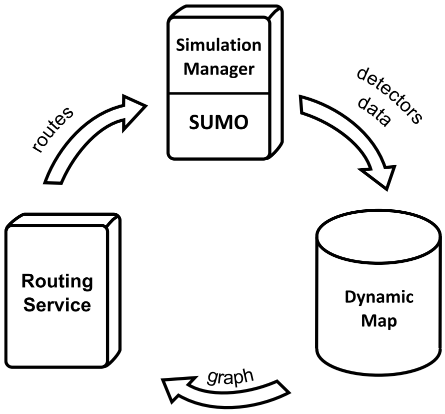


Figure 4. Information flow between SUMO and the routing service

Road sensors. We collect traffic data in the following way:

- Aerial induction loop provide us with average speed values. The loops are placed on every road.
- Multi entry/exit detectors are put on every crossroad; they are a source of travel times through a crossroad in every relation.
- SUMO can dump data periodically to XML files and we do it every 15 minutes.
- Data collection may be performed in one of the two modes:
  - Offline mode that only records data from simulation to files and allows a later "replay," i.e., setting the values in the dynamic map;
  - Online mode that requires both SUMO and control layer services be running simultaneously.

Vehicle sensors. Additionally, we support a simulated GPS Tracker sensor. Such a sensor collects raw GPS vehicle positions and instant speed values, and delivers them to a tracker service. Data are collected from a set of selected cars (possibly, from all cars), stored locally, and delivered in larger packs (for improved performance). A number of parameters can be configured. GPS data are supplemented with some debugging information, which allows to perform on-line analysis of the tracker service's correctness.

Traffic control. Traffic control will be performed using two cooperating mechanisms:

- The routing service, guiding individual vehicles and applying alternative paths to distribute load. We assume that traffic prediction (Małowidzki et al. 2013b) will enable such distribution and possible jamming of "best" routes will be eliminated.
- Traffic lights control, performed by high-level traffic optimization and control algorithms; this remains to be implemented.

Clock synchronization and performance. SUMO itself is faster than real time. It may be slowed down by our integration layer (and additional processing involved). Clock synchronization is important for most control mechanisms to work properly. It is easy to artificially slow the simulation (make 1 second of simulated time equal to 1 second of wall time by introducing artificial delays). However, a coordination in case simulation is slower than real time may pose some problems, although our current experience suggests SUMO will be fast enough.

We are going to simulate tens of thousands of vehicles and the routing service may become a bottleneck (at present, a typical request takes some 3-4 seconds to execute, with most of the time spent on graph construction). Fortunately, as the simulation will be limited to a single city, we plan (if necessary) to build and store the graph in memory, which should definitely improve performance.

7. EXAMPLE

The following example demonstrates the effectiveness of car routing using our service. Two selected cars use different routes (Figure 5.): v1 drives along the shortest route (the red one, shown at the top) computed by SUMO while v2 employs the routing service, which is supplied with current traffic data, to get the fastest route (the green, bottom one). As a result, v1 enters a severe traffic jam and is considerably delayed; v2 takes a lightly loaded detour and arrives to a destination much earlier.

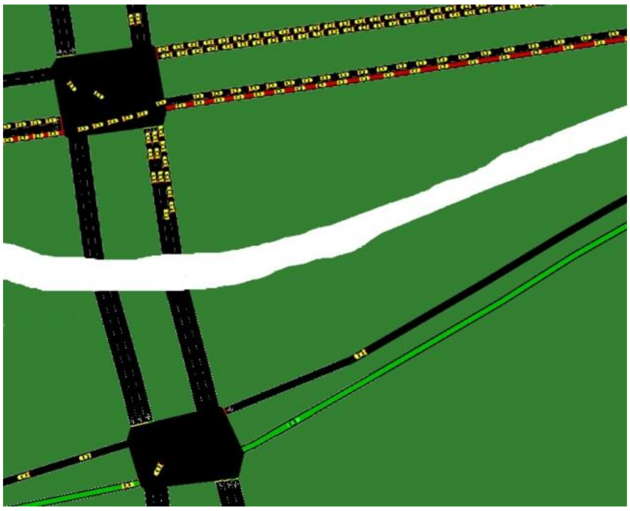


Figure 5. The jammed (upper) and the lightly loaded (bottom) routes (fragments shown)

8. CURRENT EXPERIENCE

Running simulations using an artificial graph is simple. Unfortunately, a real-world map conversion is a source of numerous problems, especially in case of a map that is not precise enough or is missing important road data. Conversion tools are either imperfect or not documented sufficiently (for example, the conversion between geographical coordinates and SUMO positions is quite tricky). Serious problems are caused by "micro-crossroads," that is, very short road segments between crossroads: They often cause a car to be unable to leave such a segment, clogging the crossroad forever. A common case is a deadlock of two opposite car directions, both trying to take a left turn and blocking each other. (These problems may be caused by map inaccuracies or simulator bugs.) In order to successfully run our simulation on the OSM map for Warsaw, we had to perform a simulation, analyze deadlock points, correct the map, and repeat these steps multiple times. That proved to be a time consuming and tedious work.

Apart from the problems mentioned above, we find running simulations in SUMO as relatively straightforward and enjoying. The strong feature of SUMO is its GUI front end, which allows to visualize moving cars in a nice manner and enables a full control over the simulation. It is uncomplicated to analyze the simulation and identify problems (such as, e.g., deadlocked crossroads). The main API, TraCI, is also functional enough; it allows, among other things, to route cars and control the traffic lights. Thanks to the fact that TraCI clients may be written in Python, we could easily write simulation scripts with desired logic and interfaces to our services (the routing service, the dynamic map, the GPS tracker, and, in future, the traffic control service).

9. SIMULATION SCENARIOS

Simulation scenarios we plan to perform include the following ideas:

Evaluation of control mechanisms. First of all, we are going to evaluate the influence (and efficiency) of additional, more complex and more intelligent control features to the overall traffic system performance. Thus, we are going to enable subsequent control mechanisms and compare results for the following cases (from simplest to most complex):

- routing service based on a static map only;
- routing service provided with current dynamic traffic data;
- routing service provided with both current dynamic traffic data and traffic prediction;
- as above, with load balancing enabled;
- as above, with traffic lights control enabled (a full scenario).

Reliable vs. irresponsible drivers. Simulated “drivers” will be offered a number (probably, two or three) alternative routes, with the preferred one advised by a load balancing function. We could compare two cases, when all drivers select the recommended route or when some of them “know better” and do not obey.

Full data about current traffic situation vs. “unsurveyed areas.” We could be able to compare the case when precise data are available for all roads and the case when only main roads are monitored (or, even worse, some sensors fail and report erroneous values). Additionally, the performance of GPS Tracker component could be compared with accurate data from higher-level sensors (i.e., directly from SUMO).

Traffic prediction accuracy. We could check what happens when traffic has been predicted perfectly and what happens if, for an unknown reason, actual traffic differs significantly from the forecast.

Modeling unexpected events. We plan to model accidents, intended traffic jams or sudden road closures and observe how the control layer copes with such a situation.

Modeling privileged vehicles. SUMO does not support privileged vehicles directly but some limited experiments are still possible. For example, we could try to control the traffic lights along a privileged vehicle’s route in order to assure a green light at every crossroad.

Note that most of the above scenarios compare the case of an “ideal” world (cooperating drivers, full information available, no unexpected events) with scenarios when something goes wrong, which often happens in the real world.

10. RELATED WORK

During related work review, we were mostly interested in (possibly) large-scale urban traffic simulations. Uppoor et al. (2013) generate a synthetic (although realistic) dataset of 24-hour car traffic for a 400-km^2 area around the city of Koln; the dataset could be then employed in other studies (e.g., research on wireless networks with on-board terminals in moving vehicles). They use OSM as map data and SUMO as the simulation tool. Another example includes simulating Tel Aviv Metropolitan Area in MATSim (Bekhor, Dobler, and Axhausen 2010) in order to compare and match traffic flows (of the original traffic model for Tel Aviv and the flow computed in MATSim). Balmer, Nagel, and Raney (2004) perform a large-scale, 24-hour microscopic traffic simulation for Switzerland (for the whole country). Garcia-Nieto, Alba and Olivera (2011) report on the usage of SUMO for finding successful cycle programs of traffic lights for large urban areas. Ben-Akiva and Davol (2002) present a case study of simulations employed for traffic model calibration for an area near Stockholm.

In addition to the above-mentioned SUMO and MATSim, there are a number of other urban traffic simulation tools. For example, MAINSIM (Dallmeyer and Timm 2012), which as able to simulate not only vehicles but bikes and pedestrians as well (although work on similar features in SUMO is ongoing (Krajzewicz et al. 2012)). DIESIS (2008) contains a categorized list of transportation systems simulation tools.

11. SUMMARY

The Insigma’s goals related to traffic optimization and control are ambitious and require that appropriate simulation environment be prepared first. We have developed an advanced routing service and have successfully integrated it with SUMO. Future work will include planned control layer components and large-scale simulations. We hope we will be able to report valuable results.

ACKNOWLEDGMENTS

The work has been co-financed by the European Regional Development Fund under the Innovative Economy Operational Program, INSIGMA project no. POIG.01.01.02-00-062/09.

REFERENCES

Balmer, M., Nagel, K., Raney, B. 2004. Large–scale multi-agent simulations for transportation applications. *Journal of Intelligent Transportation*

Systems: Technology, Planning, and Operations 8, 4 (2004), 205–221.

Bedi, P., Mediratta, N., Dhand, S., Sharma, R., Singhal, A., 2007. Avoiding Traffic Jam Using Ant Colony Optimization - A Novel Approach. *Conference on Computational Intelligence and Multimedia Applications*.

Behrisch, M., Bieker, L., Erdmann, J., Krajzewicz, D. 2011. SUMO - Simulation of Urban MObility: An Overview. *SIMUL 2011, The Third International Conference on Advances in System Simulation*, 2011.

Bekhor, S., Dobler, C., Axhausen, K. W. 2010. Integration of activity-based with agent-based models. An example from the Tel Aviv model and MATSim. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT) (2010).

Ben-Akiva, M. E., Davol, A. 2002. *Calibration and Evaluation of MITSIMLab in Stockholm*. Transportation Research Board Meeting, January 2002.

Dallmeyer, J., Timm, I. J. 2012. MAINSIM - MultimodAl INnercity SIMulation. *35th German Conference on Artificial Intelligence* (KI-2012).

DIESIS. 2008. D2.3 Report on available infrastructure simulators. Design of an Interoperable European federated Simulation network for critical InfraStructures (DIESIS) project report, 2008.

Garcia-Nieto, J., Alba, E., Olivera, A. C. 2011. Enhancing the Urban Road Traffic with Swarm Intelligence: A Case Study of Córdoba City Downtown. *Intelligent Systems Design and Applications (ISDA)*.

Google Maps Developer Documentation: https://developers.google.com/maps/documentation/

Góralski, W., Pyda, P., Dalecki, T., Batalla, J. M., Śliwiński, J., Latoszek, W., Gut, H. 2011. *On Dimensioning and Routing in the IP QoS System*. Journal of Telecommunications and Information Technology, nr 3, p. 21-28.

Hart, P. E., Nilsson, N. J., Raphael, B. 1968. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*. SSC4 4 (2): 100–107.

Janowski, L., Kozłowski, P., Baran, R., Romaniak, P., Glowacz, A., Rusc, T. 2012. *Quality assessment for a visual and automatic license plate recognition*. Multimedia Tools and Applications.

Krajzewicz, D., Erdmann, J., Behrisch, M., Bieker, L. 2012. Recent Development and Applications of SUMO – Simulation of Urban Mobility. *International Journal on Advances in Systems and Measurements*, vol 5 no 3 & 4, 2012.

Małowidzki, M., Bereziński, P., Dalecki, T., Mazur, M. 2012. Advanced Road Traffic Service Demonstrator. MCC'2012, Gdańsk, Poland.

Małowidzki, M., Dalecki, T., Bereziński, P., Mazur, M. 2013a. *Traffic Routes for Emergency Services*. Accepted for EUROSIM'2013.

Małowidzki, M., Mazur, M., Dalecki, T., Bereziński, P. 2013b. *Route Planning with Dynamic Data*. Accepted for MCC'2013.

MATSim: Agent-Based Transport Simulations: http://www.matsim.org/

OpenGTS™ - Open GPS Tracking System: http://opengts.sourceforge.net/

OpenLayers: http://openlayers.org

OpenStreetMap: http://www.openstreetmap.org/

SUMO: Simulation of Urban Mobility: http://sumo.sourceforge.net/

Uppoor, S., Trullols-Cruces, O., Fiore, M., Barcelo-Ordinas, J. M.. 2013. Generation and Analysis of a Large-scale Urban Vehicular Mobility Dataset. *IEEE Transactions on Mobile Computing*, 2013.

AGENDA-BASED BEHAVIOR IN PEDESTRIAN SIMULATION

Eric Kolstad$^{(a)}$, John M. Usher$^{(b)}$

$^{(a,b)}$Dept. of Industrial & Systems Engineering, Mississippi State University
P.O. Box 9542, Miss. State, MS 39762

$^{(a)}$ ewk6@msstate.edu, $^{(b)}$ usher@ise.msstate.edu

ABSTRACT

In order to construct a simulation that aptly characterizes pedestrian interactions in large-scale transportation facilities, it is necessary to consider the means to represent the requisite goals and activities of interest to specific individuals that act as primary influences in their navigation choice and other decision-making processes. As part of the Intermodal Simulator for the Analysis of Pedestrian Traffic (ISAPT), we have implemented an objective-based task agenda for pedestrians with priorities that are evaluated relative to factors such as resource availability, travel cost, relative level of need and estimated time to completion. Time-variant sets of such pedestrians, in turn, are configured to represent larger population groups.

Keywords: pedestrian traffic simulation, agent-based, task agenda, route planning.

1. INTRODUCTION

The broader decision-making processes of a pedestrian, as well as their momentary behaviors, are influenced by a number of factors, starting with low-level assessments based on collision avoidance and movement towards a waypoint target. When modeling interactions between pedestrians within a working facility (e.g., an airport), the practical choices pedestrians make in determining their course of action is highly dependent on their current goals and needs, relative to specific value judgments. These must be assessed by individuals dependent on resource availability, anticipated costs to utilize, and their current environmental conditions within the model.

The active pedestrian populations in a facility will correspond with one or more transportation sources, each of which may have several associated entrance regions. Upon initial arrival and at successive stages thereafter, pedestrians will review their current set of objectives and determine a prioritized course of action in route-based and conceptual terms. Their working knowledge of the available set of resource locations associated with the tasks involved (potentially augmented via information sources), along with the presently known state of the dynamic model conditions affects the relative ordering and prioritization of these activities.

This paper presents the approaches currently employed by ISAPT to model varied population-based groups, where each pedestrian maintains a unique agenda and periodically re-evaluates their goals in accordance with available time and resources. The paper describes the components which enable system definitions along with agenda-based behavioral response – which drive the emergent system dynamics of the simulation.

2. TASK-BASED AGENDA

ISAPT enables dynamic specification of the respective characteristics of a group of individuals within a set of one or more *populations*. Each pedestrian population is active over a specified time range during the simulation run, in accordance with real-world events. The attributes of individual pedestrians may either be pre-defined in a separate data file, or generated randomly via rule-basic logic and distributions that assign characteristics such as age, gender, personal needs, entry location, and agenda tasks. Timing intervals of pedestrian arrivals and departures to a facility may be configured to represent varied modeled intermodal sources such as vehicular traffic or light rail and correspond to different schedule-based observations (e.g. morning, mid-day, evening) – where a certain set of flights is available – or vary in accordance with holiday events et al.

The definition of a *task-based* agenda is intended to augment more realistic simulation by allowing each person to maintain a set of intended activities under active consideration. Higher-level planning processes must take into account the effective agenda list that includes a subset of potential resource-based tasks along with basic personal needs a person may want to satisfy (e.g. hunger, thirst, curiosity, restroom use). Each individual's agenda is generated upon their initial entry into the system – either as part of population-based generation, or as specified within trial data – for a given pedestrian ID (along with other attributes such as entry point and personal characteristics). The working agenda thus contains a list of objectives a pedestrian wants to accomplish during their visit, where each can be satisfied via one or more resource nodes located in the facility model. Up to 30 activities may be assigned to an agenda (per visit) from the list of possible activity types.

Tasks are configured with one of several *activity types* which may be travel-based activities (e.g. ticketing, baggage check, security, gate arrival, system exit), activities related to basic requirements (hunger, thirst, restroom, info et al.) or user-defined types that enable additional resources akin to those in a given transportation facility (such as a network access point). Each task is assigned an associated level of *need* (0..100) from a specified value distribution. This need represents a relative amount of service required, where a given resource has varied capacity to restore resource levels for up to five different needs, following an overall service time distribution. As an example, a vending machine would restore less hunger (or potentially thirst) needs than a visit to food vendor or restaurant. Provisions exist to accommodate a broad range of agenda affected by visitor type (e.g., traveler, non-traveler, worker), observed crowd-based flow and/or those related to certain time spans or known transportation modes.

Additional task attributes relate whether it is a `required` task (i.e. must be accomplished before exiting the simulation), marked *primary* (vs. *secondary* by default) in importance, whether it is part of a subset that must be done in sequential order, and if there are constraints on the system time(s) it can be performed. Tasks can also be marked as `procToNextReqd` (to immediately proceed to next required task marked as such), `noSecondary` (for a *primary* task that must be pursued before considering those of secondary priority), `timeFirstAvail` and `timeLastAvail` (system time range of resource availability, e.g. for GATEs or shops). Note that only *primary* tasks may be marked as *required*.

Pedestrians' current needs requirement levels – in conjunction with estimated travel and wait time to matching resources – determine the *prioritization* of tasks as the pedestrian plans their ongoing route. The impact of these factors is discussed further in Section 4. As travelers tend to have discretionary time prior to a flight's departure, they need not focus entirely on the most vital tasks and can choose to utilize other resources within a facility while en route. Beyond simply choosing to wait in a seating area, for example, pedestrians may consider less immediate concerns and choose activities such as shopping or obtaining a snack when reasonably sufficient time remains.

The generation of agenda tasks assigned to individuals within a population is enacted via XML-based definitions in the main experiment setup file. Each definition contains a named reference and task type (satisfied with corresponding system resources marked in the 3D facility model, i.e. HUNGER, THIRST, ENERGY, RESTROOM, CURIOSITY, INFO, TICKET_COUNTER, TICKET_KIOSK, BAG_CHECK, LUGGAGE_PICKUP, CAR_RENTAL, ARRANGE_TAXI, SECURITY, [departure] GATE, SYSTEM_EXIT, along with any custom user-defined types). The resource need level offered by a resource is specified using one of several distribution types and associated parameters.

Figure 1 shows a section of XML which illustrates how an end user can specify activities that will be assigned to pedestrians observed at a one entrance location, in conjunction with group characteristics for a certain population. When a task is selected for the pedestrian, that activity is assigned to their active agenda. Each task's settings can take on values defined in another section of code – which may be based on a variable – and/or randomly determined.

The first code segment establishes a macro-based activity variable definition for an information enquiry (INFO) task, which can be referenced as "`info_rnd`". When specified as a variable setting `<alt>`ernative in one of the system prototype definitions, this definition may in turn be referenced to enact part of the agenda settings by one or more populations. This code assigns a need level from a truncated normal distribution with a mean of 50, variance of 10, and lower and upper limits of 30 and 70, respectively. This activity is denoted as *required*, of *primary* task importance, and to be pursued in preference to any secondary task. There is no specific constraint on the range of time it can be performed. Once this task has been accomplished, however, the pedestrian will turn their attention to the next primary task in the agenda sequence.

The next section of code defines the pedestrian attributes and activities that may appear on the agenda of pedestrians that are instantiated at "`enter_node`" `D3`. The first assigned activity is enacted via the `info_rnd` definition noted earlier. Its chance of being assigned is 100%; therefore, every pedestrian in the population has an INFO activity added to their agenda. This is followed by a statement that directly defines a personal attribute flag indicating that the pedestrian does not currently have a ticket (`has_ticket` is `false`).

The statement thereafter randomly sets a *utility* variable value that can be used as a basis to choose whether the pedestrian will visit the kiosk, counter, or both as a part of their check-in activities. The actual check-in activities are then assigned by enacting [previously coded] activity definitions to their agenda depending on what the `utility1` variable was set to, via three select statements that follow. This approach readily permits the assignment of one of a set number of tasks based on discrete probabilities.

Once check-in activities are defined, another statement sets the `has_luggage` variable to true or false using the stated discrete probabilities, resulting in 70% of the pedestrians having luggage to check while the remaining 30% do not. If necessary (according to what `has_luggage` is set to), the activity for visiting the baggage check area will similarly be assigned or not.

Each pedestrian is now assigned a series of potential activities requiring satisfaction of personal needs (e.g., hunger) according to probabilities (i.e. 33% chance to receive each need-related task, 67% not to).

The actual level of each need is determined via prior `<activity_alt>` definitions. The final activity to be assigned to the agenda, in this example, is the mandatory need (100% chance) for them to pass through security.

```
<activity_alt>
    <alt name="info_rnd"  taskType="INFO"
     taskLevel="normal, 30.0, 50.0, 70.0, 10.0"
     primary="T" required="T" noSecondary="T"
     procToNextReqd="T" timeFirstAvail="0.0"
     timeLastAvail=""/>
 …
</activity_alt>
 …
<proto name="activities_entry1">
  <select var="enter_node"> <alt chance="100"
  value="D3"/> </select>

  <select var="activity"> <alt chance="100"
  assign="info_rnd"/> </select>

  <select var="has_ticket"> <alt chance="100"
  value="false"/> </select>

  <select var="utility1">
     <alt chance="10,20,70" value="kiosk_only,
     counter_only,kiosk_then_counter"/>
  </select>
  <select var="activity" based_on="utility1">
     <alt option="kiosk_only"
     assign="kiosk_rnd"/>
  </select>
  <select var="activity" based_on="utility1">
     <alt option="counter_only"
     assign="counter_rnd"/>
  </select>
  <select var="activity" based_on="utility1">
     <alt option="kiosk_then_counter"
     assign="kiosk_rnd"/>
     <alt option="kiosk_then_counter"
     assign="counter_rnd"/>
  </select>

  <select var="has_luggage"> <alt
  chance="70,30" value="true,false"/> </select>

  <select var="activity"
  based_on="has_luggage">
     <alt option="true"
     assign="bag_check_rnd"/>
  </select>

  <select var="activity"><alt chance="33,67"
  assign="hunger_rnd, none"/> </select>
  <select var="activity"> <alt chance="33,67"
  assign="thirst_rnd, none"/> </select>
  <select var="activity"> <alt chance="33,67"
  assign="energy_rnd, none"/> </select>
  <select var="activity"> <alt chance="33,67"
  assign="restroom_rnd, none"/> </select>
  <select var="activity"> <alt chance="33,67"
  assign="curiosity_rnd, none"/> </select>

  <select var="activity"> <alt chance="100"
  assign="security_rnd"/> </select>
</proto>
 …
```

Figure 1: XML activity spec example.

Use of XML files allows ISAPT to create an extensive variety of agendas. Table 1 shows an example of an agenda for a pedestrian departing on a flight. Although the agenda shown contains eight tasks, for a traveler it could be contain as many as 30 (an ISAPT system constraint). For instance, a traveler with only two tasks assigned may have already checked-in online before arrival and have no luggage to check, thus needing only to pass through security and reach their departure gate. The example agenda in Table 1 includes a secondary priority CURIOSITY task. The stronger the "need" for this task the more likely this pedestrian will be to explore available displays, visit shops or exhibits, sit by the window, or explore of parts of the facility that are marked as satisfying some level of curiosity. As with curiosity, both the pedestrian's need to satisfy their thirst and visit the restroom are not absolutely required and therefore will only be performed if extra time exists in their schedule.

On the simulation level, what drives activity assignment for a given pedestrian is their membership in one of several named *population groups* introduced to the simulation - where pedestrians marked as part of a certain group will be assigned their individual characteristics and agenda tasks in a similar manner to what has just been illustrated, resulting in. potential activities with needs level set via specified distributions et al. Several populations may be active within the system simultaneously, where each produces associated pedestrians across a certain span of time.

Table 1: Example pedestrian agenda list

Need level	primary	required	no second ary	proc. to next	Resource type
80	x	x	x	x	INFO
70	x	x		x	TICKET_COUNTER
100	x	x			BAG_CHECK
100	x	x			SECURITY
65					CURIOSITY
30					THIRST
10					RESTROOM
100	x	x			GATE

Figure 2 shows an example that establishes two populations within a simulation scenario entitled "facility test". The first population group, *pop_A*, will release 45 pedestrians into the system starting at clock time 0.0 with an interval between releases that follows an exponential distribution. The specific attributes and agenda for pedestrians in this population arise from their collective set of named `<proto>`types that are applied within the statements illustrated in Figure 1. The system will continue to introduce pedestrians to the system from *pop_A* until either all 45 have entered or the end time (`releaseT1`) of 6000.0 seconds is reached. *pop_B* is constructed in a similar way, although it different characteristics and will start its release later in the simulation.

```
<population name="pop_A"  count="45"
  releaseT0="0.0"  releaseT1="6000.0"
  release_distrib="exponential, 2.1, 97.0">
    <proto apply="activities_entry1"/>
    <proto apply="gender_group"/>
    <proto apply="airline_set1"/>
    <proto apply="luggage_general"/>
    <proto apply="activities_general"/>
</population>

<population name="pop_B"  count="70"
  releaseT0="2400.0"  releaseT1="8500.0"
 release_distrib="exponential, 1.43, 50.0”>
    <proto apply="activities_entry2"/>
    <proto apply="gender_group"/>
    <proto apply="airline_set2"/>
    <proto apply="luggage_general"/>
    <proto apply="activities_general"/>
</population>

<scenario name="facility_test">
    <population source="pop_A"/>
    <population source="pop_B"/>
</scenario>
```

Figure 2: Configuration of two pedestrian populations

3. TASK-PLANNING

In order to define the navigational structure and connectivity with available resources, the ISAPT system first takes as input a 3D model of the facility to be simulated that defines the architecture and layout. A set of interconnected *nodes* provide the basis for conceptual route planning within the 3D model, where adjacent nodes typically have incoming and outgoing links to neighbors on a directional graph. Each node consists of a physical location and extent (along with navigational bounds) and similarly acts as a waypoint for route decision making and coarse movement (see Figure 3). These nodes may be given additional properties that allow them to: 1) act as *resources* that provide a *service* pedestrians require, 2) enforce entry requirements, occupancy limits, etc., 3) effect line-formation changes to the graph, 4) maintain data for purposes of statistical analyses and user-directed pedestrian observations (Usher and Kolstad 2011). This set of node-based resources forms the basis for behavioral choices when pedestrians reach a decision point (e.g. a navigation branching point, or simply an upcoming node along the route), where they will review their present course of action relative to knowledge of the current system state and time remaining. The blue circles in Figure 3 represent connected nodes in proximity to an airport ticketing area with queue lines leading to kiosk and counter service resources.

While task prioritization schemes vary greatly across pedestrian models and related simulations, they share a primary goal in *routing*, in that their objective is to determine efficient paths from one arbitrary location to another. Simulation, computer gaming and robotics applications must assess the 3D model space in terms of its navigation potential, incorporating some means to represent space as a set of inter-connected destinations.

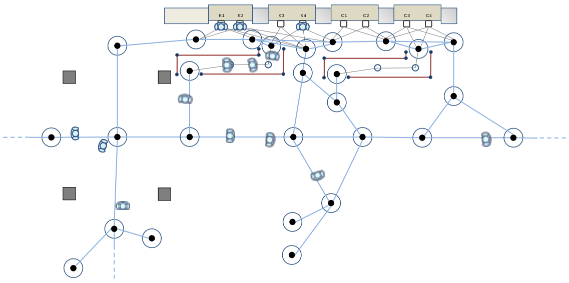


Figure 3: Portion of a connected node network model

In terms of path planning and traversal, determination of routes and respective travel cost/benefit in a simulated environment is largely dependent upon model representation. Grid-based pedestrian models (such as Kirchner et al. 2003) divide space into a set of uniform *grid cells* with inherent adjacency – thus a cell-to-cell route with minimal cost may be generated via iterative graph traversal e.g. the grid path maps of (Shao and Terzopoulos 2005), with perceived path *value* potentially influenced by prior traffic across those cells. 3D model space may also be evaluated to form a more general *navigation mesh* (O’Neill 2004) representing only the navigable regions of the model, which can be partitioned into a set of variably-sized polygons with shared boundaries that may be traversed. ISAPT is among the systems that employ a *waypoint graph* (Liden 2002) for path-finding, where navigable space is populated by nodes whose directional links have associated traversal costs (e.g. in terms of time and/or distance). Among well-known methods to find an optimal path among graph-connected nodes, a generalized form of Dijkstra’s algorithm (Knuth 1977) is utilized by ISAPT in determining shortest directionally-linked routes to all resource nodes of potential use for a given task.

When a pedestrian reaches navigational proximity to an upcoming graph node (per the orange-shirted traveler reaching the white outer node extent in Figure 4) the active (not yet completed) tasks on their agenda will have their current optimal routes assessed using that node as an origin point, considering paths directed towards any/all available resources that could satisfy the type of needs for each task. This re-planning may also occur in accordance when the pedestrian becomes aware of updates to current resources’ availability and/or anticipated wait times (including new opportunities nearby), changes in path connectivity, or updated information with regard to time constraints (e.g. the pedestrian may have just completed a task that took longer than expected and now find themselves behind schedule). This results in a time-based cost-value judgment (i.e. the associated cost weighted vs. level of current need, described in the next section) as to which potential destination works best. Therefore, the specific resource server node (and its route) the pedestrian aims for to achieve a given activity task can change as the simulation progresses. Once each activity has its most direct route determined and an associated acquisition value is computed, the pedestrian selects which activity and route to proceed with – potentially the same one as the present task in mind – and continues on their way.

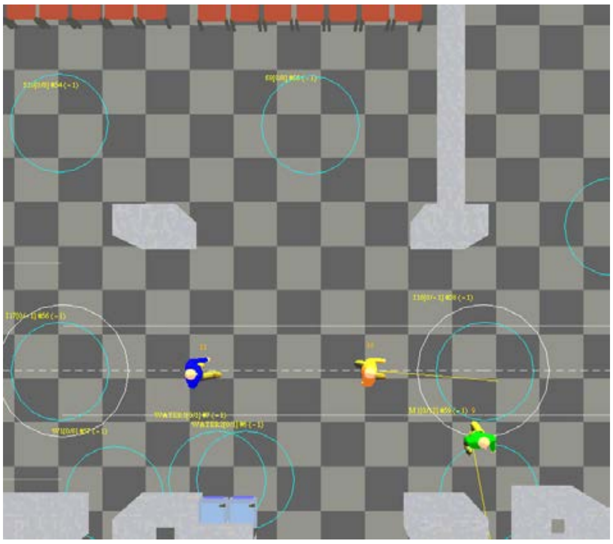


Figure 4: Pedestrians traveling on central corridor route, considering resources nearby (via perceived benefit).

Although pedestrians consider their set of agenda tasks in the context of a prioritized list, this list simply acts as a basis for the decision-making process overall. In order to affect more complex behavior where individuals, for example, may choose to pursue an activity conveniently en-route, the primary/secondary status of each task is used to help guide consideration of which to pursue, taking into account its cost-value assessment. In accordance with observations of pedestrians in-situ, certain rule-based decision processes may be inferred as to how to manage tasks that differ in these respects. In the next section we will discuss prioritization strategies implemented in ISAPT e.g. where a pedestrian might choose to change their present course of action, or decide how to spend their spare time waiting for a flight via exploration of a secondary task such as visiting shop-related resources.

4. TASK PRIORITIZATION

When the initial activities are assigned to a pedestrian – and at every decision point they encounter thereafter – all active (i.e. not yet completed) tasks in the current set are evaluated. All resource server nodes *relevant* to a task [within range] are examined in terms of availability, estimated cost, and how well they satisfy the need. Though tasks are initially added to the agenda list in the order they are specified during population generation, there is no default requirement that tasks be performed in a specific order. Certain tasks may however be marked as requisite to complete in order prior sequence to others (per the `procToNextReqd` tag noted in Section 3). Research suggests that the conceptual tasks a person has will generally be re-considered on a habitual basis (Chen 2004) and that specific actions taken can enact a shifting of priorities on an agenda list. Effective changes in task scheduling (Joh et al. 2001) may occur when activities are added, completed, or change in accordance with temporal or spatial shifts in the environment (Bladel et al. 2009) which may enact an impulsive change to the currently planned task.

Upon reaching the next *decision point* the pedestrian will conceptually reflect upon their current (not yet completed) set of agenda tasks and re-evaluate them. In addition, changing system conditions, such as a resource node becoming available, may also trigger re-evaluation of a pedestrian's current tasks. If changing conditions are observed in their nearby environment that impact the pedestrian's estimates in reaching and/or utilizing given resource nodes – thus altering their perceived acquisition costs – agenda re-planning is triggered. This assessment takes into account the tasks' relative importance, the resource node's aptness for the task, and overall time required.

In many cases there will be one or more available resources for a given task that can satisfy it (to varying extents) in addition to resources that may be presently in-use but are worth waiting for. Resource availability may be observed in terms of node occupancy and anticipated wait time - to the extent that the resources are within "visual" proximity with respect to the pedestrian's current location. Estimated cost is computed as a time-based measure in terms of shortest travel distance (expressed as travel time) and time to acquire the resource (including estimated processing and queue wait time if a line exists). When an ISAPT node has the ability to restore multiple resource needs, these will be satisfied within the overall processing time. For purposes of task planning, the best-scored *available* resource (which may be currently *in-use*) that would satisfy task objectives is noted for each activity along with the optimal route path to that resource.

As a prototype expression inclusive of these factors:

$$\begin{aligned} Cost\text{-}Value = R_{importance} * \left(\frac{\min(R_{need}, N_{restore})}{100.0}\right)^{1.2} \\ * \left(T_{travel} + T_{queue\_wait} + T_{process}\right) \\ \left[* R_{primary} \text{ for primary tasks}\right] \end{aligned}$$

where $R_{importance}$ is the system priority weight for that type of task, R_{need} is the pedestrians current level of need, and $N_{restore}$, the amount a given resource node can resolve. T_{travel}, T_{queue_wait}, $T_{process}$ represent estimated travel time (at current speed), anticipated wait time in a queue and/or until the resource is free, and the typical service time at that node, respectively. $R_{primary}$ can be optionally set to enact greater preference for *primary* tasks independent of available time. Here, the overall time requirement weighs heavily vs. relative reward, akin to real-world considerations. The power of 1.2 is an initial value based on informal experimental trials.

Further experimentation and analysis is needed to determine an appropriate value and the sensitivity of system operation to changes in this number, along with modifications for observed preference with regard to time and/or distance.

After the next task has been selected (including its resource and route), if reason exists, the pedestrian's travel to that task may potentially be pre-empted prior to reaching the intended resource. For instance, if they pass near a water fountain that has become available and happen to be quite thirsty. The exception is when system conditions effectively "lock in" the task (e.g. when the pedestrian's agenda calls for tasks to be performed in-sequence or they have progressed partway through a queue line).

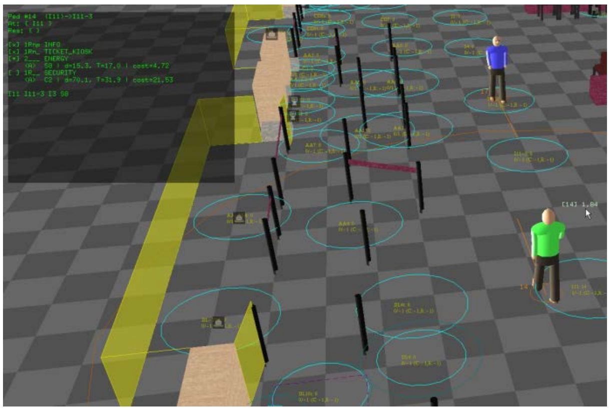


Figure 5: A pedestrian's working activity list after visiting a ticketing counter (heading towards seats)

While the pedestrian continues on their path and keeps track of their ongoing agenda list of tasks not yet completed (as seen in Figure 5 above), the decision logic that enacts the effective choice of task to be pursued must take into account some considerations beyond simply the raw cost-value assessment itself. Even though *primary* tasks are of more immediate concern, if an acceptable amount of spare time exists that would allow all required [and/or primary] tasks to be completed prior to facility departure, any secondary tasks that appear viable may also be considered for inclusion while en-route to the original task. For instance, a pedestrian might stop to get a snack or drink of water en-route to the ticketing counter or prior to entering a security zone on the map.

With these considerations in mind, the overall logic of pedestrians' activity choice can be roughly summarized as shown in Figure 6. Unless a pedestrian is occupied at a resource server, committed to a particular activity, or required to proceed with their next most immediate task, they will give priority to a certain course of action based on cost analyses of the available options. Primary tasks on the agenda will be reviewed if there are accessible resource nodes that at least partly restore the respective resource type. Certain nodes may not be reachable due to conditional requirements (e.g. a ticketing counter or boarding gate limited to passengers of a certain airline carrier and/or flight number),

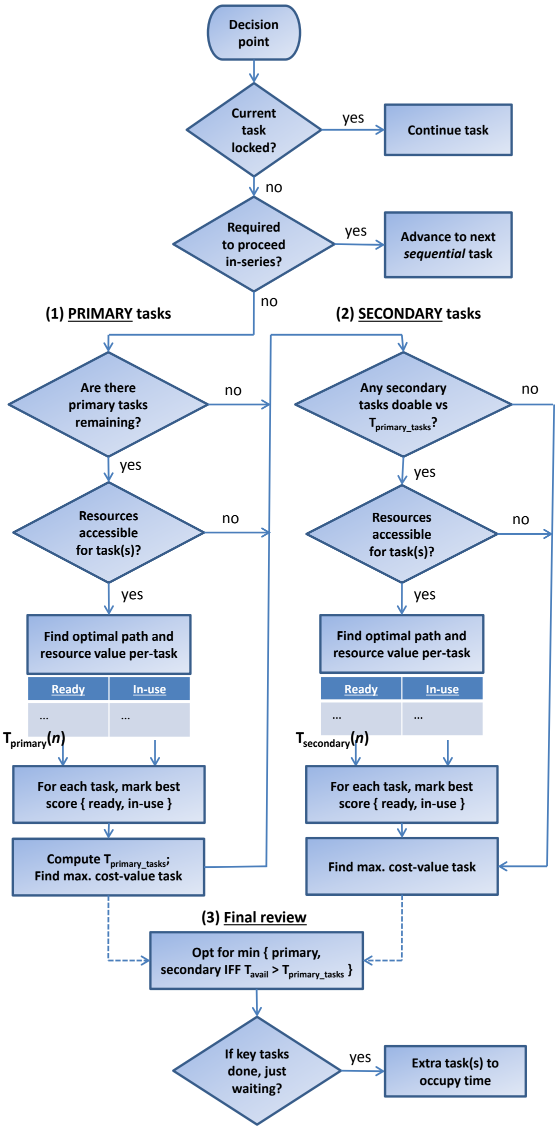

Decision point
Current task locked?
yes
Continue task
no
Required to proceed in-series?
yes
Advance to next *sequential* task
no
(1) PRIMARY tasks
(2) SECONDARY tasks
Are there primary tasks remaining?
no
Any secondary tasks doable vs $T_{primary\_tasks}$?
no
yes
yes
Resources accessible for task(s)?
no
Resources accessible for task(s)?
no
yes
yes
Find optimal path and resource value per-task
Ready
In-use
...
...
Find optimal path and resource value per-task
Ready
In-use
...
...
$T_{primary}(n)$
$T_{secondary}(n)$
For each task, mark best score { ready, in-use }
For each task, mark best score { ready, in-use }
Compute $T_{primary\_tasks}$; Find max. cost-value task
Find max. cost-value task
(3) Final review
Opt for min { primary, secondary IFF $T_{avail} > T_{primary\_tasks}$ }
If key tasks done, just waiting?
yes
Extra task(s) to occupy time

Figure 6: Task decision flowchart

temporarily blocked corridor regions, and other interruptions of graph connectivity. An optimal path is determined for each accessible resource and its cost-value computed relative to the resource needs level(s) that can be restored. In accordance with real-world pedestrian behavior, resource nodes are considered viable whether they are currently available or still being utilized – and periodically reassessed. The highest cost-valued task at that point becomes the working task choice.

As secondary tasks are viewed as optional, they will *not* be considered unless time estimates show time remaining beyond that necessary for all primary tasks' completion prior to the pedestrian's facility exit. Each secondary task will be checked for resource accessibility, followed by cost analyses, with a secondary task choice outcome if viable. While the cost scoring metric is the same for both primary and secondary tasks, an open primary task receives more immediate (and potentially weighted) consideration in the decision process.

Finally, where a pedestrian is not yet ready to exit the facility (e.g. when waiting to board a flight at their departure gate) yet has no other primary or secondary tasks remaining, they may opt either to continue waiting or add one or more time-occupying basic needs tasks of their choice (i.e. ENERGY (which may suggest taking a seat), CURIOSITY, HUNGER, THIRST, or RESTROOM) with randomly assigned levels. Incomplete tasks that remain on the agenda will take precedence otherwise. More detailed behavior in this case is not currently modeled in ISAPT.

5. SUMMARY/CONCLUSIONS

The ISAPT system facilitates structured definition of varied pedestrian populations for large-scale facilities, where individuals possess an array of characteristic personal attributes and activity interests that can vary in accordance with flight schedules, time-of-day variations, arrival source patterns and so forth – along with socio-demographic crowd distributions – as relevant trial data and/or larger research study trends may suggest. The associated tasks and priorities assigned to pedestrians within the mixed active population(s) allow users to explore the impact of these factors on resource usage and overall flow within the modeled facility.

As a key component of the simulation model, we have implemented a task-based agenda approach that allows flexible consideration of activity lists, where an individual may periodically re-assess their course of action in accordance with value judgments that reflect an ongoing reasoned choice of activities provided limited time to accomplish them. In sharing objectives with approaches that attempt to optimize working agenda lists' order via cost-benefit analyses applied to a collective task sequence in context of the surrounding environment (e.g. Hoogendorn and Bovy 2004), such strategies might be also applied to determine initial task order and/or to gauge current precedence for primary agenda tasks within ISAPT. However, incorporating less structured consideration of ongoing agenda while maintaining active preference logic has potential to provide a more free-ranging view of current agenda tasks, particularly where constraints exist but individuals may have larger periods of free time and/or be more apt to meander while exploring their surroundings.

ACKNOWLEDGEMENTS

This project was sponsored by the U.S. Department of Transportation (Grant No. DTOS59-06-G-00041).

REFERENCES

Bladel, K.V., Bellemans, T., Janssens, D., & Wets, G., 2009. Activity Travel Planning and Rescheduling Behavior: Empirical Analysis of Influencing Factors. *Journal of the Transportation Research Board*, 2134, 135-142.

Chen, C., Garling, T., & Kitamura, R., 2004. Activity Rescheduling: Reasoned or Habitual? *Transportation Research Part F*, 7(6), 351-371.

Hoogendoorn, S. P. & Bovy, P. H. L., 2004. Bovy, Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B*, vol. 38, no. 2, pp. 169-190.

Joh, C.H., Arentze, T.A., & Timmermans, H.J.P., 2001. Activity scheduling and rescheduling behavior. *GeoJournal*, 53(4), 359-371.

Kirchner, A., Namazi, A., Nishinari, K. and Schadschneider, A., 2003. Role of Conflicts in the Floor Field Cellular Automaton Model for Pedestrian Dynamics. *2nd International Conference on Pedestrians and Evacuation Dynamics*, 51-62.

Knuth, D.E., 1977. A Generalization of Dijkstra's Algorithm. *Information Processing Letters*, 6 (1): 1–5.

Liden, L., 2002. Strategic and Tactical Reasoning with Waypoints. AI Game Programming Wisdom. Charles River Media, Hingham, MA, 211-219.

O'Neill, J., 2004. Efficient Navigation Mesh Implementation. *Journal of Game Development*, vol.1, no.1, 71-90.

Shao, W., and Terzopoulos, D., 2005. Environmental Modeling for Autonomous Virtual Pedestrians. *Proceedings of the 2005 SAE Symposium on Digital Human Modeling for Design and Engineering*, Iowa City, Iowa, USA.

Usher, J.M., and Kolstad, E., 2011. Indoor Pedestrian Navigation Simulation Via a Network Framework. *Proceedings of the 23rd European Modeling and Simulation Symposium*, 247-253. Sept 12-14, Rome, Italy.

ARDUINO PLATFORM AND OBJECT-ORIENTED PROGRAMMING APPLIED TO AUTONOMOUS ROBOTS FOR DETECTION OF PIPELINE

Robson da Cunha Santos[(a)], **Marcelo Silva**[(b)], **Gerson B. Alcantara**[(c)], **Julio C. P. Ribeiro**[(d)], **Iorran M. de Castro**[(e)]

[(a)] Fluminense Federal Institute, Coordinator and Professor - Engineering and Automation Control
164 km Amaral Peixoto Road, Brazil and Estácio de Sá University, Professor - Engineering Petroleum, General Alfredo Bruno Gomes Martins Highway, s/n - Braga - Cabo Frio / RJ, Brazil
[(b)(c)] Estácio de Sá University, Coordinator and Professor - General Alfredo Bruno Gomes Martins Highway, s/n - Braga - Cabo Frio / RJ, Brazil
[(d)(e)] Estácio de Sá University, Scientific Initiation Student, General Alfredo Bruno Gomes Martins Highway, s/n - Braga - Cabo Frio / RJ, Brazil

[(a)] profrobsons@yahoo.com.br, [(b)] msc.marcelosilva@gmail.com, [(c)] gerson.alcantara@gmail.com, [(d)]iorranpt@gmail.com, [(e)] juliopribeiro@hotmail.com

ABSTRACT

The aim of this work is to Demonstrate a prototype that inspects pipelines buried through an autonomous robot. The inspection aims to follow the route of the pipeline, capturing its coordinates (x, y, z), storing and sending data through the Arduino platform, electronic boards and specific programming. For data storage we used a Micro SD Card Shield and in order to sending the data we used the General Packet Radio Services, allowing the data traffic integrating mobile with internet. The electromagnetism was used for that the robot accompany the pipeline buried through the its magnetic field. The coordinates captured by the vehicle are compared with the coordinates of setup project. After the capture of the coordinates, some decisions may be taken: monitoring the pipeline that may be positioned in slopes; regions with high pluviometric index and risks of possible oil spills. The project uses innovative technologies, accessible and low cost.

Keywords: Inspect pipelines, Autonomous robot, Specific programming, Arduino platform.

1. INTRODUCTION

Nowadays, Brazil has installed approximately 22000km oil and gas pipelines on land (Monitor Mercantil 2010). According to the regulation of ANP Technical 2/2011, is generic designation duct installation consists of pipes connected together, including Components and Accessories, for the transport or transfer of fluids between the frontiers of geographically distinct operating units. The ducts are composed of pipelines and mostly buried about 2 feet deep.

The pipelines are constructed for transporting substances, often dangerous, both in gaseous state to the liquid state. The transport of these substances through pipelines is considered not only the safest, but also, for large quantities, the most practical and economical, even if confronted with road transport or rail. Nowadays, there is a growing preference for this mode of transport, prompting the pipeline network grow quickly and consistently (Cardoso 2004). It is recognized that there is a potential risk of accidents classified as grave. As natural precaution for all pipelines should be developed risk analysis, evaluation of consequences and prepared their plans.

1.1. Studies of Techniques Inspection

In Brazil and in the world, the technique used to inspect pipelines depends on the particular fault that you want to find. There are defects that may be recognized externally to the pipeline. In the case of pipelines located in the external surface inspection can be performed visually. With this purpose, the translation of inspection through the external wall of the duct, for a exhaustive analysis, there is not a problem of great difficulty. In the atmosphere, the technologies of non destructive metallographic tests, broadly used in industrial fields, are applied with relative ease. However, when the external inspection occurs in places with water (rivers, lakes and seas) is complicated in that it increases the depth where the pipeline is located. For this type of problem, one of the most promising solutions is the development of Autonomus Underwater Vehicle (AUVs), sailing in deep sea without physical connection to a vessel or platform surface (Avia 2000)

The AUV presented can be utilized for military purposes, inspected areas with possibilities of existence and pumps for commercial purposes to finish locations for future installations of oil platforms and pipeline inspection. Thus, it is able to follow the trail of submarine pipelines by acoustic and magnetic sensors.

Figure 1: Autonomu Underwater Vehicle

As previously mentioned, the inspection can occur internally or externally to the pipeline. However, the main problem of internal inspection of pipelines is not really a technical failure analysis, but the difficulty in accessing the area of interest.

In the next figure, the article text demonstrates a robot with camera and mats for visual inspection of pipelines. This type of robot is in common use in the inspection of sewer networks, cleaning air conditioning ducts and nuclear plants. The device with high degree of autonomy, known as DAVID, is not a good example of inspection robot internal buzzing with wheels (FERASOLI et al .., 1999). This type of robot is able to move by means of wheels driven by electric motors, has a structure suitable mechanical pipes of circular cross section, and can be altered according to constraints imposed by the environment.

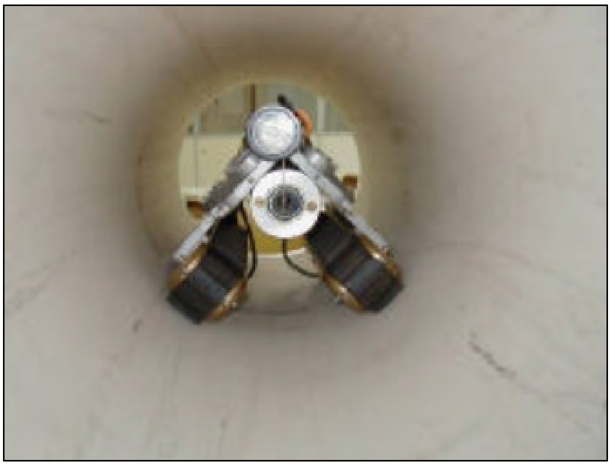


Figure 2: Example Robot for Internal Inspection of pipelines with treadmills

Another robot model that uses suction cups to clin'g to inspect the walls. This is the case SADIE robot that climbs walls with the aid of his paws, designed in England for non-destructive testing of welding a nuclear reactor (Luk et al., 2003).

However, there is a lot more complex cases of internal inspection, occurs when the flow in the pipeline can not be interrupted. In this case, the most used technique of internal displacement in ducts for this condition is the impulsion of inspection tools making use of the same fluid transported. This type of inspection tools is called Pipeline Inspection Gauge-PIG (Cardoso 2004). The PIGs collect information on the walls of the ducts through sensors and require no cord, since they use the energy of the fluid to move. The following are some models PIGs used in internal inspection of pipelines in the oil industry.

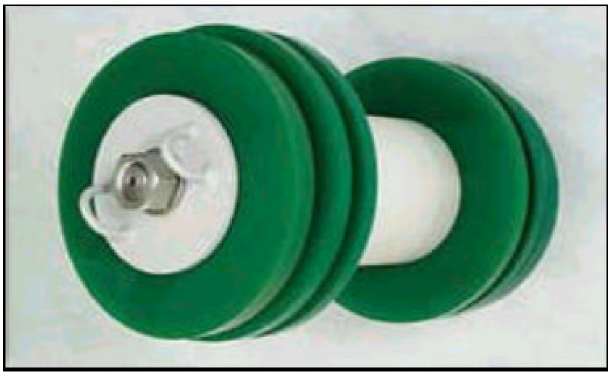


Figure 3: PIG of a Foam

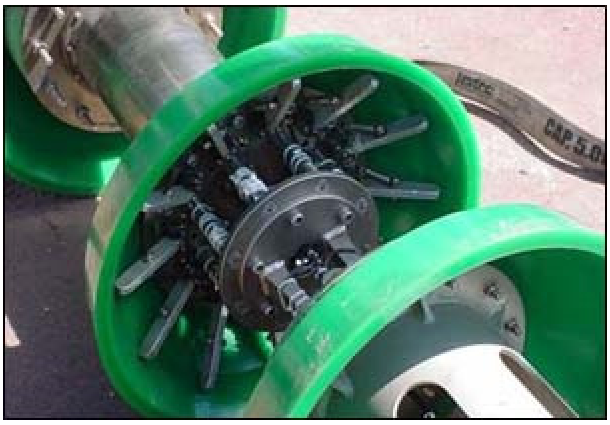


Figure 4: PIG Geometric

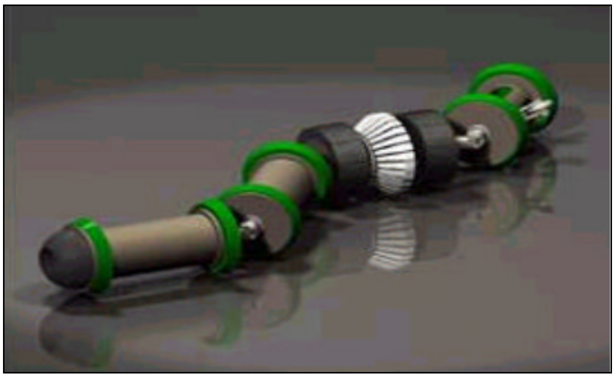


Figure 5: PIG Magnetic

1.2. Histories Duct in Brazil

Today, Brazil has kilometers of pipelines installed in regions with different geological characteristics, presenting several unstable areas called "high impact" as slums, urban agglomerations and varied underground or aerial crossings of roads and rivers (A Tribuna 2013). "Duct is the generic installation consists of pipes connected together, including Components and Accessories, for the transport or transfer of fluids between the borders of geographically distinct operating units." (ANP 2011).

In several regions of the country, pipelines are subjected to stresses imposed by earth movements. To ensure the structural integrity of the ducts installed in these areas, it is necessary to examine all areas and to map unstable and studying the movement of the soil mass. The movements usually involve trawling extensive areas and feature a slow speed. In general are difficult to detect by visual inspection risk areas (Santos 2013).

The National Agency of Petroleum, Natural Gas and Biofuels (ANP) published in several newspapers to inspect pipelines that crisscross the country (A Tribuna 2013).

To improve the operational safety of the pipeline, new technologies are being developed to detect unstable areas and to estimate its effect.

The inspection of pipelines must be observed by all range extension, as irregularities that give rise to abnormal mechanical stress on the pipes or likely to endanger the existing facilities. Some conditions can be cited, such as erosion, earth movement, landslide, vehicular traffic and / or heavy equipment on track, growth of vegetation, drainage system deficiency range, fires, invasion of track by third parties, conducting works nearby or interfering with range (buildings and explosions), deficiency in its demarcation and warning signs, outcrop duct, crossing streams with apparent duct, subjected to currents of water or erosion processes may create a risk to the pipeline (Sampaio 2012).

2. THE MOTIVATION

The main motivation of the proposed article was tedious due to existing teaching methods in technical and higher education in Brazil. Thus, it was necessary a search for innovations, in order to attract the interest of new students. Technology arising from globalization has brought many benefits to the population, however, had no such technology as its main focus new methods to improve the education system. The project in question is intended to motivate these students, improving their performance and transforming learning into a moment of pleasure and not just an obligation. These improvements were made possible by the tools selected for the project development that facilitate student understanding. An innovative tool, flexible, with an environmental vision and low cost.

2.1. Development Platform: Arduino

The Arduino platform or simply Arduino (Banzi 2011) is a platform that was built to promote physical interaction between the environment and the computer using electronic devices in a simple form and based on free software and hardware.

Considering a more succinct form, the platform consists of a circuit board with inputs and outputs to a microcontroller AVR development environment and the boot loader already recorded in the microcontroller. The microcontroller is composed of a microprocessor, memory, and peripheral input / output and can be programmed for specific functions, for example, different control and automation machines.

Studies show that there are many other platforms built for microcontrollers, but the Arduino has excelled on the world stage for ease of programming, versatility and low cost. Even for those who want high-level interactions, the Arduino has fulfilled expectations.

Besides all the adjectives preceding the development platform Arduino still differs from the others in the market by being a multiplatform development environment. Thus, its applicability is feasible in various types of operating systems, as well as being open source, where anyone has the possibility to download the tool, and is based on the Processing programming IDE, a development environment easily used.

Figure 6: Platform of Development Arduino
Source: Arduino.Cc

According Banzi, the Arduino is a physical computing platform open source, based on a simple plate of inputs and outputs, which can be used to develop interactive objects independent or connected to software, such as Flash, Processing , Java and etc.. Another concept on one of the main advantages of the platform Arduino, according to the thoughts of McRoberts, is: The biggest advantage over other platforms Arduino microcontroller development is the ease of use, people who are not technical area can quickly learn the basics and create their own projects in a relatively short time interval.

2.2. Java Platform

Nowadays there are on the market various types of programming languages, where each has a specific characteristic.

Java is a programming language object oriented very differently from conventional languages, which are only compiled or interpreted, it is hybrid, ie, compiled and interpreted by a virtual machine. The methodology of object-oriented programming is highly suitable for software development for large-sized companies, however, it is also suitable for school projects because it is simple and succinct. It also introduces concepts of modularity and reusability and approaches which enable the programmer to visualize your project running as a collection of cooperating objects that communicate via messages (Deitel 2010).

3. THE DEVELOPMENT

The project goal was to create a shopping autonomous remotely controlled, using the knowledge gained in scientific initiation, following several steps. In the next picture can be seen the evolution of students' knowledge.

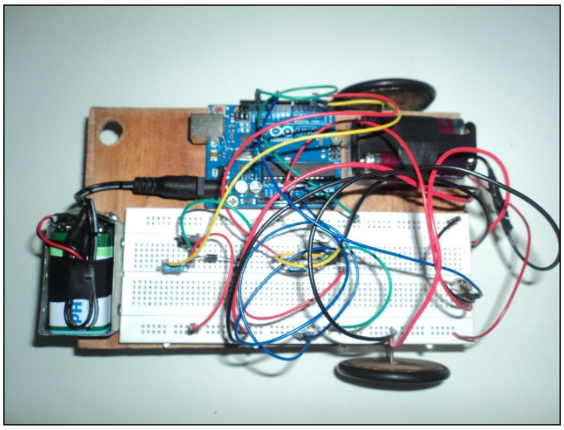


Figure 7: First Prototype of Stand

The beginning of the activities related to the use of robotics as a tool for teaching support occurred through the explanation on topics related to the research object. The weekly group meetings undergraduates, can bring greater knowledge, involving concepts about arduino, electronics, robotics and computer systems, as well as the description of the advent of technology in support of industrial process and society.

The central idea of the weekly meetings was to awaken the sense of awareness of the students about the practical usefulness in curricular subjects in teaching computer science, electronics and graphic interfaces.

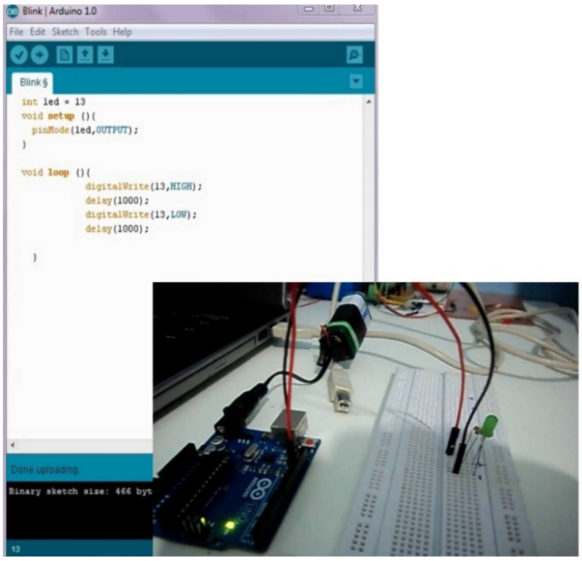
3.1. The First Steps

A small representation of card use of components mounted on a breadboard display for ease of programming and immediate results.

The configuration is intended to turn on and off a small LED from the output of the Arduino and programming itself.

After acquiring greater knowledge and studies in the area, the group got the Proteus software that has a simulation environment for electronic circuits and ISIS program to design printed circuit Ares professional. This software is used to simulate microprocessors, schematic capture and printed circuit board.

in the following figures can be seen the evolution of the first steps through the images, the software interface design and the chosen card that had to fabricate.



```
int led = 13
void setup (){
  pinMode(led,OUTPUT);
}

void loop (){
            digitalWrite(13,HIGH);
            delay(1000);
            digitalWrite(13,LOW);
            delay(1000);

  }
```

Done uploading.

```
Binary sketch size: 466 byt
```

Figure 8: Arduino Programming and Representation in Breadboard

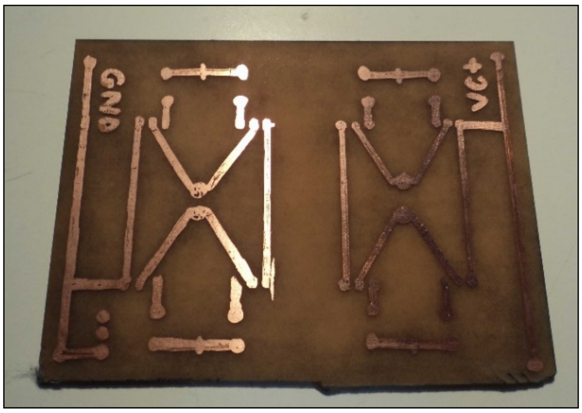


Figure 9: Plate Manually Manufactured

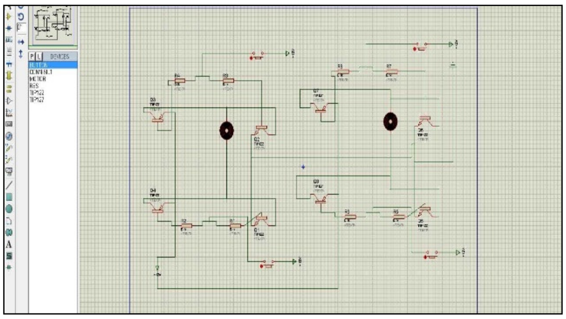


Figure 10: Interface Software Proteus

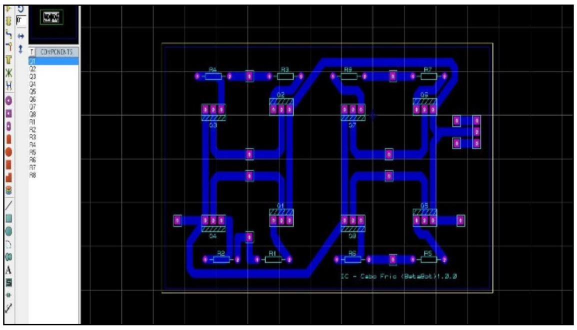


Figure 11: Interface Software Proteus

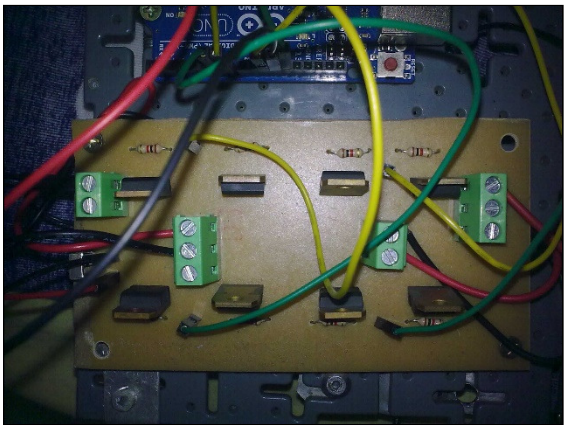


Figure 12: Made Plate and Applied to Model

3.2. New Components

From the limitations of the first prototypes, new devices have been researched and placed in the cart.

3.2.1. Micro Shield SD Card

The model Arduino UNO has 32k of flash memory, which are intended to 2k bootloader ATmega328P microcontroller. With this limitation it becomes unrealistic to to store data on the platform, since its capacity is greatly reduced (Banzi 2011). However, the simplest way to solve the problem is to use a shield Micro SD card, for the purpose of expanding the storage capacity of the card. Thus, the prototype would have a greater storage capacity of the captured data, the coordinates x, y and z.

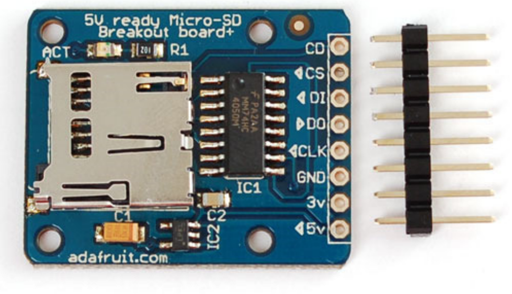
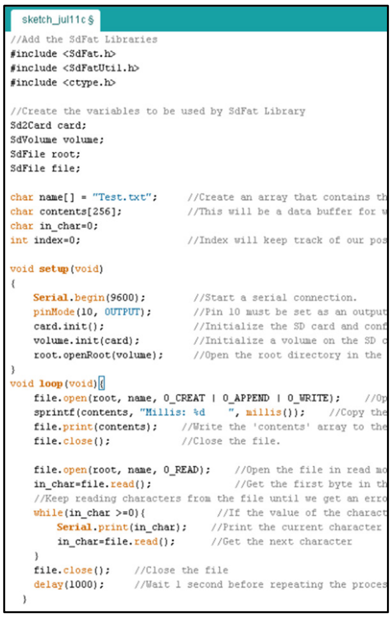


Figure 13:Micro SD CARD Model

The facility of handling of the Micro SD Card Shield make it widely used in the market. In the prototype was only necessary to save the files, because reading it was made by a external program. In the following figure is shown in the code language and the Arduino special facility to save and read files in *. Txt.



```
//Add the SdFat Libraries
#include <SdFat.h>
#include <SdFatUtil.h>
#include <ctype.h>

//Create the variables to be used by SdFat Library
Sd2Card card;
SdVolume volume;
SdFile root;
SdFile file;

char name[] = "Test.txt";      //Create an array that contains th
char contents[256];            //This will be a data buffer for w
char in_char=0;
int index=0;                   //Index will keep track of our pos

void setup(void)
{
    Serial.begin(9600);         //Start a serial connection.
    pinMode(10, OUTPUT);        //Pin 10 must be set as an output
    card.init();                //Initialize the SD card and conf
    volume.init(card);          //Initialize a volume on the SD c
    root.openRoot(volume);      //Open the root directory in the
}
void loop(void){
    file.open(root, name, O_CREAT | O_APPEND | O_WRITE);    //Op
    sprintf(contents, "Millis: %d    ", millis());    //Copy the
    file.print(contents);    //Write the 'contents' array to the
    file.close();            //Close the file.

    file.open(root, name, O_READ);    //Open the file in read mo
    in_char=file.read();              //Get the first byte in th
    //Keep reading characters from the file until we get an erro
    while(in_char >=0){            //If the value of the charact
        Serial.print(in_char);    //Print the current character
        in_char=file.read();      //Get the next character
    }
    file.close();    //Close the file
    delay(1000);     //Wait 1 second before repeating the proces
  }
```

Figure 14: Basic Programming - SD CARD

3.2.2. General Packet Radio Services

The initial proposal for the acquisition of the data obtained in the path traveled by the vehicle were restricted to collections performed manually, where user in question would have to remove the memory card micro shield installed so he could make reading the data via an external computer . This acquisition method was quite hard, causing future problems if the prototype needed to operate on a larger scale.

Studies were initiated in order to meet this new barrier. Important to note that one of the purposes of this project is to bring convenience and ease for inspection of pipelines. A technology widely used in the market today is the General Packet Radio Services (GPRS), whose purpose is to enable packet data traffic to the cellular network can be integrated internet. The GSM system with integrated GPRS named 2.5G generation, having been an important development for mobile data communication. GPRS allows transfer rates around 40 kbps (Rulik 2003).

Currently there are many GPRS Shields for Arduino platform, however, tests were performed based on GPRS SIM900, a module to operate in the mobile network GSM / GPRS cellular, capable of performing all the functions of a conventional device as: Make and receive calls, send and receive SMS, to connect to the internet through GPRS connection, and sending FAX. It is based on SIM900 chip low power and size summarized, SIMCOM the manufacturer, and is widely used for telemetry, remote sensing and actuation. The module is fully controlled and configured via AT commands (Sverzut 2005).

Figure 15: GPRS SIM900 Model

With the coordinates once saved the Shield Micro SD Card, GPRS SIM900 send the file. Txt for the external program treats the data obtained. Below a simple approach, where an SMS is sent to the desired message.

3.3. New Models

With developments and new research, a new cart is designed. The main objective was to detect metal buried about 10cm from the surface. For that detection was possible was initially used a coil in the rear of the prototype and a circuit for detecting plate with metals.

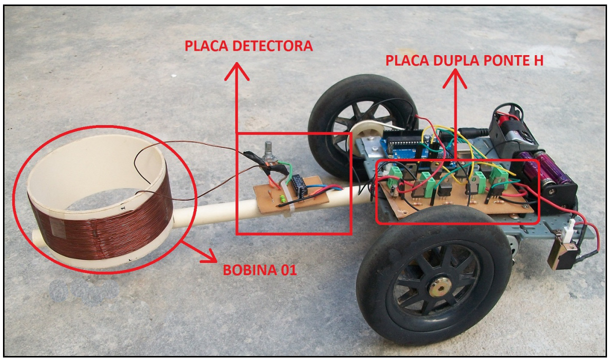


Figure 16: Prototype with a Coil

A board dual H-bridge was installed in the prototype to make it possible to change direction. The board dual H-bridge consists of an electric motor current (DC) is a common type of electric motor. Its main feature is that it has two electrical terminals, one positive and one negative. If an electric current travels in the normal motor shaft rotates to one side. If the current is reversed, the shaft rotates to the other side. Therefore with this card, we can reverse the direction of rotation of the shaft simply by reversing the polarity of the electrical terminals, and thus reversing the flow.

The prototype with only one coil was not sufficient to follow the route of the pipeline buried. Thus, a new project for a new prototype was implemented with For all the foregoing, we conclude first that technological innovations and form of teaching, objectifying an improvement the interaction between the student and the content available, directly inflicts the use of the same. If present technologies are currently utilized in the right way, a way to work in favor of the population, it tends to generate numerous and incalculable benefits. Once it possible to carry out projects far beyond the boundaries imposed by the old barriers, it becomes much easier the emergence of new ideas, where these same students will be developing fruit, resulting eventually in the labor market. two coils and the sum of the magnetic fields would direct the cart to keep the pipeline.

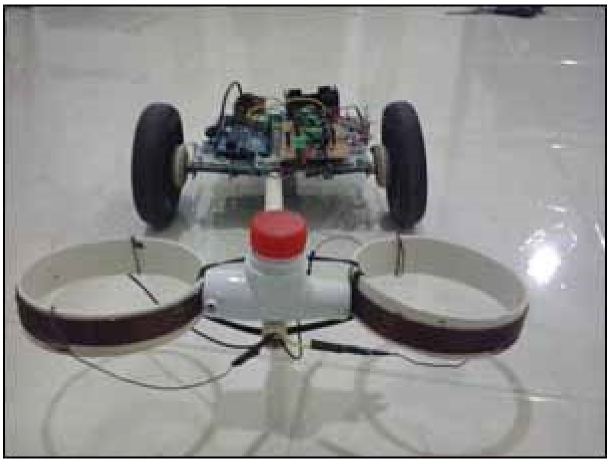


Figure 17: Prototype with two Coils

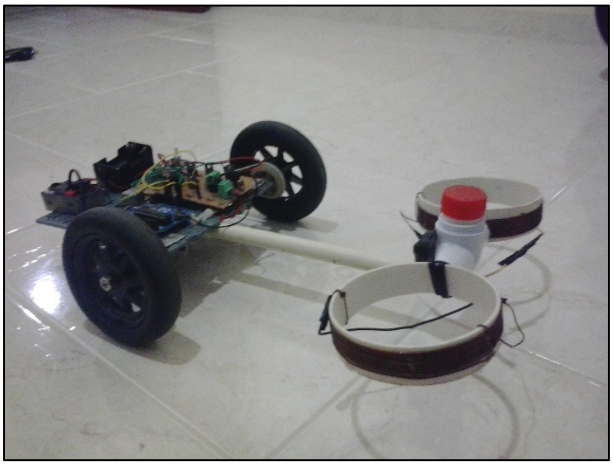


Figure 18: Prototype with two Coils

The project has a sustainable view and thus the construction of the prototype uses recyclable objects and equipment.

4. CONCLUSIONS

For all the foregoing, we conclude first that technological innovations and form of teaching, objectifying an improvement the interaction between the student and the content available, directly inflicts the use of the same. If present technologies are currently utilized in the right way, a way to work in favor of the population, it tends to generate numerous and incalculable benefits. Once it

possible to carry out projects far beyond the boundaries imposed by the old barriers, it becomes much easier the emergence of new ideas, where these same students will be developing fruit, resulting eventually in the labor market.

It is also remarkable that there are numerous such technologies with applicability, such as: Houses automated mechanical arms to aid in several areas among other projects. The technologies are there to be used, just to educators innovate in the how to teach, thereby rousing the creativity in each student.

All the Arduino platform, both in software and in hardware, it is open source, which greatly facilitates its use and dissemination. In other words, there is an endless number of libraries and tutorials available on the web for many applications.

These factors allow us to emphasize the simplicity of using the Arduino platform as data acquisition and automation, coupled with the relatively low cost and good results, contributing significantly to make the didactic laboratory.

All technologies presented led the group the Scientific Initiation to develop a prototype of an autonomous robot for the detection of buried pipelines. Considered proportionalities in size, the following figure can demonstrate that the group is on the path of developing a tool that can be useful in the petroleum industry.

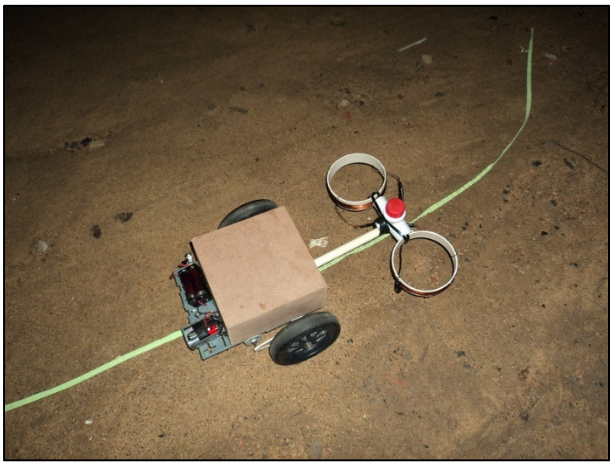


Figure 19: Prototype with Improved Protection

Studies have led students to prepare a route burying an iron tube 3 "in diameter about 10cm from the ground. The initial coordinates were determined, typed into a text file and passed to a spreadsheet, soon after, the students put the autonomous robot at the starting position to verify its efficiency. At the beginning, some adjustments were necessary, but the following graph can be proved that the robot came very close to the expected result.

In the following graphic can be seen the original trajectory and the layout of the autonomous robot. Some considerations such as the output of the robot's margin of error of GPS should be revised in future prototypes.

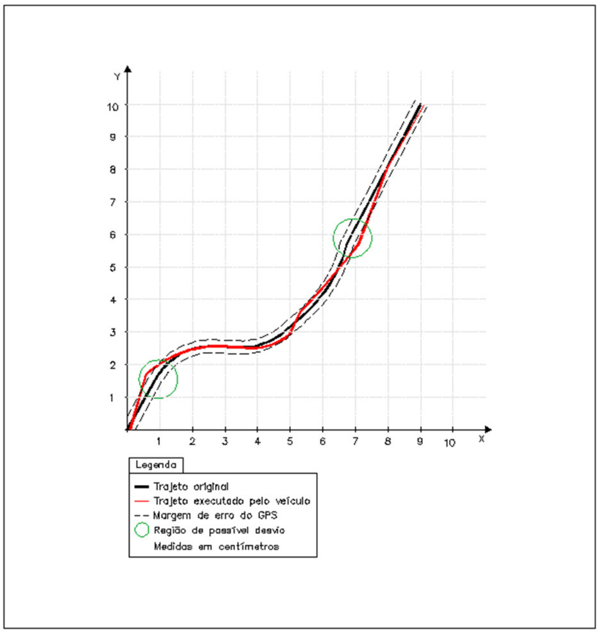


Figure 20: Graph showing the routes

If the graph showed a pathway of a pipeline suffering efforts soil, and the robot could find possible locations ranging from the initial position, responsible for the pipeline could take some decisions. Such a decision would be to install a containment barrier to prevent breakage thereof and oil spill.

REFERENCES

ANP, Regulamento Técnico de Dutos Terrestres para Movimentação de Petróleo, Derivados E Gás Natural (RTDT) -2/2011.

Arduino, Página oficial da plataforma – disponível em http://arduino.cc – accessed May 21, 2013.

Avia, D., Diego, M., Oliver, G., Ortiz, A., Proenza, J., "RAO: A Low - Cost AUV for Testing", 2000, In: Proceedings of the MTS/IEEE Oceans'2000 Conference, p. 397- 401. Set

Bantz, C.R., 1995. Social dimensions of software development. In: J. A. Anderson, ed. Annual review of software management and development. Newbury Park, CA: Sage, 502–510.

Banzi, M., "Primeiros passos com arduino", book ,Vol, No 1.,Dezembro/2011, NovaTec.

Bruzzone, A.G., Longo, F., 2005. Modeling & Simulation applied to Security Systems. Proceedings of Summer Computer Simulation Conference, pp. 183-188. July 24-28, Philadelphia (Pennsylvania, USA).

Cardoso, L. C. S.- Logística do Petróleo - Transporte e Armazenamento - I.S.B.N.: 8571931011 – Ed. Érica, 142pag.

Carvalho, G. M. B.; Valério, M.; Medeiros, J. S. "Aplicação de técnicas de sensoriamento remoto e geoprocessamento na identificação da erosão dos solos na Bacia do rio Aracoiaba - CE". In: VII

Simpósio Brasileiro de Sensoriamento Remoto. Curitiba, 1993

Deitel, Paul And Deitel, Harvey – Java: como programar; tradução Edson Furmankiewicz – 8ª. Edição – São Paulo - Pearson Prentice Hall, 2010.

Ferasoli, H. F., Franchin M. N., Rillo M., "Robôs Móveis com Alto Grau de Autonomia para Inspeção de Tubulações", In: Proceedings IV SBAI , pp. 457-462, São Paulo,1999

Jornal O Estado de São Paulo, "ANP faz inspeção de segurança em dutos" – Disponível em: http://www.istoedinheiro.com.br/noticias/112097_ANP+FAZ+INSPECAO+DE+SEGURANCA+EM+DUTOS, accessed June 18, 2013.

Jornal A Tribuna.Com.Br – "ANP faz inspeção em dutos para garantir que regras sejam cumpridas" – disponível em http://www.atribuna.com.br/noticias.asp?idnoticia=182027&idDepartamento=8&idCategoria=0 , accessed April 18, 2013.

Luk, B., Cooke, D., Galt S., Collie A., Chen, S., Intelligent legged climbing service robot for remote maintenance applications in hazardous environments, In: Journal of Robotics and Autonomous Systems , 2003.

Mcroberts, M.et al, "Arduino Básico", book , Vol, No 1.,Setembro/2011, NovaTec.

Mercer, P.A. and Smith, G., 1993. Private view data in the UK. 2nd ed. London: Longman.

Rulik, O., *at al* – "GPRS Networks" - ISBN-10: 0470853174, Ed. John Wiley, 2003.

Sampaio, R. – Inspeção em Dutos – Apostila do curso de Tecnologia em manutenção Industrial – 2012.

SILVA, O. C., 2003, Petróleo: Noções sobre exploração perfuração, produção e microbiologia, Rio de Janeiro, Interciência.

Sverzut, J.U, Redes GSM, GPRS, EDGE e UMTS – Evolução da Terceira Geração (3G) , Editora Érica , 2005 – Ed Érica - ISBN-10: 8536500875, 452pag.

EVALUATION OF THE AGRIBUSINESS CHAIN OF PANELA PROCESS USING HIERARCHICAL ANALYSIS: CASE STUDY COLOMBIAN ANDEAN REGION

Gabriela Leguizamon [(a)], Nelson V Yepes Gonzalez [(b)], Maria V Cifuentes [(c)]

[(a)] Physical Engineer, Master in Industrial Engineering; professor of Industrial Engineering faculty of the Antonio Nariño University
[(b)] Industrial Engineer, Master in Design and Project Management, professor of Industrial Engineering faculty of the Antonio Nariño University
[(c)] Matematico, Magíster in Sciences-Mathematics, student of Industrial Engineering in Antonio Nariño University, professor in mathematics in the National University of Colombia

[(a)]gleguizamon@uan.edu.co, [(b)]neyepes@uan.edu.co, [(c)] macifuentes@uan.edu.co

ABSTRACT

The diferent sectors, which boost the developing of Colombian Economic, are searching to become competitive in globalized environment. The Panela sector is the second Colombian rural agro-industry, under Coffe Production, as well as a developing support for differents regions of the Country and its productive chain is characterized by its dynamism and by its various public and private actors. Achieving sector eficience and productivity involves to work with many variables and decision criteria that are submitted to uncertainly conditions. Analytic hierarchy process (AHP) is a technique for organizing and analyzing complex decisions, and it is used in order to figure out the main influential factors in competitiveness, which have to be in the focus of Panela agro-industry.

Keywords: Productive Chain, Panela (jaggery, gur), Analytic hierarchy process (AHP).

1. INTRODUCTION

Panela production is one of the most important farming activities for Colombian economy due to different reasons such as: its significative participation in Gross Domestic Product (GDP), which is 7,3% farming; another reasons are the big amount of land dedicated to cane cultivation (249.384 hectares), rural employment generation (about 25 million annual work part-time jobs and 120.000 permanent jobs – Osorio, 2004) and finally, because Panela production joins, approximately, 350.000 people that represents 12% of Colombian economically active rural population. To study factors getting involved in Panela agro-industry competitivity, is related to physical, economic and politic conditions, or even to production factors evolution (Zimmermann and Zeddies 2002).

There are three variables that guide analysis in competitive environment: in the first place, we have the normative environment of Panela sector, which allows to contrast among current applied politics with the purpose of encouraging Panela production; secondly, is organizational environment, whose function is to identify organization and integration mechanism in order to promove competitivity in the specific sector; and last but not least, the third variable is productive environment, which is crucial to determine economical and social relevance and competitivity level in the product production and marketing (Castellanos, et al 2010). The objective of this three-variable analysis is to provide elements that facilitate the study of the organizational and institutional environment where productive activities have place.

Eficience and productivity search in enterprises is promoting the implementation of supporting metodologies for decision making in industrial sectors and in Colombian regions, so that competitivity is boosted particularly, in scenarios with multiple variables or multiple selection criteria (Berumen and Llamazares 2007).

These precepts are the starting point of the studio, as well as identify the specific weight of each determinant sector factor; this situation implies the use of current methodologies to make decisions, such as multicriteria evaluation (MCE). This study has as objective to identify the primordial alternatives for enhancing sector competitivity through Analytic hierarchy process (AHP).

Analytic hierarchy process (AHP) is a multicriteria metodology for complex decisions making, and it was developed by Thomas L. Saaty (1977, 1980). AHP has been applied succesfully since its creation in many studies, as an useful and assistive instrument for strategic problem-solving. For instance:

1. In the evaluation of risk factors for farming (Toledo, Engler and Ahumada 2011).
2. In performance evaluation for archive management (Gomes 2012).
3. In increasing competitivity environments (Berumen 2007).
4. In the Delphi method used to measure projects complexity (Ludovic 2010).
5. In the management of intellectual capital assets, in particular, a TIC services industry aplication (Calabrese, Acosta and Menichini 2013).

6. In to determine intangible priority factors for technology transfer adoption (Lee, Kim, Min and Joo 2011)
7. In Corea competitivity as a developer of hydrogen-based energy technology (Seong, Yong Jand Jong 2007).
8. In the use of TRIZ and AHP to develop innovative design for automation of manufacture systems (Li 2009).
9. In multi-dimensional evaluation of oraganizational performance: integration of BSC and AHP (Veronese, Carneiro, Ferreira da Silva and Kimura 2012).

The AHP that is based on pairwise comparison, uses a hierarchical scale model for decision problem, which has a general objective, a group of alternatives and a group of useful criteria to link the identified alternatives with the goal (Vidal 2011).

For the developing of the current study, some referents were considered such as the prospective agenda of searching and technological development for productive Panela chain and its agro-industry in Colombia, the results of the annual panela poll 2012 (by Federación Nacional de Panaleros (Fedepanela)) and interviews with actors of the Chain: farming producers, processors and marketers. The studied population was one of the largest producers of Panela: Hoya del Río Suárez, a territory composed of 13 municipalities localized between Santander and Boyaca departments (Table 1).

Table 1: Hoya del Rio Suarez, Colombia –Incoder, 2012

Departamento	Municipio	Área (ha)	% Área
Boyaca	Chitaraque	14738	7,6
Boyaca	Moniquira	21075	10,9
Boyaca	San Jose de Pare	7348	3,8
Boyaca	Santana	6962	3,6
Boyaca	Togul	10807	5,6
Santander	Barbosa	4505	2,3
Santander	Chipata	9537	4,9
Santander	Guavata	7817	4,0
Santander	Guepsa	2769	1,4
Santander	Puente Nacional	25589	13,2
Santander	San Benito	5411	2,8
Santander	Suaita	27983	14,5
Santander	Velez	48655	25,2
Total general		193198	100,0

The profile physiographic of the Rural Development of Hoya del Rio Suaez consists of a structural denudative type Mountainous landscape, which covers a surface of 134 551 hectares, representing almost 70% of the area. The largest part of the lands of ADR, (82 404 hectares), corresponds to Class VI (they allow the development of certain annual crops under semi-intensive schemes) that are mainly concentrated in the municipalities of Moniquira, Santana, San José de Pare, Chitaraque and Toguí (Romero 2012). (Figure 1)

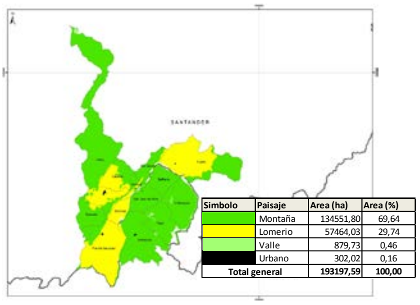


Figure 1. Physiography Hoya del Rio, Incoder 2012

In Hoya del Río Suárez, Bocadillo and Panela production represents the main economical activity in the región. That activity involves 13 municipalities that together make the sustainability font to many families for more than three generations. Currently, there are 128 Bocadillo Factories and more than 1276 sugar mills (for Panela production) that add value to the 14.000 Guayaba hectares and to the 46.000 Cane hectares, respectively (Gómez 2011) .

The AHP model processing for Panela productive chain, is made through hierarchical model structuring, which identifies the goal to accomplish, criteria, sub-criteria and alternatives; lately, it priorizes hierarchical process elements and makes the binary comparisons among the elements using weights assignment, in that way the AHP sets the ranking of alternatives according to the weights assigned. Finally, a summary of results is elaborated with a sensibility analysis to determine the inconsistency index of the established model. To develop the mentioned process, the used tool is Expert Choice 8.0 (EC), which is a program useful to eliminate conjectures in decisions making, is based on AHP as well as uses a hierarchy to organize thought and intuition in a logical way. This hierarchical approach allows that the user analyzes all options in order to get an effective decision-making. The EC program may compare tangible with the intangible factors, for instance, “Project costs” opposite to “Project viability”, besides tolerating uncertainty and allowing review so as to individuals and groups are able to address the problem with all their concerns (Expert Choice 1993).

The paper is organized following the next scheme: First section contains AHP model justification of applying it to Panela sector competitivity; the second one has the AHP model application since the conformation of hierarchical structure until sensibility analyisis and numeric results are discussed.

Finally, conclusions are submmitted and future research is considered.

2. JUSTIFICATION

With the opening of international markets, industry sectors and regions are required to be competitive and innovative. Competitiveness refers to enterprises capacity to compete and, base don its succes, to earn market share, to increse its benefits and to grow. (Berumen 2007). But such a competitiveness is not achieved without **economic sustainability** that refers to tha sector ability to generate income based on comparative and competitive advantages of products;

social sustainability, which refers to that income generated by Panela sector might be enough to guarantee an adecquate life style for producers; and finally, **environmental sustainability** means that agricultural activity should preserve environment (Leibovich 2009).

Using analysis hierarchical process (AHP) aims to help establish priorities for decision making. Besides, ranking strategic issues, assignating budgets for urgent situations, building farsightedness and managing complex projects are essential for Panela guild capabilities developing. The AHP allows to guide the key factors priorization of competitiveness, and in that way to accomplish economic, social and environmental sustainability.

One important advantage of the AHP use is its ease. It can work with processes uncer uncertainty and with subjective information, because the AHP gives priority to the criteria that are based on experience and on intuition in a logical way. Perhaps, the AHP most important advantage is in the developing of the hierarchy itself, which compels decider to considerate consciously and to justify criteria pertinancy (Nydick 1992). In conclussion, the AHP, by Thomas Saaty, is a powerful tool employed in decisión making when multiple purposes affects decision.

5. The AHP and ANP application

5.1. Hierarchical Structure

Method has four stages: a problema presentation, a criteria and alternatives evaluation through estimating the inconsistency index of the model, after an alternatives evaluation is done and lately, alternatives are hierarchized. Hierarchy is not only structurally efficient because it allows to represent a system, but also functionally, as soon as it is useful to control and transmit information via the System (Eraslan and Dağdeviren 2010).

Firstlty, it is identified the goal wanted by the decision making hierarchical model, "Prioritize the best decision alternatives for strengthening the competitiveness of Panela sector".

Secondly, the decision criteria and subcriteria that will allow to get the competitiviness of Panela agro-industry chain are identified and by using GO-CART method (Hernandez 2006) it was allowed to plan the serarch of secondary external data from previous studies obtained by experts, public and prívate entities, studies such as: agro-industrial value chains (Bisang, Anlló, Campi and Albornoz 2009); research tendencies, technologic developing and marketing in Panela agribusiness (Castellanos, Torres and Flórez 2010); business clusters zoning and organization for Panela cane (Abaunza 2012); the microeconomics of competition of Cane cluster in Colombia (Dueñas, Morales, Nanning, Noriega and Ortiz 2007); the basis of the agreement of Panela agro-industry chain development (IICA 2001); Panela producer profitability affectation because of the implementation of environmental and health standards (Llano 2012); sugar cane competitiveness: a Kenana Sugar Company study, Sudan (Emam 2010); Becoming enterprise of Panela sector, as a developing of productivity and competitiveness factor (Perez 2011); agroindustrial Panela chain in Colombia: A global insight of its structure and dynamics 1991-2005 (Martinez 2005).

Undoubtedly, it is evident that many issues affect competitiveness, so it is necessary to consider a wide amount of variables and indicators. The Institute for Management Development - IMD uses about 331 criteria organized in four main classes: economic performance, governmental eficience, enterprises eficience and infraestructure (Lopez 2009). For our purposes, decisional analysis allows teh identification of six fundamental criteria: Economic, Productive, Logisctic, Environmental, Social and Marketing, where three actors are participacting: (I) Producers (Cane Farmers / wholesalers and retailers); (II) Processsors (Big, medium and small mills); (III) Marketers (Big, medium and small marketers). Their performances depend on each one interests.

The competitiviness subcriteria in economic, marketing, productive, logistical, environmental and social aspect, for each actor, are:

1. Producers subcriteria. In economic issue: Financing to develop agricultural activities, developing of technology transference to implement new productive processes in Panela Cane production, productive planification as an important element in Panela chain to avoid product oversupply. In productive issue: Enhancing skilled labor capacity to agricultural activities, decrease in harvesting costs and crop renovation. In Marketing: expansion of marketing channels and improving market prices. In Logistical issue: enhancing distribution channels and improving transport systems since cultivation as far as processing zone or mill. In environmental issue: hygienic and health conditions according to regulations and to a sustaintable, organic and not environmental-alterative agricultural developing. In social issue: developing of skilled labor, trade union and an effective fraudulent practices control (the dishonest use of sugar cane for Panela production, when Sugar price drops).
2. Processors Cane (Mills): In economic issue Financing to develop agricultural activities, developing of technology transference and productivity increase; In Marketing: Expansion in productos diversification, decrease in production costs and enhancement in marketing margins, diversification of productive activities; in producrive issue: Mills

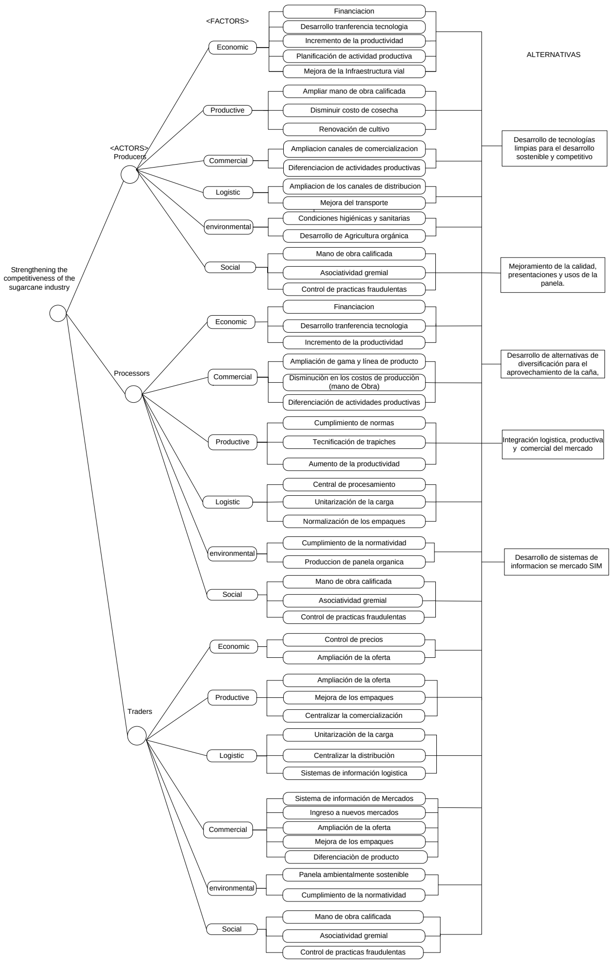
modernization, productivity increase and regulation fulfillment; in Logistic issue: presentation of the product in units, packaging standardization and processes centralization; in Ambiental one: environmental regulation fulfillment and organic Panela production. In social one: to fulfill with the same requirements that producers.

3. Competitiviness subcriteria defined for marketers are: in Economic issue: Prices control and expansion of exportation offer; in Marketing one: Product Differentiation, packaging improvement, Offer and Marketing channels extension, Marketing Information Systems for Panela productive chain; in Productive issue: Offer growth, packaging improvement and marketing centralization; in Logistic issue: Logistic information systems improvement, product availability in individual units and distribuition centralized In. Environmental one: Regultion Fulfillment and environmentally sustainable products marketing.

Once we defined subcriteria, let's determinate strategic developing alternatives that have been made by competitiveness agenda and technologic development of the Chainy:

1. Development of clean technologies for susteineable and competitive sector growing (DCT)
2. Quality presentations and Panela uses improvement (QPP)..
3. Diversification alternatives developing to take advantage of Panela cane (DAD)
4. Logistical and comercial integration integration of Market (LCI)
5. Marketing Information systems development (MIS).

In that way, hierarchical net model is structured having in count interactions of those elements that affect Panela supply chain.



<FACTORS>
ALTERNATIVAS
Strengthening the competitiveness of the sugarcane industry
<ACTORS> Producers
Processors
Traders
Economic
Financiacion
Desarrollo tranferencia tecnologia
Incremento de la productividad
Planificación de actividad productiva
Mejora de la Infraestructura vial
Productive
Ampliar mano de obra calificada
Disminuir costo de cosecha
Renovación de cultivo
Commercial
Ampliacion canales de comercializacion
Diferenciacion de actividades productivas
Logistic
Ampliacion de los canales de distribucion
Mejora del transporte
environmental
Condiciones higiénicas y sanitarias
Desarrollo de Agricultura orgánica
Social
Mano de obra calificada
Asociatividad gremial
Control de practicas fraudulentas
Economic
Financiacion
Desarrollo tranferencia tecnologia
Incremento de la productividad
Commercial
Ampliación de gama y línea de producto
Disminuciòn en los costos de producciòn (mano de Obra)
Diferenciación de actividades productivas
Productive
Cumplimiento de normas
Tecnificación de trapiches
Aumento de la productividad
Logistic
Central de procesamiento
Unitarización de la carga
Normalización de los empaques
environmental
Cumplimiento de la normatividad
Produccion de panela organica
Social
Mano de obra calificada
Asociatividad gremial
Control de practicas fraudulentas
Economic
Control de precios
Ampliación de la oferta
Productive
Ampliación de la oferta
Mejora de los empaques
Centralizar la comercialización
Logistic
Unitarizaciòn de la carga
Centralizar la distribuciòn
Sistemas de información logistica
Commercial
Sistema de información de Mercados
Ingreso a nuevos mercados
Ampliación de la oferta
Mejora de los empaques
Diferenciaciòn de producto
environmental
Panela ambientalmente sostenible
Cumplimiento de la normatividad
Social
Mano de obra calificada
Asociatividad gremial
Control de practicas fraudulentas
Desarrollo de tecnologías limpias para el desarrollo sostenible y competitivo
Mejoramiento de la calidad, presentaciones y usos de la panela.
Desarrollo de alternativas de diversificación para el aprovechamiento de la caña,
Integración logistica, productiva y comercial del mercado
Desarrollo de sistemas de informacion se mercado SIM

Figure 1: Red hierarchical process

In table 2, it is represneted decision subcriteria and its importance in productive chain structure:

Table 2. Subcriteria or essential factors and its importance.

Essential Factors	Relevance
Financial	Profitability and economic sustainability affectation of the chain, to log in credits with productive purposes. An efficient credit and to al lis primordial in sector competitiveness recuperation perspective.
Reduction in production costs	Inputs supply affects in production cost as well as in the need of promoving its appropiate and rational use , so that it is an important part of structural elements to arrange among Panela producers.
Product adulteration	Pressure by sugar melter people "Derretideros" over Panela supply in domestic market.
Productive activities differentiation	Diversification in product (panela) use, in another sectors such as cosmetics, enegy drinks, powdered panela or bioethanol.
Product Variability	Panela producers posibility of having another alternatives to process their cane.

Essential Factors	Relevance
Packaging improvement	To improve primary and secondary packaging systems in final product until it arrives to consumer.
Qualified labor	Shortage of skilled labor for harvesting and post-harvesting activities.
Marketing, transporting and distribution channels extension	Opening of marketing channels, logistical corridors and distribution systems that minimize product brokering.
Logistical integration	To integrate chain actors in market, production and logistics aspects.
Road infraestructure	Improving primary, secondary and terciary road fence for product transporting. Development of logidtical corridors.
Centralizing distribution	The development of comercialization and distribution centers with efficient transporting, storage and product-trazability operations systems.
Price control	Reference price definition and control based on econometric studies.
Organic Panela production	The development and promotion of organic powdered panela production to access international markets (Korea)
Crop renovation	The decision to renew crops requires important aspects that must be presented: 1. To Analyze crops age for strains renew if they are very old, because it causes crop production goes down in each new sowing. 2. To make the decision to renew crops taking into account the resistance of certain varieties in the country to diseases such as Coal, Ustilago scitaminea and Roya, Puccinia melanocephal.
Marketing Information Systems	Marketing Information systems that allow the Access to all the chain actors, like "a set, whose components are interrelated that meets (or obtain), process, store and distribute information to support decision making and organization control".
Technological transference	To transfer knowledge and technologies for the entire production chain.
Regulations Fulfillment	Mills adequacy is important not only for improving quality and acceptance of the product, but also because the 779 resolution, which was published in 2006 by the Ministry of Social Protection, stablished technical regulations that health requirements. Aditionally, Panela Mills must be certified in Good Manufacturing Practices and they must be enrolled in INVIMA.
	Panela Mills are compeled to accomplish an adequate water, air and another renewable resources management according to 1594 of 1984, 1791 of 1996, 948 of 1995 decrees. To give continuity and forcefulness to polluting practices eradication, such as the use of wood, tires and chemicals that threaten human health.
Associativity	Producers are scattered in almost all Colombian Regions, they dont work together and they have mostly had unfavorable prices in comparison with production costs, which arise from inefficient production and marketing systems.

One objection received in this regard is from Johnson (1979) who noted that if the hierarchy is incomplete, the weights can be distorted, therefore Epstein and King (1982) incorporated the possibility of structuring decision process through a hierarchy and the differences of information on each level of the hierarchy, should be represented by introducing distortions in the valuations of its elements. According to Saaty, the problem is the availability of information, not the method (Zahedi 1986).

.

5.3. Priority Establishment

The AHP methodology implies that decider has to indicate his preference or priority for each decision alternative. Given the information about the relative importance and preferences, it is used the mathematical process called synthesis for summarizing information and providing a priority hierarchy of alternatives according to a global preference that is built since weights and initial criteria given by decider. The AHP uses a scale of 1-9 to rate the relative preferences of both elements.

To fill the matrix, you must first understand the meaning of each value Saaty scale presented in Table 3:

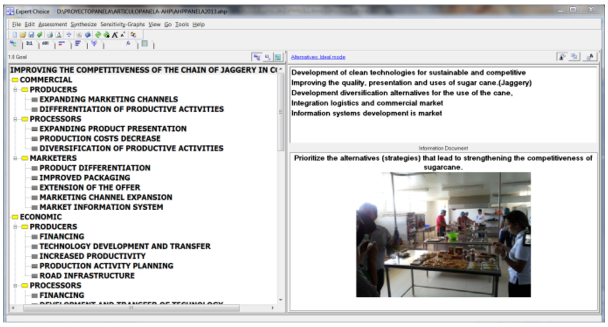
Table 3. Saaty scale

Importance/ preference	Intensity	Meaning
1	**Equally Important** Equal of different to...	Comparing two elements, and there is no difference between them. The Decider doesn't prefer any of them.
3	**Moderate Importance** Slightly more important or more prefered than...	Comparing two elements, first one is slightly more important or more prefered than the other.
5	**Strong Importance** More important or more prefered than...	Comparing two elements, first one is considered more important or more prefered thant the second one.

Importance/ preference	Intensity	Meaning
7	**Very Strong Importance** Much more important or much more prefered than...	Comparing two elements, first one is considered much more important or much more prefered that the second one.
9	**Extreme importance** of one element. Absolutely much more important or prefered than…	Comparing two elements, first one is absolutely much more important or prefered than the other.

The scale established by Mr. Thomas L. Saaty, was the development of studies based on experimental activities and it uses a scale with nine elements, in where different grades or levels are showed and which allow to discriminate ***relation intensity*** among the elements belonging to a set. In that way, comparisons and measurements are achieved, so technique is initially adjusted to the ***Homogenization Principle of Measure Theory***, particularly, when working variables or factors of great variability and diversity in the study that is being conducted.

In the first instance, it should be to accomplish an assessment of criteria importance in relation to their contribution to the achievement of the goal, then, for each criterion might determine what is the relative importance of the attributes that depend on it. The assessment process should continue with the appreciation of the importance of the alternatives respect to each of existing and valued attributes. Frequent use of AHP-Expert Choice software allows hierarchical model estructuration, through the goal, criteria, subcriteria and decision alternatives introduction (Figure 2).


Expert Choice D:\PROYECTOPANELA\ARTICULOPANELA-AHP\AHPPANELA2013.ahp
File Edit Assessment Synthesize Sensitivity-Graphs View Go Tools Help
1.0 Goal
Alternatives: Ideal mode
Information Document

IMPROVING THE COMPETITIVENESS OF THE CHAIN OF JAGGERY IN C(
- COMMERCIAL
  - PRODUCERS
    - EXPANDING MARKETING CHANNELS
    - DIFFERENTIATION OF PRODUCTIVE ACTIVITIES
  - PROCESSORS
    - EXPANDING PRODUCT PRESENTATION
    - PRODUCTION COSTS DECREASE
    - DIVERSIFICATION OF PRODUCTIVE ACTIVITIES
  - MARKETERS
    - PRODUCT DIFFERENTIATION
    - IMPROVED PACKAGING
    - EXTENSION OF THE OFFER
    - MARKETING CHANNEL EXPANSION
    - MARKET INFORMATION SYSTEM
- ECONOMIC
  - PRODUCERS
    - FINANCING
    - TECHNOLOGY DEVELOPMENT AND TRANSFER
    - INCREASED PRODUCTIVITY
    - PRODUCTION ACTIVITY PLANNING
    - ROAD INFRASTRUCTURE
  - PROCESSORS
    - FINANCING

Development of clean technologies for sustainable and competitive
Improving the quality, presentation and uses of sugar cane.(Jaggery)
Development diversification alternatives for the use of the cane,
Integration logistics and commercial market
Information systems development is market

Prioritize the alternatives (strategies) that lead to strengthening the competitiveness of sugarcane.

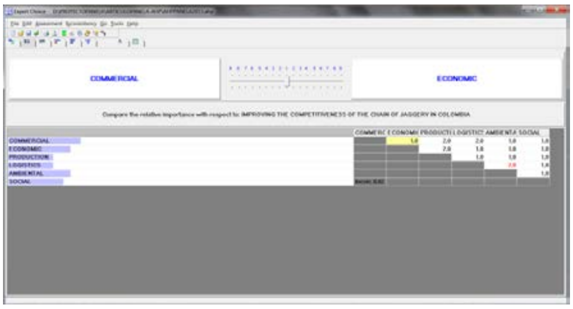
Figure 2.Hierarchical Model Introdution.Expert Choice Software.

Comparison matrix between criteria, uses Saaty scale base don parwise comparison. So it prioritizes criteria with respect to the goal.

Let Z be an matric n x n and Pij the element located in the position (ij) inside of Z, for i=1.2.3…n and j=1.2.3…..n, Z is a parwise comparison matrix of n alternatives.If Pij is the preference grade of alternative i against alterntaive j, when i=j, we will have Pij=1 due to the fact the alternative is compared against itself.

$$Z = \begin{bmatrix} 1 & P_{12} & \dots & P_{1n} \\ P_{21} & 1 & \dots & P_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ P_{n1} & P_{n2} & \dots & 1 \end{bmatrix} \qquad (1)$$

In this case, in first level, comercial, economic, productiv, logistical, environmental and social criteria were compared parwise (Pij) and it resulted that all the criteria has the same relevance, except commercial and economic ones are a little more important over productive criterion, as well as environmental criterion has a greater weighting than logistical one.



Compare the relative importance with respect to: IMPROVING THE COMPETITIVENESS OF THE CHAIN OF JAGGERY IN COLOMBIA

	COMMERC	ECONOMI	PRODUCTI	LOGISTIC	AMBIENTA	SOCIAL
COMMERCIAL		1.0	2.0	2.0	1.0	1.0
ECONOMIC			2.0	1.0	1.0	1.0
PRODUCTION				1.0	1.0	1.0
LOGISTICS					2.0	1.0
AMBIENTAL						1.0
SOCIAL	[illegible]					

Figure 3. First Level Criteria comparison

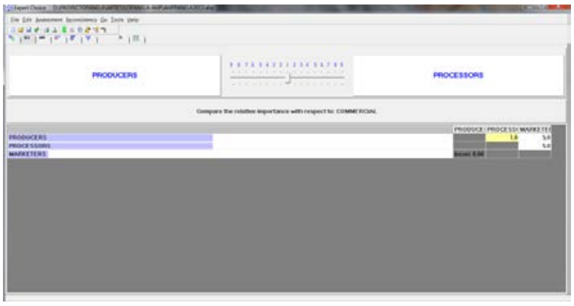
Comercial, economic and productive criteria are the most important issues inside the supply chain, having in account current competitiveness conditions. With prices and costs established for 2010 in regions and the required investment to implement health and environmental regulations, Panela production in sugar mills comes across as financially non-viable (Llano, Duarte and Moreno 2012), so environmental criteria weight will be less within the chain than the other ones.

Following with comparison matrix, another important property is that Pij.aij=1, that is to say:

$$Z = \begin{bmatrix} 1 & P_{12} & \dots & P_{1n} \\ 1/P_{12} & 1 & \dots & P_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1/P_{1n} & 1/P_{2n} & \dots & 1 \end{bmatrix} \qquad (2)$$

The last property is due to Reciproc judgments axiom: *If Z is a parwise comparison matrix, $P_{ij}=1/P_{ji}$. Where compared elements of the same level and that have hierarchical dependence.*

Hierarchical second level consists of comparing chain actors (producers. processers and marketers) with respect to economic, commercial, productive, logistical, environmental and social criteria. The weight of each actor over criteria has been established asreciproc relation, so importance/preference has been estimated as (1), without preference for any actor. For the study, everyone will have equal hierarchical importance.



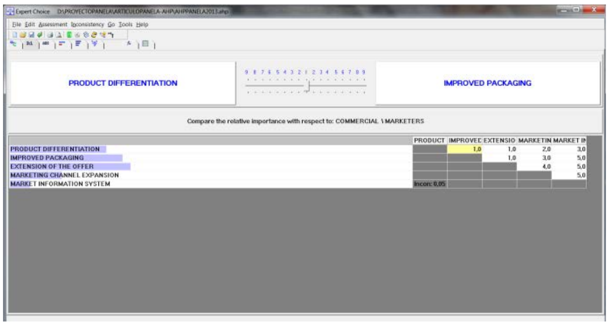
PRODUCERS — PROCESSORS

Compare the relative importance with respect to: COMMERCIAL

	PRODUCE	PROCESSI	MARKETE
PRODUCERS		1,0	5,0
PROCESSORS			5,0
MARKETERS	Incon: 0,00		

Figure 4.Parwise comparison between actors for factors

Third hierarchical level consists in comparing subfactors for each actor, with regard to criteria (commercial, economic, productive, logistical, environmental and social). As an example, it is the comparison and priorization of subfactors (product differentiation, packing improvement, product offer enhancement, marketing channels expansion and marketing formation systems implementation) for marketers in commercial competititveness structure.


Expert Choice D:\PROYECTOPANELA\ARTICULOPANELA-AHP\AHPPANELA2013.ahp
File Edit Assessment Inconsistency Go Tools Help
9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9

PRODUCT DIFFERENTIATION — IMPROVED PACKAGING

Compare the relative importance with respect to: COMMERCIAL \ MARKETERS

	PRODUCT	IMPROVEI	EXTENSIO	MARKETIN	MARKET IN
PRODUCT DIFFERENTIATION		1,0	1,0	2,0	3,0
IMPROVED PACKAGING			1,0	3,0	5,0
EXTENSION OF THE OFFER				4,0	5,0
MARKETING CHANNEL EXPANSION					5,0
MARKET INFORMATION SYSTEM	Incon: 0,05				

Figure 5.Parwise comparison between actors and subfactors, for each competitiveness factor.

5.4. Priorization of Factors

Once comparison matrix is already filled out, priorities might be calculated. Traditional AHP uses the eigen-valor method. Let´s considere a coherent matrix with the know priorities p_i. So that comparison between alternatives i and j is given through p_i / p_j, which multiplied for priority vector p, produces:

$$\begin{array}{|c|c|c|c|} \hline P_1/p_1 & P_1/p_2 & \dots & P/P_n \\ \hline P_2/P_1 & P_2/P_2 & \dots & P_2/P_n \\ \hline \dots. & \dots. & \dots & \dots\dots. \\ \hline P_n/P_1 & P_n/P_2 & \dots & P_n/P_n \\ \hline \end{array} = n \begin{bmatrix} P1 \\ P2 \\ \dots \\ Pn \end{bmatrix} \qquad (3)$$

If matrix is enougly consistent, the transitivity rule (4) is satisfied for all the comparisons Pij

$$Pij = Pik.Pkj \qquad (4)$$

Where P identifies priorities vector; in corresponds to matrix Z dimensión.

Then in (3) it is an eigenvector problem. Priorities calculation is exact for a consistent matrix (5).

$$Zp = np \qquad (5)$$

The aggregation of the results of pairwise comparison to make the prioritization of the factors. are defined in Table 4. by reference to the structure Zangeneh study results (2009). It was determined that the most important criterion for experts (0.206), is commercial one. This fact is reflected in the present time of agribusiness, where producer, processer and marketer concern is focused on product position in the Market. As a strategy it is presented the integration of Panela cooperatives or associations that can compete with new Panela producers (Delgado 2009). Another justification that provides a higher score to commercial criteria, it´s presented with the need to streng then sector sustainability from a suitable commercial management, which looks for solving two problems: the domestic price of food decrease and the diminution of domestic consumption (Rios 2013).

In the second comparison level, the earmaked weight for each actor (productors, processors and marketers) is the same (0.333): Producers could not have a greater preponderance than neither processers nor marketers. Currently, the strategy has to be a kind of “wins-wins” for every actors involved in the chain.

In the third hierarchical level, subcriteria weights are compared with regard to the actors. We could say that **productive activities differentiation** (0.833) for producers (0.333) will be more significative in **commercial issue** (0.206) in order to develop and deepen research to get new applications for the use of sugarcane different to Panela production. As an example, it is the study on the alternative use of molasses sugar cane waste to synthesize rigid polyurethane foam (ERP) for industry (Vega, Delgado, Sibaja and Alvarado 2007), or production of feed for animal breeding or dairy production agribusiness diversification into industry sucrochemistry seems to be a very interesting option to face oil depletion (Viniegra 2007). In the current conditions, production costs decrease for processors (0.674) would have greater importance in commercial aspect (0.206) in order to expand the marketing marginsof the product. The average cost of producing a kilo of panela, in the 26 departments and 350 municipalities producers. including the social and placed in the hold of the mill is $ 2200 Kilo (Colombian Pounds - COP). Someplacessold it at below cost, e.g. in Cundinamarca. Nariño and Cauca, today the price is about $ 1400 a kilo (COP) (Ramirez 2013). On the other hand, **supply expansion** (0.304) will be more important for **marketers**, for being able to enter new international market niches, preferably, with special emphasis on sugarcane products that have global merchandise: companies involved in Marketing, presentations and driven prices, imports and exports. guidance on trends, new products made from the juice sugarcane as well as Panela, characterization and potential markets for the principal panela products nationally and internationally (Castellanos, Torres and Flórez 2010).

In the second hierarchical level, economic criterion has the secod place (0.184), where producers and processers prioritize financing schemes (0.473 y 0.528). One option for applying suitably health and enironmental regulations in Panela infrastructure could be a subsidized credit, as a special line of ICR (*Incentivo de capitalización rural* – Rural capitalization incentive), however the amount of resources to guarantee a complete re-adequacy is at least $0.6 billions (COP), assuming an incentive of 40% (Llano, Duarte and Moreno 2012). For marketers in the economic issue, it is principal Price control to avoid its destabilization with in productive chain generated by the product oversupply and fraudulent practices. First of all, Panela and sugar are competitive or substitutable goods in both production and consumption, beacuse they come from the same plant species (cane) and being both daily sweeteners.At the second level of hierarchy is environmental criteria (0.184), where producers gives equal importance for **maintaining a hygienic and basic sanitation system in crops** (0.500) and for **developing an organic agriculture** (0500). for the processor sub-criterion of highest priority is compliance with environmental regulations (0.857) and the commercializador. the sale of sustainable organic panela with the environment (0.875); for processers, the most important subcriterion is the **fulfillment of the environmental regulations** (0.857) as far as for marketers it is the **environmentally Sustainable sale of organic Panela** (0.875). Adopting good practice is the operative baseline of the different mills, in as much as: it involves the application of many different processes in order to avoid consumer health risks, it reduces costs generated by a poor quality caused by mishandling, and finally, good practice increase customer satisfaction that results in increased sales (Guerrero and Luengas 2011).The environmental Panela guide becomes a reference tool and basic orientation that contains the methodological and general procedures panelera development activity,under a integrated environmental management approach (Fonseca 2002). The third level of hierarchy is occupied by social criteria (0.163) where sub-factors such as the development of skilled labor for the entire chain, the union associativity and the control of fraudulent practices have all the same weight (0.333).To develop skilled Manpower, the sugar cane sector board that is located in Villeta – Cundinamarca, has contributed to the development of occupational competency standards and qualifications.

Productive criterion is the fourth within the hierarchy (0.131), where for producers is fundamental the renewal of crops (0.577), for processors is to increase productivity (0.731) and for marketers, to centralize Marketing (0.758). The national government is promoting the initiative to develop a central of cane juices for the Hoya del Rio Suarez where should be concentrated all the cane production from the region and in that way to eliminate intermediation. On this same level, it is thelogistical criteria (0.131) where for producers the most importance focuses on expanding Commercialization channels (0.833), while for processors will be to centralize the process (0.726) and for marketers, to have a market information system, which were safe, available and reliable (0.709).

Table 4. Weights for criteria and subfactors or specific criteria for each actor

Criterion	Weight	Actors	Weight	Specific criteria for each actor	weight	Overall = (1) x (2) x (3)
Level (1)		**Level (2)**		**Level (3)**		
1. Commercial	0.206	1.1.Producers	0.333	1.1.1. Marketing channels expansion	0.167	0.011
				1.1.2. Productive activities differentiation	0.833	0.057
		1.2. Processors	0.333	1.2.1. Expantion in the product presentation	0.226	0.016
				1.2.2. Production cost´s decrease	0.674	0.046
				1.2.3. Productive activities diversification	0.101	0.007
		1.3.Marketers	0.333	1.3.1.Product Differentiation	0.235	0.016
				1.3.2.Improved packaging	0.280	0.019
				1.3.3.Offert´s extent	0.304	0.021
				1.3.4. Marketing channels expansion	0.129	0.009
				1.3.5. Market information system	0.052	0.004
2. Económic	0.184	2.1.Producers	0.333	2.1.1.Financing	0.473	0.029
				2.1.2.Development and Technology´s transfer	0.124	0.008
				2.1.3.Productivity increase	0.267	0.016
				2.1.4.Production planning	0.041	0.003
				2.1.5.Infrastructure line	0.095	0.006
		2.2.Processsors	0.333	2.2.1.Financing	0.528	0.032
				2.2.2. Development and Technology´s transfer	0.333	0.020

				2.2.3.Productivity increase	0.140	0.003
		2.3.Marketers	0.333	2.3.1.Price control	0.857	0.053
				2.3.2.Offert´s extent	0.143	0.009
3. Productive	0.131	3.1.Producers	0.333	3.1.1.Increase labor quality	0.081	0.0S04
				3.1.2.Reductionharvestingcosts	0.342	0.015
				3.1.3.Renewal of crops	0.577	0.025
		3.2.Processors	0.333	3.2.1.Factory´s upgrade	0.188	0.008
				3.2.2.Regulatory compliance	0.081	0.004
				3.2.3.Productivity increase	0.731	0.032
		3.3.Marketers	0.333	3.3.1.Offert´s extent	0.091	0.004
				3.3.2. Improved packaging	0.151	0.007
				3.3.3.Centralizing marketing	0.758	0.033
4.Logistic	0.131	4.1.Producers	0.333	4.1.1.Distribution channels expansion	0.833	0.036
				4.1.2.Improving transportation systems	0.167	0.007
		4.2.Processors	0.333	4.2.1.Centralize processing	0.726	0.032
				4.2.2.Burden unitization	0.102	0.004
				4.2.3.Packing standardization	0.172	0.008
		4.3.Marketers	0.333	4.3.1. Market information system	0.709	0.031
				4.3.2. Burden unitization	0.179	0.008
				4.3.3.Centralize distribution	0.113	0.005
5.Ambiental	0.184	5.1.Producers	0.333	5.1.1.Basic hygiene and sanitation	0.500	0.031
				5.1.2.Organic agriculture	0.500	0.031
		5.2.Processors	0.333	5.2.1.Regulatory Compliance	0.857	0.053
				5.2.2. Production and export of organic panela	0.143	0.053
		5.3.Marketers	0.333	5.3.1. Sustainable organic panela	0.875	0.054
				5.3.2. Regulatory Compliance	0.125	0.008
6. Social	0.163			6.1. labor quality	0.333	0.054
				6.2. Associativity trade-union	0.333	0.054
				6.3. Control fraudulent practices	0.333	0.054

5.5.Synthesis

The last step is to synthesize the local priority of each criterion for determining the global priority. The historical approach AHP (called late distributive) adopts an additive aggregation (6) through a weighted sum of the priorities (Table 5), let’s see the formulation:

$$Pi = \sum_{j=1}^{m} Wj \,.\, Lij \quad ; 1 \leq i \leq n \tag{6}$$

Where n is the amount of alternatives, m is the amount of criteria, Pi is the global priority of alternative i, Lij is the local priority of alternative i with regard to criterian j and Wj is the j-criterion weight.

Table 5. Summary of results

Pi Level	Criteria and Sub-Criteria	*Alternatives Lij*					*Pi*
		DCT	QPP	DAD	LCI	MIS	
2	**1. Commercial**	0.194	0.175	0.207	0.364	0.060	0.206
3	1.1.Producers	0.154	0.124	0.346	0.332	0.044	0.069
4	1.1.1. Marketing channels expansion	0.300	0.111	0.293	0.221	0.076	0.011
4	1.1.2. Productive activities differentiation	0.125	0.127	0.356	0.354	0.038	0.057

3	1.2.Processors	0.261	0.195	0.101	0.392	0.052	0.069
4	1.2.1.Expansion in the product presentation	0.135	0.427	0.069	0.314	0.054	0.015
4	1.2.2. Production cost´s decrease	0.307	0.115	0.115	0.413	0.050	0.046
4	1.2.3.Productive activities diversification	0.233	0.207	0.079	0.426	0.055	0.007
3	1.3.Marketers	0.059	0.334	0.042	0.387	0.178	0.069
4	1.3.1.Product Differentiation	0.050	0.189	0.056	0.531	0.174	0.016
4	1.3.2.Improved packaging	0.042	0.591	0.042	0.183	0.142	0.019
4	1.3.3.Offert´s extent	0.074	0.268	0.034	0.451	0.174	0.021
4	1.3.4. Marketing channels expansion	0.070	0.276	0.038	0.474	0.142	0.009
4	1.3.5. Market information system	0.077	0.145	0.034	0.244	0.500	0.004
2	**2. Economics**	0.203	0.193	0.285	0.202	0.117	0.184
3	2.1.Producers	0.207	0.111	0.453	0.180	0.049	0.109
4	2.1.1.Financing	0.198	0.108	0.451	0.180	0.063	0.052
4	2.1.2.Development and Technology´s transfer	0.429	0.250	0.154	0.130	0.036	0.014
4	2.1.3.Productivity increase	0.132	0.075	0.561	0.192	0.040	0.029
4	2.1.4.Production planning	0.268	0.087	0.570	0.050	0.026	0.004
4	2.1.5.Infrastructure line	0.144	0.060	0.498	0.265	0.033	0.010
3	2.2.Processors	0.233	0.454	0.040	0.210	0.064	0.046
4	2.2.1.Financing	0.234	0.505	0.042	0.158	0.060	0.024
4	2.2.2. Development and Technology´s transfer	0.139	0.505	0.035	0.254	0.067	0.015
4	2.2.3.Productivity increase	0.449	0.139	0.039	0.303	0.071	0.006
3	2.3.Marketers	0.142	0.087	0.038	0.274	0.460	0.029
4	2.3.1.Price control	0.145	0.056	0.038	0.233	0.528	0.025
4	2.3.2.Offert´s extent	0.124	0.271	0.037	0.517	0.051	0.004
2	**3. Productive**	0.172	0.247	0.190	0.280	0.111	0.131
3	3.1.Producers	0.278	0.139	0.485	0.058	0.040	0.044
4	3.1.1.Increase labor quality	0.248	0.133	0.517	0.067	0.035	0.004
4	3.1.2.Reductionharvestingcosts	0.320	0.125	0.445	0.068	0.042	0.015
4	3.1.3.Renewal of crops	0.258	0.148	0.504	0.051	0.039	0.025
3	3.2.Processors	0.148	0.386	0.048	0.348	0.071	0.044
4	3.2.1.Factory´s upgrade	0.131	0.272	0.051	0.522	0.024	0.008
4	3.2.2.Regulatory compliance	0.481	0.300	0.066	0.096	0.057	0.004
4	3.2.3.Productivity increase	0.115	0.425	0.045	0.331	0.084	0.032
3	3.3.Marketers	0.091	0.216	0.037	0.434	0.222	0.044
4	3.3.1.Offert´s extent	0.152	0.495	0.043	0.235	0.075	0.004
4	3.3.2. Improved packaging	0.049	0.506	0.047	0.262	0.136	0.007
4	3.3.3.Centralizing marketing	0.092	0.125	0.034	0.492	0.257	0.033
2	**4.Logistic**	0.113	0.246	0.066	0.361	0.214	0.131
3	4.1.Producers	0.152	0.221	0.087	0.495	0.045	0.044
4	4.1.1.Distribution channels expansion	0.155	0.224	0.079	0.499	0.043	0.036
4	4.1.2.Improving transportation systems	0.138	0.206	0.127	0.471	0.057	0.007
3	4.2.Processors	0.101	0.343	0.047	0.388	0.122	0.044

4	4.2.1.Centralize processing	0.107	0.371	0.045	0.342	0.135	0.032
4	4.2.2.Burden unitization	0.111	0.278	0.042	0.529	0.040	0.004
4	4.2.3.Packing standardization	0.070	0.259	0.057	0.497	0.116	0.008
3	4.3.Marketers	0.085	0.175	0.066	0.200	0.474	0.044
4	4.3.1. Market information system	0.085	0.175	0.066	0.200	0.474	0.031
4	4.3.2. Burden unitization	0.094	0.338	0.076	0.320	0.171	0.008
4	4.3.3.Centralize distribution	0.079	0.094	0.094	0.612	0.120	0.005
2	**5. Environmental**	0.405	0.286	0.171	0.080	0.055	0.184
3	5.1.Producers	0.383	0.123	0.397	0.049	0.047	0.061
4	5.1.1.Basic hygiene and sanitation	0.359	0.072	0.464	0.057	0.047	0.031
4	5.1.2.Organic agriculture	0.407	0.174	0.330	0.042	0.047	0.031
3	5.2.Processors	0.511	0.264	0.049	0.122	0.054	0.061
4	5.2.1.Regulatory Compliance	0.520	0.254	0.045	0.130	0.051	0.053
4	5.2.2.Production and export of organic panela	0.462	0.323	0.068	0.072	0.075	0.009
3	5.3.Marketers	0.320	0.469	0.078	0.068	0.065	0.061
4	5.3.1.Sustainable organic panela	0.330	0.462	0.070	0.071	0.068	0.054
4	5.3.2. Regulatory Compliance	0.251	0.524	0.135	0.047	0.043	0.008
2	**6. Social**	0.289	0.209	0.093	0.308	0.101	0.163
3	6.1. labor quality	0.261	0.189	0.171	0.189	0.189	0.054
3	6.2.Associativity trade-union	0.067	0.226	0.069	0.566	0.073	0.054
3	6.3.Control fraudulent practices	0.540	0.211	0.039	0.169	0.041	0.054
Overall priority of each alternative Pi		***0.229***	***0.232***	***0.167***	***0.262***	***0.110***	***1***

Global prioritization of results gives as the first alternative for decider, to choose Logistical and Commercial Integration Strategy (LCI) (0.262), followed by the alternative of quality, presentations and Panela uses improvement panela (QPP) (0.232); in third place, we have the development of clean technologies for sustainable and competitive sector growing (DCT) (0.229): in fourth place, it is the developing of diversification alternatives to take advantage of sugar cane (DAD) (0.167) and finally it is the Marketing Information systems development (MIS) (0.110).

5.6.Consistency

The third step will be to check the trial´s consistency.If R was a matrix completely consistent, then the λ_{max} will be equal to n.However, the decider will have some inconsistencies in his trials and is a great idea take an measure of the inconsistency´s degree of the trial made by the decider, because if you have not been careful with the ratings, the vector of priorities or weights obtained may be unrepresentative.

The consistency may be measured by the consistency index (IC), that has the following expression.

$$IC = \frac{\lambda_{max} - n}{n-1} \qquad (7)$$

This measure can be used toimprove the consistency of trials when compared with the appropriate number in the table No 3, that have the random consistency index (IA):

Tabla No.3.Random consistency index (IA) in fuction on the dimension of the matrix (n)

n	1	2	3	4	5	6	7	8
RI	0	0	0.525	0.882	1.115	1.252	1.341	1.401
n	9	10	11	12	13	14	15	16
RI	1.452	1.484	1.513	1.535	1.555	1.570	1.583	1.595

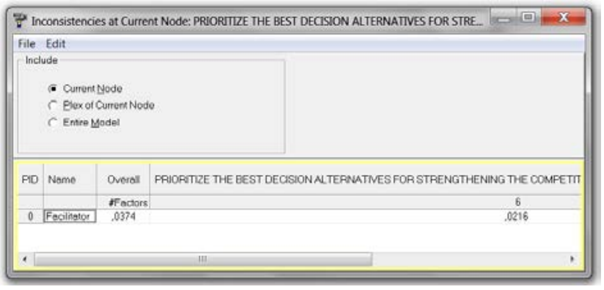
The Random consistency index (IA) is defined as the average random consistency index obteined by simulating 100,000 reciprocal matrices generated randomly using the scale of Saaty (1/9.1/8......1...... 8. 9).

If we calculate the ratio of the consistency index (IC) and the random consistency index (IA), we can be calculated the consistency ratio (RC).

$$(9) \quad RC = \frac{IC}{IA}$$

Now, if RC = 0, the matrix is consistent, but if RC ≤ 0.10the matrix R has an inconsistency admisible, which means that it is considered consistent and the weight vector obtained is accepted as valid. But if RC > 0.10, the inconsistency is unacceptable and is necessary to recheck the trials.

For our case, the inconsistency ratio is 0.0216 as show in the figure 5, this indicates that the consistency obtained is acceptable, because RC ≤ 0.10


Inconsistencies at Current Node: PRIORITIZE THE BEST DECISION ALTERNATIVES FOR STRE...

File Edit

Include

● Current Node
○ Plex of Current Node
○ Entire Model

PID	Name	Overall	PRIORITIZE THE BEST DECISION ALTERNATIVES FOR STRENGTHENING THE COMPETIT
		#Factors	6
0	Facilitator	,0374	,0216

Figure No.6. Inconsistency Index

5.7. Sensitivity analysis

A last step in AHP development, is to accomplish a sensibility analysis, a procedure that confirms results sturdiness and reducing random risk. Analysis consists in varying weight values and observating numerically and graphically how these changes affect the other weights and alternatives prioritization.

Analyzing sensibility, priorities can be changed in order to observe how alternatives prioritization would change. Expert Choice software presents five possibilities to do sensibility analysis. In figure 7, there is one of the methods use for changing dynamically objectives or criteria priorities, to establish how these changes impacts the prioritization of alternatives.

By increasing the economic criterion on a scale of 9 (Extremely important), it can be seen how the strategic decision alternatives vary, in this case becomes more vital to promote the development strategy of clean technologies for sustainable and competitive development of the productive chain.

By increasing the economic criterion on a scale of 9 (Extremely important).We can see how the strategies of alternatives decision vary (Figure 7), in this case becomes more vital to promote the development strategy of clean technologies for sustainable and competitive development of the productive chain.

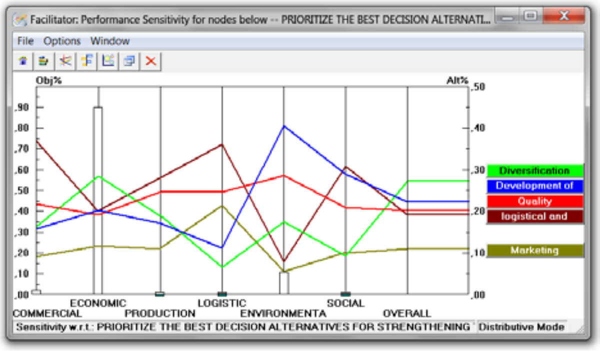


Figura 7 Sensitivity analysis when the objective economic is the most important

If we choose to give greater priority to environmental objective (Figure 8), the best alternative is the development of clean technologies and as a second alternative will be the improve the quality, the use and the presentation of the product.

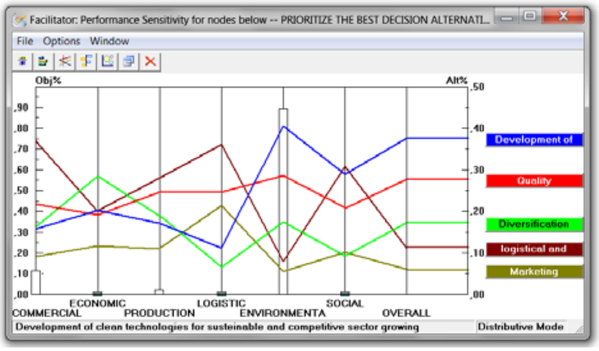


Figure 8 Sensitivity analysis when the environmental objective is the most important

When commercial purpose predominates (Figure 9) and this takes a higher value on the scale (9), the decision strategy that prioritizes is logistics and commercial integration, followed by improving the quality, presentation and usage of the product.

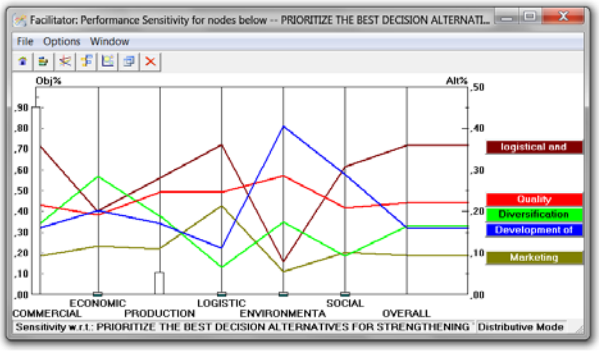


Figure 9 Sensitivity analysis when the commercial purpose predominates

Similarly, if the priority is focused on the product, the priority strategy will be the logistical and commercial integration (Figure 10).

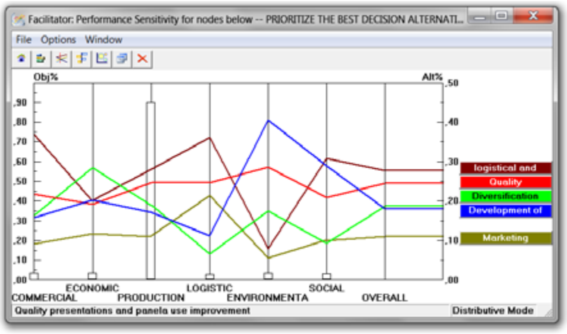


Figura 10. Sensitivity analysis when thepriority is focused on the product

If we prioritize the social objective (Figure 11), the highest alternative decision will be the strategy of commercial and logistical integration and secondly the development of clean technologies.

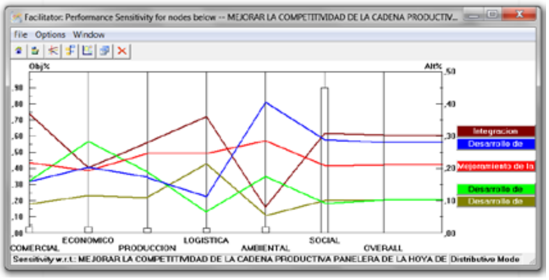


Figure 11 Sensitivity analysis when the social objective is the most important

The integration of regional logistics information resources is the most effective breakthrough for the integration of regional resources, but most of the information platforms that have established are respective and incompatible between the enterprises, so that every one is an "Information Island", which is not conducive to information sharing across enterprises throughout the regions. The integration of regional logistics information resources can make information flow smoothly across regional enterprises, so that logistics information become one of bridges among regional enterprises and provide effective services for regional enterprises (Wu and Shangguann 2012).

Authors like Gimenez and Ventura (2003) have competitive advantages derived from the integration in the supply chain, namely, the relationship between external integration and results in terms of cost of service, cost of transport, cost of ordering process , breaks in inventory and provisioning time (Marques, Molina and Vallet 2009).

The studies of Stank, Keller & Daugherty (2001) and Gimenez & Ventura (2003 and 2003b) share a common aim: to analyse the impact of internal and external integrationon performance. The integration-performance models of these authors included also a relationship between the levels of internal and external integration. All of them found that these levels of integration are positively correlated. This suggests that they positively influence each other (Gimenez 2004).

It is increasingly difficult for rural areas to meet the challenges of globalization only through vertical linkages. Beyond the need to overcome the disadvantages of demographic deficit, the size and the number of rural enterprises. etc., the presence of rural territories on the global stage, including politics, requires skills of dialogue, exchange and transfers to other territories (Farrell 2001).

6. CONCLUSIONS

In multi-criteria decision analysis there are really positive aspects. Some of the factors that favor its use is, for instance, that the AHP is a technique that offers an axiomatic theory. The participation of the actors involved in Panela production chain was of vital importance, however, it must be remembered that not only actors appreciation is important, but also experts opinion must be taken into account, not only to determine a priori actors needs, but also to display the best scenario projection.

Technological transformations and new consumer requirements have modified demand patterns towards a greater diversification facilitating new processes and products appearance, such as competition among agribusiness enterprises has based not solely on price, coming to the fore competitive factors as quality, design and product differentiation (Lopez Macias, 2007 and Boucher) .

AHP is an useful tool in multi-criteria decision making, where many actors involved. This can be very used by Panela associations and by the State in projects priorization, which are favorable for the sector.

The relevance of Commercial and Logistical integration strategy is based on that the scenarios, where union consolidation structures occurs, generate better results for the competitive structure of the chain, in this aspect can be cited as an example the agroindustry Doña Panela Ltda, which has successfully integrated all production factors and to have a place within national and international market with variety of products (Cadena 2004).

In areas where agendas converge, transport and trade facilitation measures need to be deepened to allow for further coordination and gains from cooperation. Continued emphasis on key processes regarding the development and harmonization of border crossings and the regulation of diverse transport modalities is of particular importance. Furthermore, the agenda for the expansion of productive integration and intra-regional logistics services must support both national and subnational organizations in order to fully achieve the economies of agglomeration necessary to reap the most benefits from these costly reforms (Guerrero, Lucenty and Galarza 2010).

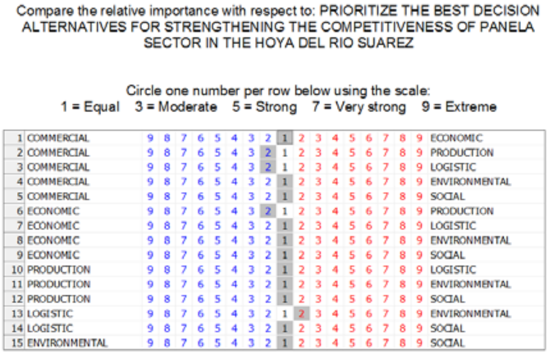
Future research should seek to identify models of logistical and commercial integration that contribute to supply chain strengthening, and using logistical simulation models to determine the best logistical and commercial integration model for the supply chain.

APPENDIX

Appendix A. Instrument for the analysis

Evaluation of the categories of objectives keeping in mind the overall shery management goal to sustain viable sheries in the long run. The tables are read horizontally where each row is a single comparison for you to evaluate. The value of one means that both criteria are equivalent, while selecting a value along the scale means that a particular criteria is more important than the other. Higher numbers correspond with increasing importance, i.e.

3.moderately important, 6.strongly important, 9.extremely more important



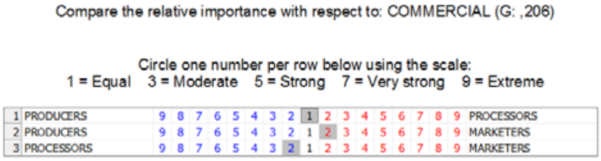
Compare the relative importance with respect to: PRIORITIZE THE BEST DECISION ALTERNATIVES FOR STRENGTHENING THE COMPETITIVENESS OF PANELA SECTOR IN THE HOYA DEL RIO SUAREZ

Circle one number per row below using the scale:
1 = Equal 3 = Moderate 5 = Strong 7 = Very strong 9 = Extreme

1	COMMERCIAL	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	ECONOMIC
2	COMMERCIAL	9	8	7	6	5	4	3	[✓ 2]	1	2	3	4	5	6	7	8	9	PRODUCTION
3	COMMERCIAL	9	8	7	6	5	4	3	[✓ 2]	1	2	3	4	5	6	7	8	9	LOGISTIC
4	COMMERCIAL	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	ENVIRONMENTAL
5	COMMERCIAL	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	SOCIAL
6	ECONOMIC	9	8	7	6	5	4	3	[✓ 2]	1	2	3	4	5	6	7	8	9	PRODUCTION
7	ECONOMIC	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	LOGISTIC
8	ECONOMIC	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	ENVIRONMENTAL
9	ECONOMIC	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	SOCIAL
10	PRODUCTION	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	LOGISTIC
11	PRODUCTION	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	ENVIRONMENTAL
12	PRODUCTION	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	SOCIAL
13	LOGISTIC	9	8	7	6	5	4	3	2	1	[✓ 2]	3	4	5	6	7	8	9	ENVIRONMENTAL
14	LOGISTIC	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	SOCIAL
15	ENVIRONMENTAL	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	SOCIAL

Appendix B

Example of comparision between the actors of the production chain



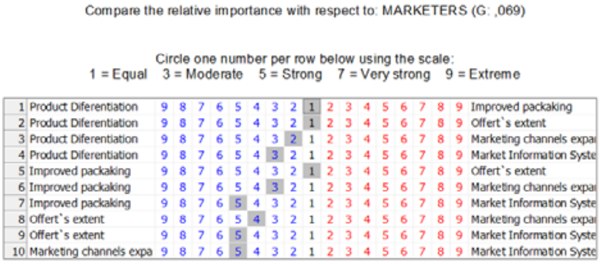
Compare the relative importance with respect to: COMMERCIAL (G: ,206)

Circle one number per row below using the scale:
1 = Equal 3 = Moderate 5 = Strong 7 = Very strong 9 = Extreme

1	PRODUCERS	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	PROCESSORS
2	PRODUCERS	9	8	7	6	5	4	3	2	1	[✓ 2]	3	4	5	6	7	8	9	MARKETERS
3	PROCESSORS	9	8	7	6	5	4	3	[✓ 2]	1	2	3	4	5	6	7	8	9	MARKETERS

Appendix C

Example of the comparision between the subfactors and factors in the chain



Compare the relative importance with respect to: MARKETERS (G: ,069)

Circle one number per row below using the scale:
1 = Equal 3 = Moderate 5 = Strong 7 = Very strong 9 = Extreme

1	Product Diferentiation	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	Improved packaking
2	Product Diferentiation	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	Offert`s extent
3	Product Diferentiation	9	8	7	6	5	4	3	[✓ 2]	1	2	3	4	5	6	7	8	9	Marketing channels expa
4	Product Diferentiation	9	8	7	6	5	4	[✓ 3]	2	1	2	3	4	5	6	7	8	9	Market Information Syste
5	Improved packaking	9	8	7	6	5	4	3	2	[✓ 1]	2	3	4	5	6	7	8	9	Offert`s extent
6	Improved packaking	9	8	7	6	5	4	[✓ 3]	2	1	2	3	4	5	6	7	8	9	Marketing channels expa
7	Improved packaking	9	8	7	6	[✓ 5]	4	3	2	1	2	3	4	5	6	7	8	9	Market Information Syste
8	Offert`s extent	9	8	7	6	5	[✓ 4]	3	2	1	2	3	4	5	6	7	8	9	Marketing channels expa
9	Offert`s extent	9	8	7	6	[✓ 5]	4	3	2	1	2	3	4	5	6	7	8	9	Market Information Syste
10	Marketing channels expa	9	8	7	6	[✓ 5]	4	3	2	1	2	3	4	5	6	7	8	9	Market Information Syste

ACKNOWLEDGMENTS

The researchers are grateful to experts, producers, traders and processors of Hoya del Rio Suarez, for his cooperation in the investigation, The National Federation of Panela makers (FEDEPANELA) for the information provided and the Science, Technology and Innovation Vice – Rector (VCTI) of the Antonio Nariño University, for funding the research project. The views expressed are those of the authors.

REFERENCES

Abaunza, C.A., Forero, C.A., Garcia, G.O., Carvajal G. H., 2012. *Zonificación y organización de clúster empresariales para las cadenas de caña panelera.frutales y papa criolla en Cundinamarca*. Colombia. Corpoica. xx p. ISBN: 978-958-740 Available from: http://www.corpoica.org.co/sitioweb/Archivos/Publicaciones/Cluster_para_evaluacion_de_tierras.pdf [accessed 15 July 2013]

Berumen, S.A. & Llamazares, R.F., 2007. *La utilidad de los métodos de decisión multicriterio (como el AHP) en un entorno de competitividad creciente*. Cuadernos de Administración. 20 (34). 65-87. Available from: http://www.redalyc.org/articulo.oa?id=20503404 [accessed 4 June 2013]

Bisang, R., Anlló, G., Campi, M., Albornoz, I., 2009. *Cadenas de valor en la agroindustria*. Cepal. Cap IV.p 219-272. Available from: www.eclac.org/publicaciones/xml/7/38557/CapituloIV.pdf [accessed 16 July 2013]

Cadena, D., Acuña, J., 2004.*"La agroindustria de la panela en la región de la Hoya del Rio Suarez, Bajo el enfoque de desarrollo regional y competitividad"* Universidad Industrial de Santander (UIS), Available from: http://repositorio.uis.edu.co/jspui/bitstream/123456789/8377/2/112751.pdf [accessed 18 July 2013]

Calabrese, A., Costa, R., Menichini T., 2013.*Using Fuzzy AHP to manage Intellectual Capital assets: An application to the ICT service industry* .Original Research Article. Expert Systems with Applications. Volume 40. Issue 9. July 2013. Pages 3747-3755. Available from: http://www.sciencedirect.com/science/article/pii/S095741741201322X [accessed 10 June 2013]

Castellanos, D.O., Torres, P.L., Flórez M.D., 2010. *Agenda prospectiva de investigación y desarrollo tecnológico para la cadena productiva de la panela y su agroindustria en Colombia*. Available from: http://www.minagricultura.gov.co/archivos/cadena_productiva_panela.pdf [accessed 15 June 2013]

Delgado, L.A., 2009. *"Propuesta para el redireccionamiento administrativo de la microempresa familiar panelera "caña gecha" en el municipio de la peña (cundinamarca)"*. Universidad de la Salle. Available from: http://repository.lasalle.edu.co/bitstream/10185/3194/1/T11.08%20D378pr.pdf [accessed 25 June 2013]

Dueñas, R., Morales, A., Nanning, C., Noriega, S., Ortriz J P., 2007. *Microeconomics of competitiveness of the sugar cane cluster in Colombia*.Harvard Business School. Boston. Massachusetts Available from: http://www.isc.hbs.edu/pdf/Student_Projects/Colombia_SugarCaneCluster_2007.pdf [accessed 05 June 2013]

Emam, A.A., 2010. *The Competitiveness of Sugar Cane Production: A Study of Kenana Sugar Company. Sudan*. Sudan University of Science and Technology.Faculty of Agricultural Studies Department of Agricultural Economics.Journal of Agricultural Science.Vol. 3.No. 3. ISSN 1916-9752. Available from: http://www.ccsenet.org/journal/index.php/jas/article/view/8361 [accessed 19 June 2013]

Eraslan, E., Dağdeviren M., 2010. *A Cognitive Approach for Performance Measurement in Flexible Manufacturing Systems using Cognitive Maps*.Cognitive Maps. Karl Perusich (Ed.). ISBN:

978-953-307-044-5.InTech. Available from: http://www.intechopen.com/books/cognitive-maps/a-cognitive-approach-forperformance-measurement-in-flexible-manufacturing-systems-using-cognitive-m [accessed 14 June 2013]

Expert choice version 8.0. Computers & Mathematics with Applications. Volume 25. Issue 8. April 1993. Page 117. Copyright © 2013 Elsevier Ltd. Available from: http://ac.els-cdn.com/089812219390179Y/1-s2.0-089812219390179Y-main.pdf?_tid=397e7e3e-ee68-11e2-b730-00000aab0f26&acdnat=1374014398_be2f5eca4f7109b4f711fc9123879077 [accessed 10 Juny 2013]

Farrel, G., 2001. *La competitividad de los territorios rurales a escala global.“innovación en el medio rural”*.cuaderno de la innovación nº 6 – fascículo 5 observatorio europeo leader. Febrero 2001. Available from: http://ec.europa.eu/agriculture/rur/leader2/rural-es/biblio/local-global/comlocalglobal.pdf [accessed 10 July 2013]

Fonseca, S.E., 2002. *Guia ambiental para el subsector panelero*. Ministerio del Medio Ambiente. Sociedad Colombiana de agricultores de Colombia (SAC).Federacion nacional de Paneleros (Fedepanela). Available from: http://www.panelamonitor.org/media/docrepo/document/files/guia-ambiental-para-el-subsector-panelero.pdf [accessed 10 July 2013]

Gimenez, C., 2004. *Logistics integration processes in the food industry*.Research Group in Business Logistics.Institutd`estudis territorials, UniversitatPompeuFabra. Available from: http://nir.upf.edu/joomla/images/pdf/publicacions/workingpapers/IET%20working%20paper%20014.pdf [accessed 25 July 2013]

Gomes, L.F., Autran, M., Andrade, R.M., 2012. *Performance evaluation in assets management with the AHP*. Pesqui.Oper.[online]..vol.32. n.1. pp. 31-54. Epub Mar 08. 2012. ISSN 0101-7438. Available from: http://dx.doi.org/10.1590/S0101 [accessed 10 June 2013]

Gomez, P.E., Silva, F.A., 2011. *Proyecto “Diseño y Desarrollo de un Plan de Marketing Territorial como estrategia de fortalecimiento del Desarrollo Local en 3 regiones de Colombia (Complejo Cenagoso de la Zapatosa. Hoya del Rio Suarez. Zona Norte del Valle del Cauca)”*.Available from: http://www.adel.org.co/archivos/LBL2HRS.pdf [accessed 10 June 2013]

Guerrero, P.K., Lucenti, Galarza, S., 2010. *Trade Logistics and Regional Integration in Latin America and the Caribbean.* ADBI Working Paper 233. Tokyo: Asian Development Bank Institute. Available from: http://www.adbi.org/files/2010.08.02.wp233.trade.logistics.latin.america.caribbean.pdf [accessed 02 July 2013]

Guerrero, C., Luengas, E., 2011. P*lan de manejo ambiental para el sector panelero en la Vereda Melgas.municipio de Chaguaní. Cundinamarca*. Universidad Militar Nueva Granada. Available from: http://www.umng.edu.co/documents/10162/745281/V3N2_4.pdf. [accessed 13 July 2013]

Hernandez, R., Fernandez C., Baptista P., 2006. *La información secundaria.* Metodologia de la Investigación. Mc. Graw Hill. ISBN: 9789701057537 Available from: http://www.mcgraw-hill.es/bcv/guide/capitulo/8448199251.pdf. [accessed 13 July 2013]

IICA 2001.*Bases para un acuerdo de desarrollo de la cadena agroindustrial de la panela.* Colección de documentos IICA. Serie Competitividad.Secretaría Técnica. Fedepanela.Available from: http://repiica.iica.int/docs/B0126E/B0126E.PDF [accessed 18 June 2013]

Lee, S., Kimb, W., Min, K.Y., Joo, O.K., 2012. *Using AHP to determine intangible priority factors for technology transfer adoption.* Available from: http://ac.els-cdn.com/S0957417411017015/1-s2.0-S0957417411017015-main.pdf?_tid=ba314f7e-e031-11e2-b903-0000aab0f6c&acdnat=1372451675_c1152276aad48844b7c2749698e86b6e [accessed 15 June 2013]

Leibovich, J., Laura, E., 2009. *Competitividad del sector agropecuario colombiano.*Availablefrom: http://www.compite.com.co/site/wp-content/uploads/informes/2008-2009/Agropecuario-(agricultura).pdf [accessed 10 June 2013]

Li, T., 2010. *Applying TRIZ and AHP to develop innovative design for automated assembly systems.* The International Journal of Advanced Manufacturing Technology,January 2010, Volume 46, Issue 1-4, pp 301-313 Available from: http://link.springer.com/article/10.1007%2Fs00170-009-2061-4#page-1.[accessed 10 June 2013]

Llano, M., Duarte, S.H., Moreno C.A., 2012. *Afectación de la rentabilidad al productor panelero por la implementación de la normatividad sanitaria y ambiental.* Contraloria General de la Republica Available from: http://186.116.129.19/c/document_library/get_file?folderId=75297808&name=DLFE-46852.pdf [accessed 18 June 2013]

Lopez, A. M., Méndez, J.J., Dones M., 2009. *Factores clave de la Competitividad regional*: *Innovación e intangibles.* Available from: http://www.n-economia.com/presentaciones/pdf/amlopez_jjmendez_mdones_jun09.pdf .[accessed 19 June 2013]

Ludovic, A.V., Marle F, Bocquet J.C., 2010. *Measuring project complexity using the Analytic Hierarchy Process.* Original Research Article International Journal of Project Management. Volume 29.Issue 6. August 2011. Pages 718-727

Available from: http://www.sciencedirect.com/science/article/pii/S0263786310001092 [accessed 07 June 2013]

Marques , A., Molina, X., Vallet, T., 2009. *Influencia de la integración logística en los resultados logísticos de las organizaciones,* Cuadernos de Estudios Empresariales vol. 19, 175-20, ISSN: 1131-6985 Available from: http://dialnet.unirioja.es/servlet/articulo?codigo=3283731 [accessed 15 July 2013]

Martinez, H.J., Ortiz, L., Acevedo X., 2005. *La cadena agroindustrial de la panela en colombia una mirada global de su estructura y dinamica 1991-2005*. Ministerio de Agricultura y Desarrollo Rural. Observatorio Agrocadenas Colombia. Documento de Trabajo No. 57. Available from: http://201.234.78.28:8080/jspui/bitstream/123456789/436/1/2005112163343_caracterizacion_panela.pdf [accessed 19 June 2013]

Nydick, R.L., Hill R.P., 1992. *Using the analytic Hierarchy Process to structure the supplier selection procedure.* International Journal of Purchasing and Materials Management; Spring 1992; 28. 2; ABI/INFORM Global. pg. 31. Available from: http://www77.homepage.villanova.edu/robert.nydick/documents/Vendor%20Selection.pdf.[accessed 19 June 2013]

Osorio C.G., 2007. *Manual Técnico: Buenas Prácticas Agrícolas -BPA- y Buenas Prácticas de Manufactura -BPM-en la Producción de Caña y Panela.*Available from: http://www.fao.org.co/manualpanela.pdf.[accessed 12 July 2013]

Perez, M.T., 2011. *La empresarización del sector panelero.factor de desarrollo de la productividad y competitividad.* Programa de Productividad y Competitividad Agropecuaria del Huila. Available from:http://huila.gov.co/documentos/agricultura/CADENAS%20PRODUCTIVAS/INFORME%20DE%20GESTION%20CA%C3%91A-PANELA%202011.pdf [accessed 19 June 2013]

Ramirez, X., 2013. *La Industria panelera pide al Gobierno precio de sustentación de $2.200 por kilo. Diario la Republica.* Available from: Http://www.larepublica.co/economia/industria-panelera-pide-al-gobierno-precio-de-sustentaci%c3%b3n-de-2200-por-kilo_38969 **[**accessed 5 july 2013**]**

Rios, J.A., 2013. *Falta de tecnificación pone en aprietos a los paneleros*. Available from: http://www.laopinion.com.co/demo/index.php?option=com_content&task=view&id=424804&Itemid=32 [accessed 18 July 2013]

Romero, C.M., 2012. *Area de desarrollo rural de la hoya del rio Suarez.componente físico biotico.* Incoder 2012. Available from: http://www.fao.org.co/manualpanela.pdf. [accessed 14 June 2013]

Rudas, G., Forero, J., 1995. A*groindustria panelera en Colombia» Pequeña producción y relaciones interempresariales.* Cuadernos de Desarrollo Rural N" 35. Santafé de Bogotá. 1995 páginas: 7-17. Availablefrom: https://www.google.com.co/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CCoQFjAA&url=http%3A%2F%2Frevistas.javeriana.edu.co%2Findex.php%2FdesarrolloRural%2Farticle%2Fdownload%2F3303%2F2508&ei=WRLzUej8K4jY8gTRv4GoDA&usg=AFQjCNGAqR4rPmO8wpXMJ0iM2C_3JolNTw&sig2=hk_eq6q6Q_V87of7_1Ul6Q

Saaty, T.L., 1990. *How to make a decision: The analytic hierarchy process Original Research Article European Journal of Operational Research*. Volume 48.Issue 1. 5 September 1990. Pages 9-26

Seong, k.L., Yong J.Y., Jong W.K., 2007. *A study on making a long-term improvement in the national energy efficiency and GHG control plans by the AHP approach*. Original Research Article Energy Policy. Volume 35.Issue 5. May 2007. Pages 2862-2868. Available from: http://www.sciencedirect.com/science/article/pii/S030142150600365X [accessed 10 June 2013]

Toledo, R., Engler, A., Ahumada, V., 2011. *Evaluation of Risk Factors in Agriculture: An Application of the Analytical Hierarchical Process (AHP)* Methodology. Chilean J. Agric. Res. [online]. 2011. vol.71. n.1 [citado 2013-07-15]. pp. 114-121. Available from: <http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-58392011000100014&lng=es&nrm=iso>. ISSN 0718-5839. http://dx.doi.org/10.4067/S0718-58392011000100014 [accessed 6 June 2013]

Vega, B.J., Delgado M.K., Sibaja B.M., Alvarado A.P., 2007. *Uso alternativo de la melaza de la caña de azúcar residual para la síntesis de espuma rígidas de poliuretano (ERP) de uso industrial. Tecnología. Ciencia. Educación.* julio-diciembre. 101-107. Available from: http://www.redalyc.org/articulo.oa?id=48222207 [accessed 5 July 2013]

Veronese, B.A., Carneiro, J., Ferreira da Silva J., Kimura H., 2012. *Multidimensional assessment of organizational performance: Integrating BSC and AHP Original.* Research Article. Journal of Business Research. Volume 65.Issue 12. December 2012. Pages 1790-1799. Available from: http://ideas.repec.org/a/eee/jbrese/v65y2012i12p1790-1799.html [accessed 12 June 2013]

Vidal, L.A., Marle, F., Bocquet, J.C., 2011. *Using a Delphi process and the Analytic Hierarchy Process (AHP) to evaluate the complexity of projects Original Research Article*. Expert Systems with Applications. Volume 38.Issue

5. May 2011. Pages 5388-5405. Available from: http://ac.els-cdn.com/S0957417410011607/1-s2.0-S0957417410011607-main.pdf?_tid=a53b26e0-e035-11e2-ba4d-00000aacb360&acdnat=1372453362_67d4234a4afc6339733e5297ba1c723c [accessed 15 June 2013]

Viniegra G., 2007.*Alternativas para el uso de la caña de azúcar*.Universidad Autónoma Metropolitana.. Iztapalapa. Aviablefrom: www.foroconsultivo.org.mx/eventos_realizados/.../dr_viniegra.pdf

Winston, W.L., 1991.*Toma de decisiones con objetivos multiples. Investigación de operaciones.aplicaciones y algoritmos*.In: PWS-kentPublising Company.Grupo Editorial Iberoamerica. S.A de C.V..eds. Investigacion de operacione.Aplicaciones y algoritmos. Mexico.Pag 792

Wu, H, Shangguann, X., 2012. *Regional Logistics Information Resources Integration Patterns and Countermeasures Original* Research Article Physics Procedia, Volume 25, Pages 1610-1615. Available from: http://www.sciencedirect.com/science/article/pii/S1875389212006992 [accessed 15 June 2013]

Zahedi, F., 1986. T***he Analytic Hierarchy Process: A Survey of the Method and Its Applications*** Interfaces. Vol. 16, No. 4 (Jul. - Aug., 1986), pp. 96-108

Zangeneh, A., Jadid, S., Rahimi-Kian, A., 2009. *A hierarchical decision making model for the prioritization of distributed generation technologies: A case study for Iran.* Original Research Article. Energy Policy, Volume 37, Issue 12, December 2009, Pages 5752-5763. Available from: http://www.sciencedirect.com/science/article/pii/S0301421509006296 [accessed 20 June 2013]

Zimmermann, B., Zeddies, J., 2002. *International competitiveness of sugar production*. 13th International Farm Management Congress. Wageningen. The Netherlands. July 7-12.2002 Department of Farm Management, University of Hohenheim, Stuttgart, Germany.D-70593 Stuttgart. Aviable from: http://www.ifmaonline.org/pdf/congress/Zimmermann_2.pdf. [accessed 15 Juny 2013]

AUTHORS BIOGRAPHY

Gabriela Leguizamon. completed her M.Sc. in Indistrial Engineering, at Los Andes University (Colombia), She received his degrees in Physics Engineer, at National University in 2007. She has been a university professor for the University Antonio Nariño and Open and Distance National University in Colombia and Research Professor of the organizations Management Group.

Nelson V Yepes.completed his M.A.Sc. in Design, Mangament Project at the University International Iberoamerican, Unini (Puerto Rico). He received his degrees in Industrial Engennering in Colombia in 1995. He has worked as an industrial engineer for logistics companies, has been a business consultant and university professor for the University Antonio Nariño in Colombia, is a professor of masters in projects for Fundaciòn Universitaria Iberoamericana (FUNIBER) and Research Professor of the organizations Management Group .

Maria Victoria Cifuentes. completed her M.Sc. in Mathematics, at National University (Colombia), She received his degrees in Mathematics, in the same university in 2009. She has worked as assistant teacher for National University and as an economic evaluator for VQ IngenieraLtda. in a special contract to Ecopetrol S.A.Currently, she is a ninth-semester student of industrial engineering of Antonio NariñoUniversity in Colombia, and a PhD student in Science – Mathematics of National University.

IMPACT OF PROJECTION SYSTEMS FOR VEHICLE SIMULATORS ON SYMPTOMS OF SIMULATOR

Grzegorz Gudzbeler[(a)], Andrzej Urban[(b)]

[(a)]Police Academy in Szczytno
[(b)] Police Academy in Szczytno

[(a)]g.gudzbeler@wspol.edu.pl, [(b)]a.urban@wspol.edu.pl,

ABSTRACT

Authors present a comparative study aimed at answering the question about quality of projection systems designed for vehicle simulators. They decided to make an attempt to give the preliminary answer to a question on which studied projection systems causes lesser degree of symptoms of simulator sickness during training. For the purposes of examinations two test platforms were prepared. One was equipped with a screen with a cylindrical projection system, the second with "on screen" projection system. This paper presents the results of comparative tests carried in consortium by Police Academy in Szczytno, Poland, as a part of scientific project "Building simulator of driving privileged vehicles in typical and extreme situations".

Keywords: simulation, visualization, projection systems, simulation sickness, cylindrical view, on screen

1. INTRODUCTION

Simulator's disease is a condition characterized by a number of symptoms in extreme conditions: nausea, vomiting, pallor, and increased sweating. They occur in humans under conditions of exposure to virtual or real visual motion stimuli, associated or not with kinetic stimuli. Those incentives are not physiological for a human and they are not adapted with humans. This definition is a broad concept encompassing: simulator, motion, air, maritime, automotive and space sickness, etc. The negative impact of the virtual environment of simulator to humans is therefore undesirable. For the first time the phenomenon was studied by Miller and Goodson (1958, 1960), who had symptoms that occur as a result of training in a simulator called motion sickness, because of the similarity of most of the symptoms of this disease to balance disorders. According to factors that affect the human body we can divide simulators on those where only exclusively kinetic incentives are used (Coriolis sample), with kinetic and visual incentives (simulators with visual stimuli on mobile platforms) and with only visual incentives (stationary simulators with visual stimulation). Due to this a simulator sickness name is more associated with the device on which symptoms can be occurred than with the phenomenon itself. Therefore, if the negative impact of the virtual environment simulator, regardless of the nature of the stimulus, we most often use name simulator sickness in relation to the set of symptoms occurring as a result of training on the simulator. Simulator disease is characterized by a rich and diverse symptomatology depending on the degree of its advancement. It often starts as stomach discomfort, bodily warmth, headache, dizziness and/or drowsiness, then proceeds to stomach distress, then nausea and vomiting. There are increasing nausea, often accompanied by symptoms of hypersensitivity to unpleasant taste and olfactory sensations, loss of appetite, headache, anxiety, adding to the ataxia and spatial disorientation. Due to the intensive tracking a virtual image at all times during examination the feeling of fatigue is common with blurred vision. In extreme cases the disease may be associated with violent vomiting, fatigue, apathy, and drowsiness and reduced mental capacity for concentration and muscle activity. Different configurations of symptoms in individual participants depends on the sensitivity of individual acting stimulus, the nature of the stimulus, the level and duration of action. Currently there is no conclusive statement about causes and prevention for simulator sickness. In the simulation of mobile objects, there are three main methods of visualization: on a helmet, on windows of a vehicle and on an external display.

.

2. TEST PLATFORMS SPECIFICATION

This paper presents the results of comparative tests carried out in consortium by Police Academy in Szczytno, Poland, as a part of scientific project "Building simulator of driving privileged vehicles in typical and extreme situations". The study involved a standard and widely used system of cylindrical projection and the increasingly popular projection system called "on screen".

For the purposes of examinations two test platforms were prepared. One equipped with a screen with a cylindrical projection system, the second with "on screen" projection system.

Simulator with cylindrical projection system.

- Cabin - of intercity bus Autosan A1012T Leader

- Screen (cylindrical: with radius R = 4.1 m and a height h = 3.75 m, angles of sight from a point of view of the driver: angle width: vfov = 180 deg, angle height hfov = 50 deg)
- Projection system (four projectors Projectiondesign F22 SX +, 1400 x 1050 resolution, brightness - 2100 ANSI lumens, contrast ratio: 2500:1, type of matrix: DLP)

That made the projection system provided an angular resolution in front of the driver's sight - 2.9 arc minute / pixel.

Simulator with on screen projection system.

- Cabin – Mercedes Acros truck.
- Screen - “on screen” - stuck projection foil to all front and side windows allowing view using the rear projection type, rear windows were completely blacked out.
- Projection system:

- 3 ultra-short throw projector Mitsubishi WD380U-EST serving front and left window (brightness: 2800 ANSI lumens, resolution: 1280 x 800, contrast ratio: 3000:1, type of matrix: DLP),

- 1 projector Panasonic PT-LB1E displays the image on the right window (brightness: 2200 ANSI lumens, contrast ratio: 500: 1, resolution: 1024 x 768, type of matrix: LCD).

That made the projection system provided an angular resolution in front of the driver's sight - 2.1 arc minute / pixel.

Photos 1-2 show the simulator with a cylindrical projection system.

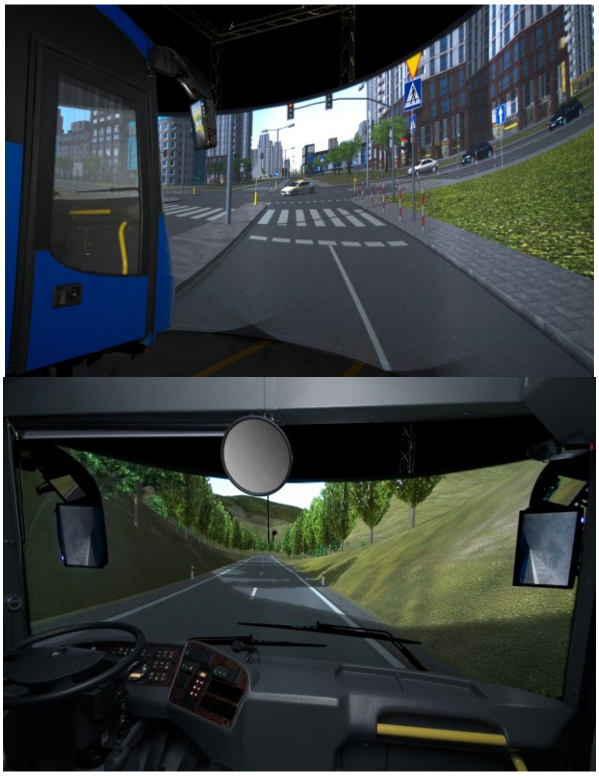


Fig. 1. Simulator with a cylindrical screen - a view of the cabin and the screen with displayed image, from outside and from inside of the cabin.

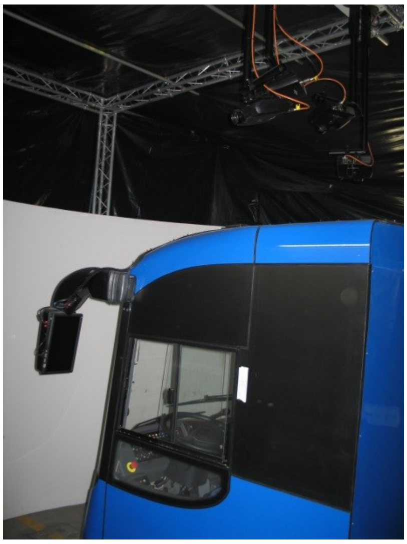


Fig. 2. Simulator with a cylindrical screen - visible cabin, cylindrical screen and projection system.

Photos 3 and 4 show the test stand with "on screen" rear projection system.

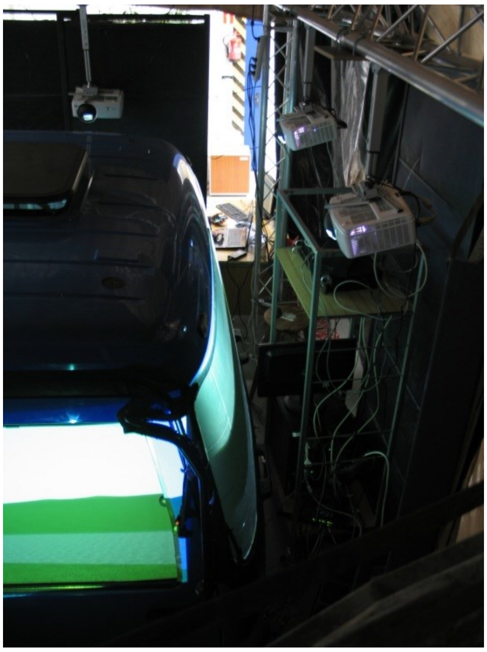


Fig. 3. The test simulator with "on screen" projection - visible cabin with "on screen" screens and projectors that support front and left side windows

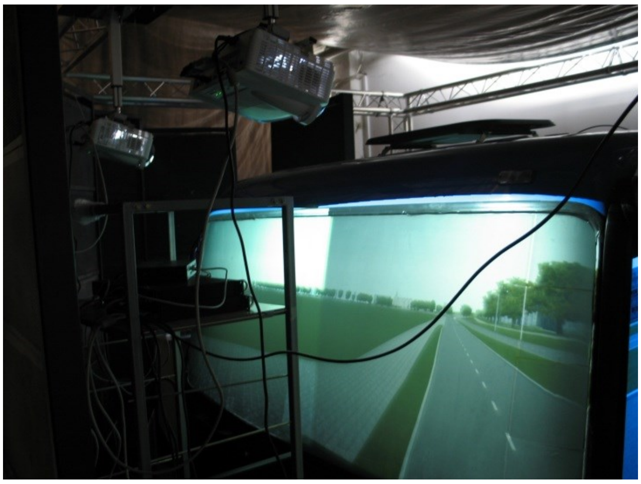


Fig. 4. The test simulator with "on screen" projection - visible cabin with "on screen" screens on the windscreen and windscreen supports projectors.

3. STUDY RESULTS

The study was performed on 15 individuals who have not previously practiced on simulators. Number of participants in the experiment is not easy to determine and depends on many aspects and especially on the purpose of the evaluation. Generally, the more participants, the research is more accurate. In preliminary tests, it seems reasonable to involve a homogeneous group of participants of similar age and experience. In ISO 16 673 standards, sufficient number of participants is 10. Taking this into account, it means that examination of 15 participants is sufficient, the acceptable minimum is 10 people.

Performed on these participants, a preliminary study has not identified diseases of their eye. Research on simulators with "on screen" and the cylinder projection system was performed at an interval of 10 days.

Results of research conducted on the simulator with cylinder projection system

The study carried out before training

A. Interview

The interview with all participants indicated no disturbance, which could have an impact on training on the simulator. 3 people have symptoms of asthenopia negative (age-related abnormal accommodation, causing problems with reading without correction glasses).

B. Concerned ophthalmological examination

1. The study of eye diseases. Refraction survey using computer autorefractometer in 11 participants showed a visual impairment that does not exceed + / - 1.5 D, the refractive state, which usually does not require a spectacle correction. 2 participants the defect was -2.0 D, with a -3.75 / -3.5 D, and a -5.0 / -5.5 D. People with these defects are not excluded from training because, according to the rules they may have a driving license.

2. The study of visual acuity. The visual acuity of the right and left eye in 10 participants ranged from 0.8-1.0. In 5 participants it was within the limits 0.5-0.6. So in any of the subjects, there was no reduction in visual acuity, which disqualifies them from driving.

3. Examination of the tear film with non-invasive test with a disruption of the tear film (NIBUT)) and the stability of the tear lipid layer films checked with Tearscope camera. NIBUT study showed normal values in all 15 participants (> 10 sec.). Examination of the lipid layer showed no abnormally thin in 13 participants (values A-C). In 2 of participants thickness of the lipid layer was thinned (E)

4. Examination of the binocular vision – Worth test. The study showed normal binocular vision in all participants.

5. The study of stereoscopic view - "Fly" test. Very good stereoscopy (Grade 8-9) occurred in 14 participants. A small reduction in stereoscopic occurred in 1 patient (grade 5).

6. The study of eyes setting - "cover test". In this study, there was no stability problems at the position of both eyes during their alternating covering (no small-angle strabismus and latent strabismus).

The study carried out after a training

A. Interview

12 persons after a training session on the simulator did not provide any information about visual disturbances. Three people gave out information about small disturbances in the form of "a strange image", "light disturbances when turning" and "strange impressions associated with non-motion simulator". These symptoms can be classified as a first degree of simulator sicknes in Chilow classification.

B. concerned ophthalmological examination

1. The study of visual acuity. The study showed no difference in visual acuity compared with state before training on the simulator. Small differences of 0.1 are within the error limits of the method.

2. Examination of the tear film with non-invasive test with a disruption of the tear film (NIBUT)) and the stability of the tear lipid layer films checked with Tearscope camera. NIBUT test showed no prolongation of the tear film break in all the participants. The study of the lipid layer showed no changes in its thickness in 14 participants. In 1 person was a small thinning of the layer thickness of 1 degree, but it was located within the normal range.

3. Examination of the binocular vision – Worth test. The study showed no changes in binocular vision in all participants after training.

4. The study of stereoscopic view - "Fly" test. After training on the simulator, there was no reduction in stereoscopy in all subjects.

5. The study of eyes setting - "cover test". In this study there was no change in the position of both eyes after a training session on the simulator.

Results of research conducted on the simulator with "on screen" projection system.

The study carried out before training

A. Interview In an interview in all participants there were no abnormalities that could have an impact on training on the simulator. 3 people have symptoms of asthenopia (age-related abnormal accommodation, causing problems with reading without correction glasses)

1. The study eye diseases. Refraction survey using computer autorefractometer in 11 participants showed a visual impairment does not exceed + / - 1.5 D, the refractive state, which usually does not require a spectacle correction. 2 participants the defect was -2.0 D, with a -3.5 / -3.25 D, and a -5.25 / -5.25 D. People with these defects are not excluded from training because, according to the Polish rules may they have a driving license (category A and B) in accordance with the Minister of Health regulation from 15 April 2011.
2. The study of visual acuity. The visual acuity of the right and left eye in 10 participants ranged from 0.8-1.0. In 5 participants it was within the limits 0.4-0.7. So in any of the subjects, there was no reduction in visual acuity, which disqualifies them to drive motor vehicles (Polish driving license category A and B) in accordance with the Minister of Health regulation from 15 April 2011.
3. Examination of the tear film with non-invasive test with a disruption of the tear film (NIBUT)) and the stability of the tear lipid layer films checked with Tearscope camera. NIBUT test showed normal values in 15 participants(> 10 sec.). Examination of the lipid layer showed no abnormally thin in 11 participants (the AC). In 4 of them thickness of the lipid layer was thinned (DE value).
4. Examination of the binocular vision – Worth test. The study showed normal binocular vision in all participants.
5. The study of stereoscopic view - "Fly" test. Very good stereoscopy (Grade 8-9) occurred in 14 participants. A small reduction in stereoscopic occurred in 1 patient (grade 5).
6. The study of eyes setting - "cover test". In the study, there was no evidence of impaired the stability of the position of both eyes during their alternating covering (no small-angle strabismus and latent strabismus).

The study carried out after a training

A. Interview

10 persons did not provide any visual disturbances after a training session on the simulator with an "on screen" projection system. 5 people reported the disorder in the form of "breathing", “a strange image,” light nausea ," twisted image "and dizziness. These symptoms can be classified as 1 degree in 4 participants, and in one case as a second stage of simulator sickness in Chilow classification.

B. concerned ophthalmological examination

1. The study of visual acuity. The study showed no difference in visual acuity compared with state before training on the simulator. Small differences of 0.1 are within the error limits of the method.

2. Examination of the tear film with non-invasive test with a disruption of the tear film (NIBUT)) and the stability of the tear lipid layer films checked with Tearscope camera. NIBUT test showed no prolongation of the tear film break in all the participants. The study of the lipid layer showed no changes in its thickness in 14 participants. In 1 person was a small thinning of the layer thickness of 1 degree, but it was located within the normal range.

3. Examination of the binocular vision – Worth test. The study showed no changes in binocular vision in all participants after training.

4. The study of stereoscopic view - "Fly" test. After training on the simulator, there was no reduction in stereoscopy in 14 participants. In one person were reduced stereoscopic range of 1 degree (from 5 to 4).

5. The study of eyes setting - "cover test". In this study there was no change in the position of both eyes after a training session on the simulator.

CONCLUSION

Why simulator sickness was more common in the simulator with the "on screen" projection system than a cylinder? It seems that this is due to the proximity of the screen. There is an analogy to the occurrence of symptoms when watching movies in 3D. Watching the three-dimensional films at the cinema rarely causes simulator sickness because the screen is far from the spectators. The introduction of 3D technology for television meant that the symptoms began to be felt much more often. It is estimated that it may occur in 10-20% of people watching 3D TV. They sit closer to the screen, so the probability of that feeling is greater.

The eye study showed that training on simulators with "on screen" and the cylinder projection systems does not cause changes in the organ of vision in participants with the form of a deterioration in visual acuity of the tear film, the state of binocular vision, stereoscopic view and eyes settings. Subjectively experienced symptoms of simulator sickness occurred less frequently after a training session on the simulator with a cylindrical screen than on the "on screen". It seems that the better tolerance of training on the simulator with cylinder screen is due to a greater distance from the screen.

REFERENCES

Lazzaro, N., 2009. The Four Keys to Fun: Designing Emotional Engagement and Viral Distribution without Spamming Your Friends, *ACM SIGCHI 2009 Procedings*, Palo Alto (USA).

Eckman, P. 2007. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication* *and Emotional Life*, NY: Henry Holt and Company LLC., New York (USA).

Gudzbeler, G., Urban, A., Nepelski, M., 2010. A prototype simulator of police operations in crisis situations, *24 th European Conference on Modeling and Simulation 2010,* Kuala Lumpur (Malaysia).

Blaauw, G.J., 1982. Driving experience and task demands in simulator and instrumented car: a validation study, *Human Factors 24(4)*, pp. 473-486.

Weir, D., 2010. Application of a driving simulator to the development of in-vehicle human-machine-interfaces, *IATSS Research (International Association of Traffic and Safety Sciences)*, vol. 34, pp. 16-21.

Crundall, D., Underwood, G., 1998. The effects of experience and processing demands on visual information acquisition in drivers, *Ergonomics*, vol. 41, pp. 448-458.

AUTHORS BIOGRAPHY

Cpt. Grzegorz Gudzbeler, PhD

is lecturer in Police Academy in Szczytno, Department on Internal Security, Poland. He received his bachelor degree in mathematics from University in Bialystok, master degree in informatics from University of Computer Sciences and Economics in Olsztyn, Poland and doctor's degree in National Defense Academy in Warsaw. He is author of many publications in topics of computer modeling and simulations, cybercrime and technical support for managing major events and crisis situations. He was engaged in project "Preparing Polish Police for Euro 2012".Now is contractor in projects „Building Prototype Simulator of Police Operations in Crisis Situations", "Building simulator of driving privileged vehicles in typical and extreme situations", and was engaged in preparing Polish Police for Euro Cup 2012.

Prof. Andrzej Urban, PhD, Eng

is Proffesor in Police Academy in Szczytno, Poland. He received engineer degree on University of Science in Gliwice , Poland and doctor degree on National Defense Academy in Warsaw, Poland. He is author of many publications in topics of computer modeling and simulations, crime prevention through environmental design and technical support for managing major events and crisis situations. He is manager of projects „Building Prototype Simulator of Police Operations in Crisis Situations" and "Building simulator of driving privileged vehicles in typical and extreme situations".

SAFETY STORAGE ASSIGNMENT IN AS/RS

Davoli Giovanni[(a)], Govoni Andrea[(b)], Gallo Sergio A.[(c)], Bortoli Erika[(d)], Melloni Riccardo[(e)]

[(a) (b) (c) (d) (e)]Affiliation Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy

[(a)]giovanni.davoli@unimore.it , [(e)]riccarco.melloni@unimore.it

ABSTRACT

Picking time reduction has been the traditional perspective for warehouse optimization. When structural safety is considered, optimization of warehouse operations should be read even in terms of load mass distribution. In many practical cases the static safety of the system is related to mass distribution and barycenter highness, for example when a vehicle hits the structure or even in the extreme case of an earthquake. To investigate the problem a simulation model is developed with AutoMod™ software. The model developed simulates a generalized AS/RS warehouse where the single physical location is managed. To evaluated the performances of the AS/RS a simulation experiment is completed. The aim of the experiment is to investigate the impact of different storage policies on intrinsic structural safety and performances in term of picking time and comparing the results with the variety caused by different factors such as: shuttle speed and warehouse filling rate.

Keywords: AS/RSs, warehouse, seismic safety, mass distribution.

1. INTRODUCTION

Automated Storage and Retrieval Systems (AS/RSs) recorded a significant increase in the last decades, which can be explained by savings in labor costs and floor space, increased reliability and reduced error rates in picking operations (Roodbergen and Vis, 2009).

The use of AS/RSs is widespread in many different industrial contest and some examples can be found even for heavy load applications. For example in a tile manufacture the single load can exceed the weight of 1.000 kg. For all these heavy duty AS/RSs also the aspects about safety are very important. In many practical cases the static safety of the system is related to mass distribution and barycenter highness, for example in the case of a collision between a vehicle and the warehouse structure or, moreover, in the extreme case of an earthquake.

The study of seismic behavior of structure with vertical irregularities in terms of mass, stiffness and strength is mainly focused on building. Past studies indicate that mass irregular distribution has little effect on seismic behavior for building (Khoshnoudian and Mohammadi, 2008), (Magliulo, Ramasco and Realfonzo, 2001) (Al-Alì and Krawinkler, 1998) but these studies were focused on civil building while there are only few examples of studies carried on industrial facilities and, to the best of the authors knowledge, no example at all on AS/RSs facilities.

The reduction of order retrieval time has been the traditional perspective for warehouse optimization only some authors, for example Heragu (2005) used total warehouse cost as objective function. Picking performances are related to the storage assignment policy, which allocates items in convenient locations (Ashayeri and al., 2002) Traditionally we have different strategies to allocate items in an AS/RS. In manufactures AS/RSs can use: a basic “First Free” strategy without any allocation optimization, a ABC strategy where each rack is divided into A, B and C zones organized in columns or in rows or with a more complex clustering strategy. The effectiveness of a strategy usually is evaluated in term of picking time. When safety aspects are considered, optimization of warehouse operations should be read in order to integrate in the evaluation model different aspects as already proposed for “green” factors (Meneghetti, 2010) (Prada and al., 2013).

In 2012 during the earthquake in Emilia region of Italy some AS/RSs facilities suffer serious damage (YouReport, 2012). Till a full comprehension of the influence of mass distribution on AS/RSs will be given, a model able to compare different storage policies even for structural safety aspects could be useful to support management on the tradeoff between time based performances and safety related aspects.

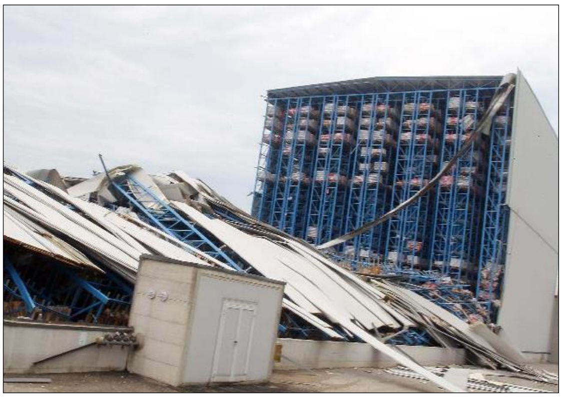


Figure 1: an AS/RS after the earthquake in Emilia – Italy 2013 (ANSA: Italian National Associated Press)

2. PURPOSE

The purpose of the paper is to provide a performance evaluation model for AS/RSs storage policies able to consider also static safety aspects. The work, grounded on a discrete event simulation model, provides a comparison between three different common allocation strategies. The proposed Key Performance Indicators (KPI) will enable practitioners to evaluate AS/RSs allocation strategy considering structural safety aspect too.

3. METHODOLOGY

To investigate the problem a simulation model of the AS/RS is developed according with the standard Bozer and White (1984) model, using AutoMod™ software. SciLab (scilab.org, 2013) open source platform is adopted to generate random initial item allocation set and retrieval orders.

The model developed simulates a single shuttle single command crane AS/RS where the single physical storage location is managed. A full description of possible different AS/RSs configurations is provided in the recent paper of Azzi and al. (2011).

3.1. Simulated system

The simulated system is formed by 2 rack served by a single shuttle crane and each rack is composed by 10 rows and 45 columns for 900 locations. Rack total length is fixed equal to 45 meters and total highness to 20 meters. Simulation model always implements FIFO rule to choose the item to pick.

An initial random items allocation is provided than for each day a picking list is random generated. For each item a random quantity is generated using a "uniform" distributed function. The minimum is always zero and the maximum value is set according with item A/B/C classification. The A/B/C classification is quite different from the classic 80/20 standard, this to better fit real operative conditions. At the end of the day the retrieved quantity is restored in the AS/RS for each item.

3.2. Key Performance Indicators (KPI)

The aim of the experiment is to investigate the impact of different storage policies on intrinsic structural safety and performance in term of picking time. To evaluate the picking performances and the safety related aspect two main KPI are defined:

- KPI1: average time to complete a picking task [sec];
- KPI2: the average barycenter highness [m], KPI2a refers to rack (a) and KPI2b t rack (b).

The aim is to quantify how a safety oriented allocation strategy effects picking time performances.

3.3. Model validation

Simulation model validation is provided to investigate outputs stability under the designed experimental conditions (Davoli et al. 2012). The simulation length has been defined according to the result of mean square pure error (MSPE) analysis and five replications have been used to perform the error analysis; over a number of five replications no significant differences were observed. Simulation stability is reached within simulation period of 10 days.

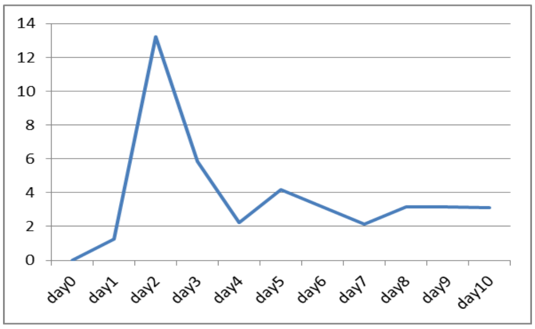


Figure 2: MSPE for KPI1 (task mean time)

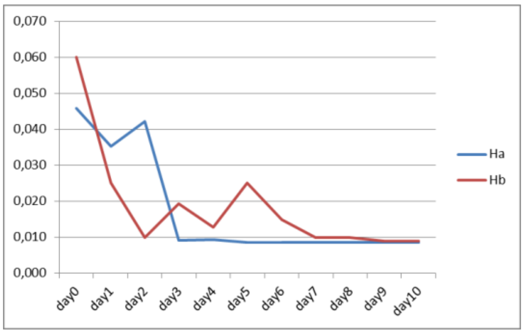


Figure 3: MSPE for KPI2a (Ha) and KPI2b (Hb)

3.4. Design of the Experiment (DOE)

The aim is not to develop a predictive response model but to demonstrate that a simulation approach can be useful and to investigate a case study representative of heavy load AS/RSs. The experiments consider different storage polices, different system fill rate and different speed set of the shuttle.

Table 1: Overview of experimental settings of the four investigated factors

Factors			
Storage Policy	Simple FirstFree	Columns A/B/C	Rows A/B/C
System fill rate	10% free space	30% free space	50% free space
Speed rate (V_h/V_v)	2	1	0.5

- Storage policy, the considered storage polices are: "First Free" where the location search is performed deck by deck; "Columns A/B/C" where the search is performed deck by deck within the columns reserved for the specific class; "Rows A/B/C" where the search is performed column by column within the rows reserved for the specific class.

- System fill rate, the considered initial random allocation for the items are: “Almost full” where the 10% of locations are free, “Half full” where the 30% of locations are free, “Half empty” where the 50% of locations are free.
- Speed rate, three shuttle crane speeds set are considered: V_h=2 m/s, V_v=1 m/s, rate=2; V_h=1.5 m/s, V_v=1.5 m/s, rate =1 and V_h=1 m/s, V_v=2 m/s, rate =0.5.

A full factorial experiment with three levels is used in this paper. Three factors and three levels give 3^3 = 27 combinations and thus 27 separate experiments were conducted. The three settings for the three factors are shown in Table 1. All the other parameters of the model are fixed at the value are presented described in Table 2.

Table 2: Model Fixed parameters set

Parameters	
N° items	10
“A class” items	2
“A class” demand	40%
“B class” items	3
“B class” demand	30%
“C class” items	5
“C class” demand	30%
Average n° picking list tasks	150

Table 3: Simulation Results

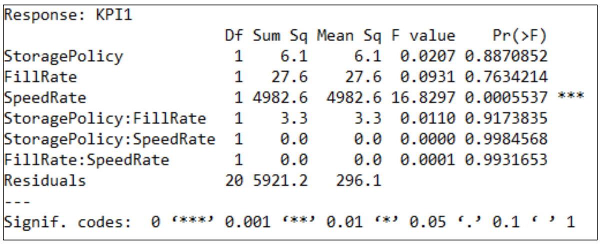
Results					
Storage Policy	Free space [%]	Speed rate	KPI1 [SKU/h]	KPI2a [m]	KPI2b [m]
Columns ABC	10	2	38	9,21	9,11
Columns ABC	30	2	39	7,91	7,65
Columns ABC	50	2	39	5,73	5,66
Columns ABC	10	1	36	9,21	9,11
Columns ABC	30	1	37	7,91	7,65
Columns ABC	50	1	37	5,73	5,66
Columns ABC	10	0,5	65	9,21	9,11
Columns ABC	30	0,5	67	7,91	7,65
Columns ABC	50	0,5	68	5,73	5,66
First Free	10	2	37	9,71	9,73
First Free	30	2	37	7,70	7,50
First Free	50	2	38	5,71	5,69
First Free	10	1	35	9,71	9,73
First Free	30	1	35	7,70	7,50
First Free	50	1	36	5,71	5,69
First Free	10	0,5	63	9,71	9,73
First Free	30	0,5	64	7,70	7,50
First Free	50	0,5	66	5,71	5,69
Rows ABC	10	2	38	9,48	9,45
Rows ABC	30	2	39	9,38	9,25
Rows ABC	50	2	38	9,09	9,11
Rows ABC	10	1	36	9,48	9,45
Rows ABC	30	1	37	9,38	9,25
Rows ABC	50	1	36	9,09	9,11
Rows ABC	10	0,5	64	9,48	9,45
Rows ABC	30	0,5	66	9,38	9,25
Rows ABC	50	0,5	65	9,09	9,11

4. FINDINGS

The results of the experiments are presented in Table 3. KPI1 is presented as Stock keeping Unit (SKU) for hour and KPI2 as meter. KPI2b, referred to rack b, is always lower than KPI2a, that because logic searching function always begin from rack b, in any case the differences are always negligible.

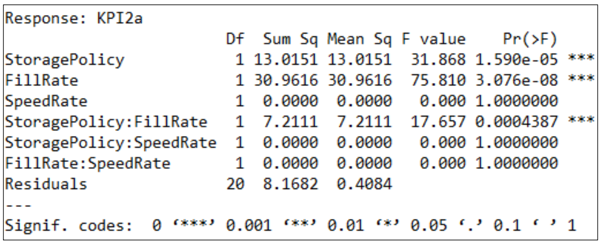
4.1. ANOVA TEST

The design and analysis of experiments have been conducted using the open source software R (r-project.org, 2013). The ANOVA tables for all KPI are presented in Figure 3, 4 and 5. Pr-values emphasized with (*) indicate variables significant on a better than 0.05 level. The ANOVA test reveals that the considered factors have different impact on KPI1and KPI2, while relevant factors have the same effects on KPI2a and KPI2b.



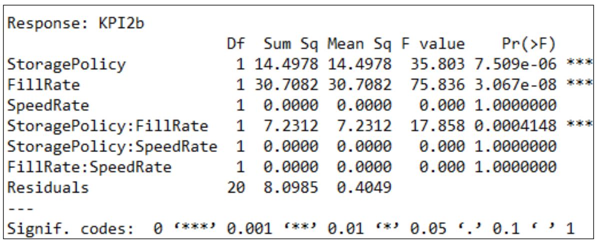
```
Response: KPI1
                         Df Sum Sq Mean Sq F value    Pr(>F)    
StoragePolicy             1    6.1     6.1  0.0207 0.8870852    
FillRate                  1   27.6    27.6  0.0931 0.7634214    
SpeedRate                 1 4982.6  4982.6 16.8297 0.0005537 ***
StoragePolicy:FillRate    1    3.3     3.3  0.0110 0.9173835    
StoragePolicy:SpeedRate   1    0.0     0.0  0.0000 0.9984568    
FillRate:SpeedRate        1    0.0     0.0  0.0001 0.9931653    
Residuals                20 5921.2   296.1                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: Anova test result for KPI1



```
Response: KPI2a
                         Df  Sum Sq Mean Sq F value    Pr(>F)    
StoragePolicy             1 13.0151 13.0151  31.868 1.590e-05 ***
FillRate                  1 30.9616 30.9616  75.810 3.076e-08 ***
SpeedRate                 1  0.0000  0.0000   0.000 1.0000000    
StoragePolicy:FillRate    1  7.2111  7.2111  17.657 0.0004387 ***
StoragePolicy:SpeedRate   1  0.0000  0.0000   0.000 1.0000000    
FillRate:SpeedRate        1  0.0000  0.0000   0.000 1.0000000    
Residuals                20  8.1682  0.4084                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: Anova test result for KPI2a



```
Response: KPI2b
                         Df  Sum Sq Mean Sq F value    Pr(>F)    
StoragePolicy             1 14.4978 14.4978  35.803 7.509e-06 ***
FillRate                  1 30.7082 30.7082  75.836 3.067e-08 ***
SpeedRate                 1  0.0000  0.0000   0.000 1.0000000    
StoragePolicy:FillRate    1  7.2312  7.2312  17.858 0.0004148 ***
StoragePolicy:SpeedRate   1  0.0000  0.0000   0.000 1.0000000    
FillRate:SpeedRate        1  0.0000  0.0000   0.000 1.0000000    
Residuals                20  8.0985  0.4049                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 6: Anova test result for KPI2b

4.2. DISCUSSION

The results indicate that, for the studied system, the key factor to maximize the AS/RS throughput is the rate between horizontal and vertical shuttle speed, or better shuttle speed itself is the key factors. Despite to the fact that the “Columns ABC” policy always guarantees the best throughput, the chosen storage policy, traditionally known as a key factor, in this specific contest is almost irrelevant, and this is clearly supported by the ANOVA test results.

The results indicate that barycenter highness is strongly related to warehouse fill rate and to storage policy. If the firs result is quite obvious the second result is relevant, moreover the ANOVA test reveals the existence of a combined effect between fill rate and storage policy. “Rows ABC” policy in particular presents a higher barycenter position even when the fill rate is 50%. “Column ABC” and “First Free” storage policies both present almost the same result for what concern barycenter highness. But the two storage polices generate different load distribution even though the barycenter highness s almost the same, this is a limit of the chosen KPI2 used to quantify “mass irregular distribution”.

5. CONCLUSIONS

The work shows that a simulation model is useful to investigate mass distribution in an AS/RS. Despite to the fact that there is no deeper study that investigate the relationship between mass irregular distribution and structural safety of AS/RSs, the recent earthquake, occurred in Emilia, Italy, suggest to consider this aspect while choosing the storage policy, especially for heavy duty AS/RSs. Moreover, the system studied in this paper reveals that in specific conditions of physical dimension, shuttle speed and A/B/C classes features adopting a more conservative storage policy doesn’t reduce significantly the AS/RS potential.

6. FURTHER WORKS

The present paper shows the limits related to the use of simply barycenter highness to quantify mass distribution. More sophisticated KPIs should be developed to measure irregularities in mass distribution, this activity should be carried on together with seismic AS/RSs behavior studies.

A statistical analysis about heavy duty AS/RSs should be carried on to investigate the real working condition of these industrial facilities especially about: system fill rate, shuttle speed, A/B/C classes features and adopted storage policies. The result of this study will be useful to understand if the adoption of a conservative, structural safety oriented, storage policy will reduce significantly AS/RSs throughput.

REFERENCES

Al-Alì, A.A.K., Krawinkler, H., 1998. *Effects of Vertical Irregularities on Seismic Behavior of Building Structure*, Report No. 130. Ed. The John A. Blume Earthquake Engineering Center at Stanford University CA (USA).

Ashayeri, J., Heuts, R. M., Valkenburg, M. W. T., Veraart, H. C., Wilhelm, M. R., 2002. *A geometrical approach to computing expected cycle times for zonebased storage layouts in AS/RS.* International Journal of Production Research, 40(17), 4467–4483.

Azzi A., Battini D., Faccio M., Persona A., Sgarbossa F., 2011. *Innovative travel time model for dual-shuttle automated storage/retrieval systems*. Computers & Industrial Engineering, 61, 600–607.

Bozer, Y. A., White, J. A., 1984. *Travel-time models for automated storage/retrieval systems*. IIE Transactions, 16(4), 329–338.

Davoli, G. Nielsen, P., Pattarozzi, G. Melloni, R., 2012. “Practical considerations about error analysis for discrete event simulations model.” *Proceedings of the International Conference on Advances in Production Management Systems*, September 24-26, (Rhodes – Greece).

Khoshnoudian, F., Mohammadi S.A., 2008. Seismic response evaluation of irregular high rise strustures by modal pushover analysis, *Proceedings of the 14th World Conference on Earthquake Engineering*, October 12-17, Beijing (China).

Heragu, S. S., Du, L., Mantel, R. J., & Schuur, P. C., 2005. *Mathematical model for warehouse design and product allocation*. International Journal of Production Research, 43(2), 327–338.

Magliulo, G., Ramasco, R., Realfonzo, R., 2001. Sul comportamento sismico di telai piani in c.a. caratterizzati da irregolarità in elevazione. *Proceedings of the 10° Convegno Nazionale “L’ingegneria Sismica in Italia”*, September 9-13 (Potenza-Matera – Italy) 2001

Meneghetti, A., 2010. Sustainable storage assignment is AS/RSs, *Proceedings of the International Conference APMS 2010,* October 11-13, Cernobbio (Como – Italy);

Prada, L., Garcia, J., Calderon, A., Garcia, J.D., Carretero, J., 2013. *A novel black-box simulation model methodology for predicting performance and energy consumption in commodity storage devices.* Simulation Modelling Practice and Theory, 34, 48–63.

Roodbergen, K.J., Vis, I.F.A., 2009. A survey of literature on automated storage and retrieval systems. *European Journal of Operational Research*, 194, 343-362.

YouReport, 2012. Available from: http://www.youreporternews.it/2012/terremoto-crolla-capannone-ceramica-a-santagostino/ [accessed 15 July 2013].

AUTHORS BIOGRAPHY

Dr. Eng. G. Davoli received MS and Ph.D. from University of Modena and Reggio Emilia (Italy) in 2005 and 2009, respectively. He is a Lecturer at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy since April 2009. His research interests include: BPR and lean production practices, discrete event simulation, supply-chain, stocks management and logistic problems. He is a Fellow of IFIP Working Group 5.7 Associates.

Eng. A. Govoni received MS from University of Modena and Reggio Emilia (Italy) in 2008. He is a

Lecturer at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy since April 2009. His research interests include: safety science, discrete event simulation, supply-chain, stocks management and logistic problems.

Dr. Eng. S.A. Gallo received MS and Ph.D. from University of Naples – Federico II (Italy) in 1993 and 1998, respectively. He is a Contract Professor at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy since April 2005. He had applied as lecturer at the University of Naples - Federico II 1998 to January 2005. His research interests include: projects management practices, discrete event simulation, scheduling and logistic problems.

Eng. E. Bortoli received BS from University of Modena and Reggio Emilia (Italy) in 2013. Her research interests include: discrete event simulation, stocks management and logistic problems.

Prof. Eng. R. Melloni received MS and Ph.D. from University of Bologna (Italy) (Italy) in 1984 and 1991, respectively. He is a Full Professor at Department of Engineering “Enzo Ferrari”, University of Modena and Reggio E., Modena, Italy since February 2005. He had applied as associate professor at the University of Modena and Reggio Emilia (Italy) from November 2001 to January 2005. He had applied as lecturer at the University of Parma (Italy) from April 1991 to October 2001. His research interests include: safety system management, projects management, BPR and lean production practices, discrete event simulation, scheduling, supply-chain and logistic problems.

SOLVING SMALL TSP ACCORDING TO THE PRINCIPLE OF MINIMUM ACTION

Diego D'Urso[(a)], Marco Cannemi[(b)]

[(a)]Dipartimento di Ingegneria Industriale, Università degli Studi di Catania
[(b)]Dipartimento di Ingegneria Industriale, Università degli Studi di Catania

[(a)]ddurso@dii.unict.it, [(b)] mcannemi@dii.unict.it

ABSTRACT

To solve the well-known Travelling Salesman Problem (TSP), many solutions based on combinatorial optimization, heuristic and meta-heuristic have been proposed. However, in managing business processes, a few times we attend to the real time optimization of picking routes either inside a warehouse or within materials or waste recovery distribution systems. This study proposes a new algorithm which is based on the analogy between TSP and conduction heat transfer; in particular, the application of the principle of minimum action to the heat transfer of a flat plate, which is coincident with the physical domain, over which the TSP points stress, helps identifying the order sought. The algorithm has been implemented in an Excel® spreadsheet; the quality of solutions which have been found is midway between the nearest neighbor algorithm and a genetic one; data processing time appears suitable for logistic processes management.

Keywords: TSP, principle of minimum action, space filling curves, unsteady state conductivity heat transfer.

1. INTRODUCTION

Given a space and a set of points to visit, the Travelling Salesman Problem (TSP) consists in finding the shortest path that enables to visit only once all the points and to return to the starting point. The minimum path has a more general meaning that comprehends the path at the least cost.

Formally, the TSP can be described as the search for the minimum of the function that represents the length of the route described above when varying the sequence in which the points are visited:

$$min(F(N,\pi(i))) = \sum_{i=1}^{N} \mathrm{d}_{i,\pi(i)} \qquad (0)$$

where the elements of the matrix di, j (i, j = 1 .. N) are the mutual distances between N points to be visited and π(i) is the permutation of the sequence in which points are visited.

The TSP has applications in many fields: material handling, order picking, vehicle routing related to materials distribution systems, both direct and reverse logistics (the latter case is of particular and current interest in the waste management); although the optimization of these activities should always take into account constraints that can greatly limit the search for the minimum of the cost function (0) (binding capacity), management or organizational improvement, often, can significantly lead to the reduction of such restrictions. Similarly job scheduling and machinery sequencing can be solved by using the TSP methods of solution. Further applications of the problem can be counted as part of communication networks, statistics, psychology and biostatistics.

The extraordinary proliferation of studies on the TSP and its applications in science led to several methods of solution. Table 1 tries to summarize the most popular:

Table 1: Main TSP methods of solution

Method	Algorithm
Linear programming	Cutting plane (Dantzig, Fulkerson, and Johnson 1954)
Linear and integer programming	Branch and bound (Land and Doig 1960)
Local research (tour construction)	Sweep (Gillet and Miller 1974) Nearest neighbor (Rosenkrantz, Stearns, Philip and Lewis 1977) Nearest insertion Farthest insertion (Rosenkrantz, Stearns, Philip and Lewis 1977)
Local research (tour improvement)	K-opt (Rego and Glover 2002; Croes 1958; Lin 1965) Lin-Kernighan (Lin and Kernighan 1973)
Meta-heuristics (local research)	Simulated Annealing (Kirkpatrick, Gelatt and Vecchi 1983) Termodynamical Approach (Cerny 1985) Tabu search (Glover 1989) Genetic Algorithms (Grefenstette, Gopal, Rosimaita, and Gucht 1985); Homaifar, Guan and Liepins 1993)
Meta-heuristics (multi-agents)	Ant colonies (Dorigo and Di Caro 1999, Dorigo, Maniezzo and Colorni 1996, Dorigo and Gambardella 1997)
Approximated (graph based)	Minimum Spanning Tree (Pizlo, Stefanov, Saalweachter, Li, Haxhimusa and Kropatsch 2005) Spacefilling Curves (Platzman and Bartholdi 1989)

Such methods, however, are usually dedicated to solve large dimension TSP or very complicated problems that we can consider as belonging to the project management knowledge; some methods of resolution approach the TSP with hundreds of thousands of points. On the other hand, the same methods are often more difficult to apply in the management of business processes; for example picking from storage or vehicle routing in local distribution usually show routes whit a smaller number of points to visit (N<100) but with a high frequency of evaluation.

A set of techniques that makes constraint programming a technique of choice for solving small (up to 30 nodes) traveling salesman problems has been presented in literature for TSPs transportation problems that either come from “real” transportation problems (e.g., with trucks) or from moving mechanical parts (Caseau and Laburthe 1997).

We can assert that TSP is so fascinating that, in some cases, has become a game and a challenge rather than solving a real problem. Indeed much research focuses on finding most suitable operators for applications or on solving large-scale problems. However, rarely research addresses the performance of different operators in small- or medium-scale problems. In addition, the differences between small- and medium-scale TSPs on suitable GA design are studied (Liu and Kroll 2012).

This paper proposes, therefore, to implement a new method to solve the TSP in order to make it flexible when a change of boundary conditions is requested (i.e. the geometric domain, the number of points of the routes, the types of routes) and easy to use in management of logistics processes.

In the first part of the paper a brief literature review is reported; then the proposed model is described; it is based on the principle of the minimum action which is applied to heat transfer in unsteady state conditions. In the second part of the paper the model is applied to solve the TSP having to visit randomly generated points in the range [10 ... 30]; the results are finally compared with those obtained by the application of algorithms which are, at least in perspective, easily implementable in the same simulating environment (nearest neighbor algorithm and a genetic algorithm).

2. METHODOLOGY

The first principle of thermodynamics is applied to an elementary control volume under the following assumptions:

- the medium is composed of a fixed solid whose thermo-physical properties aren’t time dependent;
- changes in volume, due to changes in temperature, are negligible if compared to the same volume;
- internal heat sources, described by $\dot{q}$(x,y,z) function, represent the energy generated per unit of volume and time.

Given the limited variation in volume, mechanics work exchanged by the elementary volume is negligible and the change of the internal energy is equal to the heat which is exchanged with the nearest neighbors in the unit of time: dU = dQ.

So the internal energy variation is only a function of temperature and internal sources:

$$dU = (\rho c \frac{\partial T}{\partial t} + \dot{q}(x,y,z))dxdydzdt \qquad (1)$$

As regards the heat exchange, it is assumed to be only conductive; so the balance of heat flows along each direction allows writing the equation (Fig.1):

$$dQ=(q_x-q_{x+dx})dydzdt+(q_y-q_{y+dy})dxdzdt+(q_z-q_{z+dz})dxdydt \qquad (2)$$

$$q_x = - k \frac{\partial T}{\partial x} \qquad (3)$$

$$q_{x+dx} = - [k \frac{\partial T}{\partial x} + \frac{\partial\left(k \frac{\partial y}{\partial x}\right)dx}{\partial x}] \qquad (4)$$

where dx, dy and dz are the elementary volume sizes, ρ is the density of the material, c the specific heat, T is the temperature, k is the thermal conductivity and t is the time.

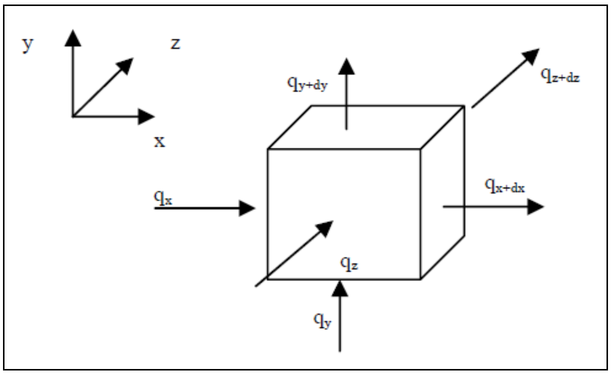


Figure 1: Balance of heat conduction flow inside the elementary volume

Assuming thermal conductivity k as a constant, the above mentioned heat balance, which is the application of the first principle of thermodynamics, allows deriving the general conduction equation (Ozisikin 1980):

$$a\nabla^2 T + \frac{q}{\rho c} = \frac{\partial T}{\partial t} \qquad (5)$$

where: a = k/(ρc) is the so-called heat diffusivity of the material.

Let us consider, now, the thermodynamic system showed by figure 2; it is indefinitely along the z dimension, so it may be considered as a flat, square plate whose side measurers are both L.

Boundary conditions for the figure 2 system are now defined: temperature is fixed and equal to Ta along border edges; temperature is fixed and equal to T_{fix} for a given set S of points $P_i(x_i, y_i)$, which belong to that flat-square plate; the constancy of temperature at a point is equivalent to assume that heat exchanging from the same point to the outside world can happen with infinite intensity; finally if there are no internal heat sources, the general conduction equation applied to the thermodynamic system and its boundary conditions are the following:

$$\nabla^2 T = \frac{\partial T}{\partial t}; \forall\, x, y$$

$$\begin{aligned} &T(0,y) = T_a; \\ &T(L,y) = T_a; \\ &T(x,0) = T_a; \\ &T(x,L) = T_a; \\ &T(P_i) = T_{fix}\ \forall P_i \in S. \end{aligned} \qquad (6)$$

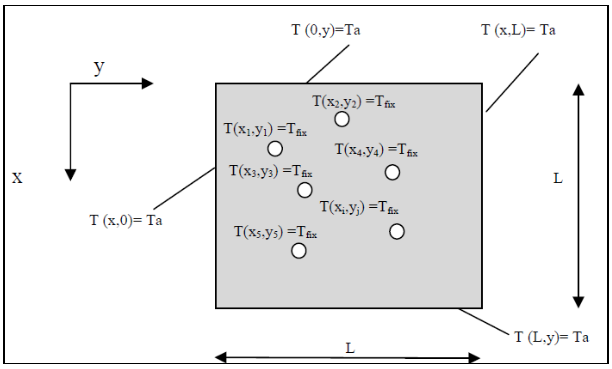


Figure 2: Boundary conditions of the flat-square plate system

If $T_{fix} \neq Ta$, the above thermodynamic system, after a thermal transient condition, reaches a state of thermal equilibrium; this occurs because heat flow which is transmitted through the border edges is virtually equal to that exchanged with the outside environment through points Pi ∈ S.

So the equation (5), after the thermal transient condition, turns into the Laplace equation.

The system of equations (6) can be numerically integrated by using the finite differences method.

The flat plate system can be discretized by using the finite element method; likewise heat conduction system of equation (6), which is composed of a partial differential equation, can be integrated by using finite differences formulae as below reported:

$$\frac{\partial T}{\partial t} \cong \frac{T_{i,j}^{k+1} - T_{i,j}^{k}}{\Delta t} \qquad (7)$$

$$\frac{\partial^2 T}{\partial x^2} \cong \frac{T_{i,j+1}^{k} - 2T_{i,j}^{k} + T_{i,j-1}^{k}}{\Delta x^2} \qquad (8)$$

$$\frac{\partial^2 T}{\partial y^2} \cong \frac{T_{i+1,j}^{k} - 2T_{i,j}^{k} + T_{i-1,j}^{k}}{\Delta y^2} \qquad (9)$$

Putting equations (7), (8) and (9) in (5), together with the Δx=Δy condition, leads to the finite differences equation of conduction for the each finite control volume:

$$T_{i,j}^{k+1} = T_{i,j}^{k}\left(1 - 4a\frac{\Delta t}{\Delta x^2}\right) + a\frac{\Delta t}{\Delta x^2}\,(T_{i-1,j}^{k} + T_{i+1,j}^{k} + T_{i,j+1}^{k} + T_{i,j-1}^{k} \qquad (10)$$

The equation (10) is recursive; the calculation process is stable if and only if it satisfies the following criterion (Ozisikin 1980):

$$0 \leq \left(1 - 4a\frac{\Delta t}{\Delta x^2}\right) \qquad (11)$$

The criterion (11) defines the maximum value to the time bucket which can be used to simulate the thermal transient condition once upon thermo-physic properties of the material are chosen.

Figure 3 shows the geometry of the finite element discretization in order to highlight the relevant measures.

The system of equations (6), numerically integrated by using the finite differences formulae, can be encoded within an Excel® spreadsheet.

Figure 4 shows the temperature distribution T(x, y) which the software application can perform for the thermodynamic system at the end of thermal transient condition; the mathematics model takes into account a set S that counts 10 points; the temperature of each Pi points is T_{fix}= 0 °C. The temperature of the border edges is T_a = 20 °C. The thermo-physical properties of the pseudo material are imposed to unit values (Δx=Δy=1 m; c=1 J/kg°C; ρ=1 kg/m3). The size of the flat plate was set to L = 20 Δx.

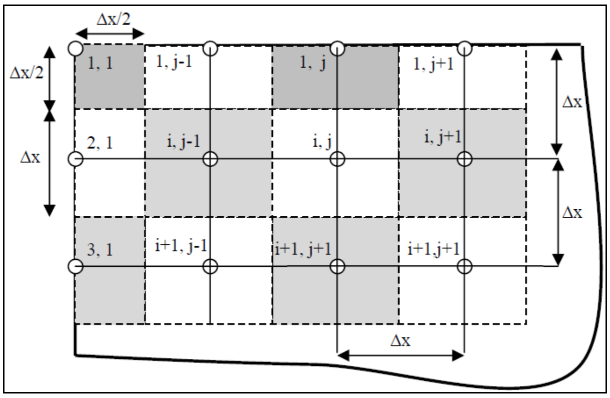


Figure 3: Finite element discretization of the flat plate

The system of Figure 2, is now seen under the light principle of the minimum action (Landau and Lifshitz, 1971). It guarantees that any dynamic or thermodynamic system evolves by minimizing a functional: we can call it energy (or action). The temperature distribution of figure 4 is the result of a thermal transition that leads to the configuration of minimum internal energy. The temperature distribution on the flat plate (the shape of the isotherm curves) shows how the heat is flows.

The latter thermodynamic system exchanges energy and organizes its temperature distribution as a function of its shape, of the temperature imposed along its border edges, of the one imposed in the points $P_i \in S$ and, in particular, of points allocation. The thermodynamic system must bring energy from its border edges to the points P_i which may dissipate outside this heat flow; this last process happens in a manner which adheres to the second principle of thermodynamics; the above mentioned flat plate has to solve a problem that is similar to the TSP. The analogy is therefore established among points to be visited and the sources of internal heat of the system and between the cost of the travelling salesman path and the internal energy level.

Let us consider now the isotherm curve T_{iso}=14 °C (Fig. 3); this is the first closed curve which encircles the set of points Pi; after projecting the points Pi on the isotherm curve T_{iso} the sequence in which they appear on the isotherm (Fig. 5) gives the solution of the problem. We can think about the isotherm T_{iso} = 14 °C as a spacefilling curve which is drawn by the thermodynamic system.

The solution to the TSP may, therefore, be obtained by calculating the steady state thermal condition of a flat plate whose sizes are that of the logistic domain of the TS problem, having a border edges at a constant temperature and heat sources placed on the points to be visited. Once the calculation of the thermal transient condition is performed, which has no computational difficulties, the problem turns in the research of T_{iso} curve and in the projection of points P_i on such curve. The existence of this isotherm is guaranteed by the nature of heat conduction (equation 6), which is in fact an equation of Laplace.

The system of thermal loads and the position of the points Pi in the simulated domain alter the temperature distribution at the thermal balance; so the choice of Tiso curve must be tuned in order to solve the problem.

This behavior seems a drawback of the novel methodology, which has, anyway, the possibility of increasing the number of points to be visited without having to modify the thermodynamic model, but only having to increase the above mentioned projection process according to a directly proportional law.

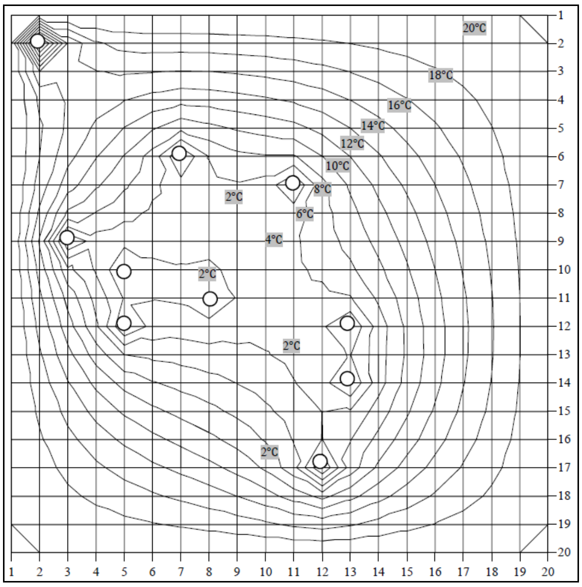


Figure 4: Temperature distribution on the flat plate. P1(2;2), P2(6;7), P3(7;11), P4(12;13), P5(14;13), P6(17;12), P7(11;8), P8(12;5), P9(10;5), P10(9;3) (tsim=120 s; Δt=0,2 s)

Figure 6 shows the end of the thermal transient condition which follows on from a different system of boundary conditions: internal heat sources are continuously placed in the flat plate ($\dot{q}$(x,y) = cost) and the points Pi are kept at fixed temperature Tfix. Also in this scenario, the N points Pi have to be regarded as a target of heat generation; the system reaches again a thermal equilibrium because of the equivalence between heat generated and dispersed. The equations of heat conduction once were integrated numerically by using the finite differences method.

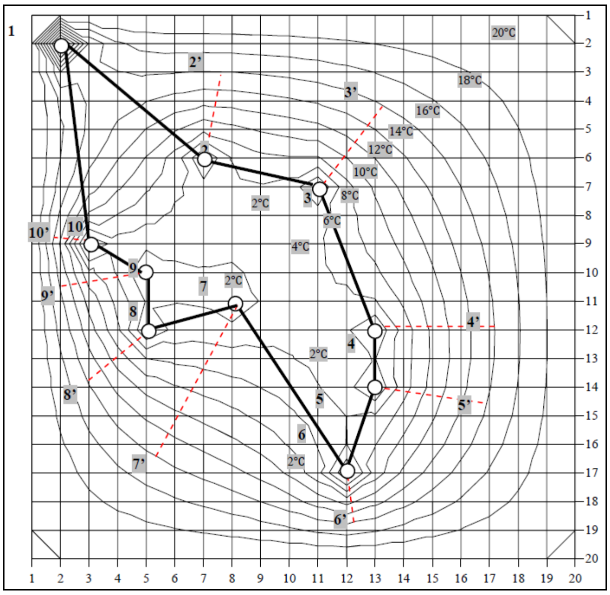


Figure 5: Pi projection process along isotherm curve Tiso=14°C

This time the T_{iso} isotherm curve encircles points Pi only if also the edge of the plate is taken into account; this also shows an approximation whose solution passes through an expansion of the domain in which the conduction thermal transient is calculated; under the above mentioned boundary conditions, the value of isotherm curve is T_{iso}=16 °C (Figure 6). The process of projection of the points P_i along the isotherm curve gives the same final result of the previous one.

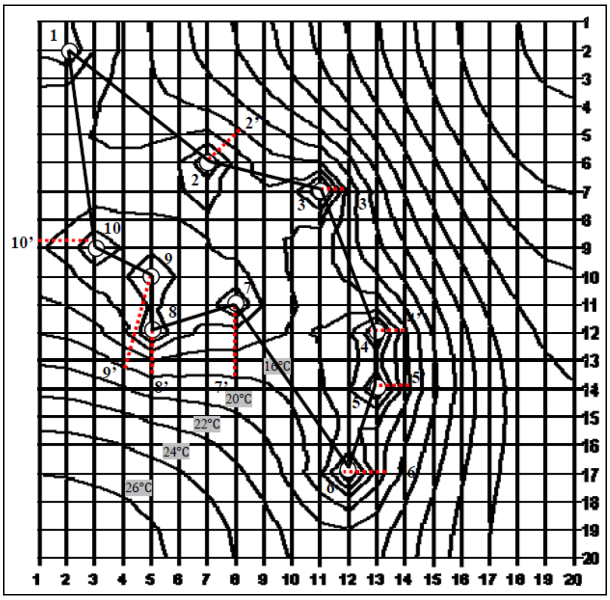


Figure 6: Flat plate temperature distribution with continuous heat source and Pi points projection along T_{iso}=16°C. $P_1(2;2)$, $P_2(6;7)$, $P_3(7;11)$, $P_4(12;13)$, $P_5(14;13)$, $P_6(17;12)$, $P_7(11;8)$, $P_8(12;5)$, $P_9(10;5)$, $P_{10}(9;3)$ (t_{sim} = 120 s; Δt = 0,2 s)

3. RESULTS

The thermodynamic model, previously showed, has been coded in an Excel® spreadsheet; the simulating model can be easily switched from one set of boundary conditions, such as fixed temperature on the border edges and at the points P_i, to another. The software environment allows the circular calculation that enables time driven simulating processes. Once calculated the temperature distribution at thermal equilibrium (a period of simulation t_{sim}=120 s was more than enough), the identification of the T_{iso} curve, on which to project the points of TSP, has been simply obtained by the following relationship: T_{iso} = 3/2 T_m; where T_m is the average temperature of the thermal field.

Although the latter relationship is rough and a more sophisticated check can be encoded in order to find the first closed isotherm around the P_i points to visit, the tests were in most cases fulfilled at the first iteration. The projection of the points P_i on the first closed isotherm consists in calculating for each P_i which is the nearest point belonging to the isotherm curve. To this aim it was codified a routine, by using standard Excel® function; it allows to split the thermal domain of figure 5 in a region warmer than T_{iso} and, consequently, the remaining one; once this partition is performed the elements (cells) of the border are serially numbered in order to establish the rule by which a point is before or after another. Table 2 shows the summary of tests performed for growing number of points to visit. The tests, as many as a hundred for each value of number N, were performed by choosing randomly N-1 points; the first point, that has coordinates P(2,2), has been imposed as a point of departure and arrival of the routes. The results have been compared with those obtained by a genetic algorithm and the nearest neighbor one. The quality of solutions found is midway between that of the solutions found by the two latter; the data processing time appears suitable for business process managing. The implemented algorithm seems open to many improvements particularly as regards the projection of the points P_i on the T_{iso} curve; its computational simplicity and the principle on which it is based ensures positive developments in the next research.

It has to be noted that the coding environment (Excel®) enjoys the favorable property WYSIWYG that allows easier modification when the boundary conditions change; there is evidence also that the same environment on one hand can be integrated with the interfaces of information systems, through barcode and RFID technology; on the other hand, it appears increasingly shared with powerful software applications such as Matlab®; all of the above features are considered of great value in order to support operations management (i.e. order picking; vehicle routing).

Table 2: Results summary (Δx=Δy=1m; L=20 m); the novel algorithm is called thermal space filling curve (Thermal SFC). (*Genetic algorithm is performed by Matlab®)

	Algorithm	N=10	N=20	N=30
TSP point density (N/L^2)		2,5%	5,0%	7,5%
Average value of specific tour lenght, Lm =L_{tour}/N	Nearest neighbor	6,03	3,01	3,29
	Thermal SFC	5,70	2,85	3,02
	Genetic	5,50	2,75	2,78
Standard deviation of specific length tour (Lm)	Nearest neighbor	0,70	0,35	0,36
	Thermal SFC	0,68	0,34	0,23
	Genetic	0,50	0,25	0,14
Data processing time (s)	Nearest neighbor	≈ 1	≈ 1	≈ 1
	Thermal SFC	≈ 20	≈ 25	≈ 30
	Genetic	≈ 15	≈ 30	≈ 360
Software size (kB)	Nearest neighbor	100 500	100 500	100 500
	Thermal SFC	1,11	1,11	1,11
	Genetic*			

4. CONCLUSIONS

The TSP is one of the lines of research most frequently beaten and, despite the passage of time, it is yet very fascinating. The numerous methods of resolution of TSP are usually dedicated to solving the problem in a large scale and with the highest number of points to visit; the complication of these methodologies has the consequence of their difficult implementation in business processes.

A novel algorithm for solving the TSP of symmetric kind, therefore, it has been proposed; the algorithm is based on the analogy with conduction heat transfer and, in particular, it is the result of the application of the principle of minimum action to the thermal transient of a flat plate; this plate coincides with the logistic domain in which the TSP is defined; within this domain Pi points to visit have to be considered as sources of heat. When the thermal transient condition is elapsed, the resulting temperature distribution contains some closed isotherm curves which can be taken in to account as spacefilling curves; projecting P_i point along one of these isotherm enables to determine the solution sequence of the problem.

Tests carried out on routes with increasing number of point to visit [10, 20, 30] show a quality of solution which is a midway between that of the solutions found by genetic algorithm and by nearest neighbor one. Data processing time appears useful to business process managing (order/batch picking, vehicle routing, machinery sequencing).

The novel model has the advantage of being encoded in a widely distributed electronic environment (Excel®), with ergonomic features useful for the easy correction or amendment; it doesn't suffer, if not in a linear manner, the combinatorial complexity of the problem when the number of points to be visited increases.

ACKNOWLEDGMENTS

A special thank to Professor Gaetano Giaquinta (Università degli Studi di Catania), who teaches how to be consciously adhering to the principle of minimum action.

REFERENCES

Caseau, Y., Laburthe, F., 1997. Solving Small TSPs with Constraints, International Conference on Logic Programming, Citeseer

Cerny, V., 1985. A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm. *Journal of Optimization* Theory and Applications, 45, 41-51.

Croes, G. A., 1958. A method for solving traveling-salesman problems. *Operations Research*, 6(6), 791-812,.

Dantzig, G., Fulkerson, D., Johnson, S., 1954. Solution of a large-scale traveling salesman problem. *Operations Research*, 2, 393-410.

Dorigo, M., Di Caro, G., 1999. Ant Colony Optimization: A New Meta Heuristic. *IEEE Evolutionary Computation, CEC99. Proceedings of the 1999 Congresson*, 2, pp. 1470-1477.

Dorigo, M., Gambardella, L.M., 1997. Ant Colonies for the Traveling Salesman Problem. *BioSystems*, 43, 73-81.

Dorigo, M., Maniezzo, V., Colorni, A., 1996. The Ant System: Optimization by a colony of cooperating agents Systems. *Man and Cybernetics, Part B. IEEE Transactions*, 26 (1), pp. 29-41.

Gillet, B.E., Miller, L.R., 1974. A heuristic algorithm for the vehicle dispatch problem. *Operations Research*, 22, 340–349.

Glover, F., 1989. Tabu Search - Part I. *ORSA Journal on Computing* 1(3), 190-206.

Grefenstette, J., Gopal, R., Rosimaita, B., Gucht, D.V., 1985. Genetic Algorithms for the Traveling Salesman Problem. *Proceedings of an International Conference on Genetic Algorithms and their Applications*, pp. 160-168.

Homaifar, A., Guan, S., Liepins, G.E., 1993. A New Approach to the Traveling Salesman Problem by Genetic Algorithms. *Proceedings of the 5th International Conference on Genetic Algorithms*, pp. 460-466, Morgan Kaufmann.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by Simulated Annealing. *Science*. New Series 220 (4598), 671-680.

Land, A., Doig, A., 1960. An automatic method for solving discrete programming problems. *Econometrica*, 28, 497-520.

Landau, L.D., Lifshitz, E.M., 1971. The Classical Theory of Fields. *Addison-Wesley*.

Lin, S., 1965. Computer solutions of the traveling-salesman problem. *Bell System Technology Journal*, 44, 2245-2269,.

Lin, S., Kernighan, B., 1973. An effective heuristic algorithm for the traveling-salesman problem. *Operations Research*, 21(2), 498-516.

Liu, C., Kroll, A.; 2012. On designing genetic algorithms for solving small- and medium-scale traveling salesman problems, *Proceeding of SIDE'12 Proceedings of the 2012 international conference on Swarm and Evolutionary* Computation, 283-291, Springer-Verlag Berlin, Heidelberg©.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6), 1087-1092.

Ozisikin, M.N., 1980. Heat conduction, *John Wiley and sons, Inc.*, (New York, U.S.A.).

Pizlo, Z., Stefanov, E., Saalweachter, J., Li, Z., Haxhimusa, Y., Kropatsch, W.G., 2005. Adaptive Pyramid Model for the Traveling Salesman Problem. *Workshop on Human Problem Solving*.

Platzman, L.K., Bartholdi, J.J., 1989. Spacefilling Curves and the Planar Travelling Salesman Problem. *Journal of the Association for Computing Machinery*. 36(4), pp. 719-737.

Rego, C., Glover, F., 2002. Local search and metaheuristics. *In Gutin and Punnen* , pp. 309-368.

Rosenkrantz, D.J., Stearns, R.E., Philip, I., Lewis, M., 1977. An analysis of several heuristics for the traveling salesman problem. *SIAM Journal on Computing*, 6(3), 563-581.

RECOGNIZING CHARACTERISTIC PATTERNS IN DISTORTED DATA COLLECTIONS

Tomáš Kocyan [(a)], Jan Martinovič[(b)], Pavla Dráždilová[(c)], Kateřina Slaninová[(d)]

[(a,b)]VŠB - Technical University of Ostrava,
IT4Innovations,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[(c,d)]VŠB - Technical University of Ostrava,
Department of Computer Science,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[(a)]tomas.kocyan@vsb.cz, [(b)] jan.martinovic@vsb.cz, [(c)]pavla.drazdilova@vsb.cz, [(d)]katerina.slaninova@vsb.cz

ABSTRACT

Many models and artificial intelligence methods work with the inputs in the form of time series. Generally, success of many of them strongly depends on ability to successfully manage input data, which often contains repeating similar episodes (patterns). If these patterns are recognized, they can be used for instance for indexing, prediction or compression. These operations can also be very useful for improving the already existing model performance and accuracy. Our effort is to provide a robust mechanism for retrieving these characteristic patterns from the collections that are subject of various distortions. The whole process of our pattern recognition consists of receiving the episodes, their clustering into the groups of similar episodes and deriving the representatives of each cluster. These representatives will be used for further indexing collections. This paper is focused on the last step of this process – receiving the representatives of concrete clusters using Dynamic Time Warping method.

Keywords: dynamic time warping, clustering, pattern recognition, time series

1. INTRODUCTION

Processing and analyzing time series data is very important task in many domains, especially in modeling and simulations. In this domain, time series data is often used as one of simulation inputs, or can be produced as one of the simulation outputs. For this purpose, it is appropriate to be able to manage this type of data, e.g. describe the data nature, search in data in reasonable time, or to recognize characteristic patterns in collection. If such patterns are recognized, then they may be used for instance in data compression, for prediction or for indexing large collections. Time series analysis covers the methods for analysis of time series data with a focus on extraction of various types of information like statistics and other characteristics of the data. However, the problem arises for data collections that are a subject to different types of distortions, because the patterns can differ in time, shape or amplitude. In these cases, the classic methods for pattern recognition can fail.

During time series processing, it is common that a time series is divided into a large amount of smaller parts named episodes, which are interconnected or partially overlapped (Keogh, Chu, Hart, and Pazzani 2004) and which are important for further processing. For example, interconnected outputs of hydrological models, data collections from traffic monitoring of selected stretches, or long time series divided by segmentation algorithm like Voting Experts (Kocyan, Martinovic, Podhorányi, and Vondrak 2012) can be mentioned. These obtained episodes exactly belong to a previously mentioned group of distorted collections, because there are no strictly defined rules for generating the data collection (time series). Our effort is to provide a robust mechanism for retrieving characteristic patterns just from such distorted time series. In our case, the obtained patterns will be used for further creation of an index file, which will allow much faster and more accurate searching for similar episodes in large data collections. This will be used for better and faster prediction in our Case-Based Reasoning (Aamodt and Plazza, 1994) system (Kocyan, Martinovic, Unucka, and Vondrak 2009). The structure of suggested index file is shown in Figure 1, where each of the found patterns will contain its own group of similar episodes in original data collection.

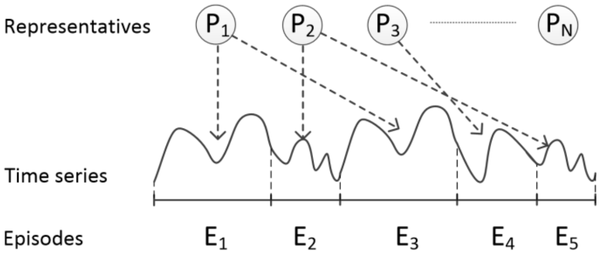


Figure 1: Collection of Representatives Pointing to Locations in Time Series

Then, once the most similar episodes in data collection will need to be found, a suitable pattern, which corresponds with the input sequence, will be searched first. Thereafter, it is possible to search in depth in a group of the selected pattern or a set of patterns, which are similar to a found episode from the input. By this way, the process of searching similar episodes will be speed up. However, there is question how to receive the patterns from distorted data collection and make the index file. Research area aimed to finding patterns, pattern mining, has been studied in several fields. Pattern mining, or pattern recognition, is a scientific discipline focused on object classification into categories or classes (Koutroumbas and Theodoridis 2008; Hand, Smyth, and Mannila 2001).

Our suggested approach is done in following manner. First of all, it is necessary to receive the particular episodes from data collection (i.e. cut the collection into the episodes). For instance, this can be done by the Voting Experts algorithm (Cohen, Adams, and Heeringa 2007) or by our unsupervised algorithm for retrieving characteristic patterns from time-warped data collections (Kocyan, Martinovic, Podhorányi, and Vondrak 2012). Once these episodes are obtained, they should be processed by a suitable clustering algorithm and divided into the clusters (Guojun, Chaoqun, and Jianhong 2007). Since each obtained cluster contains a concrete amount of similar episodes, it is suitable to select an appropriate representative, which would describe the whole cluster. Given selected representative is named pattern. Finding the representative of a cluster is defined as finding such set of representative patterns P, which describe episodes E inside these clusters by the most appropriate way. There are two basic generally known ways for finding representatives. The first approach is based on selecting one episode, which is the most accurate for a given cluster. The second approach is based on the creation of a new representative episode using the combination of episodes in the cluster.

While searching the representative, it is important to define a mechanism for comparing two episodes. In common, the Euclidean distance and other common methods for measuring the similarity between the episodes can be used. However, it is possible only while working with the undistorted episodes of the identical length. In cases where we have distorted episodes of different lengths, we need a specific algorithm which respects this requirement or an algorithm which is immune to sequence distortions. In the paper, it is described the comparison of the both approaches, and the introduction of a new approach which combines the both ways for finding representatives using Dynamic time warping method (DTW) is presented in Section 2.

The organization of the paper is following: DTW and the utilization of DTW for finding cluster representatives is described in Section 2 and in Section 3. Afterwards, in Section 4, a practical demonstration of proposed approach is presented. The paper is concluded by Section 5, in which obtained results of suggested approach are discussed and the future work is outlined.

2. DYNAMIC TIME WARPING

Recently, finding a signal similar to a signal generated by computers, which consists of accurate time cycles and which achieves a determined finite number of value levels, is a trivial problem. A main attention is focused more likely on the optimization of searching speed. A non-trivial task occurs while comparing or searching the signals, which are not strictly defined and which have various distortions in time and amplitude. As a typical example, we can mention measurement of functionality of human body (ECG, EEG) or the elements (precipitation, flow rates in riverbeds), in which does not exist an accurate timing for signal generation. Therefore, comparison of such episodes is significantly difficult, and almost excluded while using standard functions for similarity (distance) computation. Examples of such signals are presented in Figure 2. A problem of standard functions for similarity (distance) computation consists in sequential comparison of opposite elements in both episodes (comparison of elements with the identical indexes).

DTW is a technique for finding the optimal matching of two warped episodes using pre-defined rules (Muller 2007). Essentially, it is a non-linear mapping of particular elements to match them in the most appropriate way. The output of such DTW mapping of episodes from Figure 2 can be seen in Figure 3. This approach was used for example for comparison of two voice patterns during an automatic recognition of voice commands (Rabiner 1993). The main goal of DTW method is a comparison of two time dependent episodes X and Y, where $X = (x_1, x_2, \dots, x_n)$ of the length $n \in \mathbb{N}$ and $Y = (y_1, y_2, \dots, y_m)$ of the length $m \in \mathbb{N}$, and to find an optimal mapping of their elements. A detailed description of DTW including particular steps of the algorithm is presented in (Muller 2007).

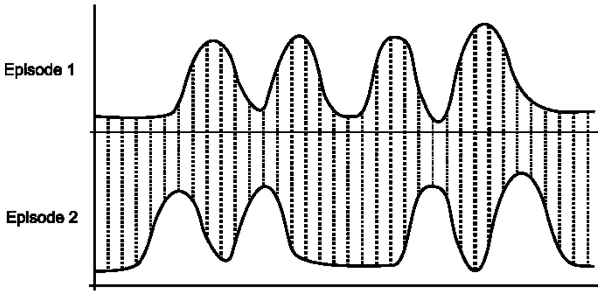


Figure 2: Standard Metrics Comparison

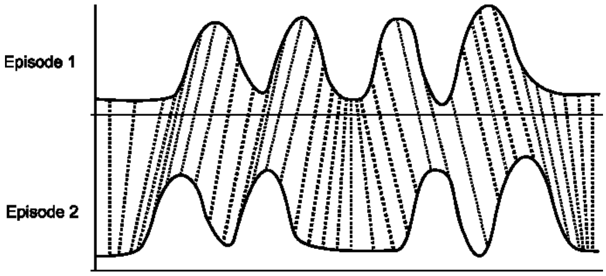


Figure 3: DTW Comparison

3. USING DTW FOR FINDING CLUSTER REPRESENTATIVE

In cases, where it is necessary to gain the most suitable representative of the set of similar episodes, we need to find an algorithm appropriate to a given domain. Sometimes it is possible to use simple average of episodes X and Y, which means that for a representative episode R is valid, that:

$$R_i = \frac{X_i + Y_i}{2}, \forall i = 1, \dots, o, where\ o = |X| = |Y|. \qquad (1)$$

However, this approach is not sufficient in cases, where we have data with distortion. Examples of such episodes are presented in Figure 4a and 4b. If only we used simple average presented in Equation 1, we would achieve an episode showed in Figure 4c. As we can see, this episode absolutely is not a representative and all the information about the episode course is loosed.

As we can see from Figure 4, it is necessary to find a more appropriate algorithm for domains which yield to distortion. The algorithm should be immune to such distortions. This paper is focused on using DTW for finding a representative of set of similar, but distorted episodes.

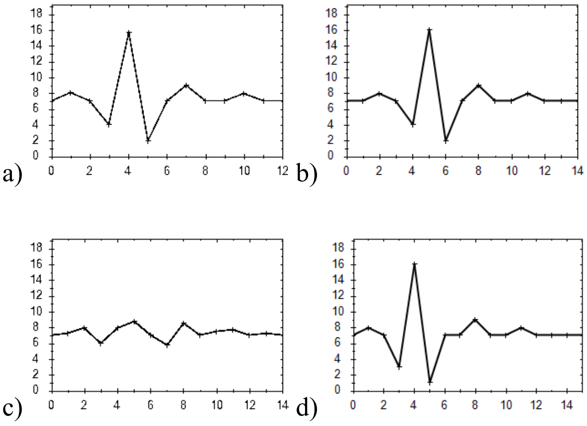


Figure 4: Sample Figure Caption

3.1. Finding Representative for Episode Couples

The approach for finding a representative of two episodes X and Y by finding the optimal mapping of two episodes using DTW was described in Section 2. In this method, the most important is obtained warping path $p^* = (p_1, p_2, \dots, p_L)$, which allows to find a representative. The approach for finding such representative is described in Algorithm 1. The output of presented algorithm applied on episodes in Figure 4 is presented in Figure 4d.

```
Algorithm 1: Searching Representative of a Pair
Input:       Episodes $X$ and $Y$
Output:      Representative episode $R$
Steps:
1. Compute $DTW(X,Y)$ for episodes $X$ and $Y$; obtain
   warping path $p^*$.
2. Initialization: $R$ is a representative episode for
   episodes $X$ and $Y$, $q = 1$ gives a position in $R$,
   $l = 2$ gives a position in warping path $p^*$.
3. Value in the first position in $R$ is determined as
   average of values in the first positions of episodes $X$
   and $Y$, ie. $r_1 = \frac{x_1 + y_1}{2}$.
4. if $l \leq L$ then for couple of the subsequent points of
   warping path $p_l$ and $p_{l-1}$ perform:
       if $(p_l - p_{l-1}) = (1,1)$ then:
           - $q = q + 1$;
           - A new $r_q = \frac{x_{n_l} + y_{m_l}}{2}$ is inserted into $R$;
       else if $(p_l - p_{l-1}) = (0,1)$ or $(1,0)$ then
           - no item is inserted into $R$;
       end if
   • $l = l + 1$
   • Repeat Step 3.
5. end if
6. Output of the algorithm is representative episode $R$
   of length $q$.
```

Algorithm 1 finds a representative common for two episodes, where both episodes have the same importance. It finds such episode, which is the most similar to the both two episodes. If it is necessary, a one of the episodes may be preferred by adding a weight $w \in (0, \infty)$ and by adjusting a computation of element r_1 and r_q by Equation 2:

$$r_1 = \frac{x_1 * w + y_1}{w+1}, r_q = \frac{x_{n_l} * w + y_{m_l}}{w+1}. \qquad (2)$$

The impact of adding a weight on achieved representative R for episodes X and Y is following:

- $w = 1$: episodes are equal
- $w \in (1, \infty)$: episode X is preferred
- $w \in (0,1)$: episode Y is preferred

3.2. Finding Representative for Set of Episodes

Algorithm 1 can be applied only on two episodes. However, this is often insufficient in common practice; we need to find a representative for the whole set of episodes in most cases. Given a collection C with generally N episodes $C = (e_1, e_2, \dots, e_N)$. The question is, how the presented approach applies on generally N episodes. A first solution is based on an approach, in which a representative is found step by step by finding particular representatives for episode couples. More precisely, the first step consists of finding representative R_{1-2} for the first two episodes e_1 and e_2. Then, representative R_{1-2-3} is found for a new obtained episode R_{1-2} and for episode e_3. Then, such approach is used for the rest of episodes in the cluster.

However, our experiments showed that this approach is not as much suitable as it could be. It is strongly dependent on the order of particular episodes in collection. The solution is to find an approach that would be immune to the order of elements in an episode. Our proposed approach which solves this problem is presented in Algorithm 2.

```
Algorithm 2: Searching Representative of a Set
Input:      Collection $C$ of $N$ episodes
Output:     Representative episode $R$
Steps:
1. Initialization: $N$ is count of input episodes, $u$=1 is
   level of collection; $C^1 = C$ is the first level of
   collection; $M = N - u + 1$ is count of processed
   episodes in level $u$.
2. Create from collection $C^U$, which consists of
   episodes $\{e_1, e_2, \dots, e_N\}$ distance matrix $D^U \in \mathbb{R}^{M \times M}$,
   where particular matrix elements are defined
   as $d_{ij}^u = DTW(e_i^u, e_j^u)$, i.e. matrix elements are
   created by values of reciprocal mapping of particular
   episodes.
   Calculate sum for each row $r_i^u$ in matrix $D^U$ and
   select a row with the lowest sum value. Find row
   $r_{min}^u$ where $\sum_{j=1}^{M} d_{min,j}^u = min_{\forall i=1,\dots,M}(\sum_{j=1}^{M} d_{ij}^u)$.
   The found row refers to the episode, which is
   selected as the most similar to the others in the
   current collection, and which could be declared as
   representative $R^U$ of the collection for $u$-th level.
3. Remove representative $R^U$ from the current
   collection and create $(N - u)$ new episodes by
   application of method for searching representative
   from couple $(R^U, e_i^u)$, described in Section 3.1. This
   algorithm can be modified by adding weight
   (preference) to one of the episodes, which can prefer
   (or discriminate) the importance of the
   representative $R^U$.

   if $M > 2$ then:
       - $u = u + 1$;
       - $M = M - 1$;
       - Repeat from Step 2 for remaining
         $(N - u)$ episodes;
   else if $M = 2$ then:
       - Select a representative from the two
         episodes as a representative of the
         whole original set of episodes $C$;
   end if
```

The presented approach is not restricted only to using DTW as a method for the expression of episode similarity. Of course, DTW could be replaced by any other indicator, for example Euclidean distance or statistical indicators for time series (Mean Absolute Error, Mean Percentage Error, Root Mean Square Error etc.). In such cases, it is necessary to adapt steps 2 and 4 of Algorithm 2, where instead of finding a representative for the episodes couple by DTW is necessary to use (weighted) average of two compounded episodes. Section 4 describes both two approaches with a visual comparison of the impact to a found representative.

4. EXPERIMENTS

In this section, a practical demonstrations of previously introduced methods are presented. First of all, the step by step example for better understanding of proposed algorithm will be demonstrated. Then, several outputs of the algorithm will be showed.

It must be noted that meaning and usage of DTW method is closer to a human judgment and perception of similarity than a machine definition of physical distance. For this reason, it is hard or almost impossible to perform a numerical evaluation for the following outputs (Berndt and Clifford, 1994), so the results will be presented only visually.

4.1. Step by Step Example

Consider we have two episodes $X = (1,\ 1, 4, 1, 10, 1)$, $Y = (1,\ 6, 1, 1, 10, 10, 1)$ and we want to find their mutual representative. First of all, a distance matrix (see Table 1), accumulated distance matrix (see Table 2) and warping path $p^* = \{(1,1),\ (2,1),\ (3,2),\ (4,3),\ (4,4),\ (5,5),\ (5,6),\ (6,7)\}$ have to be found according the steps listed in (Muller 2007). Visualization of found mutual episodes' mapping can be seen in Figure 5.

1	0	0	9	0	81	0
10	81	81	36	81	0	81
10	81	81	36	81	0	81
1	0	0	9	0	81	0
1	0	0	9	0	81	0
6	25	25	4	25	16	25
1	0	0	9	0	81	0
X/Y	**1**	**1**	**4**	**1**	**10**	**1**

Table 1: Distance Matrix

1	187	146.5	75.5	66.5	83	2
10	187	146.5	66.5	71	2	42.5
10	106	65.5	30.5	57.5	2	83
1	25	12.5	17	2	42.5	17
1	25	12.5	11	2	62	17
6	25	12.5	2	21.5	17	42
1	0	0	9	9	90	90
X/Y	**1**	**1**	**4**	**1**	**10**	**1**

Table 2: Accumulated Distance Matrix

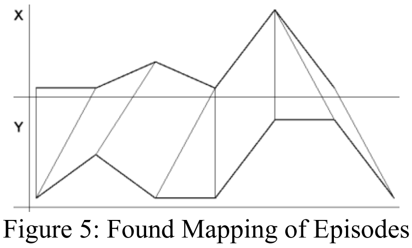


Figure 5: Found Mapping of Episodes

Now, the process of searching a representative can start. The first element of the representative is determined as $r_1 = \frac{x_1+y_1}{2} = \frac{1+1}{2} = 1$. Then, we move to the next pair in the warping path. Since $(p_2 - p_1) = (2,1) - (1,1) = (1,0)$, no new element is added to the representative episode. However, the next step $(p_3 - p_2) = (3,2) - (2,1) = (1,1)$ causes an addition of a new element r_2, where $r_2 = \frac{x_3+y_2}{2} = \frac{4+6}{2} = 5$. In the same way, the rest of the representative R is constructed:

$(p_4 - p_3) = (4,3) - (3,2) = (1,1) \rightarrow r_3 = \frac{1+1}{2} = 1$
$(p_5 - p_4) = (4,4) - (4,3) = (0,1) \rightarrow nothing\ to\ add$
$(p_5 - p_4) = (5,5) - (4,4) = (1,1) \rightarrow r_4 = \frac{10+10}{2} = 10$
$(p_6 - p_5) = (5,6) - (5,5) = (0,1) \rightarrow nothing\ to\ add$
$(p_7 - p_6) = (6,7) - (5,6) = (1,1) \rightarrow r_5 = \frac{1+1}{2} = 1$

The final found representative is then specified as $R = (1, 5, 1, 10, 1)$.

4.2. Demonstrations of Algorithm Outputs

As it was mentioned earlier in Section 4, it is almost impossible to numerically evaluate the success of the algorithm. For this reason, the following samples of output will be demonstrated only graphically and each of the result will be visualized in the following manner: the first row contains episodes, which were used as the input to the algorithm, whereas the second row consists of outputs for the different approaches. The first output is always the average of input episodes (defined in Equation 1), the second output is obtained from the proposed approach described in Section 3.2, and in the third case the Euclidean distance instead of DTW is used.

The first input dataset was a set of similar signals (see Figure 6), which shapes resemble ECG records. The signal is ended with tiny swings. As we can see from the second row of the episodes in Figure 6, the average of values from the both episodes absolutely degraded the signal information; the shift of signal peaks and drops was smoothed nearly to one level. Also usage of Euclidean distance did not provide sufficient results, which did not differ from averaged outputs much. On the other way, usage of DTW method for finding representative fully depicted a character of the signal and brought the most accurate results.

The next set of episodes contains signals with the three peaks mutually shifted in time, while each of them had a variable duration (see Figure 7). It was supposed that the representative would have a curve with the three evident peaks. It is obvious from the results, that even though the Euclidean distance worked much better, the loss of information was still noticeable.

The last input dataset represents the situation, in which the signal consists of two waves - one in a positive and one in a negative part (see Figure 8). These waves were deformed in time, while they were spread or shrunk in X axis. Although the other methods achieved seemingly the best results, the distortion was evident again. The output representative did not contained as high amplitudes as the input waves, did not have smoothed waves and did not detect the constant segments, which were distorted.

The most important advantage of the proposed solution is the fact that the Algorithm 1 in combination with DTW is able to process even episodes with different lengths. This is very difficult while using other methods. In these cases it is necessary to shrink the episodes into the identical length, which of course cause the loss of information. Using DTW, we are able to process such episodes with different lengths without any loss of information. In Figures 9 and 10, there are presented outputs from proposed algorithm applied on episodes with different lengths.

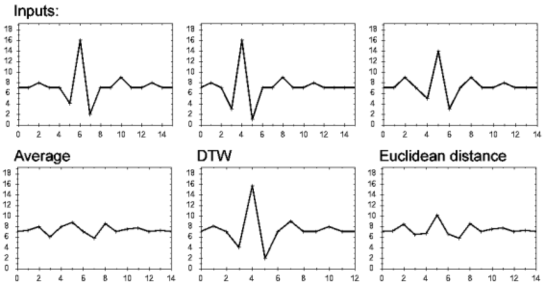


Figure 6: First set of inputs and outputs

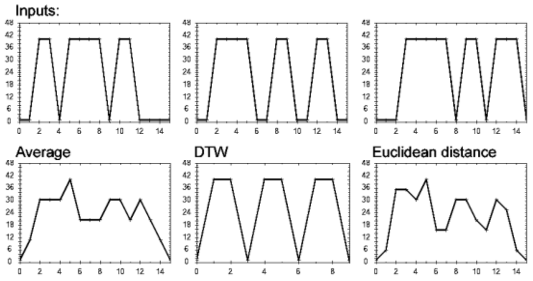


Figure 7: Second set of inputs and outputs

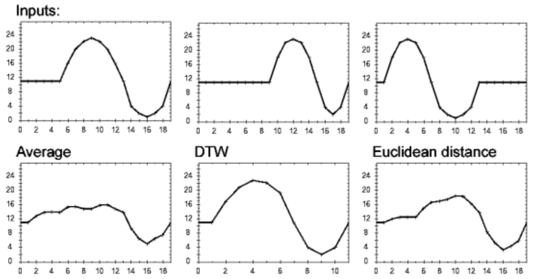


Figure 8: Third set of inputs and corresponding outputs

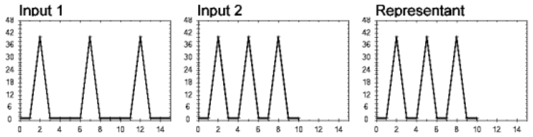


Figure 9: First set of inputs with variable lengths

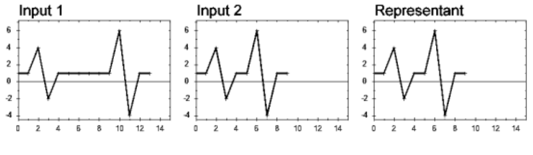


Figure 10: Second set of inputs with variable lengths

Observing the value of the weight parameter and its influence onto the resulting representative is also very interesting. If the weight parameter is not set (respectively if the weight is set to 1:1 – it means that the importance of the current level representative R^U and other episodes is equal), searching representative looks like as in the Figure 10. If the weight is set to 10:1

(the R^U is strongly preferred), the episodes in the next level will be strongly influenced by R^U as it is shown in Figure 11. On the other hand, if the weight is set to 1:10, the influence of R^U is minimal (in substance the R^U almost ignored) and the episodes in the next level are almost unchanged (see Figure 12).

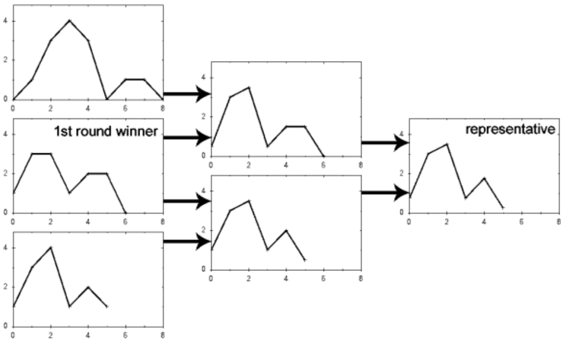


Figure 10: Weight parameter set to 1:1

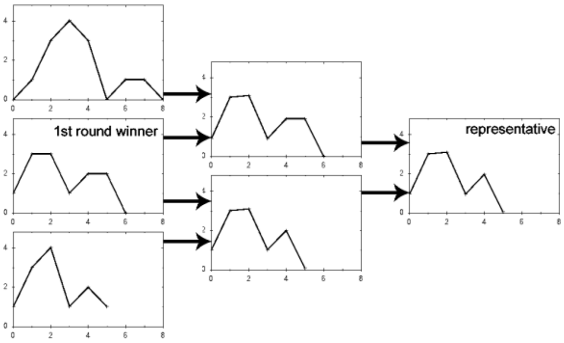


Figure 11: Weight parameter set to 10:1

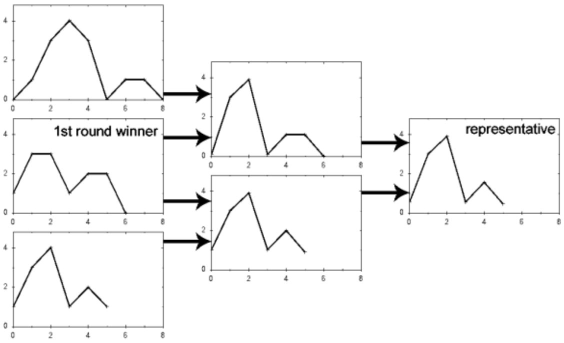


Figure 12: Weight parameter set to 1:10

5. CONCLUSION AND FUTURE WORK

The real application of proposed algorithm showed that it is able to find a representative not only from the set of typical episodes, but also from their distorted variants. The tested input datasets consisted of signals with changed amplitudes, and which were distorted by time shifting. The proposed solution was compared with conventional methods, in which much worse success was obvious.

Further work will be focused on creation of index file, which structure was defined in Section 1, and which visual representation was presented in Figure 1. The aim is to create a sufficiently robust mechanism, which will be able to find all the similar episodes to the selected pattern in data collection during the shortest time. Furthermore, these found episodes will be used for a prediction using the Case-Based Reasoning method. This method requires a suitable mechanism that is able to extract the most similar patterns from the input.

ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the grant project of the Technology Agency of the Czech Republic under number TA01021374, titled New Technologies for Environmental Protection against Negative Effects of Natural Mass Transfer.

REFERENCES

Aamodt, A., Plaza, E., 1994. *CBR: Foundational issues, methodological variations, and system approaches,* AI COMMUNICATIONS, 39-59.

Berndt, D., Clifford, J., 1994. *Using Dynamic Time Warping to Find Patterns in Time Series*, In KDD Workshop (1994), pp. 359-370.

Cohen, P. R., Adams, N., and Heeringa, B, 2007. Voting Experts: An Unsupervised Algorithm for Segmenting Sequences. *In Journal of Intelligent Data Analysis*.

Evans, W.A., 1994. Approaches to intelligent information retrieval. *Information Processing and Management*, 7 (2), 147–168.

Guojun, G., Chaoqun, M., Jianhong, W, 2007. Data Clustering: Theory, Algorithms, and Applications. *ASA-SIAM Series on Statistics and Applied Probability*.

Hand, J., Smyth, P., Mannila, H, 2001. *Principles of Data Mining,* MIT Press, ISBN: 0-262-08290-X.

Keogh, E., Chu, S., Hart, D., Pazzani, M., 2004. *Segmenting Time Series: A Survey and Novel Approach,* Data mining in Time Series Databases, World Scientific, 1-21.

Kocyan, T., Martinovic, J., Podhorányi, M., Vondrak., I., 2012. Unsupervised algorithm for retrieving characteristic patterns from time-warped data collections. *Proceedings of The 11th International Conference on Modeling and Applied Simulation*.

Kocyan, T., Martinovic, J., Unucka, J., Vondrak., I., 2009. FLOREON+: using Case-Based Reasoning in a system for flood prediction. *Proceedings of ZNALOSTI 2009*.

Koutroumbas, K., Theodoridis S., 2008. *Pattern Recognition, 4th Edition.* Academic Press, ISBN:9781597492720 .

Liao, T. W., 2005. Clustering of time series data—a survey. *Journal of Pattern Recognition*, Volume 46, Issue 9, 2391-2612.

Muller, M., 2007. Dynamic Time Warping. *Information Retrieval for Music and Motion*, Springer, ISBN 978-3-540-74047-6, 69-84.

Rabiner, L., Biing-Hwang Juang, 1993. *Fundamentals of Speech Recognition,* Prentice Hall, ISBN: 0130151572.

ECONOMIC ANALYSIS OF A LOW-COST LAPAROSCOPIC SIMULATOR: A DESCRIPTIVE STUDY

M. Frascio[a], M. Sguanci[b], F. Mandolfino[c] , A. Testi[d] , M. Parodi, G. Vercelli[e]

[a]DISC - Department of general surgery
[b]DISC - Department of general surgery
[c]DISC - Department of general surgery
[d]DIEM - Department of economics and quantitative methods
[e]DIBRIS – Department of Informatics, Bioengineering, Robotics and Systems Engineering

[a]mfrascio@unige.it, [b]marcosguanci@yahoo.it, [c]fcmandolfino@gmail.com [d] testi@economia.unige.it
[e]gianni.vercelli@unige.it

ABSTRACT

Simulation plays a basic role in medical education. At School of Medicine in Genoa a prototype of virtual reality simulator for videolaparoscopy is now in development phase. This simulation platform includes haptic interface for force feedback and autostereoscopic displays for glasses-free 3D rendering.

Aim of this study is to present the simulation as a training method in medicine for laparoscopic surgery. The analysis also studies the economic impact of a market-ready, low cost, multi-user virtual reality simulation platform engineered starting from prototype currently under development.

Videolaparoscopy application is analyzed. The investigation is based on 2011-2012 data obtained from queries to "Datawarehouse Sanitario" of Regione Liguria. Secondly, benefits provided by the introduction and use of prototype of a virtual reality simulation platform prospectively and quantitatively examined.

The processing relates the possible reduction in total length of hospital stay due to the introduction of prototype simulator and the consequent reduction of laparotomic cholecystectomies.

Keywords: training, skill, laparoscopic surgery, simulator, haptic feedback

1. BACKGROUND

Simulation plays a basic role in medical education. In Italy there are many "physical" simulators and few virtual reality simulators for videolaparoscopy training.

University of Genoa includes an Advance Simulation Center in which a virtual reality simulator prototype for videolaparoscopy is now in developing.

This prototype, eLaparo4D, has been designed as a simulation platform allowing an immersive training space for videolaparoscopic surgery. In particular elaparo4D is a low-cost training space which integrates haptic devices with realistic surgery tools and 3D rendering with physically deformable 3D CG models of the human internal organs.

The simulator is based on a client/server layered architecture in order to act as a sort of data gateway: the hardware is interfaced with the physics 3D engine to obtain real-time performances, with an HTML5-based 3D output visual interface in order to integrate the tracking of operating sessions within a custom training platform.

1.1 The literature review

Despite in Italy the spread is limited, evidences on the clinical and medical virtual reality simulators have been widely documented.

Some research has focused on basic laparoscopic techniques. A study carried out by "University of Michigan analyzed how training with virtual reality simulator (LapMentor) can improve the performance of organizing some general laparoscopic techniques.

Two groups of surgeons have performed a specific training for six basic laparoscopic procedures (exercises for the operation of the camera, the eye-hand coordination, for cutting and grasping, for suturing, for the "electrocautery and for moving objects). One group used the simulator, while the "other, control, has followed the training course provided for by "University. After each subject had reached the level of proficiency required, the two groups have carried out a series of exercises on anesthetized pig. Such exercises were then evaluated by two experts. For each type of exercise, the running time of the tests was significantly lower for the group that had carried out the training on the simulator. Similarly, the precision in the fulfillment of the procedures performed was superior for subjects using the simulator.

A similar study was done by Korean researchers. The authors used a simulator with haptic feedback to test the improvements in the training process both medical students without previous knowledge of both laparoscopic trainees with previous experience in minimally invasive surgery. The subjects performed five sessions of the simulator tests, spaced in time, in which four exercises performed laparoscopic technical base: grasping, cutting, clipping, suturing. For each test was evaluated the execution time and the level of accuracy through the measurement of some parameters. The study showed that the "use of the simulator allows rapid and significant improvements in the accuracy in the" execution of the tasks offered. Even the time spent in completing exercises decreases significantly in the progress of the sessions. In addition, the curves to improve performance and decrease the execution time are similar for both groups in question, although composed of individuals with different levels of knowledge laparoscopy.

Another study carried out by researchers in Sweden analyzed the "effect of" use of the simulator in "learning basic laparoscopic techniques considering the first 10 laparoscopic cholecystectomies performed by interns. Thirteen subjects were divided into two groups. The first group carried out the simulator training for six specific tasks at three levels of difficulty, until you reach the level of proficiency required to operate. The second group, the control, on the other hand reached the level of competence necessary to follow the training classic. For the study we used a simulator without haptic interface for force feedback (LapSim). Were then recorded and analyzed by groups of experts the first 10 laparoscopic cholecystectomies performed by each subject.

The group that has carried out the training to the simulator is the best result in a statistically significant manner, as regards the operating performance. The total number of errors committed during the operations analyzed turns out to be three times lower than the control group (figure 4).

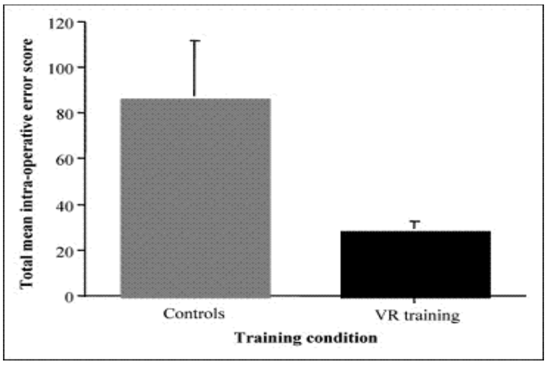


Figure 4: Number of errors in laparoscopic cholecystectomy

The surgical intervention has been divided into three distinct phases, and the ratio of the number of errors is maintained in each of these three parts. Furthermore, the total time of the "operation of the control group is greater than 58%, although this result was not considered significant from a statistical standpoint.

2. OBJECTIVE

The aim of analysis is to study the economic impact of a market-ready, low cost, multi-user virtual reality simulation platform engineered starting from prototype currently under development.

3. MATERIALS AND METHODS

3.1 The simulator system

The system is based on a nodejs application server that manages the visualisation system, the communication with hardware interfaces and the database where users' data are stored. The server technology is indeed a sort of data gateway between the several different elements, regardless they are hardware or software. The following figure (figure 1) shows how communication data are exchanged from the very low part of the system (Hardware Interfaces, bottom) to the user interface (HTML Client, top).

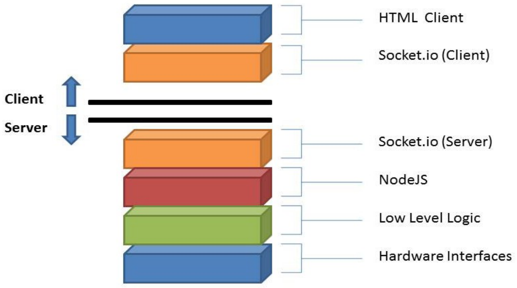


Figure 1: part of the system simulation

The user interface is a simple HTML5 web page running a Unity3D engine plugin. We run several performance tests to compare Unity3D and native WebGL, getting same results. We finally decided to adopt Unity3D engine due to its rapid development time. WebGL is a great technology but still too young to allow us working on a powerful and robust framework. The use of web pages as the main user interface allows us to be more versatile and in the future will give us the possibility, thanks to HTML5 powerful characteristics, to easily share contents in a live way with other systems. An interesting feature is, for example, having the possibility to be guided by an external supervisor, who is monitoring the training phase, while data are quickly exchanged via internet.

3.1.1 Visual and fisical modelling

As previously introduced, visual modelling is a very important aspect of the entire project. A videolaparoscopic surgery simulator needs a detailed representation of the organs and the tissues inside of the human abdomen. The meshes included in eLaparo4D are developed in Blender 3D Modelling software, and then imported in Unity3D, including textures and UV maps. Eventually, in Unity3D render materials are added to the raw meshes, to simulate the specific surface of each of the modelled tissues. In Figure 2, a screenshot of the current virtual environment is shown.

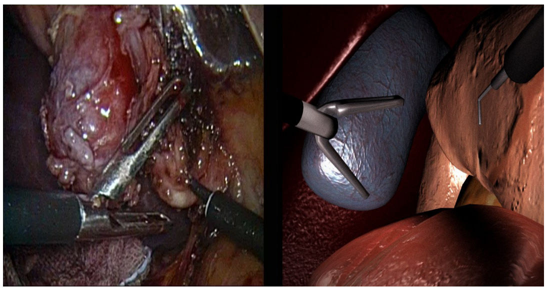


Figure 2: a screenshot from the current aspect of the virtual environment compared to a screenshot of the camera view of a real surgical operation.

As remarked by our colleagues of the Videolaparoscopy Unit of the Department of Clinical Surgery, highly specific training sessions are required to help the operator achieving a proper skill set. In an ideal scenario, medical students should have access to a complete simulator composed of several training scenes, as part of a modular and step-based training process. While the main components and controls of the simulator should be in common, each scene should focus on a very specific surgery operation, differentiating in: the zone and the organs physically manipulated (the target), the particular surgical maneuvers performed (the task), and the type of manipuli used (the means).Considering these remarks, we developed a dynamic parametric physical simulation approach, arbitrary applicable to the rendered meshes in every scene and able to avoid system overloads. Such an approach permits the creation of different scenes starting from the same set of models and interaction algorithms, easily supporting a step-based training. In detail, each 3D object in the scene carries a selectable 3 layer collider component, driving a vertex deformation script. The first layer is a simple box collider; the second one is a combination of simple shape colliders which cover, with good approximation, nearly all the volume of the object; the third is a precise mesh collider which exactly coincides with the vertex disposition of the object's mesh. In the following figure (figure 3) is possible to see the 3 different collider layer for a gallbladder model.

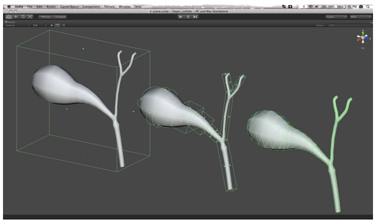


Figure 3: I.e of a collider layer for a gallbladder model

3.1.2 Feedback system

Haptic feedback is implemented thanks to the use of three Phantom Omni devices from Sensable. The first two are used as manipuli (grasper, hook or scissors) and the third one is used to move the camera within the virtual abdomen, as it happens in a real scenario. The system generates a resultant force when the user puts a manipulus in contact with a mesh, according to the executed task. Phantom devices have been chosen because reasonably low cost although precise enough for the needed level of realism. Furthermore, their stylus-like shape will permit a complete merging of the devices with the physical environment reconstruction; in particular, each stylus will be easily connected to real manipuli. Thanks to an Arduino board connected to a vibrating motor we have also included a vibration feedback. Vibration is used to enhance the realism of operations like tissue shearing (hook) and cutting (scissors).

3.2 The statistical economic analysis

In general surgery there are Numerous kind of interventions which can be performed with laparoscopic access. For some types of these, such as cholecystectomy and plastic gastro-oesophageal reflux, the advantages of "laparoscopic approach are now defined and well-established, so that the laparoscopic technique represents the gold standard for the" execution. Other interventions via laparoscopic access is by way of affirmation but still require satisfactory clinical evidence, as in the case of "appendectomy. Other interventions are still quite complicated or require a learning curve quite wide, such as surgery of the colon and rectum, liver resections and pancreatectomy, such as to be carried out only in highly specialized centers.

An analysis of the DRG used in Italy, which classify each episode of hospitalization in homogeneous groups for absorption of resources involved, showing how a single operation, cholecystectomy, showing already divided at the level of DRG surgical approach employed.

Even a "cataloging analysis of surgical procedures by ICD-9 codes reveals that the" indication of "surgical approach used is present only for a very few interventions, and only in the case of" appendectomy clearly divides the "intervention depending on the technique used. The purpose of this study was examine the use of laparoscopic approach, and the benefits present and future, in relation to cholecystectomy.

This intervention was also one of the first interventions for which guidelines have explicitly defined as the “gold standard” laparoscopic procedure for all clinical cases that do not present obvious complications or severe comorbidities.

Firstly, the current Ligurian region (north-west of Italy) environment concerning videolaparoscopy applications was analyzed in order to understand the features of future market segment of such kind of simulator. The investigation is based on 2011 data and it refers to Ligurian hospitals cholecystectomy surgical operation. Laparoscopic approach application, presence of complications and comorbidities, average length of hospital stay and splitting of hospitalization cases for classes of severity data are shown.

Data are obtained from queries to a specific regional database called “Datawarehouse Sanitario”.

The system allows an additional classification of interventions in clinical severity classes through the" use of APR – DRG.

The APR - DRG (All Patient Refined - DRG) are a classification system of "episode of hospitalization depending on the severity of the clinical condition of the patient, and allow you to review the role of the complications and comorbidities that is not fully central DRG classification . Every APR - DRG is divided into four classes that differentiate patients in relation to the complexity of care, the severity of the disease and the risk of death: the lower or absent, moderate, major, extreme. The attribution of the classes is realized through a specific software that processes data from the Hospital Discharge Data using a complex algorithm that takes into account mainly of secondary diagnoses present.

Are then extrapolated from the data warehouse, through the "use of filters in cascade, all data concerning the" cholecystectomy further divided into classes of severity. This prospective analysis is performed according to two different points of view:

- the division of the interventions by DRG
- the division of clinical interventions by classes of severity

For the purposes of processing it is assumed that, thanks to the introduction of the simulator, the percentage of interventions decrease (that are deemed inappropriate) depending of "surgical approach used in relation to the clinical conditions of the case.

4. RESULTS AND DISCUSSION

The economic analysis can demonstrate a reduction in total length of hospital stay due to the introduction of prototype simulator and the consequent reduction of laparotomic cholecystectomies estimated as not appropriate.

In Italy, gallstones affects the rate varying from 10% to 20% of the population. Cholecystectomy is now a routine surgical operation, which presents a operative mortality low, about 0.1 to 0.05%.

Even morbidity, namely on lethal complications, is very modest. It's estimated that at the national level are carried out approximately 100,000 cholecystectomies every year, so Liguria, with the interventions of 2825 2011 therefore represents a portion percentage around 2-3%.

A regional level interventions are distributed in a more or less uniform among the various Health and hospital in Liguria. For what concerns instead the fees associated to the DRG relative to cholecystectomy, both to those adopted in 2009 by the Liguria Region, both for newer ones introduced in 2013 at national level, the main differences given to the presence of complications or comorbidities.

However, a level playing field clinics, the rate is higher for the laparoscopic approach. Laparoscopic surgery (compared to laparotomy) maybe at a higher cost of "intervention in itself, because of the" use of a "high-tech equipment and instrumentation.

The intervention as the case of hospitalization has a lower cost, mainly due to the significant reduction in hospital days post-op.

The cholecystectomies in Liguria are in most cases (73%) associated with the DRG 494, namely the intervention that does not present complications or comorbidities performed by laparoscopic approach.

In the remaining percentage the greater portion of the work is associated with the DRG 49 (18%).

Finally, smaller percentages are relative to the DRG 197.

The laparoscopic technique is now widely recognized for many years to internationally as the "gold standard approach for the cholecystectomy.

Thanks to the advantages that the approach laparoscopic approach attorney, fundamentally linked to the reduction of hospitalization, currently the laparoscopic access is indicated not only in healthy adult subjects, but also in children and the elderly.

About training, the result very good about the current use of laparoscopic access shows how the surgeons much sympathetic use this approach. In an learning optical, the starting point is already possess a relevant knowledge of such techniques that will allow cost savings of the training.

This advantage is not, however, currently, quantifiable.

The decision to intervene by laparoscopy allows a decrease in hospitalization average of approximately 7.5 days in the case of presence of complications and about 3.5 days in case of absence of complications.

In both cases, both in the presence that in the absence of complications and comorbidities, the corresponding average hospital stay in lower for the complete laparoscopic approach (45-46%) than the laparotomic one.

The analysis for classes of severity with APR - DRG examines in more the specific situation from the point of view of the clinical condition of the case of hospitalization. The cholecystectomies are classified using the severity as parameter in minor (74%) and in moderate (21%).

Only the remaining 5% presents clinical conditions rather critical with greater severity if extreme. In general, the percentage of operations performed laparoscopically increases with decreasing the severity of the clinical case of hospitalization.

If compared to a reduction of inappropriate interventions and higher tariff are attending increase charges and appropriate interventions with lesser rate, would produce a savings from the Region for the remuneration of interventions cholecystectomy.

In this case, the savings could range from 14,754 € due to the change of only 6 interventions inappropriate in the corresponding appropriate, until you get to € 322,129 for the ideal situation for which all 131 inappropriate interventions are carried out by laparoscopy. The parallel processing done with classes of severity APR – DRG showed similarly positive results. In this case are considered a inappropriate interventions classified with severity moderate or minor were performed by laparotomic procedure.

The prospective study has suggested the decrease of this type of interventions.

Unlike the previous processing, in this case, equal to percentage decrease the reduction of hospital days is greater. For a reduction of 57 interventions (-2% of the total) approached by laparotomy instead laparoscopic in the absence of complications (about 27% of the interventions considered inappropriate in this process), the decrease of total hospital stay expressed in number of days was 301. The theoretical ideal situation, for which the number of interventions considered inappropriate in this analysis were reduced to zero, would result in the reduction of 1,136 days of hospitalization at regional level.

The prospective analysis on the benefits of the introduction of the simulator in Liguria suggests a possible intervention on the inappropriateness of some types of cholecystectomies.

The premise of the clinical analysis, instead, it's based on evidence, demonstrated by numerous scientific articles where the use of the simulators for learning laparoscopic techniques improves performance and efficiency of surgeons.

In Liguria the introduction of virtual reality simulator for videolaparoscopy, equipped with a highly technology reproducing a totally realistic scenario, and its use as a support for the learning in curricular path or as exercise device for surgeons, would lead to a greater use of laparoscopic approach instead of laparotomic. The surgeons, as demonstrated by several studies, would show in safer to undertake the decision to perform the surgery using the minimally invasive procedure.

In this sense would be decreased interventions carried out by laparotomy that are achieved despite the absence of complications and comorbidities or presence of a condition with low clinical severity, or most would follow the indications given by the guide lines.

From the point of view of cost savings, the benefits would be for both the Region (in the case would use the new rates for DRG-2013) for what regarding the remuneration of performance, both for the individual hospital that would reduce its cost of production for a single admission since the days of hospitalization would be much lower.

The laparoscopic simulator could include a system of evaluation of surgeons. If the high degree of realism were verified and approved, you would have a tool with which it would be possible to check in a objective skills of those who work.

It could, for example, evaluate the improvements made during the learning curve, or verify through periodic testing the degree of knowledge of the techniques of those who already work routinely. The possibilities, of course, from this point of view are innumerable. The simulator then it could also be implemented at the level technology in order to simulate operations "tailored".

Using radiological data, you could virtually recreate the endo abdomial anatomy of the patient who will be operated.

In this way the whole operation that will be carried out could be simulated, addressing and resolving any difficulties and problems in a completely safe ambient.

ACKNOWLEDGMENTS

We would like to thank We thank G. Vercelli of Dept. of Informatics, Bioengineering, Robotics and Systems Engineering, A. Testi of Dept. Of Economics and Quantitative Methods (DIEM) together with M. Camia and E. Commisso of EMAC, for their precious collaboration.

REFERENCES

Ahlberg G, Enochsson L, Gallagher A G, Hedman L, Hogman C, et al. 2007. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American Journal of Surgery,* June; (193)6: 797-804

Andreatta P B, Woodrum D T, Birkmeyer J D, Yellamanchilli R K, Doherty G M, Gauger P G, Minter R M. 2006. Laparoscopic Skills Are Improved With LapMentor™ Training. *Annals of Surgery*, June; (6)243: 854–863

Attinà G, Rulli F, Galatà G, Ridolfi C, Tucci G, Grande M. 2007. Appendicectomia laparoscopica vs tecnica tradizionale: rapporto costi-benefici. *Congresso Società Italiana di Chirurgia,* 109 Verona

Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. 1998.

Development of a model for training and evaluation of laparoscopic skills. *American Journal of Surgery,* 175(6):482–487

Dhariwal K, Prabhu R Y, Dalvi A N, Supe A N. 2007. Effectiveness of box trainers in laparoscopic training. *Journal of Minimal Access Surgery*, Apr-Jun; (3)2: 57-63

Kim T H, Ha J M, Cho J W, You Y C, Sung G T. 2010. Assessment of the Laparoscopic Training Validity of a Virtual Reality Simulator (LAP Mentor™). *Korean Journal of Urology*, Nov; (11)51: 807

Köhler L.1999. Endoscopic surgery: what has passed the test? *World Journal of Surgery*, Aug;23(8):816-24.

Lamata P, Gomez E J, Sanchez-Margallo F M, Lamata F, Antolin M, Rodriguez S, Oltra A, Uson J. 2005. Study of Laparoscopic forces perception for defining simulation fidelity *Studies in Health Technology and Informatics*, (119): 288-292

Nuzzo G, Giuliante F, Murazio M. 2008. Complicanze biliari della colecistectomia: la gestione della fase acuta delle lesioni iatrogene della via biliare principale. *Ospedali d'Italia Chirurgia*, 14:126-136

Paige J T, Kozmenko V, Yang T, Paragi Gururaja R, Hilton C W, Cohn I, Chauvin S W. 2009. High-fidelity, simulation-based, interdisciplinary operating room team training at the point of care *Surgery*, Feb; (2)145: 138-146

Zendejas B, Wang A T , Brydges R, Hamstra S J, Cook D A, 2013. Cost: The missing outcome in simulation-based medical ed ucation research: A systematic review. *Surgery*, Feb; (2)153: 160-76.

LAPAROSCOPIC SKILLS SIMULATOR: A GRADUAL STRUCTURED TRAINING PROGRAM FOR ACQUIRING LAPAROSCOPIC ABILITIES

M. Frascio[a], M. Sguanci[b], F. Mandolfino[c] , M. Gaudina[d] , E. Bellanti[e]

[a]DISC - Department of general surgery
[b]DISC - Department of general surgery
[c]DISC - Department of general surgery
[d]DIBRIS – Department of Informatics, Bioengineering, Robotics and Systems Engineering
[e]DIBRIS – Department of Informatics, Bioengineering, Robotics and Systems Engineering

[a]mfrascio@unige.it, [b]marcosguanci@yahoo.it, [c]fcmandolfino@gmail.com [d]marco.gaudina@gmail.com
[e]edoardo.bellanti@gmail.com

ABSTRACT

Aim of this study is to investigate the importance of acquiring basic and avdanced laparoscopic skills with a virtual reality low cost simulator in laparoscopic general surgery.
We considered six basic skills and five advanced skills.
The first exercises are related to the acquisition of tasks which allow students to reach basic gestures competences.
In the second phase students will perform complex drills acquiring a correct gesture in the use of specific instruments.
We developed a standardized, graduated and evidence-based curriculum.
The team designed a new software able to handle the training task, creating a virtual interface based on the concept "student - exercise – evaluation".
The results are “attended results” because the data analysis will be possible only after a period of testing of the simulator on different samples of students.
Referring to experience from other scientific groups, we expect significant results in terms of: reduction of learning time, better dexterity and ability to intervention in case of procedural errors.

Keywords: laparoscopic surgery, training, simulator, skills

1. BACKGROUND

The use of simulation in laparoscopic surgery training appears to be qualitatively effective if supported by a suitable evaluation system.
The continually increasing demand of more complex laparoscopic simulators has inspired the creation of a 4d simulator which is a physical low-cost laparoscopic training platform that reproduces the tactile feedback: eLaparo4d) integrated with a software for virtual anatomical realistic scenarios (Unity3D V 4.1).
The School of Medicine of Genoa and the Biomedical Engineering and robotic department (DIBRIS) have cooperated to create a low-cost model based on existing and brand new software.
Aim of this work is to describe the educational-training course and tools that students can use to achieve a complete mastery of surgical gestures till to the correct execution of a laparoscopic cholecystectomy task.

2. MATERIALS AND METHODS

For a correct evaluation of the training assessment, the team designed a new software able to handle the training task, creating a virtual interface based on the concept "student – exercise – evaluation".

2.1 The simulator system

The system is based on a nodejs application server that manages the visualisation system, the communication with hardware interfaces and the database where users' data are stored. The server technology is indeed a sort of data gateway between the several different elements, regardless they are hardware or software. The following figure (figure 1) shows how communication data are exchanged from the very low part of the system (Hardware Interfaces, bottom) to the user interface (HTML Client,top).

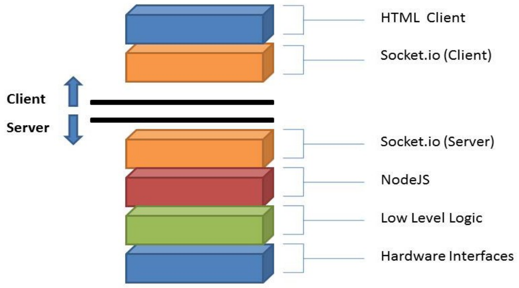


Figure 1: part of the system simulation

The user interface is a simple HTML5 web page running a Unity3D engine (12) plugin. We run several performance tests to compare Unity3D and native WebGL, getting same results. We finally decided to adopt
Unity3D engine due to its rapid development time. WebGL is a great technology but still too young to allow us working on a powerful and robust framework. The use of web pages as the main user interface allows us to be more versatile and in the future will give us the possibility, thanks to HTML5 powerful characteristics, to easily share contents in a live way with other systems. An interesting feature is, for example, having the possibility to be guided by an external supervisor, who is monitoring the training phase, while data are quickly exchanged via internet.

2.1.1 Visual and fisical modelling

As previously introduced, visual modelling is a very important aspect of the entire project. A videolaparoscopic surgery simulator needs a detailed representation of the organs and the tissues inside of the human abdomen. The meshes included in eLaparo4D are developed in Blender 3D Modelling software, and then imported in Unity3D, including textures and UV maps. Eventually, in Unity3D render shader materials are added to the raw meshes, to simulate the specific surface of each of the modelled tissues. In Figure 2, a screenshot of the current virtual environment is shown.

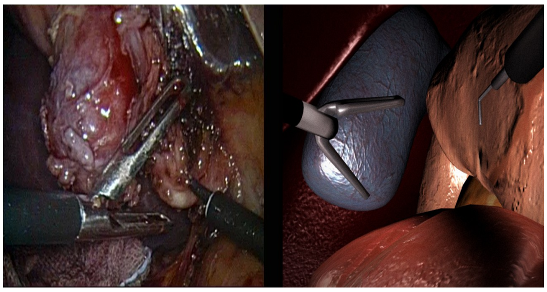
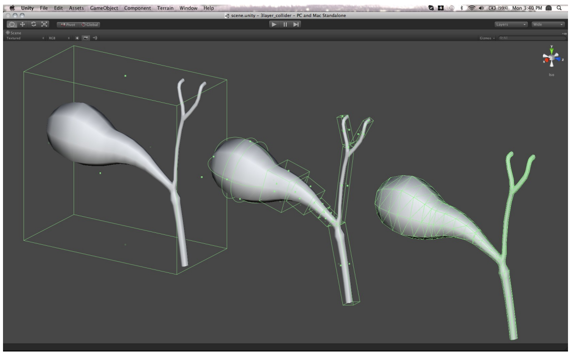


Figure 2: a screenshot from the current aspect of the virtual environment compared to a screenshot of the camera view of a real surgical operation.

As remarked by our colleagues of the Videolaparoscopy Unit of the Department of Clinical Surgery, highly specific training sessions are required to help the operator achieving a proper skill set. In an ideal scenario, medical students should have access to a complete simulator composed of several training scenes, as part of a modular and step-based training process. While the main components and controls of the simulator should be in common, each scene should focus on a very specific surgery operation, differentiating in: the zone and the organs physically manipulated (the target), the particular surgical maneuvers performed (the task), and the type of manipuli used (the means).Considering these remarks, we developed a dynamic parametric physical simulation approach, arbitrary applicable to the rendered meshes in every scene and able to avoid system overloads. Such an approach permits the creation of different scenes starting from the same set of models and interaction algorithms, easily supporting a step-based training. In detail, each 3D object in the scene carries a selectable 3 layer collider component, driving a vertex deformation script. The first layer is a simple box collider; the second one is a combination of simple shape colliders which cover, with good approximation, nearly all the volume of the object; the third is a precise mesh collider which exactly coincides with the vertex disposition of the object's mesh. In the foollowing figure (figure 3) is possible to see the 3 different collider layer for a gallbladder model.



scene.unity – 3layer_collider – PC and Mac Standalone

Figure 3: I.e of a collider layer for a gallbladder model

2.1.2 Feedback system

Haptic feedback is implemented thanks to the use of three Phantom Omni devices from Sensable. The first two are used as manipuli (grasper, hook or scissors) and the third one is used to move the camera within the virtual abdomen, as it happens in a real scenario. The system generates a resultant force when the user puts a manipulus in contact with a mesh, according to the executed task. Phantom devices have been chosen because reasonably low cost although precise enough for the needed level of realism. Furthermore, their stylus-like shape will permit a complete merging of the devices with the physical environment reconstruction; in particular, each stylus will be easily connected to real manipuli. Thanks to an Arduino board connected to a vibrating motor we have also included a vibration feedback. Vibration is used to enhance the realism of operations like tissue shearing (hook) and cutting (scissors).

2.2 The training model

The training course is divided into three main phases according to the literature review:

Acquisition of basic skills: exercises related to the acquisition of tasks which allow students to reach basic gestures competences. They could practice using probes that simulate the haptic feedback according to the kind of action.

We considered six different tasks:

1. Grasp: this task aims to evaluate the abiltity to grasp a object in a lot of position.
2. Transport: this task required the participant to grasp an object and to transport it to a target.
3. laparoscopic - focusing - navigation: This task aims to evaluate the ability to navigate a laparoscopic camera with a 30º optic. This is done by measuring the ability to identify 14 different targets placed at different sites. Each target includes a large symbol only identifiable from a panoramic viewpoint and a small symbol only identifiable from a close up viewpoint. The task starts by identifying the large symbol on the first target (i.e. 1) and then the small symbol situated next to it, which must be shown on the centre of the screen. This small symbol indicates the next large symbol to be identified. Following this order, the participant continues until the identification of the small symbol on the last target (i.e. end).
4. hand – eye – coordination (HEC): This task aims to evaluate the ability to navigate a laparoscopic camera with a 0º optic with the non-dominant hand (NDH) and to handle laparoscopic forceps with the dominant hand (DH). This is done by measuring the ability to grasp and transport six pre-defined objects to six pre-defined targets in the LASTT model, which is fitted with coloured objects (5 x 4 mm open cylinders) and coloured targets (10 x 1 mm nails). The matched targets and objects are identifiable by colour. The exercise starts by identifying a target and an object of the same colour. The object is then grasped, transported and introduced onto the relevant nail. Only when the participant has succeeded in introducing the cylinder into the matching nail is he/she allowed to continue with the next object.
5. pick – up: consists in moving a lot of object to different targets with different form.
6. ring and rail: consists in moving a ring along a twisted metal rod without applying excessive force to either the ring or the rail.

Acquisition of advanced skills: in this phase students will perform complex drills acquiring a correct gesture in the 'use of more specific instruments such as scissors, needle holder and retractors.

The main actions are classified in five classes:

1. cutting: This task required the participant to cut a circle from a rubber glove stretched over 16 nails in a wooden board. Penalty points were calculated when the individual deviated from cutting on the line. Score=time in seconds + surface of glove in milligrammes deviated from the circle.
2. knot tying: This task involved the tying of an intra-corporeal knot (two turn, square knots) in a foam uterus. A penalty was calculated to reflect the security (slipping or too loose) of the knot. Score= time in seconds + 10 when knot was slipping or loose.
3. work simultaneously: the purpose of this task is to evaluate the students' skills in handling the camera in one hand while working with an instrument in the other hand. The task involved moving three pieces of linen , one at a time, from one structure to another, which was so far from the first structure that both could not be seen in the camera at the same time.
4. touching clips: the purpose of this task consists in touching endoclips on a structure.
5. pipe cleaner: this task involved the placement of a pipe cleaner trough four small rings. A penality was calculated when a ring was missed . Score= time + the number of missed rings x 10

A progressive performance "step-by-step" of laparoscopic basic and intermediate abilities: this phase requires the simulation of surgery. The student works in a virtual surgical environment.
I.e.: five steps: cystic elements isolation, clipping, dissection, recovery, haemostasis procedures, that could the students complete a laparoscopic cholecystectomy.

For each phase common simple activities are defined to be evaluated,:

Vision/ navigation

Tools selection: the software allows the student to use several different operational devices: dissector, grasper, scissors, clip appliers).Through a graphical interface students will be able to choose the appropriate device.

Correct use of tool

Change of medical parameters: i.e. our software allows to manage the coagulation values and other conditions such as the level of pneumoperitoneum.

The educational assessment provides the application of gestural acquisition "step by step".
Students own their ID and password which allow the access to their virtual academic book and consequently to the exercises planned, updating their training through a personal academic profile.
The program manages the educational value of defining a standardized scale (Guilbert Scale) in relation to the

level of competence - mastery of a single surgical act: the scale consists of 5 points (0: no mastery ... 5: complete mastery).
The variables which can influence the evaluation are:

Score cut off: standard score benchmark below which the student will not have access to the next stage. This method was adopted both in the basic activities and advanced.
(Phase I: 30 pt; Phase II: 25 pt; Phase III: 25 pts)

Running time: each exercise is time-dependent; student will have a set time to complete the simulation. (Single Exercise Phase I: t = 180 sec, t = 240 sec phase II, phase III t = 800 sec)

Errors: quantified according to their severity (mild, moderate, severe) (slight error: - 1 pt; moderate error: -1.5 pt; serious error - 2 pt) and the difficulty related to the intervention's context (basic, intermediate, advanced) defining for each context a compensatory positive bonus value (base + 0 pt; intermediate + 0.2 pt; advanced + 0.5 pt).

3. ATTENDED RESULTS AND DISCUSSION

In the recent years the demand for laparoscopic simulators is growing up. Consequently, it becomes mandatory to make use of assessment instruments suitable to the complexity of new simulators in order to validate the training experiences even at advanced level.
In this sense, was carried out a review of the relevant scientific literature through PubMed and Medline database using the following keywords: simulation, skill, task, laparoscopic.
We provide a "scan" of the medical literature over the past 10 years in terms of simulation outcomes in laparoscopic surgery, learning efficacy years 2003-2013 (94 articles).
Ten studies were included in the review of the literature. The literature was categorized into three themes:

- internal outcomes
- external outcomes
- clinical evaluation

These works describe, in our opinion, the more specialized and main interesting learning experiences .
The aim of this study is to demonstrate the effectiveness of the suggested educational program. The model is based on the acquisition of clinical and gestural skills in laparoscopic surgery through the use of two integrated devices: the laparoscopic simulator "eLaparo4d" and the software for the virtual reality (Unity3D V 4.1).
We provides an experimental teaching to a sample of students, evaluating different variables such as the time of learning, the user satisfaction and the level of competence at the end of the clinical didactic trail. Preliminary results will be presented.
Our proposed curriculum represents an individual training program tailored to each trainee's needs and performance levels.
In this sense we established a set of benchmark criteria based on experienced surgeons' performance.
The concept of virtual training has been acknowledged for some years and a number of studies have been published on the importance of this new potentially rewarding technology.
The tasks were based on the studies of Derossis et al. The results of our study provide a firm basis for the simulation model to be implemented as mandatory in the residency training curriculum, with the laparoscopic experts' performance level as the training goal.
Simulator has just been introduced in the university curriculum.
The results are "attended results" because the data analysis will be possible only after a period of testing of the simulator on different samples of students. Referring to experience from other scientific groups, we expect significant results in terms of: reduction of learning time, better dexterity and ability to intervention in case of procedural errors.

ACKNOWLEDGMENTS

We would like to thank G. Vercelli, L. Lagomarsino, V. Zappi of Dept. of Informatics, Bioengineering, Robotics and Systems Engineering, together with M. Camia and E. Commisso of EMAC, for their precious collaboration.
This project is partially funded by POR Regione Liguria and jointly developed with the Italian biomedical company EMAC s.r.l..

REFERENCES

Campo R et al. 2012. Training in laparoscopic surgery: from the lab to the OR, *Zdrav Var* 51: 285-298

Debes A J et al. 2012. Construction of an evidence-based, graduated training curriculum for D-box, a webcam-based laparoscopic basic skills trainer box. *American Journal of Surgery,* 203: 768-775

Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. 1998. Development of a model for training and evaluation of laparoscopic skills. *American Journal of Surgery*, 175(6):482–487

Grantcharov T P, Reznick R K. 2008. Teaching procedural skills. *British Medical Journal.* 336: 1129-31

Hyltander A et al. 2002. The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surgical Endoscopy,* 16: 1324-1328

Kolkman W, Van de Put M A J, Wolterbeek R, Trimbos J B M Z, Jansen F W. 2008. Laparoscopic skills simulator: construct validity and establishement of performance standards for residency training.

Gynecological Surgery 5: 109-114

Loukas C et al. 2012. A head-to-head comparison between virtual reality and physical reality simulation training for basic skills acquisition. *Surgical Endoscopy,* 26 (9): 2550-2558

Molinas C, Campo R. 2010. Defining a structured training program for acquiring basic and advanced laparoscopic psychomotor skills in a simulator. *Gynecological Surgery,* 7: 427-435

Perrenot C et al. 2012. The virtual reality simulator dV-trainer is a valid assessment tool for robotic surgical skills. *Surgical Endoscopy,* 26: 2587-2593

Stefanidis D et al. 2010. Initial laparoscopic basic skills training shortens the learning curve of laparoscopic suturing and is cost-effective. *Journal of American College of Surgeons*, 210 (4): 436-440

Xeroulis G J et al. 2006. Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery*, 141 (4): 442-449

SIMULATING HUMAN RESOURCE CAPABILITY AND PRODUCTIVITY IN SOFTWARE PROCESS SIMULATIONS

Štěpán Kuchař[(a)], Ivo Vondrák[(b)]

VSB - Technical University of Ostrava, IT4Innovations
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[(a)]stepan.kuchar@vsb.cz, [(b)]ivo.vondrak@vsb.cz

ABSTRACT

Software processes are considered to be one of the most complex processes because they are very dependent on human behaviour, creativity and productivity. Simulation models are trying to capture and model these properties to provide more precise information about the progress and parameters of the process. This paper describes a method to describe competencies of the workers in the process using competency models. These competencies are used to simulate human resource capability and productivity that influences the duration and allocation of resources to activities in the process. This method is then integrated into the BPM Method modelling and simulation environment that is used in an experiment to compare the allocation strategies in software process simulations for a middle-sized software development company.

Keywords: Software Process Simulation, Discrete Event Simulation, BPM Method, Human Resource Productivity, Competency Model

1. INTRODUCTION

Business processes represent the core of each organization's behaviour (Madison 2005). They define a set of activities that have to be performed to satisfy the customers' needs and requirements, roles and relationships of the employees that are needed for actually performing these activities and objects that are consumed or produced by these activities (Šmída 2007). Software processes are also a special type of business processes that are highly dependent on human creativity, competencies, experience and interaction (Dutoit et al. 2006). Human-based processes tend to be more uncertain than automated processes performed by machines, because human behaviour is very complex and depends on a lot of factors, including capabilities, emotions, social status, health, etc. (Urban and Schmidt 2001). All these aspects influence the productivity of workers in the process and subsequently change the cost and duration of the process and the quality of the final product.

Unfortunately, existing process simulation models are not very concerned with accurate human resources modelling and description (Rozinat et al. 2009) and this can lead to the loss of precision in the simulation results. This paper proposes a capability and productivity model for allocating competent workers to activities during the simulation and dynamically changing duration of such activities based on their skills and abilities.

2. RELATED WORK

Numerous simulation models for software processes were created to evaluate different types of processes to help the companies with determining the best process for their needs and to help them estimate the total cost and duration of their software development projects. Several of these simulation models contain a specification of workers and their capabilities to support the allocation or estimate their productivity.

(Abdel-Hamid and Madnick 1991) proposed a system dynamics model which divided workers in the process to two groups – experienced and newly hired. Part of the experienced workers workload was to train the newly hired workers that were gradually assimilated to the experienced workers stock. This model was used for modelling the delays for the introduction of newly hired personnel to the process.

Experience and capability on a more detailed level was specified in the Generalized Stochastic Petri Net simulation model designed in (Kusumoto et al. 1997). Individual workers were modelled with their experience level in mind with three possible values – novice, standard, expert. Each activity is modelled as a loop of possible communication effort, thinking and writing/creating the result. Each of these transition durations and workloads are influenced by the experience of the worker. The activities do not have any requirements and no allocation model is specified.

(Hanakawa et al. 1998) defines a simulation model with exactly specified activity, productivity and learning models that served as a base for our own model. Each activity is divided into primitive activities and each one of these specifies the required knowledge level. The primitive activities' knowledge levels are distributed normally for each activity. The productivity model then evaluates the amount of work produced in one time step based on the required knowledge level and worker's acquired knowledge. Finally, the knowledge level uses a learning curve to calculate the increase in the worker's acquired knowledge. In (Hanakawa et al. 2002), this model is further enhanced by detailing the worker's knowledge by a cognitive map. This enhancement is similar to our competency approach, but the paper does not work with the process as the whole and only works with individual activities and individual workers, ignoring the allocation of the right resource for the right activity.

(Hanne and Neu 2004) uses a similar model as (Hanakawa et al. 1998) with each activity specified by one skill and learning curve used for changing the knowledge level of this skill dynamically during the process. Influences of several emotional factors like stress and boredom are also integrated to the model. Allocation of resources is not dealt with.

(Raunak 2009) specifies a very well thought out model for human resource modelling that includes multiple capabilities for workers and multiple requirements for activities, constraints for allocating resources of specified groups or in specified order, bidding mechanism for deciding allocated resource based on his availability and motivation, dynamic change of requirements based on the state of the process, etc. This model has a similar objective as ours, but it does not work with different competency levels and requirements, define importance of individual competencies for an activity and does not provide the means to evaluate the worker's productivity and its effect on the duration and cost of the process.

3. THE BPM METHOD

A modelling and simulation method that is able to sufficiently model human-based processes was needed to provide the simulation environment for the capability and productivity models. For these purposes we used the BPM Method (Vondrák et al. 1999). We had to enhance this simulation environment with stochastic parameters (Kuchař and Kožusznik 2010) and also with the means to share generic resources between concurrent process instances and activities (Kuchař et al. 2012a). This method defines three basic models of the process – architecture of the process, objects and resources utilized in the process and the behaviour of the process. The most important one of these models for performing simulations is the behavioural model. This model is called the Coordination model and it specifies the behaviour of the process as a sequence of activities. It also specifies what resources the activities require and which artefacts they consume and produce. Alternative flow in the coordination model is enabled by multiple activity scenarios and concurrency of the activities can also be modelled using special modelling techniques. This model can also be converted to a Petri net to provide exact semantics for performing simulations (Kuchař and Kožusznik 2010).

The Coordination model is visualized by the Coordination diagram and a simple example of this diagram is shown in Figure 1.

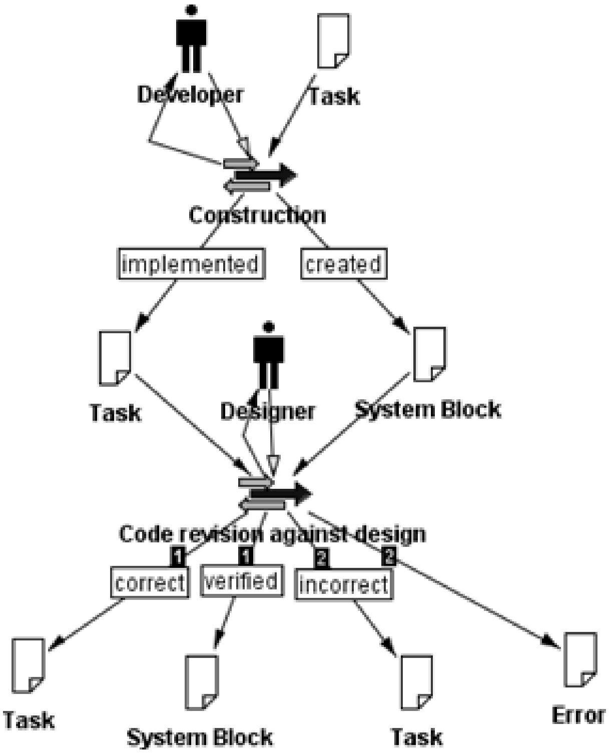


Figure 1: Part of the Coordination Diagram

This diagram describes a part of the Software Construction subprocess. The Designer and Developer active objects describe the roles of employees in the process and the Task passive object defines the task objects that serve as input for the Construction activity. System block can be constructed when the Task is created and the Developer resource is available. The yellow arrow shows that the Developer is responsible for executing the Construction activity. By completing this activity the state of the Task changes to implemented, new System block is created and the Developer is ready to implement another task.

The subsequent activity is Code verification that is performed by the Designer and consumes the implemented Task and created System block. This activity can end up in two ways. The first scenario signifies that the constructed code is correct and its outputs are marked by number 1. The second scenario shows that there were errors in the implementation and the process will continue by reporting and repairing the error. Outputs of the second scenario are marked by number 2.

4. HUMAN RESOURCE COMPETENCIES

Our proposed simulation method uses competency models to describe the capabilities of human resources to perform activities in the process. Competency models (see e.g. (Sinnott et al. 2002; Ennis 2008)) describe various competencies which are important for the process. Competencies are defined as sets of knowledge, abilities, skills and behaviour that contribute to successful job performance and the achievement of organizational results (Sinnott et al. 2002).

Competency models also describe how to measure and evaluate individual competencies. In most cases competencies are measured by a number of advancing stages where higher levels of competency include everything from their lower levels. There is no standard for how many levels a competency model should have and every model defines its own set of levels, so our method is able to work with an arbitrary competency level limit.

Competency levels of resource *r* in the process can be defined using the following vector:

$$\boldsymbol{r} = (l_{r,1}, l_{r,2}, \ldots, l_{r,n}) \tag{1}$$

where $l_{r,1}$, $l_{r,2}$, ..., $l_{r,n}$ are levels of competencies c_1, c_2, ..., c_n that have been mastered by the resource *r*.

Activities in the process also have some requirements on the resources and their competencies. The resources that fulfil these requirements are able to effectively perform the activity, but it does not always mean that resources without required competencies are not able to perform the activity at all. Such resources, which are lacking some of the required competencies, will of course have troubles with the activity, prolonging the duration and increasing the probability of faults in the activity's results. The effect of the lack or abundance of competencies on the activity is specified in section 5.

The requirements of the activity *a* can be described by the following vector:

$$\boldsymbol{a} = (\boldsymbol{rr}_{\boldsymbol{a},1}, \boldsymbol{rr}_{\boldsymbol{a},2}, \ldots, \boldsymbol{rr}_{\boldsymbol{a},\boldsymbol{m}}) \tag{2}$$

where $\boldsymbol{rr}_{a,i}$ ($i \in \{1,\ldots,m\}$) is a vector of requirements for the *i*-th resource required by activity *a* and *m* is a number of resources the activity requires. The required resources are specified by the BPM method as the active objects that are entering the activity as inputs. Each requirement vector is then specified as:

$$\boldsymbol{rr}_{\boldsymbol{a},\boldsymbol{i}} = (\boldsymbol{rc}_{\boldsymbol{a},\boldsymbol{i},1}, \boldsymbol{rc}_{\boldsymbol{a},\boldsymbol{i},2}, \ldots, \boldsymbol{rc}_{\boldsymbol{a},\boldsymbol{i},\boldsymbol{n}}) \tag{3}$$

where $\boldsymbol{rc}_{a,i,j}$ is a vector that specifies the requirements for competency c_j concerning the *i*-th resource of activity *a*. Structure of the $\boldsymbol{rc}_{a,i,j}$ vector follows:

$$\boldsymbol{rc}_{\boldsymbol{a},\boldsymbol{i},\boldsymbol{j}} = (ri_{a,i,j}, rll_{a,i,j}, rlt_{a,i,j}, rhl_{a,i,j}, rht_{a,i,j}) \tag{4}$$

where $ri_{a,i,j}$ specifies the importance of competency c_j for the *i*-th resource required by activity *a*.

$rll_{a,i,j}$ and $rhl_{a,i,j}$ represent the low and high limits for required levels of competency c_j for the *i*-th resource required by activity *a*.

$rlt_{a,i,j} \in \{strictly\,required, requested\}$ describe the type of the appropriate low requirement. The *strictly required* type defines a strict constraint on the competency level. This means that only resources that meet the condition $l_{r,j} \geq rll_{a,i,j}$ can be allocated to the activity. On the other hand, the *requested* type means that even resources with lower competency levels can be allocated if they are chosen by the allocation strategy. $rht_{a,i,j}$ follows a similar pattern only for the high requirement level.

The activity requirements have to be compatible with the resource competencies meaning they have to use the same vector of competencies, the competencies have to be defined on the same scale and have the same meaning. Exact specification of these conditions is stated in our previous work (Kuchař and Martinovič 2013).

4.1. Worker's Capability to Perform Process Activities

Whenever any activity with requirements in any process case needs to start its execution, the simulation needs to allocate one or more resources to perform this activity. These chosen resources have to be suitable for performing this activity by having all *strictly required* competency levels to fulfil the activity's requirements. On top of these hard constraints, both the *strictly required* and *requested* requirements are used to evaluate the resource's capability to perform the activity. This capability can be calculated by encoding the resource competencies and activity requirements to their comparable vector representation and evaluated in the vector space model. The capability of resource *r* for activity *a* is defined as similarity of their vector representations:

$$cap(r,a) = sim(rvect(\boldsymbol{r}), avect(\boldsymbol{a})) \tag{5}$$

where *rvect* is a function that converts a resource vector to the comparable vector representation and *avect* converts an activity vector to this representation (specific conversion and similarity evaluation is elaborated in (Kuchař and Martinovič 2013)).

5. HUMAN RESOURCE PRODUCTIVITY

The resource capability introduced in previous section specifies how skilled and knowledgeable the worker is to perform specific activity. This skill and knowledge level influences the productivity of the resource. (Hanakawa et al. 1998) also agrees with this statement and specifies a productivity model that simulates this influence. We will also use this model in our method.

The productivity model derives the worker's productivity by processing the worker's capability and the required capability to perform the activity. It uses a cumulative distribution function of the standard normal distribution to specify the productivity *P(r,a)* as:

$$P(r,a) = C_{r,a} \int_{-\infty}^{a_a(cap(r,a)-cap(r_a,a))} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \\ = C_{r,a} \Phi\Big(a_a\big(cap(r,a) - cap(r_a,a)\big)\Big) \tag{6}$$

where $C_{r,a} > 0$ is a maximum of the productivity of resource *r* for activity *a*. $a_a \geq 0$ is a level of accuracy needed to perform the activity *a* and determines the sharpness of the decline in the productivity. *cap(r,a)* is a capability of resource *r* for activity *a* (see section 4.1).

cap(r_a,a), on the other hand, is a required capability for activity *a*. This capability is derived from the capability of referential resource r_a for activity *a*. The referential resource of the activity has exactly the same competency levels as those required by the activity and therefore has exactly an average productivity while performing the activity. This average productivity corresponds to the default activity durations specified in the process model.

To simplify further calculations with productivity, it is useful to universally set the $C_{r,a}$ parameter to 2. This leads to the fact that the referential resource r_a always has a productivity of 1 regardless of the value of parameter a_a because

$$cap(r_a, a) - cap(r_a, a) = 0 \quad (7)$$

$$P(r_a, a) = 2\Phi(a_a * 0) = 2 * 0.5 = 1 \quad (8)$$

This can then be perceived as doing 1 unit of work in 1 unit of time. Resources with higher capabilities than the referential resource have higher productivity and can therefore do more units of work in 1 unit of time (e.g. if the worker's productivity is 1.25, she can do 1.25 units of work in 1 unit of time). On the other hand, resources with lower capabilities have lower productivity and therefore less units of work in 1 unit of time.

Figure 2 shows the productivity curves for different values of parameter a_a with $C_{r,a}$ set to 2.

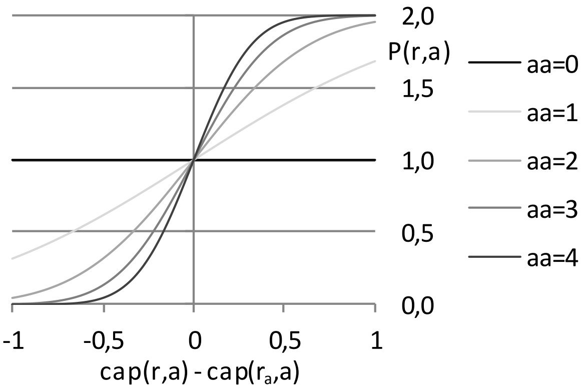


Figure 2: Productivity Curves

The plot in Figure 2 shows that the values of parameter a_a greatly influence the productivity values. Bigger values of the a_a parameter cause sharper rise of the curve, meaning that small changes in capability cause a major change in productivity. This property can be used to differentiate between specialized and universal activities in business processes, specialized activities having a steeper curve and universal activities having a flat curve. It is also useful to allow setting the the a_a parameter to 0 for fully automated activities or for disabling the influence of the productivity for other types of simulations where productivity is not required.

5.1. Productivity and Duration of Activities

The last section described the notion of workers' productivity but how does this productivity translate into the duration of the process instances? Each activity in the process is performed by some resource (either human or non-human) and can be influenced by the resource's competencies and productivity. Each activity can therefore take a different amount of time for different workers even though the amount of work for the activity does not change.

The worker's productivity specifies how much faster the worker can perform an activity. The duration of activity *a* being performed by resource *r* can therefore be defined as:

$$d_a(r) = d_a * m(r, a) \quad (9)$$

where d_a is the standard duration of activity *a* for an average worker and *m(r,a)* is a duration multiplier for resource *r* and activity *a* that is specified as:

$$m(r, a) = \frac{P(r_a, a)}{P(r, a)} \quad (10)$$

where *P(r_a,a)* is a productivity of the referential resource r_a for activity *a* and *P(r,a)* is a productivity of resource *r* for activity *a*.

The productivity of the referential resource in equation (10) has to be taken into account, because productivities of all resources are related to the referential worker (see equation (6)). And here, the choice of the $C_{r,a}$ parameter having a value of 2 pays off, because then *P(r_a,a)* is always 1 (see equations (7) and (8)) and it can be removed from equation (10) leading to:

$$m(r, a) = \frac{1}{P(r, a)} \quad (11)$$

Figure 3 shows the duration multipliers *m(r,a)* for different values of parameter a_a.

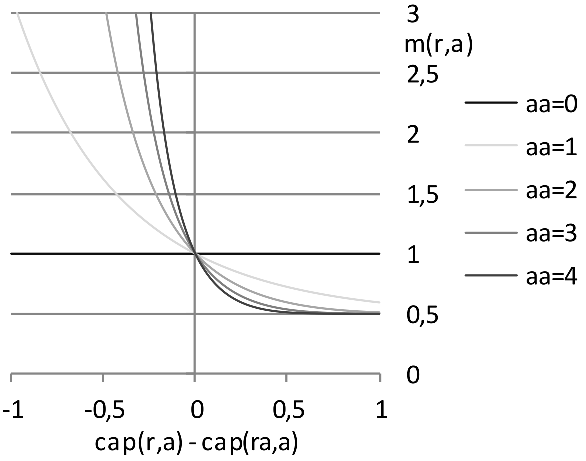


Figure 3: Duration Multipliers

5.2. Introducing Productivity into the BPM Method

The theory described in previous sections can be easily introduced into the simulation engine of the BPM Method because the simulation engine is already prepared to take worker's competencies and capabilities into account (see Kuchař et al. 2012b). The only problem is to implement the productivity enhancements into the simulation engine and link them with the probability distributions that determine the duration of activities.

The duration of activities in the BPM Method is modelled by using a normal distribution specified by two percentiles – low and high boundaries (for more specific details see Kuchař and Kožusznik 2010). These two percentiles are then used to calculate the mean value and variation of the normal distribution. Every time the activity is started, its duration is determined as a random variable from the distribution.

There are three possible ways how to influence the distribution with productivity:

1. Multiply the final duration value acquired from the distribution just at the time when the activity starts.
2. Change the parameters of the distribution just before the final duration value is acquired from the distribution.
3. Change the values of the percentiles before the distribution parameters are determined.

Concerning the normal distribution, all these three possibilities are equal in their accuracy. The first option would be the most efficient, because the basic duration without productivity effect could be evaluated before allocating any resource and then changed right at the start of the activity. But in the future, we would like to add additional distributions for durations, mainly the lognormal distribution that has better properties for modelling the duration of activities (Hanne and Neu 2004). Using the first and second option directly for other distributions could potentially skew the probability of resulting values and each distribution would need to specify the conversion method for them to preserve the distribution of the results. Therefore we chose the safest third option that can be used directly regardless of the chosen stochastic distribution because the distribution parameters are determined after the productivity change itself.

The resulting values of the low percentile $ldp_a(r)$ and high percentile $hdp_a(r)$ of activity a influenced by the productivity of resource r can be evaluated as:

$$ldp_a(r) = ldp_a * m(r, a) \tag{12}$$

$$hdp_a(r) = hdp_a * m(r, a) \tag{13}$$

where ldp_a is the low percentile for activity a, hdp_a is the high percentile for activity a and $m(r,a)$ is the duration multiplier for resource r and activity a.

6. HUMAN RESOURCE ALLOCATION

Knowing how the capability of resources influences their productivity and duration of the process is only half of the problem of simulating processes with specific resources. The second half is a proper allocation of these resources to activities in the process. In manual simulations the best approach is to allow manual allocation of workers by providing information about their capability, productivity and availability, letting the user decide which worker should be allocated to the current activity. This is unfortunately not possible for automatic simulations that have to run without user input during the simulation and have to work only with predefined settings. Because of the stochastic nature, changing conditions, activity and process instance concurrency, every instance of the process is unique and cannot be easily predicted or pre-set. On the other hand it is possible to define several allocation strategies based on the resource capabilities and availability that would find the most appropriate worker for different process conditions.

6.1. Resource Availability, Utilization and Process Waiting Time

Before defining allocation strategies, it is important to define resource availability. One resource cannot perform two activities at the same time and when he is executing some activity in the process he is unavailable to other concurrent activities. If another activity needs the same worker (e.g. one developer is needed to implement a new feature in one system and at the same time needs to repair a fault in another system), she has to perform these tasks sequentially by:

- finishing the first task and then starting the second one, or
- pausing the first task and returning to it after finishing the second one, or
- switching back and forth between these tasks.

The BPM Method is only able to model the first sequencing option and it opens a question if another worker is able to perform the task in shorter timeframe instead of waiting for the unavailable worker. Here, the workers' capabilities and productivity can be used to evaluate if it would be better to wait or to allocate another worker.

The unavailability and demand also present two interesting statistical indicators of the process that will be used in the case study – utilization and process waiting time.

Utilization is measured by simply counting up the time when the resource is performing any activity. It is an interesting result of the simulation but it is not very useful for looking at the performance of the process. Performance is not about how long one resource was doing something in the process, but rather how long did the process have to wait for unavailable resources when they were needed to perform another activity.

It is therefore important to be able to simulate and measure these process waiting times. Whenever an activity is enabled but the resource is unavailable, the BPM Method counts and notes the time needed for the resource to become available to perform the activity. Total waiting time for each resource is then computed by adding up these noted times for the appropriate resource.

6.2. Automatic Allocation Strategies

The notion of unavailability in connection with capability and productivity is very useful for determining allocation strategies. Each activity in the process should have some allocation strategy associated with it to enable running automatic simulations and performing automatic allocations of resources to these activities. Our possible strategies consider following problems and their possible solutions:

1. How high has to be the capability of the worker that can be allocated to this activity?
   (a) based on the percentage of referential resource's capability (e.g. all workers that have higher capability then 80% of referential resource's capability)
   (b) specific number of workers with highest/closest to referential/closest but lower than referential/closest but higher than referential capability
2. In what order should the capable workers be considered for allocating to this activity?
   (a) from highest to lowest
   (b) from lowest to highest
   (c) by the distance from the referential resource
   (d) by the distance from the resource in specific position from the highest/lowest capability
3. Should the activity wait for the first unavailable resource or should it allocate the first resource that is currently available?
   (a) always use the first available resource
   (b) compare unavailability and performance durations and use the resource that can finish the activity first
   (c) always wait for the first resource
4. Which worker should be requested if all capable workers are unavailable?
   (a) use the first worker that will become available
   (b) compare unavailability and performance durations and use the resource that can finish the activity first

By combining possible mentioned solutions different strategies can be created. One possible strategy would for example be to consider all workers that have higher capability then 100% capability of the referential resource, sort them from the highest to lowest capability, always use the most capable but available worker and if all considered workers are unavailable, then use the first worker that will become available.

To abbreviate the description of such strategies for further use in the experiment, we will be using a code based on the numbering used in the solution list. A code for the previous example would then be 1a(100%),2a,3a,4a.

7. EXPERIMENT

We have conceived an experiment to compare the impact of different allocation strategies after implementing the resource productivity extension to the BPM Method. This experiment is based on a simplified software process based on a real process described and modelled in a middle-sized software development company from Czech Republic. This simplified process contains 37 basic activities grouped into 7 standard subprocesses – Requirement Specification, Analysis, Design, Implementation, Testing, Deployment and Post-Deployment Support.

7.1. Configuration of the Experiment

Requirements for activities in the process were evaluated for 16 basic competencies on a 10-level scale with 8 of these competencies being further specified by 16 process instance parameters, thus creating a total of 22 specific competencies. Each activity was assigned to one of 7 roles in the process, each role containing a different number of workers based on a standard project team structure. Analysis was created and managed by 1 Analyst and the following design was elaborated by 2 Designers that were also overseeing code revisions in the Implementation phase. Implementation was performed by 6 Developers and 2 Testers did the testing of the software. Everyone was supported by 1 Administrator and supervised by 1 Project Manager. The post-deployment support of the project results was backed by web-based helpdesk software that was managed by 1 Incident Manager that ensured proper categorization and reporting of incidents. Each of these human resources in the project was evaluated for the same set of 22 competencies on a 10-level scale as activity requirements to ensure their compatibility.

We simulated this process with 12 different allocation strategies to see how the process behaves with the resource productivity extension specified in this paper. For easier comparison of the results and their impact, all activities in the process were using the specified strategy for each simulation run. In real-life simulations, each activity could specify its own allocation strategy that best suits this type of activity. Activity accuracy a_a was also specified globally for all activities and was set to 3.

The first 6 allocation strategies used in the experiment were:

1. 1a(100%),2a,3a,4a
2. 1a(100%),2b,3a,4a
3. 1a(100%),2c,3a,4a

4. 1a(100%),2a,3b,4b
5. 1a(100%),2b,3b,4b
6. 1a(100%),2c,3b,4b

The second 6 strategies were the same with the exception that 1a was set to 80% to identify how the enabling of lower skilled resources influences the process.

7.2. Results of the Experiment

To compare the impact of various allocation strategies and productivity extension on the process, we measured the total duration of the process instance. Each simulation run was done by executing 200 iterations of the simulation to weaken the impact of stochastic properties and risks in the process to enable proper statistical analysis of the results. Figure 4 shows a box plot of total process instance durations for different allocation strategies.

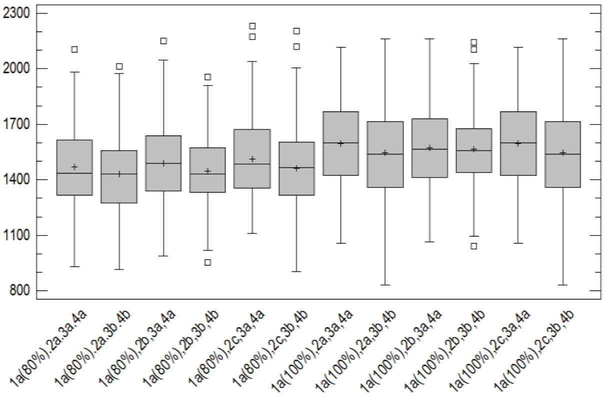


Figure 4: Process Instance Durations for Different Allocation Strategies

By using the test for normal distribution using standard skewness and kurtosis, all runs but one were deemed to follow the normal distribution. The only non-normal run was the 1a(80%),2c,3a,4a with skewness evaluated to 3.13. Looking at the box plot, it is clear that this result is skewed towards lower values. This was probably caused by frequent assignment of lower values for stochastic properties in the process (activity durations and error functions). This proved to be the case because repeated executions of this simulation run had normal skewness and kurtosis.

The first interesting result is directly visible from the plot on top of it being statistically significantly different (at the 95% confidence level). This difference is that the second half of the strategies results in a longer duration. This result was expected because these strategies work only with resources more capable than the referential resource unlike the other strategies that allocate even resources with lower capability (down to 80% of referential resource capability). This difference is also shown in Figure 5 that compares utilization of workers in the process for two representative strategies from different groups. Utilization in the 1a(100%) strategy is very high for few highly skilled workers (Designer1, Developers 0-3, Tester1) and very low for workers with lower capabilities (Designer0, Developers 4-5, Tester0) that could not be allocated to some activities in the process. In 1a(80%) the allocation is more evenly distributed and this leads to lower total duration times even though it takes longer to finish the activities for slightly lower skilled workers.

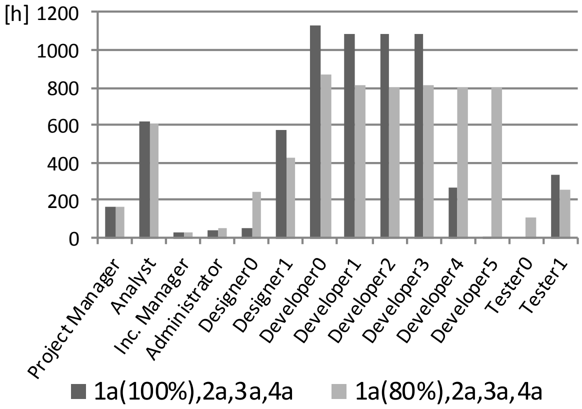


Figure 5: Comparison of Resource Utilization for Strategies 1a(100%),2a,3a,4a and 1a(80%),2a,3a,4a

There are two exceptions for the rule of 1a(100%) strategies taking longer than 1a(80%) ones. The comparison of 1a(80%),2c,3a,4a (lower skilled resources are used first and the first available resource is allocated), 1a(100%),2a,3b,4b (highly skilled resources are used first and the resource with fastest finish time for the activity is allocated) and 1a(100%),2c,3b,4b (resources close to the referential resource should be used first but in the end the resource with fastest finish time for the activity is allocated) provided statistically insignificant difference. This means that primarily allocating workers with low capabilities (and therefore lower productivity) yields similar results in this process as when using the currently fastest resource possible. Strategy 1a(100%),2b,3b,4b should also be included in this exception because it leads to the same strategy as the previous two, but it is slightly off the insignificant difference interval. This can be caused by frequent deviations of stochastic properties in the process or simply by falling out of the 95% confidence interval when deciding significance of their difference.

The second important result was already mentioned in the previous paragraph and concerns the insignificant difference inside each group of 3b,4b strategies. These choices of solutions to the third and fourth allocation questions effectively overshadow the choice for the second question. This is because the order of resources is ignored in favour of finding a worker that will manage to perform the activity first based on the availability of all resources. This hypothesis is proven by the data because all 1a(100%),2*,3b,4b strategies and all 1a(80%),2*,3b,4b strategies are not significantly different.

The same situation is with 1a(100%),2b,3a,4a and 1a(100%),2c,3a,4a that are not significantly different because sorting by the distance from the referential resource is the same as sorting from the lowest capability in this situation. This is because the referential resource is always the lowest capable resource in 1a(100%) strategies.

With expected results put aside, it is time to focus on comparing the remaining strategies with the solution to the second question in mind. Does it matter if the resources are allocated from highest to lowest or from lowest to highest or based by the distance from the referential resource? In this process, the results show that it does not matter, because all strategies that differed only in the second question solution had only insignificant differences. This is probably caused by high process waiting times of highly utilized resources (shown later in Figure 6). This means that highly utilized workers are still supplied by more work and all capable resources are unavailable when allocating resources for an activity. This distributes the work evenly because the process waits for the worker to finish his job. Every worker that finishes his work on an activity is immediately assigned to another activity without much emphasis on his capability. The second choice would be very important in processes that do not have such a high density of activities to be performed, but this is not the case with this experimental process.

This leaves us with the comparison of choosing the first available resource (3a,4a strategies) against choosing the resource that will finish the job first regardless of his availability (3b,4b strategies). It is interesting to look at the process waiting times for workers in the process to see the basic difference in these strategies (see Figure 6).

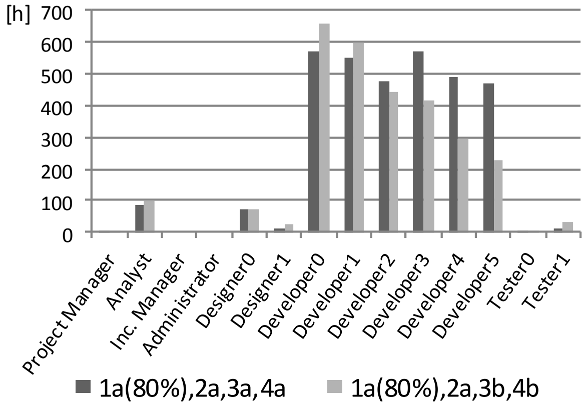


Figure 6: Comparison of Process Waiting Times for Strategies 1a(80%),2a,3a,4a and 1a(80%),2a,3b,4b

The process waiting times are fairly evenly spread in 3a,4a strategies because they always take the first available worker. At the start of the process, workers are allocated with regards to their capability but when all workers are unavailable, the first worker that finishes his activity is assigned to the waiting activity regardless of his capability (he only has to be capable enough to pass the limiting factor set by the first allocation question). On the other hand, 3b,4b strategies have a different pattern of the process waiting times. The most capable (and therefore most productive) workers have higher process waiting time than the less capable workers. This is the product of the "fastest one wins" solution which leads to prioritizing resources with high productivity. Unfortunately, this trend is not followed in the utilization of these workers that is still evenly distributed like in the 3a,4a strategies. This means that even though the more productive resources are chosen for the activity when they are still unavailable, they are allocated to other activity when they become available. This is caused by the fact that resources in the BPM Method can be chosen to several activities when they are unavailable but they can be allocated only to one activity afterwards, leaving other activities to other workers. This only delays the allocation for these activities and it leads to higher process durations. This is also mirrored in the process duration results that were expected to be better for 3b,4b strategies, but experiments showed that they are not significantly different from the 3a,4a strategies. There is only one significant difference between the 1a(*),2a,3b,4b and 1a(*),2c,3a,4a strategies meaning that allocating from the most capable resources on the "fastest one wins" bases leads to shorter durations than allocating from the less capable and taking the first available resource.

8. CONCLUSION AND FUTURE WORK

This paper presented a method for simulating software processes and its extensions for enhancing automatic simulations of human-based processes to provide additional and more precise information about the bottlenecks in the process. Such bottlenecks can be caused by insufficient number of resources in the process or even by wrong allocation of these resources. The proposed simulation solution can be used to try different allocation strategies and find out about their advantages and disadvantages in the simulated process. The productivity extension enables a few of these more complex strategies and at the same time helps to individualize the resources in the process for more transparent simulation results. These extensions are integrated to the BPM Method to provide a robust simulation environment based on the Petri nets formalism.

Integration of the presented extensions to the BPM Method still has one problem that was identified during the experiment. One highly productive unavailable worker can be chosen for several waiting activities at one time but can only be allocated to one of these activities when he becomes available. This will be solved in our future work by creating a prioritized queue for each resource that will enhance the allocation by preferring prioritized activities and streamlining the potential allocation of unavailable resources.

This paper also presented only the time aspect of the productivity and capability of resources in the project but there are additional properties that can change on the basis of resource capability. For example more capable workers make fewer errors and their capability should influence the error rate of the activity. Our future research will focus on analysing these properties and integrating them to the BPM Method.

Finally, all workers in the process simulations have their competencies pre-set and they do not change during the simulation. In some processes, it is possible to have some training activities that could provide additional competencies for the workers in the process. Even if it is not the case, people are honing their competencies by performing each activity in the process. This learning-by-doing aspect could also be introduced to the BPM Method in our future work.

ACKNOWLEDGMENT

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

REFERENCES

Abdel-Hamid T., Madnick S. E., 1991. *Software Project Dynamics - An Integrated Approach*. Prentice-Hall, Englewood Cliffs.

Dutoit A.H., McCall R., Mistrik I., 2006. *Rationale Management in Software Engineering*. Springer.

Ennis M.R., 2008. *Competency Models: A Review of the Literature and The Role of the Employment and Training Administration (ETA)*. US Department of Labor.

Hanakawa N., Morisaki S., Matsumoto K., 1998. A learning curve based simulation model for software development. *International Conference on Software Engineering* (ICSE) 1998, pp. 350-359.

Hanakawa N., Matsumoto K.-i., Torii K., 2002. A Knowledge-Based Software Process Simulation Model. *Annals of Software Engineering* 14, No. 1-4, pp. 383-406.

Hanne, T., Neu H., 2004. Simulating human resources in software development processes. Berichte des Fraunhofer ITWM 64.

Kuchař Š., Kožusznik J., 2010. BPM Method Extension for Automatic Process Simulation. *8th Industrial Simulation Conference 2010*, pp. 80-85. 7-9 June, Budapest, Hungary.

Kuchař Š., Ježek D., Kožusznik J., Štolfa S., 2012. Sharing Limited Resources in Software Process Simulations. *10th Industrial Simulation Conference 2012*, pp. 33-37. 4-6 June, Brno, Czech Republic.

Kuchař Š., Podhorányi M., Martinovič J., Vondrák I. 2012. Simulation of the Flood Warning Process with Competency-based Description of Human Resources. *The 11th International Conference on Modeling and Applied Simulation 2012*, pp. 100-105. 19-21 September, Vienna, Austria.

Kuchař Š., Martinovič J., 2013. Human Resource Allocation in Process Simulations Based on Competency Vectors. *Advances in Intelligent Systems and Computing 188*, pp. 231–240. Springer Berlin Heidelberg.

Kusumoto S., Mizuno O., Kikuno T., Hirayama Y., Takagi Y., Sakamoto K., 1997. A new software project simulator based on generalized stochastic. *Proceedings of 19th International Conference on Software Engineering*, pp. 293-302.

Madison D. 2005. *Process Mapping, Process Improvement and Process Management*. Paton Press.

Raunak M.S., 2009. Resource Management in Complex and Dynamic Environments. *Open Access Dissertations*. Paper 141.

Rozinat A., Wynn M.T., Aalst W.M.P. van der, Hofstede A.H.M. ter, Fidge C. J., 2009. Workflow simulation for operational decision support. *Data & Knowledge Engineering* 68 (9), pp. 834–850.

Sinnott G.C., Madison G.H., Pataki G.E., 2002. *Competencies: Report of the competencies workgroup, workforce and succession planning work groups*. New York State Governor's Office of Employee Relations and the Department of Civil Service.

Šmída F. 2007. *Zavádění a rozvoj procesního řízení ve firmě*. Grada Publishing, a.s.

Urban C., Schmidt B., 2001. PECS - Agent-Based Modelling of Human Behaviour. *Emotional and Intelligent II - The Tangled Knot of Social Cognition*, AAAI Fall Symposium.

Vondrák I., Szturc R., Kružel M., 1999. BPM – OO Method for Business Process Modeling. *ISM '99 Proceedings*, pp.155-163. Rožnov pod Radhoštěm, Czech Republic.

A SYSTEMS ENGINEERING APPROACH TO MODELING AND SIMULATING SOFTWARE TRAINING MANAGEMENT EFFORTS

Asli Soyler Akbas[a,b], Waldemar Karwowski[b,c]

[a] Institute for Simulation and Training
[b] Institute for Advanced System Engineering
[c] Department of Industrial Engineering and Management Systems
University of Central Florida, Orlando, FL, USA

[a,b]asli.soyler@ucf.edu, [b,c]wkar@ucf.edu

ABSTRACT

Although being directly affected by the fluctuations in complex adaptive systems such as knowledge transfer, and economy, technology training within organizations, are managed as independent projects, furthermore causing critical information, as requirements and scope changes, being failed to be shared. This existing approach, forces only current status to be used with discrete data in decision making rather than evaluating the continuous behavior of the training process integrated with possible future environmental conditions. The purpose of this paper is to initiate the design of a model for understanding the behavior of complex technology training management system (TTMS). Recognizing the process as adaptive and continuous, this paper captures the ongoing efforts to simulating training management efforts that can support organizations in critical decision making, and requirements and risk management using system dynamics and agent-based simulation designed with model-based system engineering approach.

Keywords: Training management, SysML, system dynamics, agent-based simulation

1. INTRODUCTION

Today, software solutions are used aiming to support employees with almost every task, related to their area of work. Due to vastly improving technology, stakeholders are often forced to improve the existing software packages in order to satisfy the arising needs and improve the work process. While some of the improvements stay to be as existing version upgrades, in cases where new software is selected, depending on the project scope, the training process may involve the majority of employees within that organization.

Defined as complex adaptive social systems (Morel and Ramanujam, 1999), organizations behave in motion of rotating circles, building a continuously repeating curve with three high level states as Equilibrium, Dissolution and Growth, which can be seen in Figure 1 (Marten, 2001). Besides the external triggers affecting the state of equilibrium, by changing the requirements of the tasks and their approach, adapting a new software solution that affects the majority of the organization will create an internal trigger causing the state to change to dissolution. At this state, effectiveness of transformation efforts will define the duration until the Growth state is achieved, or in worst cases, drive the system in to chaos.

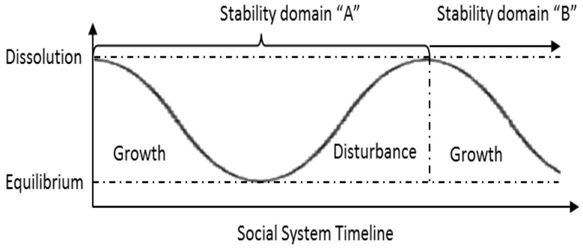


Figure 1 Social System State Adapted from Marten, G. G. (2001)

The importance of training efforts through organizational transformation has been emphasized in literature (Kezar and Ecke, 2002). However, the current body of knowledge lacks research on studies modeling and simulating technology training management system (TTMS) as complex adaptive even though the knowledge transfer within training is recognized as one (Burns and Knox, 2011)

2. BACKGROUND

The key predictors of training transfer are grouped in three as (1) immediate training climate (Kontoghiorghes, 2001), (2) Trainee personality traits and characteristics (Colquitt et al., 2000), and (3) overall organizational environment (Baldwin and Ford, 1988) such as teaching methodology, self efficacy and immediate peer support, respectively. Furthermore effect of (4) economy (Bass and Voughan, 1966, Galf and Hammour, 1993), and (5) technology improvements (Helpman and Rangel, 1999) on training knowledge transfer were noted in literature. A review of literature was conducted to further understand the individual behavior of these factors as systems interacting with training performance and their level of complexity. The review efforts were grouped in four areas as knowledge transfer-which included the two groups of training transfer key predictor factors (1 and 2)-organization, financial system, and technology dynamics.

The reviews in knowledge transfer dynamics supported existence of complexity in learning systems, a phrase introduced by Davis and Simmt (2003) describing collective classroom components, is advocated by also other researchers (Burns and Knox, 2011, Davis and Sumara, 2006). Additionally, Newell (2008) evaluated the potential benefits and challenges of accepting this theory following Davis, Simmt and Sumara's published arguments on how individual learner and teacher dynamics interacts and emerges as learning. On the other hand, organizations are accepted as "dynamic systems of adaptation and evolution that contain multiple parts which interact with one another and the environment" (Morel and Ramanujam, 1999). Furthermore, their nested structure continuously interacts with other macro and micro, systems and sub-systems, respectively (Folke & Folke, 1992). New systems may arise from emerging dynamics as part of the system, due to change processes occurring with an organization (Dooley and Van de Ven, 1999). Similarly, the large-amplitude and aperiodic fluctuations in economic variables were shown as evidence for its complex adaptive nature (Chian, 2007). Finally, technology management in organizations recognized as complex adaptive systems as it interacted with emerging and non-linear trends (McCarthy, 2003).

The characteristics of technology training management that matches the most common properties of complex adaptive systems (Bot, 2012) are as follows:

- Training management interacts with organization system and knowledge transfer variables (micro).
- Although there are techniques to support training planning often times changes in duration, cost, training performance occur.
- Training management is part of knowledge transfer system.
- Employees within an organization create a unique knowledge share structure creating a culture which emerges individual and organization's learning state (Weick , 1979)
- Training is applied in organizations in iterations, the lessons learned from each experience (outputs) feeds the following management strategy as inputs. (Armstrong, 2003)
- If started without well planning the effects of each variable and their interactions, training efforts will fail
- In training management, change in one variable, for instance organization's climate or available resources, will trigger a change in the whole system and will affect outcomes.
- Training management rely on the resource availability, depletion of any resource will trigger system's state to change to 'steady-state'.
- Training has emerged from interaction of systems such as learning, organization and technology. Through time its internal interactions derived management variables (Dooley and Van de Ven, 1999)
- Due to continuous change in it is variables such as humans and technology same management approaches will result in varying outputs.

These characteristics support our argument that instead of managing technology training using linear optimization techniques, complex adaptive systems theory processes should rather be used to understand and capture its patterns and interacting mechanisms.

3. METHODOLOGY

Today, the literature provides researchers with a great source of knowledge on studies related to technology training. Although these studies improved the understanding of the training systems, the vastly improving technology continuously been introducing new areas to the existing gap that's been already identified by the researchers (Salas and Cannon-Bowers, 2001). Furthermore, technology training management is yet to be recognized as complex adaptive in literature.

This research aims to develop a hybrid simulation model using mode-based system engineering (MBSE) approach which some of its benefits are captured as (Sage, 2009):

- Develops a unified coherent model
- Enables the realization of successful systems
- Defines needs
- Documents requirements and
- Facilitates the interoperability between people and organizations (Ramos et. al, 2012)

This approach will support efforts in representing technology training behavior and management processes, driven by the requirements arising from its interactions with other systems knowledge transfer, organization, financial, and technology. The research efforts will be grouped in four main areas.

3.1. SysML Model Development

The overall design will propose a unified coherent model (Sage, 2009) that will define and capture technology training management system by studying its structure, and behavior shaped from its stakeholder and system requirements. Systems Modeling Language (SysML), which facilitates systems engineering activities that are concerned with the whole, complexity, multi-disciplinary and holistic thinking and synthesis (Ramos et. al, 2012) and its four pillars will be used in modeling the system that are captured in Figure 2 (Hause, 2006). Traceability, will be one of the key benefits of this approach that later will allow establishing the corresponding feedback loops within the system dynamics model.

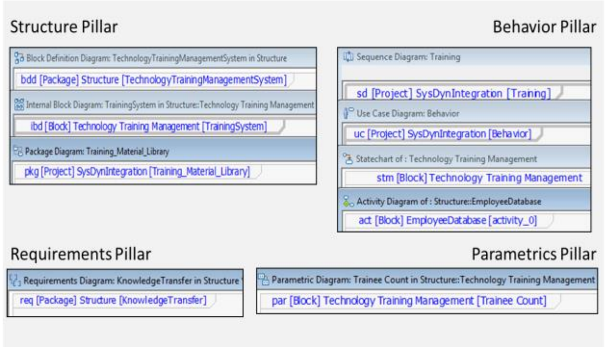

Structure Pillar
Behavior Pillar
Block Definition Diagram: TechnologyTrainingManagementSystem in Structure
bdd [Package] Structure [TechnologyTrainingManagementSystem]
Internal Block Diagram: TrainingSystem in Structure::Technology Training Management
ibd [Block] Technology Training Management [TrainingSystem]
Package Diagram: Training_Material_Library
pkg [Project] SysDynIntegration [Training_Material_Library]
Sequence Diagram: Training
sd [Project] SysDynIntegration [Traning]
Use Case Diagram: Behavior
uc [Project] SysDynIntegration [Behavior]
Statechart of : Technology Training Management
stm [Block] Technology Training Management
Activity Diagram of : Structure::EmployeeDatabase
act [Block] EmployeeDatabase [activity_0]
Requirements Pillar
Parametrics Pillar
Requirements Diagram: KnowledgeTransfer in Structure
req [Package] Structure [KnowledgeTransfer]
Parametric Diagram: Trainee Count in Structure::Technology Training Management
par [Block] Technology Training Management [Trainee Count]

Figure 2 Four Pillars of SysML Adapted from Hause, M. (2006)

3.1.1. Structure Pillar

Structure pillar diagrams-which may be considered as the building block of the overall system, as they capture and map the components and boundaries of the system - including sub-systems, their components and the possible interactions among each using Block Definition Diagram (BDD) and Internal Block Diagram (IBD), respectively, will be designed in this phase. The overall system will be composed of three subsystems that are Project Management Office (PMO), Knowledge Transfer (KT) and Software as captured in Figure 3. This structure will allow capturing factors that directly and indirectly affect the training outcomes, identified in literature while maintaining a representation of the current structure of the organization. For example, the organizational dynamics and their effects, such as peer influence behavior, will be studied under PMO sub-system.

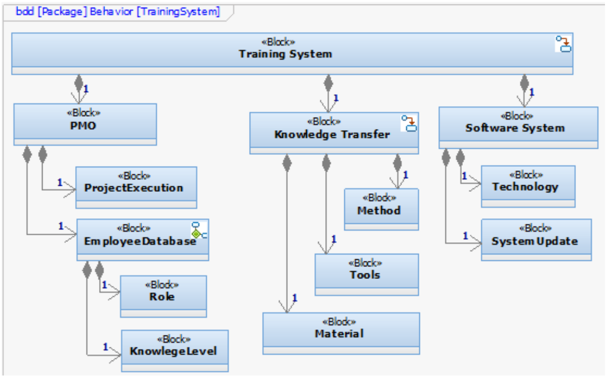


Figure 3 Block Definition Diagram (BDD)

3.1.2. Behavior Pillar

The second pillar, Behavior, consists of the diagrams that define how the model would behave within the capabilities of the designed structure. The design process in this phase will start by defining the states of the overall system, its subs-systems and their components using State-Machine Diagrams (STM). At the highest level, the TTMS is designed to have two states, serving as an “on/off” decision switch for the overall system. The states of the models are designed to be triggered by the factors such as “Identified trainee count”, “Project budget” and “Project schedule” as seen in Figure 4.

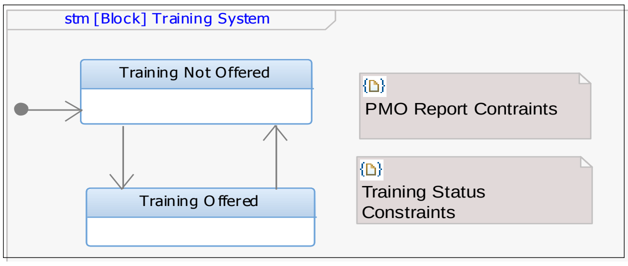


Figure 4 Highest-Level State Diagram (STM)

Later, starting from the highest level, the Use-Case (UC) diagrams will be built capturing the actions and the responsible actors, as groups of different stakeholder classes. Modeling of actors will be one of the most important steps in this section. Properties for each actor class will be filtered out according to criteria: employee type (trainer, trainee, project manager and other stakeholders), skill level (novice, advanced beginner, competent, proficient, and expert), employee position (manager, engineer), department, training attendance count and training transformation factors (demographics, personality, cognitive capabilities and so on). The Package diagram will be used to capture the class hierarchy and the attributes of each agent sub-classes. Furthermore, individual UC diagrams will be linked according the assigned “composition” relations creating traceability among all diagrams. The following step will include capturing the functions included within use-cases, occurring within and in between these components, which will be designed using the Activity Diagrams (ACT). The use of Sequence Diagrams will be decided after completion of the ACT.

3.1.3. Requirements Pillar

The Requirements pillar will allow establishing the rules of the system, rather than using long descriptive paragraphs, the requirements will be captured as short, testable statements. Later, the traceability between child and parent requirements will be captured using “Refine/Refined by” relation. Furthermore, each will be linked as conditions to states of TTMS components which will define the behavior of the system.

3.1.4. Parametrics Pillar

The parametric diagrams allow capturing and modeling constraint expressions, representing system constrains derived from the requirements. As the initial step, the structure, hierarchal relations, of the constraints, which will be represented using type-specific block, will be built using the BDD diagram from the first pillar. After establishing the parent-child relationships using "composition" type, the parameters will be assigned according to the constraint logic statement. Similar to the process followed in IBD, each constraint block will be connected using the input/output ports within the Parametric Diagram (PAR).

3.2. System Dynamics Model Development

The principles of system dynamics modeling, such as the ability to study the effects of individual variables and their interactions, provide a pragmatic and holistic nature (Romme and Dillen, 1997) that is found useful in modeling humans as social systems that are characterized by "dynamic complexity" (Senge, 1990).

The state charts for the system dynamics model will allow creating conditional rates between stock variables. For example, it will be possible to set the training rate to 0 if there are budget cuts, or similarly, have the ability to arrange it according to the training demand. The states of the components created in the SysML model will identically represent the states in system dynamics model (Figure 5).

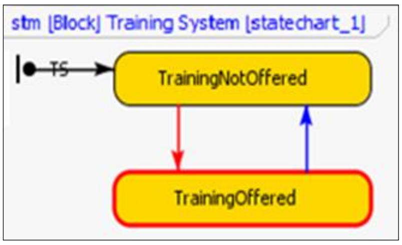


Figure 5 Highest-Level System Dynamics States

Similarly, the states of components, which have assigned constraints, such as skill acquisition and attendance will be created as stock variables and corresponding state change conditions will be used as conditions for the flow rates within the system dynamics model as shown in Figure 6.

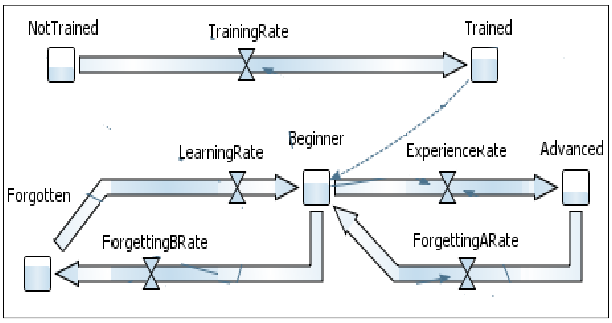


Figure 6 High Level Stock and Flow Variables

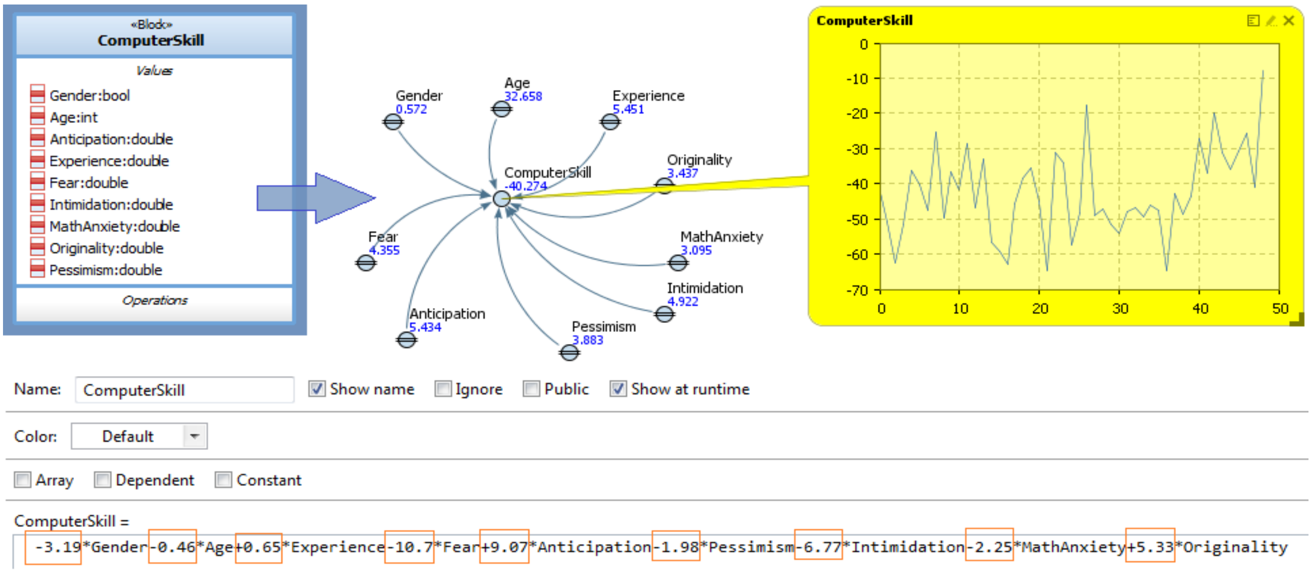
Furthermore, properties assigned to actors such as demographics, department and so on within SysML will be the variables of the system dynamics model. Their causality coefficients will be assigned according to the meta-analysis and experiment results collected from literature. Figure captures an example of an actor and the corresponding dynamics of the variable "ComputerSkill" modeled with AnyLogic which was adapted from the experiment conducted by Harrison and Rainer (1992). Additionally, the coefficients of the significant factors found in the regression analysis, indicated by rectangles, were used as factors that calculate the ComputerSkill variable. It is important to note that, rather than correlation, significant predictive factor coefficients collected from regression analyses, will be used to as weights of the variables in the system dynamics model. The variables in system dynamics are connected to one another by a cause-effect type relation thus the correlations collected from a study cannot be used as a coefficient of a variable.

3.3. Agent-Based Model Development

The limitation of system dynamics modeling, which is the missing capability to capture the properties of observed entities and the resulting effect of these differences (Bonabeau, 2002) will be supported using the agent-based model of TTMS. The hierarchy of the agents will mirror the actors previously created in the SysML model. Although stakeholders, such as software developers and project managers are involved in the overall system, they will not be created as agents. The primary reason is that their decisions may trigger a state change in simulation, but they don't directly interact with either trainee or trainer agents within the scope of this system. Secondly, even though their decisions as inputs are an interest factor in this simulation their behavior is not. Thus, the two agent classes that will be included in the system are Trainee and Trainer. Furthermore, the rules of the agents will be driven from the assigned properties of the actors in the SysML model. Each trainee agent will be assigned with five state groups as employee position (manager, engineer), department, skill level, attendance count and training transformation factors properties.

3.4. Integration of Two Simulation Models

In this phase, the classes of agents will be created and linked under the system dynamics model object. The different properties of agents will be connected with the corresponding stock variable. For example, the skill level property will be connected to the stock variables representing the acquired skill levels. Similarly, count of training attendance property will be linked to the "Not Trained", "Trained" stock variables. With this connection, additional to being able to capture the emerging behavior of the social group, it will be possible to differentiate count of employees from each department within a stock variable at any given time during simulation run.



Name: ComputerSkill ☑ Show name ☐ Ignore ☐ Public ☑ Show at runtime

Color: Default

☐ Array ☐ Dependent ☐ Constant

ComputerSkill =

```
-3.19*Gender-0.46*Age+0.65*Experience-10.7*Fear+9.07*Anticipation-1.98*Pessimism-6.77*Intimidation-2.25*MathAnxiety+5.33*Originality
```

Figure 7 System Dynamics Design Example

4. TRAINING MANAGEMENT SYSTEM

The overall version of the proposed model will be a simulation platform, which will support the decision makers of an organization in planning and testing their training management strategies. The capability to expand upon existing capabilities of both simulation methods, of system dynamics and agent-based modeling, was the aim in selecting a hybrid model for development. The final simulation model will provide decision makers three variables, which will solely focus on high level project overview, cost, duration and the overall knowledge level of the identified trainees as can be seen in Figure 8. This output will serve as a quick evaluation of the decision alternative being tested. Additionally, descriptions of arising risks, unsatisfied requirements, and contributing factors will be provided for detailed analysis.

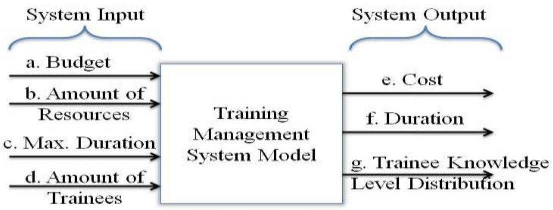


Figure 8 Black-Box Diagram of Overall System

4.1. Validation and Verification

The simulation platform efforts described in the Methodology section will be derived from the previously developed SysML model and depend on one another. As a result, each design and development phase will individually consist of system specific verification and validation processes, which will follow the methodologies captured by Kleijnen (1995).

The modeling will start from the highest-level (macro level) possible and add micro details in iterations. At each level the verification of the simulation models will include running simulations with deterministic values. Furthermore, different scenarios will be tested to check for any programming errors. The probabilistic values will be added once the behaviors of the models are verified.

At the end of phases, at first a validation test will be applied using the simple plot of the simulation output versus data collected from available literature. Immediate feedback on any inconstancies between the two samples will be evaluated. If the plot test does not show any variation to the naked eye, we will move to the hypothesis testing phase.

Each key predictive factor and their coefficients will be adapted from previous experiments published in literature. Thus, to validate a hypothesis test which will test existence of any significant difference between the correlations among two variables calculated from simulation and published correlations of each matching factor will be conducted.

5. CONCLUSION

The aim of this study was to initiate the design of a hybrid simulation framework that can support in understanding and managing technology training using model-based systems engineering approach. This research suggests that technology training management, rather than only being a process within technology transformation efforts, emerges as a component of an organization as it directly affects the outcomes and duration of achieving the state of growth while improving the organization's system stability. Furthermore, it is argued that studies using linear optimization techniques without feedback mechanisms are no longer usable due to the level of complexity involved in technology training management. Recognizing technology training management as a complex adaptive system, instead of managing efforts individually, a systems engineering approach to model its structure and behavior at an organizational level by studying its structure and behavior, driven by the requirements arising from knowledge transfer and its interactions with other complex systems was suggested. Additionally a methodology to establish a link between SysML and system dynamics and agent-based simulations was proposed.

REFERENCES

Armstrong, M. (2003). A Handbook of Management Techniques (p. 544). Sterling, VA.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41, 63–106.

Bass, B. M., & Vaughan, J. A. (1966). Training in industry: The management of learning (p. 2). Wadsworth Publishing.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. Proceeding of the National Academy of Sciences of the United States of America, 99(10), 7280–7287.

Bot, K. D. (2012). Rethinking multilingual processing: From a static to dynamic approach. Third Language Acquisition in Adulthood (p. 82). Phillidelphia, PA. Retrieved from http://books.google.com/books?id=FDYls4wgX50C&printsec=frontcover#v=onepage&q&f=false

Burns, A., & Knox, J. S. (2011). Classrooms as complex adaptive systems: A relational model. TESL-EJ, 15(1), 1–25.

Chian, A. (2007). Complex Systems Approach to Economic Dynamics. Berlin: Springer Berlin Heidelberg.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. Journal of Applied Psychology, 85(5), 678–707. doi:10.1037//0021-9010.85.5.678

Davis, B., & Simmt, E. (2003). Understanding learning systems: Mathematics teaching and complexity science. Journal for Research in Mathematics Education, 34(2), 137–167.

Davis, B., & Sumara, D. (2006). Complexity and education: Inquiries into learning, teaching, and research. Mawah, NJ: Lawrence Erlbaum Associates.

Dooley, K. J., & Van De Ven, A. H. . (1999). Explaining Complex Organizational Dynamics. Organization Science, 10(3), 358–372.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. Research Policy, 30(1), 1019–1039.

Folke, G., & Folke, C. (1992). Characteristics of nested living systems. Journal of Biological Systems, 1(3), 257–274.

Harrison, A. W., & Rainer, R. K. J. (1992). The influence of individual differences on skill in end-user computing. Journal of Management Information System, 9(1), 93–111.

Hause, M. (2006). The SysML Modelling Language. Proceedings of Fifteenth European Systems Engineering Conference.

Helpman, E., & Rangel, A. (1999). Adjusting to a new technology : Experience and training. Journal of Economic Growth, 4(1), 359–383.

Kezar, A., & Eckel, P. (2002). Examining the institutional transformation process: The importance of sensemaking, inter-related strategies and balance. Research in Higher Education, 43(4), 295–328.

Kleijnen, J. P. C. (1995). Verification and validation of simulation models. European Journal of Operational Research, 82(1), 145–162.

Kontoghiorghes, C. (2001). Factors affecting training effectiveness in the context of the introduction of new technology-A US case study. International Journal of Training and Development, 5(4), 248–260. doi:10.1111/1468-2419.00137

Levin, S. A. (2002). Complex Adaptive systems: Exploring the known, the unknown and the unknowable. American Mathematical Society, 40(1), 3–19.

Marten, G. G. (2008). Human ecology – Basic concepts for sustainable development (3rd ed.). Sterling, VA: TJ International.

McCarthy, I. P. (2003). Technology management a complex adaptive systems approach. International Journal of Technology Management, 25(8), 728. doi:10.1504/IJTM.2003.003134

Morel, B., & Ramanujam, R. (1999). through the looking glass of complexity: The dynamics of organizations as adaptive and evolving systems. Organization Science, 10(3), 278–293.

Newell, C. (2008). The class as a learning entity (complex adaptive system): An idea from complexity science and educational research. SFU Educational Review, 2(1), 5–17.

Ramos, A. L., Ferreira, J. V., & Barcelo, J. (2012). Model-based systems engineering : An emerging approach for modern systems. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 42(1), 101–111.

Romme, G., & Dillen, R. (1997). Mapping the landscape of organizational learning. European Management Journal, 15(1), 68–78. doi:10.1016/S0263-2373(96)00075-8

Sage, P. A., & Rouse, W. B. (2009). Handbook of Systems Engineering and Management (p. 847). Hoboken, NJ: John Wiley & Sons Inc.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. Annual Review of Psychology, 52, 471–499.

Senge, P. M. (1990). The fifth discipline. The art and practice of the learning organization. New York: Doubleday.

Weick, K. E. (1979). The social psychology of organizing (p. 32). New York, NY: Random House.

KNOWLEDGE ACQUISITION FOR CLOUD MANUFACTURING

Yifan Mai[a], Lin Zhang[a], Anrui Hu[a], Chenfei Lv[b], Fei Tao[a]

[a]School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
[b]Sino-French Engineering School, Beihang University, Beijing 100191, China

maiyifan@asee.buaa.edu.cn, zhanglin@buaa.edu.cn

ABSTRACT

This paper proposes a methodology for knowledge acquisition in cloud manufacturing (CMfg) system which refers to a new knowledge based manufacturing paradigm. In view of the practical needs, the proposed methodology is designed to be cross-domain. Non-automatic and semi-automatic knowledge acquisition methods were used with the assistant of the automatic one. Details over knowledge acquisition for CMfg were presented, as well as the proposed methodology. System architecture for the knowledge acquisition tool was also put forwards with analysis and illustration.

Keywords: knowledge acquisition, cloud manufacturing, system architecture

1. INTRODUCTION

Cloud Manufacturing (CMfg) refers to a new service-oriented intelligent manufacturing paradigm based on knowledge with high efficiency and low consumption. It is a confluence of multiple disciplines, such as networked manufacturing, web services, cloud computing, cloud security, high performance computing, Internet of Things and so on, which is pushing the manufacturing to be agile, virtualized, intelligent, service oriented, integrated and green (Bo Hu Li, Lin Zhang, and Xudong Chai 2010; Bo Hu Li, Lin Zhang, and Shilong Wang 2010; Fei Tao, Lin Zhang, and VC Venkatesh 2011; Lin Zhang, Yongliang Luo and Fei Tao 2012). During the lifecycle of CMfg system, knowledge is recognized as the key to carry out the intelligent cloud services. It can be classified into five parts including domain knowledge, task knowledge, case knowledge and service description knowledge as required. Concretely, it may appear in the form of service labels, service rules, service ontology, case descriptions, intelligent algorithms, and so on (Anrui Hu and Lin Zhang 2012). From the manufacturing resources' perception to its virtualization package and access, from the cloud services' description to its combination, from the cloud services' scheduling to its optimizing, from the system's fault-tolerant management and task migration to its business process management (Fei Tao, Lin Zhang, and Hua Guo 2011), all the vital parts of CMfg seem to rely on knowledge firmly. When perceiving the manufacturing resources, the system builds up mappings between various resources and their corresponding virtual resource pools while packages the resources as services. When constructing the manufacturing cloud, the system labels each cloud service with its sort, promulgator, application implement and usage before stores these messages into knowledge bases. When matching and combining the cloud services, the system works smoothly owing to the efficient pre-analysis of the users' demands and service descriptions. When dispatching the cloud services, the system makes solutions with the help of the service rules and description labels and optimizes them on the basis of some specific tactics to avoid service scheduling conflicts. During the fault-tolerant management, the system first evaluates the loads of both its physical and virtual modules and stores the results into knowledge bases. Then all the indexes are accessed and abnormal modules are migrated virtually to maintain the load balance while the system is on run. During the business process management, the system first puts some basic flow descriptions of the cloud services into knowledge bases as rules. Then it decomposes the submitted tasks into smaller ones according to the rules and optimizes the combination of the services with some corresponding algorithms. Solutions are returned to the users at last. Obviously, CMfg is a modern manufacturing paradigm depending on the application of knowledge, which is rest upon the accurate knowledge acquisition.

For half a century, knowledge acquisition has restricted the development of knowledge based systems. Many experts and scholars have done a lot of researches on this domain. For example, Cairo and his group proposed (Osvaldo Cairo 1998; Osvaldo Cairó and Silvia Guardati 2012) a graphical logic language for knowledge expression and designed a methodology for multiple-domain knowledge acquisition, which is achieved basically by setting up communications between domain experts and knowledge engineers. Tang and his team (Yuan Yan Tang, Chang De Yan, and Ching Y. Suen 1994) divided documents into geometric layer and logic layer, and mapped the information mined from the geometric layer into the logic layer to accomplish automatic knowledge acquisition. Chen (Ping Chen and Chris Bowes 2012) proposed a word sense disambiguation method based on automatic knowledge acquisition in which he used the DepScore function and GlossScore function together with the most-frequent-sense method to determine the feasible word sense in a given English context. Yu and his partners (Daren Yu, Qinghua Hu, and Wen Bao 2004) developed a way to acquire knowledge from quantitative data by merging the rough set and fuzzy clustering algorithms, which has been successfully used in the diagnosis of the vibration fault of turbine, etc. However, due to the differences in knowledge definition, these researches are always limited in some specific domains and their strategies are not available for knowledge acquisition in CMfg system.

The rest of the paper is organized as follows. Section 2 studies knowledge acquisition for CMfg and proposes a potentially feasible solution for it. Section 3 presents an architecture view for the proposed solution. Finally, conclusions are drawn in Section 4.

2. THE PROPOSED METHODOLOGY

As a new intelligent manufacturing paradigm, the realization of CMfg is bound up with knowledge acquisition. How to acquire knowledge effectively is the premise of knowledge based application. Generally, knowledge acquisition is achieved in the following three ways: non-automatic acquisition, semi-automatic acquisition, and automatic acquisition (Na Li 2009).

(1) The non-automatic knowledge acquisition is usually up to domain experts' own oral instructions.

(2) The semi-automatic acquisition adopts the man-machine interactive mode to help extracting knowledge from domain experts.

(3) On contrast to the former two ways, automatic knowledge acquisition shows strong capabilities to sum up and extract new knowledge from the system's self-learning process automatically.

Considering the practical situation of CMfg, the non-automatic and semi-automatic knowledge acquisitions are primarily introduced into the manufacturing process with the assistant of the automatic knowledge acquisition (Anrui Hu and Lin Zhang 2012). Firstly, the system acquires domain knowledge, reasoning knowledge, and task knowledge through non-automatic acquisition and stores them into knowledge bases. Then it refreshes the existing knowledge as well as gains new service description knowledge by means of semi-automatic acquisition. At last, case knowledge are achieved from knowledge both obtained non-automatically and semi-automatically in its self-learning process automatically. Knowledge storage is also setup. The basic flow of the described knowledge acquisition for CMfg is depicted in Fig. 1, which acts as a dynamic cycle.

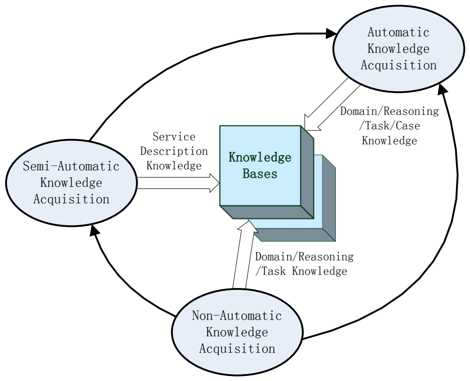

Automatic Knowledge Acquisition
Domain/Reasoning /Task/Case Knowledge
Service Description Knowledge
Knowledge Bases
Semi-Automatic Knowledge Acquisition
Domain/Reasoning /Task Knowledge
Non-Automatic Knowledge Acquisition

Figure 1: Process of Knowledge Acquisition in Cloud Manufacturing.

2.1. Non-Automatic Knowledge Acquisition

A non-automatic knowledge acquisition model for CMfg is presented in Fig. 2. It is the most crucial step in knowledge acquisition for CMfg, which is mostly oriented to domain knowledge, reasoning knowledge, and task knowledge. The traditional non-automatic knowledge acquisition tends to be adapted only to problem solving in some specific domains and have less trouble in knowledge fusion. However, CMfg is an integrated service paradigm, from which users are always looking forwards to complete manufacturing schemes. In fact, a whole manufacturing process often consists of small multiple processes, which requires varieties of professional knowledge. For instance, from design to implementation, from component manufacturing to assembly line working, from machine debugging to safe testing, from coating to after-sales services, a car manufacturing process can be rather complex. It seems that using only single domain knowledge may be likely to lose some important manufacturing details and lower the quality of the service combination, which will lead to resource wastes.

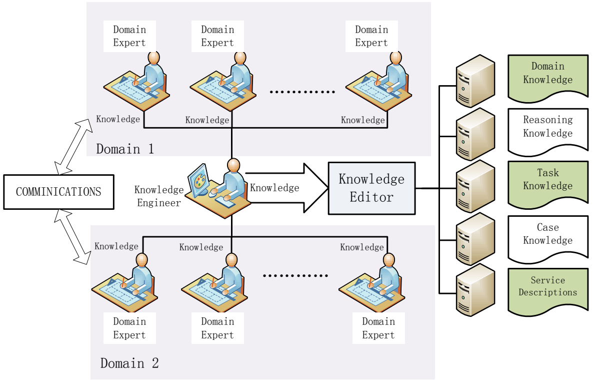


Figure 2: Non-Automatic Knowledge Acquisition in Cloud Manufacturing.

Due to the relatively closeness among knowledge from different domains, non-automatic acquisition is expected to solve the cross domain problems in CMfg. Therefore, the cloud service providers are asked to gather experts from multiple domains together to face-to-face communications and discussions. Then the knowledge engineers fuse the derived knowledge according to the discussions, and ask for the experts to evaluate the quality and feasibility of the fusion results. After that, they file the feasible knowledge in some structured forms and deposit them into knowledge bases. Seeing from the rapid evolvement of manufacturing resources, the cloud service providers should regularly repeat the above process to renew and maintain the knowledge bases and reassess the existing manufacturing knowledge as well.

2.2. Semi-Automatic Knowledge Acquisition

In the process of non-automatic knowledge acquisition, the structures and contents of the knowledge to be obtained are unknown before as well as the specific demands. Nevertheless, semi-automatic knowledge acquisition achieves the structured knowledge with clear demands from domain experts via man-machine interaction. In CMfg, semi-automatic knowledge acquisition is mainly used for getting service description knowledge. Sometimes it also helps updating the existing knowledge with certain templates. A brief architecture model is described in Fig. 3.

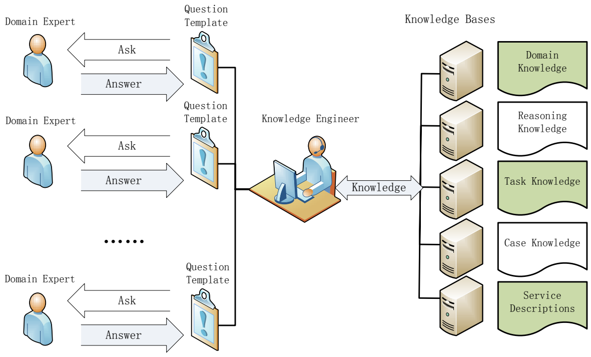


Figure 3: Semi-Automatic Knowledge Acquisition in Cloud Manufacturing.

Knowledge engineers get in touch with domain experts on line to achieve knowledge from them directly by taking prompts, showing guides and using question and answer method, and then finish the knowledge storage. To combine with the semi-automatic method in knowledge acquisition, CMfg can balance the high costs bringing by the non-automatic mode and improve the economic benefits of the whole.

2.3. Automatic Knowledge Acquisition

Automatic knowledge acquisition has been widely considered as the top level acquisition mode in knowledge acquisition as well as the most difficult bottleneck to break through. Compared with non-automatic knowledge acquisition, its difficulties are not only referring to the unclear demands for knowledge but also the ambiguity existing in the acquisition process. In CMfg, automatic knowledge acquisition is also supposed to be cross-domain. It deals with domain knowledge, reasoning knowledge, task knowledge and case knowledge primarily. A model is built up in Fig. 4.

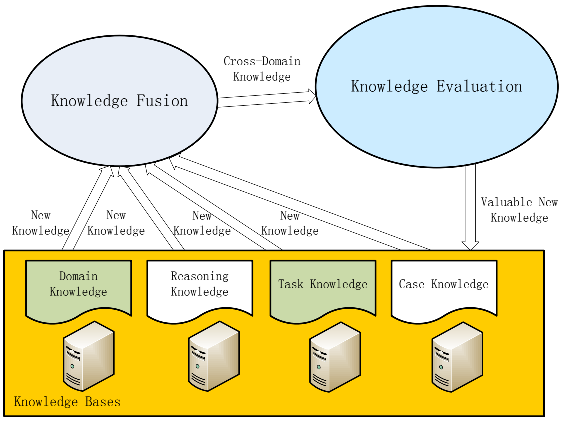


Figure 4: Automatic Knowledge Acquisition in Cloud Manufacturing.

Small knowledge bases published in the manufacturing cloud modify and refresh their existing manufacturing knowledge dynamically by certain strategies and algorithms, and then fuse the new knowledge produced during the previous process with each other. At the meantime, the system evaluates the quality of the fusion based on some assessment mechanism given by domain experts before, and pushes the valuable new cross-domain knowledge into designated knowledge bases via some dynamic correlation and fuzzy clustering methods. However, as a result of the heterogeneity and magnanimity of knowledge in various domains as well as the poor assessment rules for knowledge qualification, to fully fulfill the automatic knowledge acquisition for CMfg without employing the other two modes seems to be next to impossible. Starting from the reality, automatic knowledge acquisition is more appropriate to play an auxiliary role instead of working as the core supporting way to help improve the service efficiency of the whole.

3. THE SYSTEM ARCHITECTURE

In the paradigm of CMfg, the knowledge acquisition tool is supposed to be composed of the following four layers: the knowledge storage layer, the logic interface layer, the application interface layer and the human interaction layer. An overall architecture is viewed in Fig. 5.

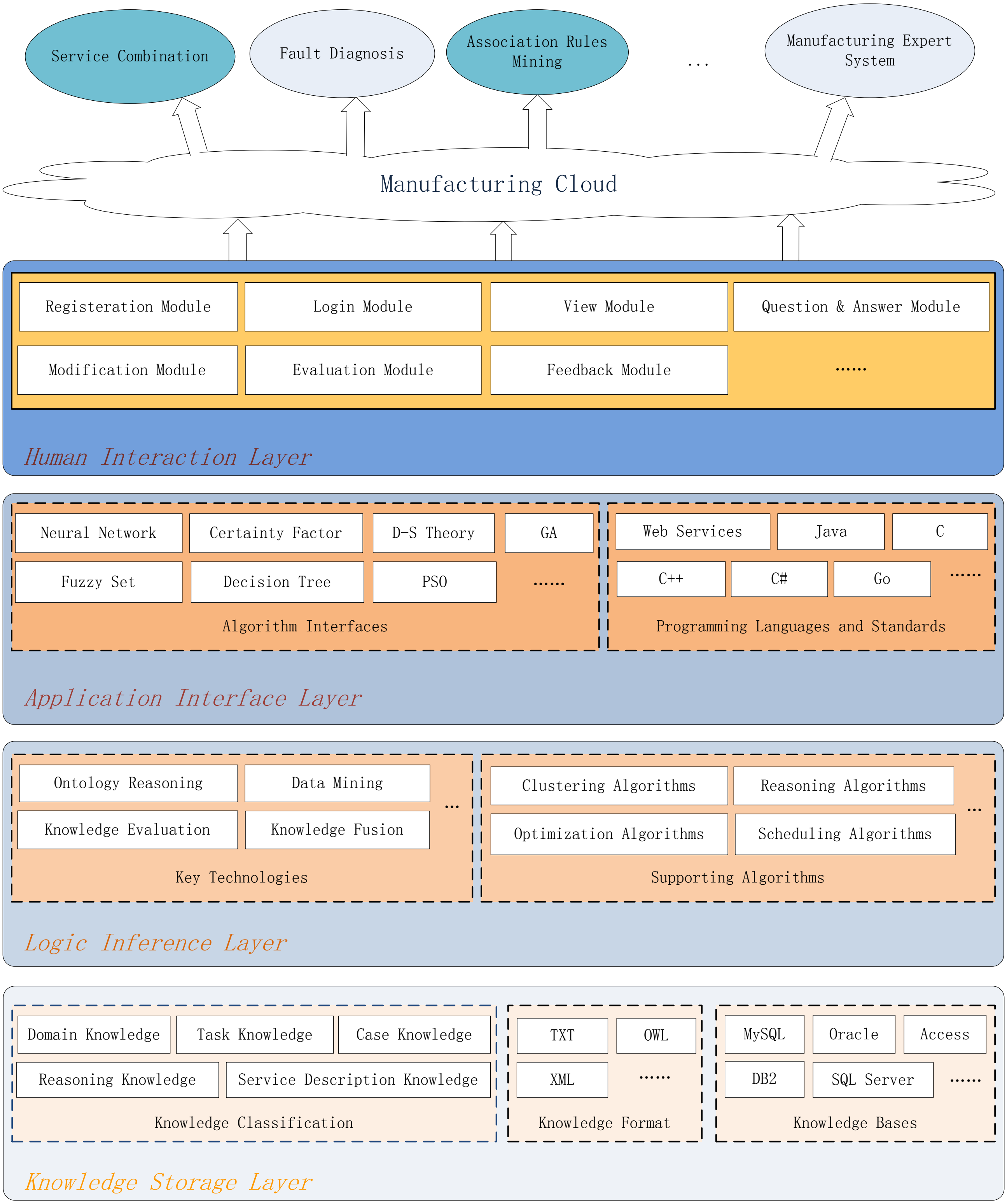


Figure 5: Architecture View of Knowledge Acquisition in Cloud Manufacturing.

3.1. Knowledge Storage Layer

The knowledge storage layer has a double-layer structure, including the index layer and the knowledge layer. The knowledge layer is composed of small knowledge bases published in the manufacturing cloud. The index layer is used to store description labels of each knowledge base and update them dynamically through some inner strategies. In the storage of knowledge, the new knowledge catches its membership notification by matching with labels in the index layer, and then knowledge bases in the knowledge layer distinguishes and receives different kinds of knowledge in the way of verifying these notification. The system updates the label layers at the same time.

3.2. Logic Reasoning Layer

The logic layer encapsulates the inference engine for knowledge acquisition in CMfg, which consists of the intelligent reasoning algorithms library, the intelligent optimization algorithms library (F Tao, D Zhao, YF Hu, and ZD Zhou 2008; Fei Tao and Lin Zhang 2012; Fei Tao, Yuanjun Laili, and YL Liu 2013; F Tao, YJ Laili, L Xu, and L Zhang 2013; YJ Laili, F Tao, and L Zhang 2013) and the ontology reasoning machine and so on. In the process of automatic knowledge acquisition, the system updates and fuses the existing manufacturing knowledge by calling some high performance algorithms and logic strategies.

3.3. Application Interface Layer

The application interface layer provides general interfaces for calling algorithms and logic strategies in the logic reasoning layer under Web service. Web service is a kind of modular component which is language and platform independent. It can be used to realize the combination of various algorithms feasibly by standardizing the input and output interfaces accompanying with their description messages. The standardization of the application interfaces may contribute to developing the performance of the knowledge acquisition in Cloud Manufacturing. Moreover, the application interface layer enhances the extensibility of the knowledge acquisition tool.

3.4. Human Interaction Layer

The human interaction layer offers I/O interfaces for both domain experts and knowledge engineers. It is mainly made up of the modification module, the question and answer module, the evaluation module and the feedback module and so on.

4. CONCLUSIONS

Cloud Manufacturing is a modern service-oriented intelligent manufacturing paradigm, which is based on knowledge. In the lifecycle of cloud manufacturing, the key to improve the quality of cloud services is the accurate knowledge acquisition. However, due to the highly heterogeneity in knowledge representation, knowledge acquisition has been extensively recognized as the kernel and bottleneck of knowledge based systems. In no doubt, it is going to be one of the most important technologies in the future development of Cloud Manufacturing.

ACKNOWLEDGMENTS

This work is partly supported by the NSFC projects (No.61074144and No.51005012) and the Fundamental Research Funds for the Central Universities in China.

REFERENCES

Anrui Hu, Lin Zhang, et al, 2012. Resource Service Management of Cloud Manufacturing Based on Knowledge. *Journal of Tongji University (natural science)*, vol.40, no.7, pp.158-166.

Bo Hu Li, Lin Zhang, and Xudong Chai, 2010. Introduction to Cloud Manufacturing. *ZTE Communications*, vol.16, no.4, pp.5-8.

Bo Hu Li, Lin Zhang, Shilong Wang, et al, 2010. Cloud Manufacturing: A New Service-oriented Networked Manufacturing Model. *Computer Integrated Manufacturing Systems*, vol.16, no.1, pp.1-7.

Daren Yu, Qinghua Hu, and Wen Bao, 2004. Combining Rough Set Methodology and Fuzzy Clustering for Knowledge Discovery from Quantitative Data. *Proceedings of the CSEE*, vol.24, no.6, pp.205-210.

F Tao, D Zhao, YF Hu, and ZD Zhou, 2008. Resource Service Composition and Its Optimal-Selection Based on Particle Swarm Optimization in Manufacturing Grid System. *IEEE Transactions on Industrial Informatics*, vol.4, no.4, pp.315-327.

Fei Tao, Lin Zhang, VC Venkatesh, et al, 2011. Cloud Manufacturing: A Computing and Service-oriented Manufacturing Model. *Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture*, vol.225, no.10, pp.1969-1976.

Fei Tao, Lin Zhang, Hua Guo, et al, 2011. Typical Characteristics of Cloud Manufacturing and Several Key Issues of Cloud Service Composition. *Computer Integrated Manufacturing Systems*, vol.17, no.3, pp.477-486.

Fei Tao, Lin Zhang, et al, 2012. Research on Manufacturing Grid Resource Service Optimal-Selection and Composition Framework. *Enterprise Information Systems*, vol.6, no.2, pp.237-264.

Fei Tao, Yuanjun Laili, YL Liu, et al, 2013. Concept, Framework and Application of Configurable Intelligent Optimization Algorithm. *IEEE Systems Journal, in press*.

F Tao, YJ Laili, L Xu, and L Zhang, 2013. FC-PACO-RM: A Parallel Method for Service Composition Optimal-Selection in Cloud Manufacturing System. *IEEE Transactions on Industrial Informatics, in press*.

Lin Zhang, YongLiang Luo, Fei Tao, et al, 2012. Cloud Manufacturing: A New Manufacturing Paradigm. *Enterprise Information Systems*, iFirst article, pp.1-21.

Na Li, 2009. *Amendment and Acquisition of Text Knowledge Based on the Ontology*. Degree Thesis of Engineering Master. China University of Petroleum.

Osvaldo Cairo, 1998. KAMET: A comprehensive Methodology for Knowledge Acquisition from Multiple Knowledge sources. *Expert Systems with Applications*, vol.14, pp.1-16.

Osvaldo Cairó and Silvia Guardati, 2012. The KAMET II Methodology: Knowledge Acquisition, Knowledge Modeling and Knowledge Generation. *Expert Systems with Applications*, vol.39, pp.8108-8114.

Ping Chen, Chris Bowes, et al, 2012. Word Sense Disambiguation with Automatically Acquired Knowledge. *IEEE Intelligent Systems*, pp.46-55.

Yuan Yan Tang, Chang De Yan, and Ching Y. Suen, 1994. Document Processing for Automatic Knowledge Acquisition. *IEEE Transaction on Knowledge and Engineering*, vol.6, pp.3-21.

YJ Laili, F Tao, L Zhang, et al, 2013. A Ranking Chaos Algorithm for Dual Scheduling of Cloud Service and Computing Resource in Private Cloud. *Computer in Industry*, vol.64, no.4, pp.448-463.

KNOWLEDGE SEMANTIC SEARCH IN CLOUD MANUFACTURING

HU Xiaohang[(a)], ZHANG Lin[(b)], HU Anrui[(a)], ZHAO Dongming[(c)], TAO Fei[(b)]

[(a,b)]School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191
Engineering Research Center of Advanced Manufacturing System of Complex Product, Ministry of Education, Beihang University, Beijing 100191, China
[(c)]Department of Electrical and Computer Engineering, University of Michigan-Dearborn, Michigan 48128, USA

[(a)]hxh1989626@163.com [(b)] zhanglin@buaa.edu.cn

ABSTRACT

In cloud manufacturing system, the distributed stored knowledge is in multiple forms and structure, and its contents are in multiple fields. In this paper, a semantic search engine method based on shared ontology is presented. The application status of ontology in semantic search is studied. In order to enhance the recall rate and precision rate, this search engine computes the semantic matching degree between user requirements and knowledge by semantic similarity computing and logical reasoning.

Keywords: cloud manufacturing, knowledge, semantic search, shared ontology, matching degree

1. INTRODUCTION

Cloud manufacturing is a kind of networked manufacturing mode (Bo Hu Li and ZHANG 2010, ZHANG and LUO 2011, TAO and ZHANG 2011), one of the key thought of it is the servitization of manufacturing resources in the whole life cycle of manufacturing (Bo Hu Li, ZHANG, and CHAI 2010). Users visit cloud manufacturing center through cloud terminal equipment and invoke appropriate services for specific business needs. When a user invokes services through cloud manufacturing center, he has to pay fees to the owners of the resources. In the cloud manufacturing mode, resources are divided into manufacturing resources and manufacturing abilities (Zhang and Luo 2010). Manufacturing resources are resources physically exist, while the configuration and integration abilities in specific production activities are called manufacturing abilities(LUO 2012), for instance, designing and simulation abilities. Manufacturing resources consist of hard resources that include manufacturing devices, computing devices and material and soft resources including data, information, knowledge, etc. As a kind of resources, knowledge plays more and more important roles in the whole life cycle of manufacturing. Cloud manufacturing is a manufacturing mode with the purpose of building a public manufacturing environment in which users can take part in every process of product designing and manufacturing (Tao, Zhang, and Luo 2011; Tao and Guo 2012; Tao and Cheng 2012). In practical application, knowledge from resources providers and consumers works in the management of enterprise business process as cloud services. In the search, match, combination and configuration of cloud services, knowledge in multiple fields must be taken into account. As a result, cloud manufacturing is a manufacturing mode based on knowledge (Hu and ZHANG 2012).

When performing complex tasks, users need knowledge related to multiple fields from cloud manufacturing. But facing big knowledge saved in cloud manufacturing, traditional search based on key words or themes cannot satisfy the requirement of accuracy. Efficient search model that understands users' meaning correctly and achieves intelligent search has to be established. Existing mature search modes of network information primarily include the following three:

1. Search engine based on key words. Most search engines, for example, Google, using this mode. It adopts distributed search mechanism which runs in multiple storage equipment at the same time through the local agent. Search results are feedback to users after unified sorting. This mode can ensure coverage, but cannot understand requirements of users in semantic level.
2. Search engine based on metadata. This method collects and selects information of a topic or in a subject scope according to certain standards. The selected resources are described and signed for users to search rapidly, for example, subject gateways. However, the requirements of users are related to production activities of complex product in cloud manufacturing, so it's difficult to meet the requirements in specific field.

3. Search engine based on document structure. Literatrue libraries use this mode in usual. Literatures are structured in certain form, the weight of key words can be defined according to the importance of every part. The knowledge in cloud manufacturing has no standard structure, so this mode does not suits.

To the non-applicability of search engine for knowledge in cloud manufacturing in present, this paper proposes semantic search based on shared ontology as the search mode in cloud manufacturing. The semantic relativities between cloud terminal users' requirements and knowledge are got through semantic similarity computing and logical reasoning.

2. KNOWLEDGE SEARCH IN CLOUD MANUFACTURING

2.1. Characteristics of knowledge in Cloud Manufacturing

The knowledge in cloud manufacturing has the following characteristics:

1. Multiple forms and structures. According to the knowledge level principle proposed by Alan Newell in 1982, knowledge modeling is on the level of concept. In other words, one of the basic rules of knowledge modeling is to focus on the concept of knowledge structure and expression ability while the representation of knowledge is ignored. As a result, knowledge is very different in forms, structures, storage modes, data structures and application interface, the interoperability of knowledge is pretty low.
2. Complex content and fields. Except for the real-time information from manufacturing site, most knowledge is edited and maintained by engineers, managers or experts. As a result, the knowledge is complex in content and fields. In cloud manufacturing, reasonable knowledge system is needed to join the complex knowledge up in to achieve semantic search.
3. Distributed storage. The knowledge in cloud manufacturing is stored in distributed devices. Considering the varied forms and complex content of knowledge, it's difficult to find suited search mode to achieve efficient distributed parallel search when the manufacturing resources are unstable in cloud manufacturing.

2.2. Knowledge Search Requirements in Cloud Manufacturing System

The concept of semantic search is proposed on the basis of semantic Web, which is aimed to find the implied relation of information through semantic computing. Semantic computing includes the following three aspects.

The first is computing the semantic distance (ZHONG, ZHU, and LI 1995) between concepts or terms using concept system to realize human-machine interaction in semantic. Second, logical relation can be applied to the reasoning process of search engine (LIU and LI 2005) in order to improve search quality. The last is to learn the user's habits and search intentions by mining historical data. In this way the personalization of search results is formed.

In view of the above characteristics of knowledge in cloud manufacturing, in order to understand users' intentions and find out knowledge highly related to requirement in semantic from big data, the semantic search engine must have the following features:

1. Reasonable resource description mechanism. A reasonable description mechanism could be set up to achieve the universality of semantic search in cloud manufacturing. The mechanism formalizes description of knowledge in different forms and structures. The search for knowledge is turned into that for knowledge description information.
2. Complete knowledge system. Because of the complexity of the users' requirements and knowledge in different fields, the semantic search engine must set up complete knowledge system including concept, instance, relation and definition, etc. The system should be expressed in language and grammar that can be identified by computer. Knowledge system is the foundation for computer to understand users' intentions.
3. High recall rate. The recall rate of search engine is the percentage of searched related information in all related information. The goal of semantic search engine is to find knowledge that related to users' requirements impliedly by semantic relativities computing. This process enhances call rate to a great extent.
4. High precision rate. The precision rate of search engine is the percentage of related information in all results. Precision rate is the fundamental measure of search performance.

3. KNOWLEDGE SEARCH MECHANISM

The operating principle of knowledge semantic search engine is shown in figure 1.

The semantic search engine in cloud manufacturing includes two main function modules, *information extraction* and *semantic computing based on ontology*. This chapter will elaborate the technical realization of these two function modules.

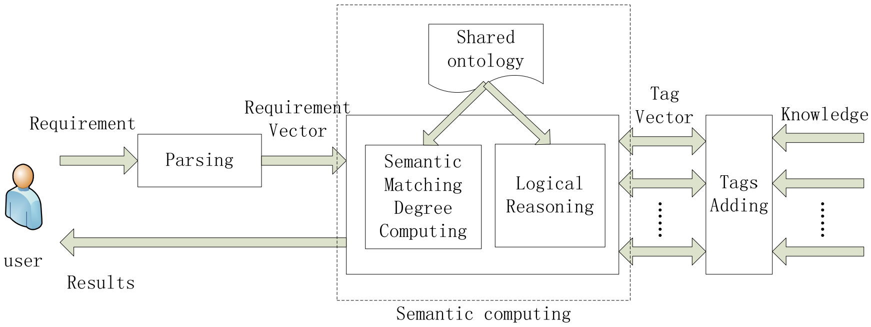


Figure 1: Search Operating Principle

3.1. Information Extraction

3.1.1. Vector Space Model (VSM)

VSM is a model that universally used in natural language processing. This model regards every document as a vector composed of characteristics. And the characteristic are endowed with a certain weight by algorithm computing. The mapping relations between documents and vectors in VSM are as table 1.

Table 1: mapping relations in VSM

documents	vectors
word	characteristic
weight	coordinate value
document	vector
document collection	vector set

Documents are represented as vectors of characteristics and the weight of them in VSM. Thus natural language processing becomes vector mathematic operation. The semantic computing between requirement and knowledge in cloud manufacturing is realized in vector space model. Thus requirement and knowledge should be in the form of vectors.

3.1.2. Parsing

As requirements are sentences or phrases generally. They should be transformed into vectors by stem segmented and stop removal in VSM before semantic computing.

One word can show a variety of forms because of inflection or derivation. Stem segmentation (WU and QIAN 2012) is the process of simplifying the forms into a common stem. Stem segmentation can be divided into two classes, algorithm-based and dictionary-based. In this paper, porter stemmer is used, which is a kind of mature algorithm-based stem segmentation. Stop words are functional words in documents, the engine will skip and process the next one when meets these words. The stop removal is often done by querying the stop words dictionary. Results by parsing are elements that compose the requirement vectors.

3.1.3. Tags Adding

As knowledge is in multiple forms in cloud manufacturing, it’s difficult to find a mechanism suits all knowledge. So it’s necessary for the engine to add formalized tags to knowledge. The relativities between requirement vectors and tag vectors are achieved by semantic computing.

Tags can be added manually or automatically.

1. The tags of knowledge in forms of image, audio or formula can be added manually by engineers in the process of servitization. The tags are published by maintainers of cloud manufacturing after being audited.
2. For the knowledge in forms of document, tags can be added automatically through TF*IDF framework which is known as feature extraction. The computing process is as formula 1.

$$w_k = \frac{W_{Tf}\left(k,\vec{d}\right)\times IDF_k}{\sqrt{\sum_{k\in\vec{d}}\left[W_{Tf}\left(k,\vec{d}\right)\times IDF\right]^2}} \qquad (1)$$

TF is the term frequency, which means the times of the key word appearing in files. This paper takes the logarithm to it to smooth the influence of word frequency jumping, as formula 2.

$$W_{Tf}\left(k,\vec{d}\right) = \log\left[1+Tf\left(k,\vec{d}\right)\right] \qquad (2)$$

IDF is inverse document frequency, which reflects the distribution of characteristics in the full set of documents. The value of IDF means the quantity of information a characteristic brings, as formula 3.

$$IDF_k = \log\left(\frac{N}{n_k}+0.01\right) \qquad (3)$$

The product of IF and IDF can be used to get the weight matrix of document collection. Take the characteristics with maximum weight as the tags of each document.

3.2. Semantic Computing based on Ontology

3.2.1. Shared Ontology Construction

Ontology is selected to be the form of knowledge system in semantic search as it's the core of semantic Web. Ontology is a kind of shared conceptual model can be recognized by computer, which abstracts objective reality in certain fields. It is the basis of semantic interoperation.

Ontology is different in modeling depth. Guarino (Guarino and Nicola 1997) regards the detail level of description as the basis of classification of ontology. The ontology with high detail level of description is named reference ontology. It takes the comprehensive concepts of object on the World Wide Web as metadata to establish fine-grained ontology model. Shared ontology is the ontology with low detail level of description, which mainly consists of concepts, relations and the hierarchical structure in certain fields. Shared ontology is independent of the specific application environment and task instances. It describes the constraints of relations instead of expressing the detail content of knowledge. The ontology example edited by protégé 4.1 is as the following figure 2.

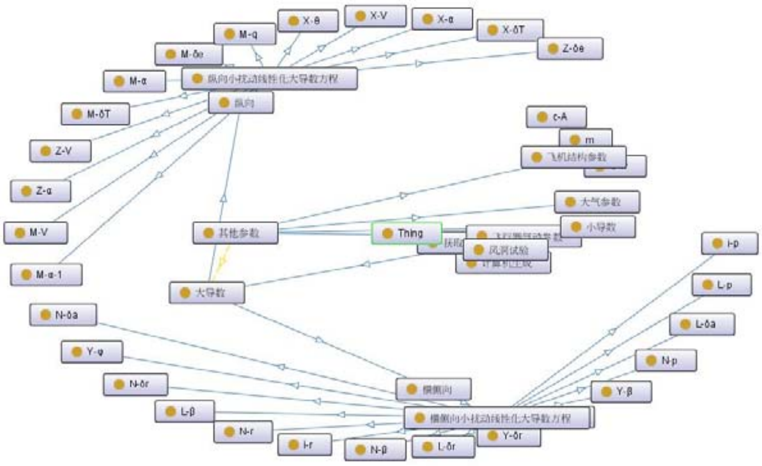


Figure 2: Ontology Example

Shared ontology plays the role of field concepts dictionary to some degree. A shared ontology has realized decoupling with manufacturing resources as soon as it is constructed. Thus it can be applied to different configuration environment and application requirements. Shared ontology makes it possible for computer to understand resources information and users' requirements when it provides the standard terminology to computer. In search engine, the shared ontology with reasonable structure and accurate content is the premise to realize efficient semantic search.

3.2.2. Semantic Matching Degree Computing

Ontology can be considered as a tree structure for semantic distance computing as its layer of structure is stable. The calculating flow is as figure 3.

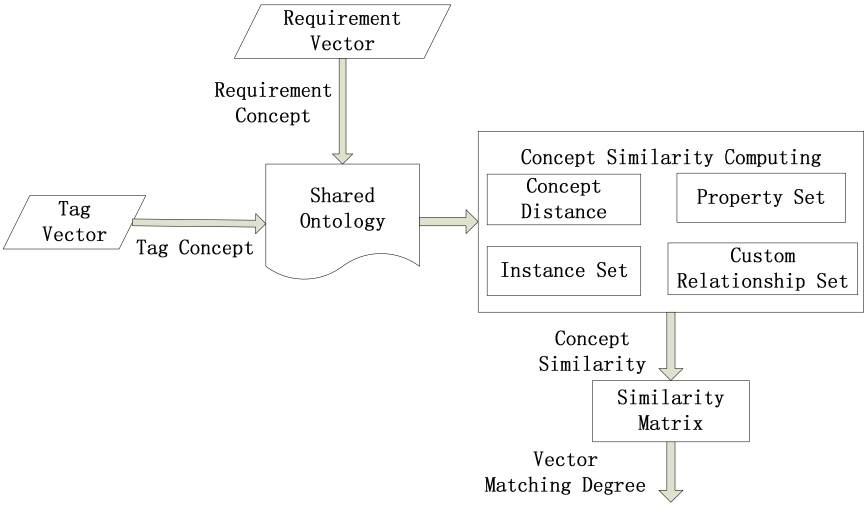


Figure 3: Semantic Relativity Calculating Flow

This paper takes shared ontology as the basis of semantic matching degree computing between requirement and knowledge tag. While concept similarity computing is the fundamental computing (WANG and MA 2007), the semantic similarity between two concepts can be got by concept similarity computing.

Concept similarity computing consists of four parts in this paper. This is a synthesis algorithm taking the distance, property and content of ontology into consideration. Concept distance can be got by analyzing the shared ontology structure. The property, instance, custom relationship set similarity can be got by computing Cartesian product of two concepts in ontology. Then we obtained the concept similarity by weighted summing the above four aspects. Every concept is one dimension of requirement vector or tag vector. The similarity matrix is generated from vectors by semantic similarity computing. Finally, semantic matching degree, which is the matching degree of two vectors, is worked out based on the matrix.

3.2.3. Logical Reasoning

With the emergence of semantic Web, a series of ontology representation language arises. In the modeling of complex knowledge, OWL has been recommended as a standard of ontology modeling by W3C since 2004. OWL has logical reasoning ability and extensibility to a certain extent. OWL is used as the description language to construct shared ontology in this paper.

The basic building blocks of OWL include class (concept), role (properties) and individual (concept or instance). The role is relation between class and class, class and individual, or individual and individual.

For instance, in figure 4, the similarity between B and C can be more than 0 if computing it by semantic distance. But A and B have the relation of *disjointWith*, according to the definition of *disjointWith*, the intersection of A and B is empty. So the similarity between C and B is 0 as C is the subclass of A.

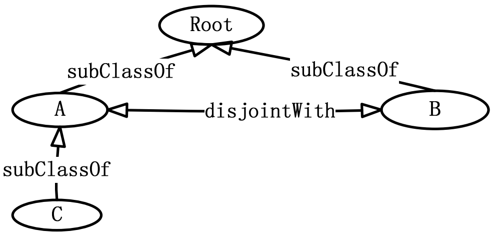


Figure 4: Ontology Logical Reasoning

This shows that logical reasoning can be used to filter the results of semantic similarity computing in ontology-based semantic search. The number of search results is decreased through logical reasoning and the results left behind suit requirements better. The precision rate of search engine is improved.

Semantic search engine sorts the search results based on the matching degree between requirement vectors and tag vectors in cloud manufacturing. Sorting results are feedback to users through cloud manufacturing center. What users need is knowledge but not its description information, so the results should be application interfaces or links

4. CONCLUSION

This paper studies the knowledge semantic search engine in cloud manufacturing. In view of the diversity, complexity and dispersibility characteristics of knowledge in cloud manufacturing, it establishes semantic search mode on the basis of knowledge system. This paper computes the semantic relativities between users' requirements and knowledge by concept semantic similarities computing based on shared ontology. In the process of computing, the engine takes use of the logic relations of OWL to make logical reasoning for requirement in order to improve recall rate and precision rate.

The next work consists of the algorithms improvement of tags adding and concept similarity computing. And prove systemically the correctness of them.

ACKNOWLEDGMENTS

This work is supported in part by NSFC project (No.61074144 and No.51005012) and the Fundamental Research Funds for the Central Universities in China.

REFERENCES

Bo Hu Li, ZHANG Lin, et al. Cloud manufacturing: a new service-oriented networked manufacturing model. *Computer Integrated Manufacturing Systems*. 2010,16(1):2-4.

L Zhang, YL Luo, WH Fan, F Tao, L Ren. Analyses of cloud manufacturing and related advanced manufacturing models. *Computer Integrated Manufacturing Systems*, 2011.11(3):458-468. (in Chinese)

F Tao, L Zhang, VC Venkatesh, YL Luo, Y Cheng. Cloud manufacturing: a computing and service-oriented manufacturing model. *Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture* (Proc IMechE Part B: J Eng Manufact), 2011,225(10):1969-1976. (8 Pages) Oct. 2011

Bo Hu Li, ZHANG Lin, CHAI Xudong. Introduction to Cloud Manufacturing. *ZTE COMMUNICATIONS*. 2010,16(4)

L Zhang, YL Luo, F Tao, L Ren, H Guo. Study on the key technologies for the construction of manufacturing cloud. *Computer Integrated Manufacturing Systems*, 2010, 16(11)：2510-2520

LUO Yongliang, ZHANG Lin, TAO Fei, etc. Key technology of manufacturing capability modeling in cloud manufacturing mode. *Computer*

Integrated Manufacturing Systems. 2012,18(7):1357-1361.

F Tao, L Zhang, YL Luo, L Ren. Typical characteristics of cloud manufacturing and several key issues of cloud service composition. *Computer Integrated Manufacturing Systems*, 2011.17(3):477-486. (in Chinese)

F Tao, H Guo, L Zhang, Y Cheng. Modelling of combinable relationship-based composition service network and theoretical proof of its scale-free characteristics. *Enterprise Information Systems*, 2012, 6(4) : 373-404 (32 Pages), Oct. 2012

F Tao, Y Cheng, L Zhang, D Zhao. Utility modeling, equilibrium and collaboration of resource service transaction in service-oriented manufacturing. *Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture* (Proc. IMechE Part B: J Engineering Manufacturing) 2012, 226(6):1099-1117 (19 Pages) Jun. 2012

A Hu, L Zhang, F Tao, Y Luo. Resource Service Management of Cloud Manufacturing Based on Knowledge. *Journal of Tongji University*, 2012, 40(7):1092-1101

Klinger T. Image Processing with LabVIEW and IMAQ Vision. *Upper Saddle River, New Jersey: Prentice Hall PTR*, 2003.

ZHONG J, ZHU H, LI Y, et al. Using information content to evaluate semantic similarity in a taxonomy. *processing of the 14th International Joint Conference on Artificial Intelligence* (IJCAI-95). Washington: ACM Press, 1995.

LIU Jin, LI Bing. Research on Logical Analysis of Web Ontology Language. *Computer Engineering*. 2005,4.

WU Sizhu, QIAN Qing, et al. Comparative Analysis of Methods and Tools for Word Stemming. *Library and Information Service*. 2012,52(15).

Guarino, Nicola. Semantic Matching: Formal Ontological Distinction for information Organization, Extraction, and Integration. *Pazienza M T, eds. Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, Springer Verlag. 1997. 139-170

WANG Liancheng, MA Qiang.The Computation of Ontology Similarity Based on Concept Value. *Computer Technology and The Application Progress*. 2007:692-694.

A MODEL BASED ON ARTIFICIAL NEURAL NETWORK FOR RISK ASSESSMENT TO POLYCYCLIC AROMATIC HYDROCARBONS IN WORKPLACE

F. Facchini, G. Mossa, G. Mummolo

Department of Mechanics, Mathematics and Management, - Polytechnic of Bari, - Viale Japigia, 182, Bari, Italy

f.facchini@poliba.it, g.mossa@poliba.it, mummolo@poliba.it

ABSTRACT

Polycyclic aromatic hydrocarbons (PAHs) are formed during incomplete combustion in different production processes; exposure to PAH-containing substances increases the risk of cancer in humans.
The environmental monitoring used to assess human exposure to airborne PAHs during work, generally involves the employment of diagnostic methods derived from analytical chemistry, characterised by an elevated cost and the use of a "trial and error" approach.

The aim of this study is to develop a decision support tool that, through the characteristic parameters of a workplace and using an artificial neural network, simulates the concentration of different species of pollutants (PAHs groups) statistically present in the environment. In this way it is possible to perform a preliminary risk assessment that, besides allowing an immediate perception of the level of risk to which workers are exposed, can undertake environmental monitoring analysis on the detection of a limited number of pollutant species, in order to reduce costs and increase the sustainability of the production system.

Keywords: model of prediction, risk assessment, environmental monitoring, decision support system

1. INTRODUCTION

There is a growing interest, both scientific and industrial, in the environmental management of production systems. Great attention is paid to activities that increase the sustainability of production systems.
In this context, there is significant interest in workplaces in which air pollutants that are highly dangerous to the health of workers are emitted by production processes. Among the various types of pollutant, Polycyclic Aromatic Hydrocarbons (PAHs) are amongst the most harmful chemical compounds to human health. The carcinogenic effect of PAHs has been deeply investigated in the past, and is nowadays well known.

Considering the large variety of production processes that result in the generation and dispersion of pollutants, there is a wide range of monitoring equipment based on technologies derived from analytical chemistry. So it is often very difficult to perform preliminary risk assessments of a specific workplace. In fact, in many cases the analytical methods used for environmental monitoring are expensive and use a "trial and error" approach.

Forecasting the concentrations of air pollutants represents a difficult task due to the complexity of the physical and chemical processes involved. Several approaches have been used, branching into two main streams: deterministic approaches, which involve numerically solving a set of differential equations, and empirical approaches, where different functions are used in order to approximate the concentrations of the pollutants depending on the external conditions (Hrust et al., 2009).
The first type of approach does not require a large quantity of measured data, but it demands sound knowledge of the pollution sources, the temporal dynamics of the emission quantity, the chemical composition of the exhaust gasses and physical processes in the atmospheric boundary layer. This crucial knowledge is often limited and also requires computational resources. Thus approximations and simplifications are often employed in the modelling process. On the other hand, applications of deterministic models are limited to a lesser extent with regard to the selection of the domain. A recent example of such an approach is the work of (Finardi et al. 2008).
In contrast, the second type of approach usually requires a large quantity of measured data collected under a large variety of atmospheric conditions. By applying regression and machine learning techniques, a number of functions can be used to fit the pollution data in terms of selected predictors. One drawback of this technique is that the model is usually confined to the area and conditions present during the collection of the measurements (Kukkonen et al., 2003). Nevertheless, this approach is generally more suitable for the description of complex site-specific relations between concentrations of air pollutants and potential predictors, and consequently it often results in a greater accuracy, when compared with deterministic models (Gardner and Dorling, 1999).

Neural network empirical approaches have been frequently used in recent atmospheric and air quality modelling studies. (Božnar et al. 1993) were the first to describe the neural network modelling of hourly concentrations of sulphur dioxide. (Gardner and Dorling 1998) gave a very informative review of the applications of artificial neural networks in science in general and, particularly, in atmospheric sciences. They emphasised the usefulness of neural networks (NN) when dealing with non-linear systems, especially when theoretical models of the system cannot be constructed. (Perez et al. 2000) developed a neural network model to predict PM_{10} hourly concentrations by fitting a function of 24-hourly average concentrations from the previous day. They found errors ranging between 30% and 60%. In order to decrease the errors, they considered noise reduction in the data, rearrangement, an increase in the learning dataset, and the inclusion of meteorological variables as input variables. They concluded that noise reduction prior to modelling is essential. A possible improvement can be achieved by explicitly taking into account the relevant meteorological variables.

Karppinen et al. published two papers addressing the development of a modelling system for predicting NOx and NO_2 concentrations in an urban environment in Helsinki. The first paper (Karppinen et al., 2000a) was related to the model development and its application in air quality prediction, as well as traffic planning. The system includes the following models: the estimation of traffic volumes and travel speeds, the computation of emissions from vehicular sources, a model for stationary source emissions, a meteorological pre-processing model and dispersion models for stationary and mobile sources. Alternatively the second paper (Karppinen et al., 2000b) presented a comparison between the predicted and measured concentrations. According to the authors, the modelling system was fairly successful in predicting NOx concentrations and was successful in predicting NO2 concentrations. They also argued that none of the methods are able to forecast the peak values, due to the under-representation of these cases within the overall dataset.

(Perez and Reyes 2006) developed a multi-layer perceptron (MLP) model to forecast the daily maxima of PM_{10} concentrations one day in advance. The same model was applied to five measuring stations in the city of Santiago, Chile. They compared values forecasted with MLP, linear and persistence models using the same input variables. They concluded that the MLP model performed well and that the relatively small differences between the linear and MLP models emphasised the importance of selecting the correct input variables.

In the scientific literature there are only a few cases related to the use of predictive models in confined spaces, such as apartments, residential buildings, workplaces, etc.

The aim of this work consists in defining a support decision tool for preliminary risk assessment in workplaces, where is possible to identify a narrow set of environmental variables, and there is a strong correlation between concentrations of pollutant in the air and the typology of the manufacturing process. For this scope an Artificial Neural Network (ANN) has been realised that is able to provide a forecast about the presence and concentration of the main pollutants released by a particular production process, in connection to different distances from the source of emission.

This paper is structured as follows: section two deals with a taxonomic clustering of the industrial processes that provide the production of similar outputs, also characterised in many cases by the same emission levels. In section three a model is proposed that, depending on the input parameters such as the total PAHs, distance and categories of workplace, can provide prediction data for the relative concentrations of different pollutants. Finally, the model results are examined in order to evaluate the reliability of the forecast generated by the model.

2. WORKPLACE: TAXONOMIC EVALUATION OF SOURCE PAHS

Polycyclic aromatic hydrocarbons (PAHs) are formed during incomplete combustion. They occur in the environment as complex mixtures of many components with widely varying toxic potencies. Several compounds of this group have been classified by the International Agency for Research on Cancer (IARC) as probable (2A) or possible (2B) human carcinogens (Boström, 2002). Due to its high carcinogenic potency and its presence in the environment, benzo(a)pyrene (Bap) is often used as an indicator of human PAHs exposure (Han, 2011). Road paving, sintering plants, and rubber production are only a few of the principal workplaces where there are industrial processes that emit high concentrations of PAHs into the environment.

The first phase of this study consists in the classification of workplaces into four categories. Each category includes industrial processes that provide for the production of an output that is similar or complementary to the same supply chain.

The categories identified are as follows:

- *Energy activities* that include all those processes related to the production of energy, including: combustion plants, oil refineries and gas, coke ovens, gasification plants and liquefaction of coal;
- *Production and processing of metals* which include: production plants of cast iron or steel, plants for the processing of metals, foundries, sinter plants and surface metal treatment plants;
- *Cooking activities* that include: industrial kitchens and restaurants;

- *Insulation activities* that include: treatments of roofing bitumen for buildings and road paving.

For all categories a series of case studies from the literature have been analysed, and for each the measures of concentration of particulate-bound PAHs emitted into the environment have been considered.

To assess the health-risks associated with PAHs exposures, it is important to know the total carcinogenic potency arising from the exposures of various PAHs compounds. In principle, the carcinogenic potency of a a given PAHs compound can be assessed according to its benzo[a]pyrene equivalent concentration (BaP_{eq}). Calculating the BaP_{eq} concentration for a given PAHs compound requires the use of its toxic equivalent factor (TEF; using benzo[a]pyrene as a reference compound) to adjust its original concentration. Among the 16 PAHs species identified as priority pollutants, we consider seven PAHs species absorbed on particulate matter, according to the carcinogenic potency factor developed by the US Environmental Protection Agency (Table 1), including: Benzo[a]anthracene (BaA), Chrysene (CHR), Benzo[b]fluoranthene (BbF), Benzo[k]fluoranthene (BkF), Benzo[a]pyrene (BaP), Indeno[1,2,3-c,d]pyrene (IND), Dibenz[a,h]anthracene (DBA).

For each case study the corresponding BaP equivalent concentration (BaP_{eq}) has been calculated using the TEF approach recommended by the US EPA.

Table 1: Benzo(a)pyrene Toxic Equivalent Factor (US EPA, 1986)

Compound	Carcinogenic potency
Benzo[a]anthracene (BaA)	0.1
Chrysene (CHR)	0.001
Benzo[b]fluoranthene (BbF)	0.1
Benzo[k]fluoranthene (BkF)	0.01
Benzo[a]pyrene (BaP)	1
Dibenz[a,h]anthracene (DBA)	1
Indeno[1,2,3-c,d]pyrene (IND)	0.1
Other 9 PAHs species	0

It has been observed that categories related to activities relating to insulation and energy are characterised by the highest values of Bap_{eq} (about 800 ng/m3). In fact bitumen is a complex hydrocarbon material containing components in many chemical forms, the majority of which are of high molecular weight (Posniak 2005). In experimental studies polycyclic aromatics with 3 to 7 fused rings with molecular weights in the range 200 to 450 have been shown to be biologically active carcinogens, in particular Benzo(a)pyrene and Dibenz[a,h]anthracene (present in a high percentage) are considered to be powerful carcinogens (Agency for Toxic Substances and Disease Registry, 2009). Moreover fumes, created when asphalt is heated, contain very small, solid, airborne particles that are easily inhaled by workers. Fumes may also contain hydrogen sulphide vapours, which are very toxic, as well as vapours generated by the solvents used to "cut" the asphalt (Burstyn 2000).

In the combustion processes, for the production of energy, the main cause of emissions to the atmosphere of particulate-bound PAHs is coal. Coal in fact contains large quantities of organic and inorganic matter. When coal burns, chemical and physical changes take place, and many toxic compounds are formed and emitted; PAHs are among those compounds. The emissions, in this case, are limited by filtering systems present along the lines of the process (Boström 2002).

The category for the production and processing of metals is characterised by the intermediate values of Bap_{eq} (of the order of 50 ng/m3). The mechanisms associated with the generation of PAHs in the high-temperature combustion process of the smelters works, followed three major pathways, including pyro-synthesis, direct emission of unburned fuel, and thermal destruction of fuel components (Tsai, 2001). The PAHs formed and released by pyrolysis in a limited oxygen supply can appear free in gaseous form and are adsorbed onto dust particles (Tsai, 2000). For the iron and steel industries, PAHs are released from coke manufacturing, sintering, iron making, casting, moulding, cooling, and steel making processes (Lin, 2008). The average emission levels are lower than the values of the preceding categories, because this process requires a lower percentage of organic matter, compared to the activities of insulation and energy.

The category that includes cooking activities is the least dangerous when compared to the other categories; in fact emissions of PAHs in this workplace are in the order of 10 ng/m3. In industrial kitchens or in the restaurants, the emission of PAHs is related to frying at high temperatures. The levels of the (PAHs) pollutants, from the process of cooking, are strongly related to cooking style, lipid content of the food, and the quantities of food cooked. Another variable that significantly affects the level of PAHs emission is the type of cooker, in fact frying on a gas stove caused significantly higher amounts of ultrafine particles compared with frying on an electric stove (Li, 2003). If on the one hand this category includes workplaces with smaller dimensions compared to the workplaces of the other categories, on the other hand the rate of natural ventilation is higher than in non-residential buildings.

3. MODEL OF PREDICTION FOR PAHS EMISSIONS

The second step in this work consists of defining the input and output parameters of the model.

The model includes indoor workplaces or those activities that imply a direct contact among the sources of emission and the worker (e.g. workers employed in road paving, manufacturing and laying of bituminous mixtures, etc.). The input variables of the model are as follows:

Total PAHs: it is a numeric variable given by the sum of the concentration (only the particulate matter phase) of all priority compounds, belonging to PAHs groups, that are the most dangerous to human health;

Distance: it is a qualitative variable that as a function of the distance between the sources of PAHs emission and the position of devices for environmental monitoring can assume three different parameters:

1. First Line: distance between the source of PAHs emission and the device for environmental monitoring is less than 2 metres;
2. Second Line: the distance between the source of PAHs emission and the device for environmental monitoring is between 2 and 10 metres;
3. Third Line: the distance between the source of PAHs emission and the device for environmental monitoring is over 10 metres.

Categories of workplace: this is a qualitative variable that consider the type of manufacturing process or activity from which the PAHs emissions are generated. This variable can assume four types corresponding to: energy activities, production and processing of metals, cooking activities or insulation activities, as already described in the previous paragraph.

The core of the prediction model consists of a system based on an Artificial Neural Network (ANN). This learning technique, that mimics the biological learning process occurring in the brain, has been used. Neural networks present a robust way to predict real-value concentrations after learning from a supplied sample set. Such networks connect a number of individual elements, each of which take a set of inputs and produce a single real number. The learning algorithm determines the numeric weights to be applied between each of these neurons to obtain the desired output. One main advantage of this technique is that it can produce good results, even when supplied with noisy and incomplete data (Aquilina, 2010).
The network architecture provides: three nodes in the input; three hidden layers consisting, respectively, of 2-12-21 neurons and seven nodes of output (Fig. 3). Each input signal is weighted, it is multiplied by the weighted value of the corresponding input line (by an analogy to the synaptic strength of the connections of the biologic neurons). The artificial neuron will combine these weighted inputs by determining their sum, and with reference to a threshold value and an activation function it will determine its output. In this case a *Gaussian distribution* is used for assigning the weights to each variable.

The ANN works with a data set identified by a sample, i.e. a subset of the population representing the phenomenon studied. To be more precise, given the ANN three types of subset of the available sample can create the forecasting model: the training set, the test set, and the validation set:

- *training set*, the group of data constituted by a sample of 60% (percent of total data) that train the network, i.e. by which the network adjusts its parameters (thresholds and weights), according to the gradient descent for the error function algorithm, in order to achieve the best fit of the non-linear function representing the phenomenon;
- *testing set,* the group of data constituted by a sample of 20% (percent of total data), given to the network still in the learning phase, by which the error evaluation is verified in order to effectively update the best thresholds and weights;
- *validation set*, the group of data constituted by a sample of 20% (percent of total data) used to evaluate the ANN generalisation, i.e. to evaluate whether the model has effectively approximated the general function representative of the phenomenon.

The network has been trained using the back-propagation routine. This typology of algorithm is used to train a network for a desired output. This method minimises the squares of the residuals (differences between desired outputs and network outputs) by modifying the network weights. It approximates the desired output using the gradient descent technique.

The validation data set has been used to monitor the alteration in the training error during the learning progress (Fig. 1) of the neural network.

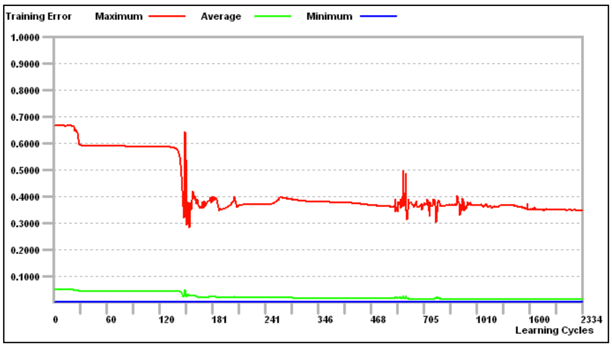


Figure 1: Alteration of the training error

In order to minimise the overtraining problem, the training phase is stopped when the mean square error (MSE) assumes values lower than 0.01.

To evaluate the accuracy of this ANN, the correlation coefficient between the measured data (training set) and the data predicted by the trained neural network (Fig. 2) has been calculated.

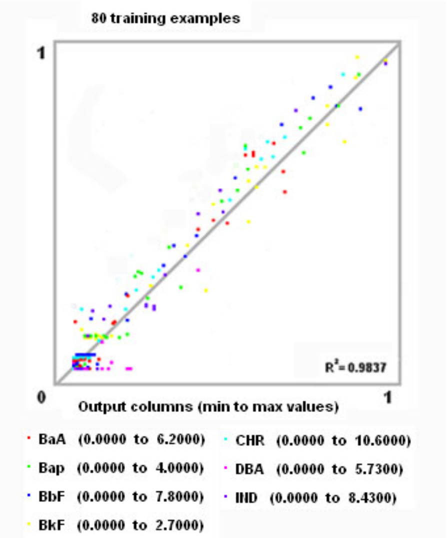


Figure 2: Forecast of the ANN (on training set)

According to US EPA guidelines, among the 16 PAHs species identified as priority pollutants, the model includes only seven PAHs species absorbed on particulate matter, that are considered powerful carcinogens (Tab. 1). The concentration, measured in the environment, of this species representing the output variables of the model (Fig. 3):

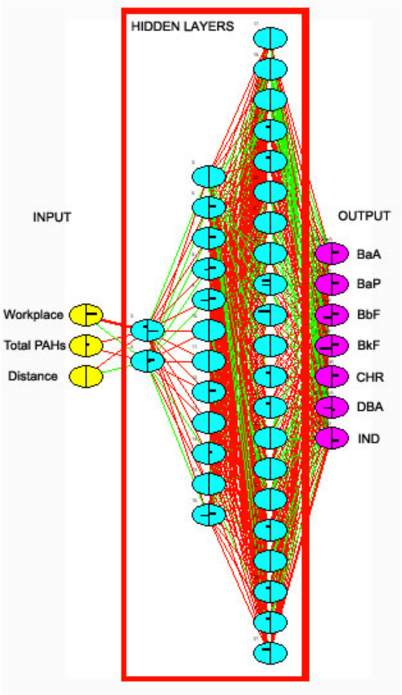


Figure 3: Scheme of the model

Based on this model it is possible to determine the level of concentration of individual PAHs pollutants, considered the most dangerous, in the environment. According to TEF defined by the US EPA, for each species of the PAHs group (Tab. 1) it is possible to calculate the concentration of B(a)p_{eq} in a specific workplace.

In this way, we have a direct perception about for the health risk assessment of workers exposed to PAHs particulate matter in the air.

3.1. Analysis of results

The model is tested using a series of data: for each category two set of data were collected for each of the three input variables that identify the distances between the sources of PAHs emission and the position of the devices for environmental monitoring.

Therefore, for each category and for every PAHs species identified as priority pollutants, the values of the concentration measured by analytical methods (shown as “real” in the following graphs) and those obtained by the prediction model (shown as “forecast”) have been compared.

For energy activities (Fig. 4) it is observed that the gap between the real and forecasted values increases for higher levels of concentration. In fact, a Mean Squared Error (MSE) of about 0.02 has been calculated for values of concentration lower than 1 ng/m3, on the other hand for values of concentration higher 1 ng/m3 a MSE of approximately 2 has been calculated.

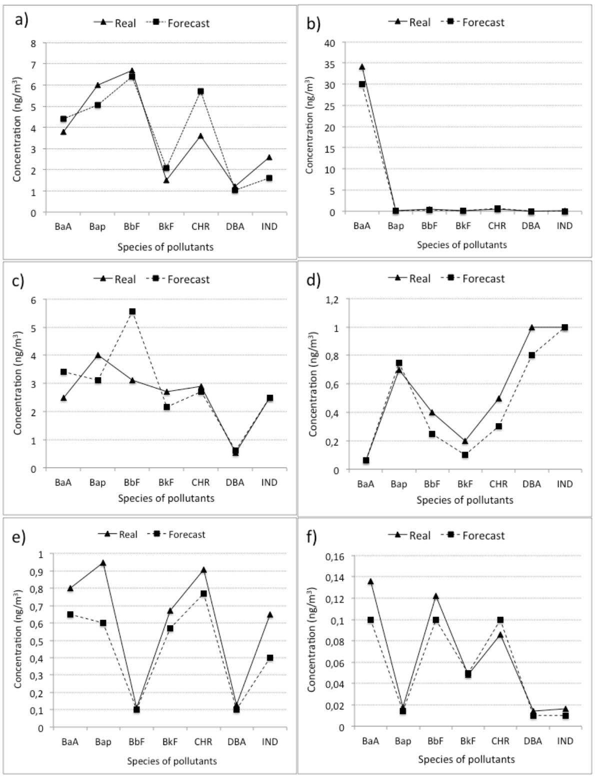


Figure 4: Comparison between real and forecasted values of concentration for single species of PAHs groups, in different scenarios of the "energy activities" category, in: first line (a-b); second line (c-d) and third line (e-f).

In order to evaluate the reliability of the forecast, the Mean Absolute Percentage Error (MAPE) has been calculated. For the first category of workplace the MAPE equals 24.65%.

The MAPE obtained for the category identified as *Production and processing of metals* amounts to 29.33%. In the following charts (Fig. 5) it is possible to assess the gap between value of the concentration measured by analytical methods (real), and the values generated by the prediction model (forecasts). Naturally these cases, like in the previous category of workplaces, are not included in the data set used for the training of the ANN.

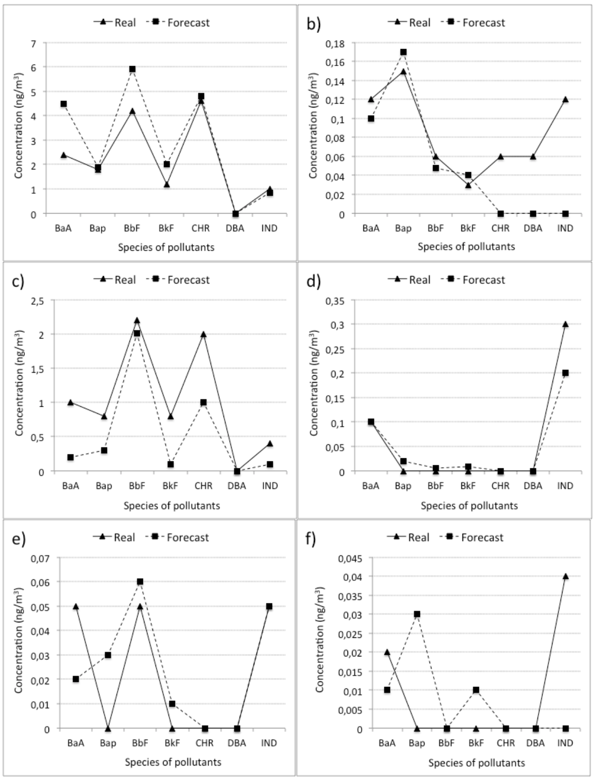


Figure 5: Comparison between real and forecasted values of concentrations for single species of PAHs groups, in different scenarios of the "production and processing of metals" category, in: first line (a-b); second line (c-d) and third line (e-f).

The evaluation of the forecasts for *cooking activities* (Fig. 6) shows that the prediction model reported a MAPE with a similar value as the other categories, in fact it calculated a MAPE equal to 23.03%.
For this category no measure of the concentration is available, with analytical methods, for a distance (between the source of PAHs emission and the device used for environmental monitoring) of over 10 metres.

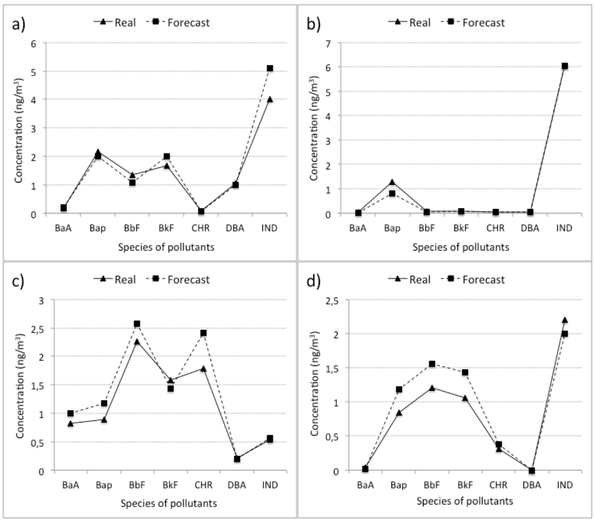


Figure 6: Comparison between real and forecasted values of concentrations for single species of PAHs groups, in different scenarios of the "cooking activities" category, in: first line (a-b) and second line (c-d).

The forecast generated by the model, for the last category, *Insulation activities* (Fig. 7), is more reliable compared to the predictions obtained for the other categories; in fact for this category of workplace a MAPE equal to 10.79% has been calculated. It is believed that this is due to the types of work process that fall in the category of insulation activities. In fact, in this category the activities are very similar to each other, all all characterised by the same temperatures, raw materials, combustion processes and boundary conditions. Indeed, this is not true for the activities belonging to the other categories whose work processes and boundary conditions may vary significantly from one working environment to another.

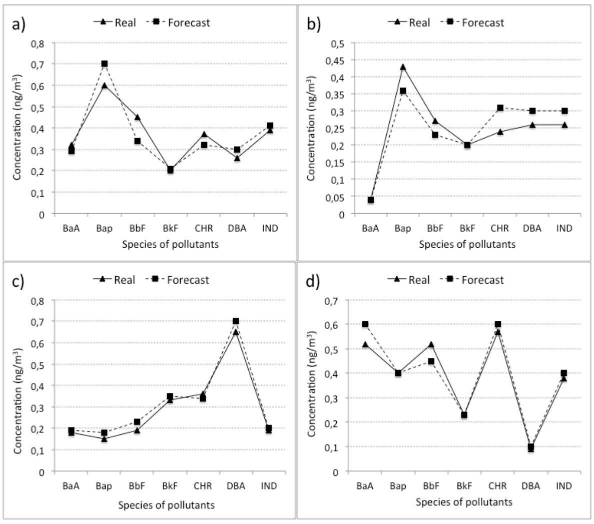


Figure 7: Comparison between real and forecasted values of concentrations for single species of PAHs groups, in different scenarios of the "insulation activities" category, in: first line (a-b) and second line (c-d).

Besides the evaluation of the MAPE for each category of workplace, the MAPE for all species of pollutants, independent of the type of industrial process (Tab. 2), has been calculated.

Table 2: Mean absolute percentage error for compounds considered most dangerous to human health

Compound	Carcinogenic potency	MAPE (%)
Benzo[a]anthracene	0.1	27.84
Chrysene	0.001	23.26
Benzo[b]fluoranthene	0.1	23.87
Benzo[k]fluoranthene	0.01	19.58
Benzo[a]pyrene	1	19.21
Dibenz[a,h]anthracene	1	14.66
Indeno[1,2,3-c,d]pyrene	0.1	25.22

It is possible to observe from the previous table that, according to carcinogenic potency determined by the US EPA (Tab. 1), the most dangerous compound to human health, among the PAHs components, are Benzo[a]pyrene and Dibenzo[a,h]anthracene (TEF equal to 1). The forecast of the model of these compounds is characterised by a MAPE of about 17% .

Calculating the $B(a)p_{eq}$for each category of workplace using the values of concentration, both real and forecasted (Fig. 8), it has been observed that a higher gap has been detected between "$B(a)p_{eq}$real" and "$B(a)p_{eq}$ forecast" for a distance, among the sources of PAHs emissions and devices for environmental monitoring, of over 2 metres (second and third line).

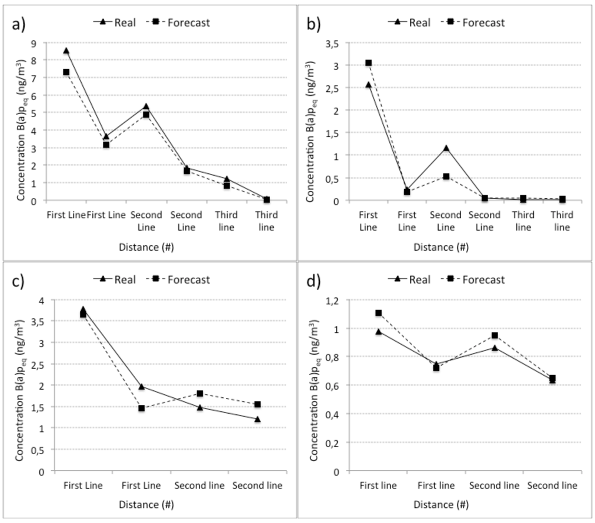


Figure 8: Comparison between real and forecasted values of concentrations of $B(a)p_{eq}$ for different distances, in cases of: Energy activities (a); Production and processing of metals (b), Cooking (c) and Insulation activities (d).

The evaluation of the MAPE, for different distances of each category, is summarised in the following table:

Table 3: Mean absolute percentage error for different distances and categories of workplace

Workplace	First Line (%)	Second Line (%)	Third Line (%)
Energy (a)	13.45	9.14	28.95
Prod. (b)	20.62	40.58	30.83
Cook. (c)	14.44	26.29	-
Insulat. (d)	8.54	6.04	-
Average	**14.26**	**20.51**	**29.89**

A lower MAPE has been calculated for "first line"; namely when the distance between the source of PAHs emission and the device for environmental monitoring is less than 2 metres. This is the most dangerous case for human health; because the level of the concentration of the pollutants to which the workers are exposed is the highest. The model, in this case, is able to ensure the best reliability of the forecast.

Significant MAPE values are due to a limited number of case studies from the scientific literature. However, the 'learning' capability of the ANN will provide more and more reliable results, provided that new training cases are available. To this end, simulated cases by specialised software (e.g. NIST [x]) could also represent a good opportunity to enrich the 'knowledge' of the ANN. Under this perspective the computer-based tool is found to be even more effective in predicting PAHs concentration values, thus avoiding or reducing time-consuming and expensive field investigations for the preliminary assessment of work environments as well as continuous monitoring.

CONCLUSIONS

The model is an efficient and economically sustainable tool: efficient because it can improve the performance with learning in time through the increase in the training data set. It is indispensable for a preliminary risk assessment and the environmental monitoring of a specific workplace.

It is economically sustainable because the model can orientate the decision making process toward the identification of a limited number of PAHs compounds (whose presence is statistically noted). In this way it is possible to reduce the costs of monitoring environmental performance with analytical techniques. In fact, if the prediction of the model does not indicate the presence of a single compound it is not necessary to measure its concentration with analytical techniques. In doing so the costs can be reduced by 20% for each compound not detected. Assuring, however, a high level of protection to the workers' health.

Moreover, performing environmental monitoring in parallel with the forecasting model and analytical methods, it is possible to reduce, in time, the sampling frequency of the analytical method. Reducing, once again, the costs of environmental monitoring and ensuring that the control of health in the workplace is efficiently and economically sustainable.

REFERENCES

Agency for Toxic Substances and Disease Registry, 2009.Case Studies in Environmental Medicine.*Toxicity of Polycyclic Aromatic Hydrocarbons (PAHs).*

Aquilina, N.J., Delgado-Saborit, J.M., Gauci, A.P., 2010. Comparative modeling approaches for Personal Exposure to Particle-Associated PAH. *Environmental Science & Technology,* 9370-9376.

Boström, C.E., Gerde P., Hanberg A., 2002. Cancer Risk Assessment, Indicators, and Guidelines for Polycyclic Aromatic Hydrocarbons in the Ambient Air.*Environmental Health Perspectives,* 451-488.

Burstyn, I., Kromhout, H., Kauppinen T., 2000. Statistical Modelling of the Determinants of Historical Exposure to Bitumen and Polycyclic Aromatic Hydrocarbons Among Paving Workers. *The Annals of Occupational Hygiene,* 43-56.

Finardi, S., De Maria, R., D'Allura, A., Cascone, C., Calori, G., Lollobrigida, F., 2008. A deterministic air quality forecasting system for Torino urban area, Italy. *Environmental Modelling & Software*, 344–355.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer per- ceptron) – a review of applications in the atmospheric sciences. *Atmospheric Environment*, 2627–2636.

Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London. *Atmospheric Environment*, 709–719

Han, I., Ramos-Bonilla, J.P., Rule A.M., 2011. Comparison of spatial and temporal variations in p-PAH, BC, and p-PAH/BC ratio in six US counties.*Atmospheric Environment,* 7644-7652.

Hrust L., Klaic, Z.B., Krizan, J., Antonic, O., Hercog, P., 2009. Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmospheric Environment,* 5588-5596.

Karppinen, A., Kukkonen, J., Elola˙hde, T., Konttinen, M., Koskentalo, T., Rantakrans, E., 2000a. A modelling system for predicting urban air pollution: model description and applications in the Helsinki metropolitan area. *Atmospheric Environment*, 3723–3733.

Karppinen, A., Kukkonen, J., Elola˙hde, T., Konttinen, M., Koskentalo, T., 2000b. A modelling system for predicting urban air pollution: comparison of model predictions with the data of an urban measurement network in Helsinki. *Atmospheric Environment*, 3735–3743.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment*, 4549–4550.

Li, C.T., Lin Y.C., Lee, W.J., Tsai, P.J., 2003.Emission of Polycyclic Aromatic Hydrocarbons and Their Carcinogenic Potencies from Cooking Sources to the Urban Atmosphere.*Environmental Health Perspectives,* 483-487.

Lin, Y.C., Lee, W.J., Chen, S.J., Chang-Chien, G.P., Tsai, P.J., 2008.Characterization of PAHs exposure in workplace atmospheres of a sinter plant and health-risk assessment for sintering workers.*Journal of Hazardous Materials,* 636-643.

Perez, P., Trier, A., Reyes, J., 2000. Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile. *Atmospheric Environment*, 1189–1196.

Perez, P., Reyes, J., 2006. An integrated neural network model for PM_{10} forecasting. *Atmospheric Environment*, 2845–2851.

Posniak, M., 2005.Polycyclic Aromatic Hydrocarbons in the Occupational Environment during Exposure to Bitumen Fumes.*Polish Journal of Environmental Studies,* 809-815.

Tsai, P.J., Shieh, H.Y., Hsieh, L.T.,Lee, W.J., 2001. The fate of PAHs in the carbon black manufacturing process.*Atmospheric Environment*, 3495-3501.

Tsai, P.J., Shieh, H.Y., Lee, W.J., Lai S.O., 2000. Health-risk assessment for workers exposed to polycyclic aromatic hydrocarbons in a carbon black manufacturing industry. *The Science of the Total Environment*, 137-150.

AGGREGATE MODEL BASED PERFORMANCE ANALYSIS OF AN EMERGENCY DEPARTMENT

I.J.B.F. Adan$^{(a)}$, E. Lefeber$^{(a)}$, J.J.D. Timmermans$^{(a)}$, A. v.d. Waarsenburg$^{(b)}$, M. Wolleswinkel-Schriek$^{(b)}$

$^{(a)}$Manufacturing Networks Group, Department of Mechanical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands.
$^{(b)}$Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ Eindhoven, The Netherlands.

$^{(a)}$[I.J.B.F.Adan,A.A.J.Lefeber]@tue.nl

ABSTRACT

In this paper we present an aggregate simulation model for a real-life emergency department, which is based on the concept of effective process times and which uses a token system to model patients claiming multiple resources simultaneously. Although it has been developed for a specific hospital, the model is flexible, and capable to describe different settings. The modeling steps, model specification and model validation are explained in detail. By using a process-based simulation language, the resulting model is transparent, intuitive and easy to use in quantitatively evaluating proposed changes in the operational processes of the emergency department.

Keywords: Aggregate modeling, effective process time, health care, simulation.

1 INTRODUCTION

Due to rising costs, the health care sector is forced to work more efficiently and to better utilize their resources. Therefore, LEAN principles have been introduced in health care. Also at the Emergency Department (ED) of the Catharina Hospital in Eindhoven (CZE) the LEAN concept has been introduced to improve operational processes (Wolleswinkel 2012). The aim is to streamline operational processes, i.e., to eliminate unnecessary operations to achieve better performance using existing resources.

To support decision making in process improvement programs, simulation has proved to be an effective tool (Brailsford 2007, Duguay and Chetouane 2007, Sinreich and Marmor 2005). A literature review on the use of simulation and modeling in the health care domain can be found in (Jun et al. 1999, Brailsford et al. 2009, Brailsford and Vissers 2011), showing evidence that simulation and modeling are growing in popularity. This approach is also followed for the CZE: we develop a simulation model for the ED, based on actual data from the electronic hospital information system (EZIS), and exploit the concept of *effective process time* (EPT), cf. (Hopp and Spearman 2008). The basic idea is that the various details of patient treatment times are not modeled in detail, but their contribution is *aggregated* into an EPT distribution, the parameters of which are directly estimated from the available data. This concept has been developed in semi-conductor manufacturing (Etman et al. 2011). Its applicability in health care modeling, in particular for an MRI department, has recently been explored in (Jansen et al. 2012), and it is further investigated in the current paper. Typical features of the ED are that (i) patients *simultaneously* require multiple resources (e.g., treatment room, nurse, physician) and (ii) nurses and physicians can spread their attention over *multiple* patients. We propose a novel *token system* to model the above mentioned features of simultaneous resource possession and multi servicing. This token system, in combination with EPTs, describes the ED at an aggregate level, suitable and sufficiently flexible to support the improvement program of CZE, and it distinguishes the current model from other, typically more detailed models proposed in the literature, cf. (Duguay and Chetouane 2007) and the references therein.

The resulting model, specified in the process-based simulation language Chi 3.0 (Hofkamp and Rooda 2012), is transparent, flexible and intuitive, and hence, in the spirit of the principles set out in (Sinreich and Marmor 2005). It can be used to investigate the capacity level needed to deliver the health care services within the target times set by the hospital management. The capacity consists of (ED-) physicians, (ED-)medical interns, (ED-)nurses and treatment rooms. Using this model, it is possible to address questions such as, for example:

- What capacity is at least required on a typical Monday to meet the target maximal waiting times?
- How much does the waiting time decrease if the number of nurses increases?

- How does the waiting time change if patients arriving by ambulance are treated with the same priority as other patients?

In the next section we will first explain the modeling steps and challenges. Then, in Section 3, the model is validated, and its decision supporting power is demonstrated in Sections 4-5.

2 MODELING

As mentioned in the introduction, we developed a simulation model of the ED of the CZE. Before we introduce this model, we first describe in more detail what happens at the ED of the CZE (and what is also typical for EDs at other Dutch hospitals). The patient flow is described using the map in Figure 1a, and the patient flowchart, shown in Figure 1b. Prior to the actual arrival, for most referred patients, it is already known that they will arrive in the near future. The paramedic or general partitioner contacts the senior nurse in order to keep him/her up-to-date. The senior nurse then adds a new entry to EZIS. New patients arrive by own transportation or by ambulance. Both patient flows register at the reception. At that time, patients are also logged in on EZIS. This means that the patient is physically present at the ED.

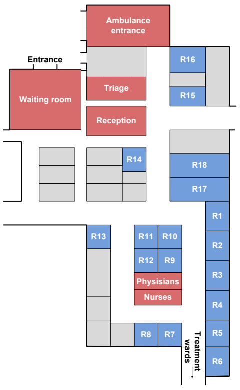


Figure 1a: Map of the Emergency Department of the CZE.

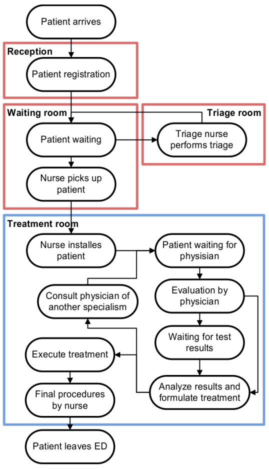


Figure 1b: Schematic view of the patient flow in an ED.

After registration, the patient will wait in the waiting room. Generally, patients go in order of arrival to the triage room to undergo triage according to the *Dutch Triage Standard*. This system uses four emergency levels: acute (red), urgent (yellow), standard (green) and non urgent (blue). When finished, the patient returns to the waiting room. If a nurse is free and a treatment room is available, the nurse picks up the longest waiting patient with the highest priority to accompany him/her to the treatment room.

Paramedics, transporting patients by ambulance, have to wait at the reception before the patient can be dropped off at a treatment room. The target maximal waiting time is 15 minutes for these patients. This patient flow is not drawn in Figure 1b. In the current situation, patients arriving by ambulance are served with priority over patients in the waiting room.

When the patient has arrived in the treatment room, the treatment process starts and the nurse ensures that the patient is installed properly in the emergency room. In some cases, the nurse already starts up a few small examinations, such as taking a blood sample. Next, a physician visits the patient for a first evaluation of the complaints, in most cases done by the medical intern. After consultation with a medical specialist, it is decided which extra examinations are needed, e.g. an X-ray. When the tests are finished and the results are reviewed, the physician determines what treatment is needed to cure the patient. If the physician is uncertain about the complaints and how to treat, then a medical specialist of another speciality is paged to examine and treat the patient. During the treatment, the responsible nurse keeps monitoring and nursing the patient when needed.

When the treatment is finished, several options are possible. A patient can go home and the nurse can schedule a follow-up appointment at the general partitioner or at the policlinic. Another option is that the patient has to stay for hospitalization, or is transported to another hospital. In those cases, the patient can only leave if a nurse from the ward or a paramedic has arrived to pick up the patient. During the delay that occurs, the treatment room stays occupied and is therefore not available for a new patient.

The above way of working at the ED of the CZE forms the basis of our model. However, we are limited by the available data in EZIS. In the remainder of this section we outline how the available data is incorporated in our model. In particular, limited or no data is available on the activities taking place while the patient is in the treatment room (e.g., number and duration of visits of the responsible nurse and physician), though the entrance and exit time of the patient in the treatment room are accurately recorded. Therefore, we will lump the treatment room process in the blue box of Figure 1b) into a single EPT distribution. The parameters of this distribution, however, depend on several patient characteristics. This calls for further lumping, which will be done by application of data mining techniques, as described in Section 2.2.

2.1 Patients arrivals and diversity

In Figure 2 we show the average number of patient arrivals from Monday to Sunday, as well as the distribution of patients over the most important specialities. In 2011, over 20 different medical specialities were consulted. In our simulation model, only the 11 most visited specialities are included. The ones that are left out have, on average, less than one patient visit per day. The included specialities are surgery, internal medicine, cardiology, orthopedics, pediatrics, lung diseases, neurology, urology, gynecology, plastic surgery and geriatrics. ¿From Figure 2 it can be observed that the number of patient arrivals on Monday and Friday are significantly higher than on the other days. Also, a significant difference in these numbers for both surgery and orthopedics can be seen. This is due to agreements between the surgeons and orthopedists: on Monday, Wednesday, and Thursday more patients are seen by the orthopedist, whereas on the other days, these patients are seen by the surgeon.

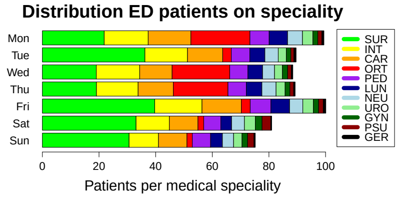


Figure 2: Patients per specialism on the different weekdays.

Though not visible in Figure 2, the arrival rate of patients strongly varies over the day, say from 1 patient per hour during the night, up to 10 patients per hour during peak office hours. We therefore model the patient arrivals as an *inhomogeneous Poisson process* (Alexopoulos 2008), with piecewise constant arrival rates during one hour, where we distinguish between ambulance arrivals and arrivals by own transportation. The hourly arrival rates come from EZIS data. To each arriving patient we assign its required speciality, according to probabilities which can be read from Figure 2.

2.2 Treatment times

Statistical analysis shows that the total treatment times of patients (while being in the treatment room) depend on several factors: medical speciality, triage color, age, type of attending physician (ED-physician or other medical specialist), the number of patients currently treated by the physician, and whether the patient requires a second consult or not. This combination of factors leads to almost 7000 treatment groups. Since the ED has been visited by 34.000 patients in 2011, that gives, on average, 5 treatment time realizations per group. Hence, it is unreliable to sample treatment times from empirical data or fitted (EPT) distributions for these groups. To cope with this problem, the data mining technique of *recursive partitioning* has been used. The package 'rpart' (Therneau et al. 2012) of the statistical software program R (Gentleman and Ihaka 2012) has been used to generate a decision tree, that specifies the partitioning. The first part of this tree is shown in Figure 3. In the root, all groups are lumped together. Then, in each decision step, the current group is split into two groups by the factor with the highest influence on the mean treatment time. The 'n' in a leaf denotes the number of patient treatment times that fit into that group. The 't' stands for the mean of the treatment times into that group. The entire tree can be found in (Timmermans 2012b,Appendix A); it consists of only 37 leaves, each containing patient treatment times from many groups. The treatment time of each group is now assumed to be Gamma distributed, fitted to the mean and variance of *all* treatment times in the corresponding leaf. As a result, many of the almost 7000 groups use the same, but now *reliably fitted*, treatment time distribution.

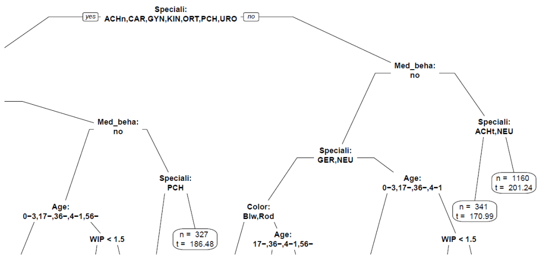


Figure 3: Part of the treatment time decision tree.

2.3 Resource capacity

Nurses, but also physicians, are capable of handling multiple patients simultaneously. To capture this 'multi-processing' feature, we adopt a *token system* to model the capacity of the physicians, nurses and triage nurses. To start the treatment, a patient claims a combination of tokens representing the resources that are simultaneously needed. Four nurse tokens are used to represent one nurse, because (s)he can treat a maximum of four patients at the same time. So each patient needs one nurse token. Moreover, the triage nurse, senior nurse and physician are modeled as respectively one, two and two or three tokens. So, the simulation model uses, for example, 20 nurse tokens to represent 5 nurses. Note that the token system describes the capacity claim by patients at an *aggregate level*: nurse and physician capacity are claimed during the treatment, but the number and duration of visits during the treatment are not modeled. In other words, nurses and physicians can spread their capacity (tokens) over multiple patients present in the treatment rooms, but we do not exactly model how and when.

Each patient demands one triage nurse token during the triage process and one nurse and one physician token during treatment. An exception is made for acute (red) patients. They demand more intensive care for the first 15 to 30 minutes of their treatment. So, all red patients claim more tokens for the first part of the treatment.

Not only the arrival rate is different for each hour of each day of the week, the model also assumes a different working roster for each weekday. The roster specifies the available capacity at each point in time during the day. For example, the staffing levels are lower at night. Also the transfer from night to morning shift is taken into account by decreasing the available capacity between 7:30h and 9:00h.

2.4 Discrete-event simulation model

For specifying our discrete-event simulation model we used the programming language Chi 3.0, see (Hofkamp and Rooda 2012), which is an instance of a parallel processes formalism. A system is abstracted into a model, with cooperating processes, connected to each other via channels. The channels are used for exchanging material and information. The model of the ED consists of a number of concurrent processes connected by channels, denoting the flow of patients or information.

The process to simulate a *single day* of the ED at CZE, is depicted in Figure 4. It consists of the generators G, the arrival process delay D, the waiting room W, the triage process T, the treatment rooms R and the exit E.

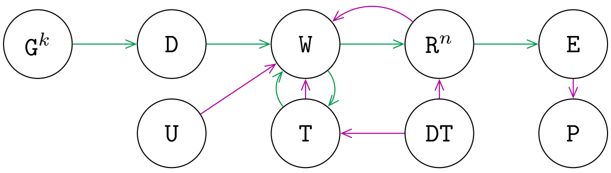


Figure 4: The process to simulate a single day. Patients are transferred using the green channels, other information uses the purple channels.

The generators G^k represent the arrival processes and create not only the individual patients with a certain arrival rate, but also their attributes (such as, e.g., speciality, triage color, age). The first generator creates patients arriving by ambulance and the second one generates patients arriving by own transportation. A new patient is sent via D to the waiting room W. The delay process D represents the time needed to register a new arrival.

The waiting room process keeps track of all waiting patients and of the availability of the recourses. It uses this information to determine when and which patient will go to the triage process or to the treatment room first. The triage process T receives patients from the waiting room and sends the patients back after a certain amount of time, required for performing the triage. A triage can start if both the triage room and triage nurse are available. When the triage is completed, the waiting room process is informed that the triage nurse and triage room are available again. Process DT is used to sample the triage time. If the treatment can start, the patient is sent to one of the treatment rooms R^n. The update process U is used to report staffing changes to process W. The waiting room W also receives information from T and R^n about their availability.

The treatment rooms are modeled individually and each room can be occupied by one patient. Treatment room R receives a patient and requires nursing and physician capacity. The treatment itself is modeled as a time delay for the patient, an occupation of nursing and physician tokens and an occupation of the treatment room. The total number of available tokens is decreased by the number of tokens required by the patient for the entire treatment time. As mentioned in Section 2.3, red patients need more capacity for the first part of their treatment. If this first part is finished, the required capacity is reduced to normal level, i.e., to one nurse token and one physician token. For patients that need a second consult, the speciality of required physician capacity is changed when the treatment is halfway. Process DT is also used to sample the treatment time. After (and possibly during) this delay, capacity is released again and process R informs the waiting room.

When the treatment is finished, the patient goes to the exit E. This represents the departure of a patient. The patient either goes home or is hospitalized. The exit process sends information to the print process P which takes care of the simulation output. Next, the print process also signals when the system is empty and a new simulation day can start.

Note that the available resource capacities (such as, number of triage and treatment rooms, (triage) nurses, physicians, and so on) are model parameters, which can be easily adapted to the situation at hand. For more details on the model and code, see (Timmermans 2012b,Chapter 3 and Appendix B).

3 MODEL VALIDATION

The model has been validated by (i) team discussion, and (ii) comparing the model output with the historical data.

The simulation structure and the results have been discussed in several team discussions. Hospital managers, the head of the ED, ED-physicians and senior nurses have been involved in these discussions. During these meetings, a software tool developed in R (Gentleman and Ihaka 2012) was used for the analysis of historical and simulation data. This tool has been developed to increase the ease of visualizing and analyzing the data. For more information, see the software package manual (Timmermans 2012a).

As a result of these discussions, most assumptions were confirmed, but the meetings also led to new insights, such as, e.g., the need of a separate stream for patients arriving by ambulance. Also, these patients get a higher priority while waiting. Another remark during the discussions was that not all treatment rooms are used equally. Some rooms are more suitable for gynaecology or otolaryngology patients and other rooms are more often used for small traumas. The latter, however, has not been taken into account in the simulation model.

As part of the validation, the historical data and simulation output are compared. As mentioned in Section 2.1, Monday and Friday are the busiest days. Therefore, the results of the simulated Mondays are used in this section to show the match between historical data and simulation results. All colored bands shown in the figures are the 95% confidence intervals.

In Figure 5a and Figure 5b, the historical and simulated average occupations are given for patients that are present in the waiting room, present in the treatment rooms and for the total number of patients at the ED. The results for the first hours of the day are different because the simulated ED begins each day empty. Starting from 9:00h, the waiting room fills to approximately five patients in the afternoon, on average. Both figures show that the waiting room is empty at the end of the day and that there are on average around 7 patients present in the treatment rooms during the day. Next, the waiting times are discussed. As can be seen in Figure 6a and Figure 6b, the distributions of the waiting times for yellow patients give a relatively close match. Similar results can be obtained by comparing other patient categories. In Figure 7, the average *cycle time factor* (Hopp and Spearman 2008) is plotted during the day. This factor is the total time a patient is present at the ED divided by the treatment time, and thus provides an indicator of logistic efficiency. If, for patients starting their treatment at time t, this factor is close to one, their waiting time is short compared to their treatment time. Figure 7 shows a good match between historical and simulated data. During peak hours the cycle time factor raises to 1,5-2.0, which is, in terms of manufacturing, a good performance.

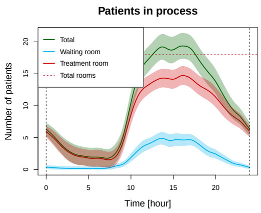


Figure 5a: Historical average occupation of patients on Monday.

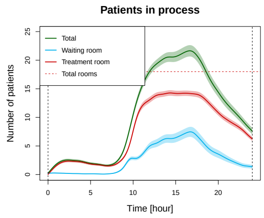


Figure 5b: Simulated average occupation of patients on Monday.

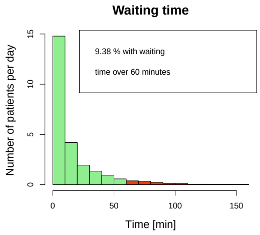


Figure 6a: Historical waiting time for yellow patients arriving on Monday.

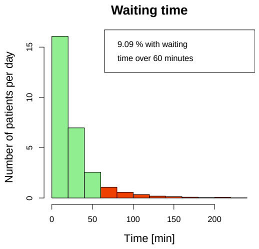


Figure 6b: Simulated waiting time for yellow patients arriving on Monday.

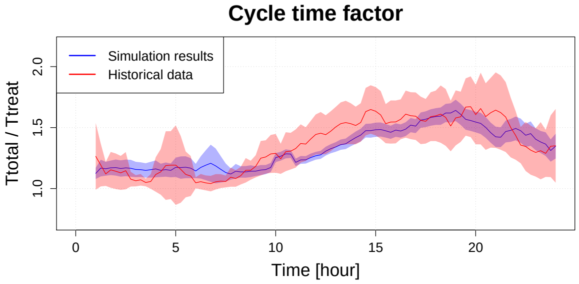


Figure 7: Cycle time factor for historical and simulated patients on Monday. The colored band represents the 95% confidence interval. The patient's cycle time factors are located according to the time their treatment starts.

4 ANALYSIS OF 2011

In addition to the simulation model, a tool for analysis of simulation output has been developed in R. The output is processed and *adaptive* plots are created using the R package *playwith*. That tool has been used to conduct the analysis in this section. By means of this tool, improvement opportunities can be evaluated, such as, for example, opportunities to reduce waiting times, to reduce the number of patients waiting or to improve the utilization of treatment rooms or nursing capacity. Here we restrict ourselves to investigating opportunities to reduce the percentage of yellow and green patients, present on Monday between 10:00h and 20:00h, that exceed the target maximal waiting time of respectively 60 and 120 minutes. This time window is chosen, because it is one of the busiest moments at the ED.

Before investigating possible improvements, first the original situation is simulated. The simulation results for Monday are shown in Figure 8. One can see that over 18% of the yellow patients exceed the target maximal waiting time. The vertical black dotted lines in the plot on the right mark the time interval of 10:00h to 20:00h. Possible opportunities for improvement are, e.g., more treatment rooms, no priority for ambulance patients, more nursing capacity, more physician capacity and treatment time reduction, or a combination of these opportunities. Be-

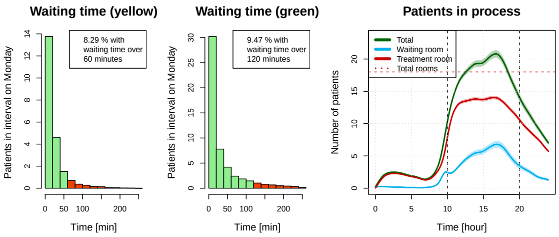


Figure 8: Unadapted simulation results on Monday.

low we only investigate the effect of treatment time reduction.

4.1 Treatment time reduction

As mentioned in the introduction, the LEAN concept has been introduced in the CZE to improve operational processes (Wolleswinkel 2012). The aim is to eliminate unnecessary operations to achieve better performance using existing resources. This approach can, for example, result in a reduction of the treatment time by 10 minutes per patient.

One potential way to achieve this reduction is to shorten the time for hospitalization. This time starts from the moment that the treatment actually finishes until the patient is picked up by the nurse of the ward. About 30% of the patients, mostly elderly, are hospitalized.

A simulation is performed in which the treatment time per patient is reduced with 10 minutes. The results are shown in Figure 9. The number of waiting patients as well as the number of occupied treatment rooms is decreased. The percentage of patients that exceed the target maximal waiting time is reduced to 6.55% and 5.49% for respectively yellow and green patients, i.e., a reduction of 21.0% and 42.0%.

If the treatment time is increased by 10 minutes per patient, an opposite result can be observed. This increase results in 9.12% and 14.76% of the yellow and green patients exceeding the target.

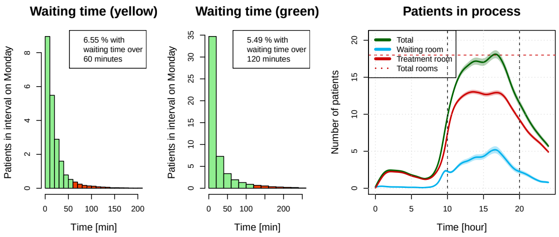


Figure 9: Simulation output for 10 minutes treatment time reduction per patient.

5 SCENARIO ANALYSIS

In the previous section, the simulation model has been used to investigate improvements based on the 2011 situation. Alternatively, by modifying the input files for the simulation, different trial scenarios can be studied, such as:

- In general, older patients have longer treatment times. What effect has an increase of ED visits by elderly patients, due to the aging population?
- What extra capacity is needed if a neighboring ED closes and the CZE ED has to partially take care of their patients?
- What if the average urgency of patients increases? For example, due to less self-referrals.
- What if more accurate triage results in less second consults and thus in a decrease of treatment times?
- What capacity of ED-physicians is needed if more patients are consulted by the ED-physician instead of the specialist of the attending medical speciality.

Here we consider the second scenario only: An increase of the arrival rate.

5.1 Scenario: Increasing arrival rate

We consider the scenario: What happens if a neighboring ED has to (temporarily) close? This closure can be caused by a MRSA-outbreak or by financial cutbacks. In this case we assume that the introduced closure results in an increase of 15% in patient arrivals. The growth of patient arrivals by 15% results in a large increase of waiting times, as shown in Figure 10. On average, there will be more than 12 patients waiting during peak hours. 22.42% of the green patients, present between 10:00h and 20:00h, exceed the target waiting time. For yellow patients, the percentage is 10.97%.

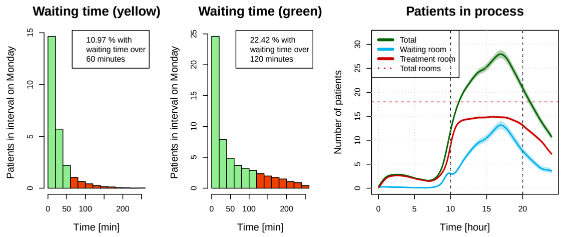


Figure 10: Simulation output for the CZE ED on Monday with a 15% growth of patient arrivals.

6 CONCLUSIONS

In this paper we developed a simulation model, and we explained modeling challenges and solutions, implementation issues and usage. The resulting process description in the simulation language Chi is transparent, flexible and intuitive, and therefore more easily accepted by its potential users. Also, the use of the visualization and data analysis capabilities of software package R appeared to be crucial to the acceptance of the model. We can conclude that the developed simulation model, equipped with the user interface developed in R, bears the promise to play an important role in the process improvement program at CZE, and possibly also at other hospitals.

Currently there is a discussion in the Netherlands on reducing the number of EDs. Simulation models, like the one in this paper, can support this discussion by *quantitatively* evaluating the (logistic) effects of proposed closing or merging of EDs, or by comparing the efficiency of EDs.

REFERENCES

Alexopoulos, C., 2008. Modeling patient arrivals in community clinics. *Omega: The International Journal of Management Science*, 36, 33–43.

Brailsford, S. and Vissers, J., 2011. OR in healthcare: A european perspective. *European Journal of Operational Research*, 212, 223–234.

Brailsford, S.C., 2007. Tutorial: Advances and challenges in healthcare simulation modeling. Proceedings of the 2007 Winter Simulation Conference, pp. 1436–1448. Washington D.C.

Brailsford, S.C., Harper, P.R., Patel, B. and Pitt, M., 2009. An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3, 130–140.

Duguay, C. and Chetouane, F., 2007. Modeling and improving emergency department systems using discrete event simulation. *Simulation*, 83, 311–320.

Etman, L.P.F., Veeger, C.P.L., Lefeber, E., Adan, I.J.B.F. and Rooda, J.E., 2011. Aggregate modeling of semiconductor equipment using effective process times. Proceedings of the 2011 Winter Simulation Conference, pp. 1795–1807. Phoenix (Arizona, USA).

Gentleman, R. and Ihaka, R., 2012. R project. `http://www.r-project.org`. [accessed 25 October 2012].

Hofkamp, A.T. and Rooda, J.E., 2012. Chi 3.0.0 documentation. `http://chi.se.wtb.tue.nl/`. [accessed 28 January 2013].

Hopp, W.J. and Spearman, M.L., 2008. *Factory Physics: Foundations of Manufacturing Management.* 3rd ed. New York: IRWIN/McGraw-Hill.

Jansen, F.J.A., Etman, L.F.P., Rooda, J.E. and Adan, I.J.B.F., 2012. Aggregate simulation modeling of an MRI department using effective process times. Proceedings of the 2012 Winter Simulation Conference, pp. 1795–1807. Berlin (Germany).

Jun, J.B., Jacobson, S.H. and Swisher, J.R., 1999. Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society*, 50, 109–123.

Sinreich, D. and Marmor, Y., 2005. Emergency department operations: The basis for developing a simulation tool. *IIE Transactions*, 37, 233–245.

Therneau, T., Atkinson, B. and Ripley, B., 2012. Package ‘rpart’ — recursive partitioning. `http://cran.r-project.org/web/packages/rpart/rpart.pdf`. [accessed 2 November 2012].

Timmermans, J.J.D., 2012a. *Documentation of the simulation model tools of the emergency department at Catharina Hospital Eindhoven.* Report MN-420703, Eindhoven University of Technology, Manufacturing Networks Group, Department of Mechanical Engineering, Eindhoven, The Netherlands.

Timmermans, J.J.D., 2012b. *Efficiency of the Emergency Department of the Catharina Hospital Eindhoven.* Thesis (MSc). Eindhoven University of Technology, Manufacturing Networks Group, Department of Mechanical Engineering, Eindhoven, The Netherlands. `http://www.tue.nl/uploads/media/2012_MN-420704_JJD_Timmermans.pdf`.

Wolleswinkel, M., 2012. In de start blokken — Tweegesprek over het meerjarenbeleidsplan. *Ons Catharien*, 2(2), 6–7. In Dutch.

WIRELESS SENSOR DEPLOYMENT DECISION SUPPORT USING GENETICS ALGORITHM AND DEVS FORMALISM

Bastien POGGI [(a)], **Jean-François SANTUCCI** [(b)], **Thierry ANTOINE-SANTONI** [(c)]

[(a ,b, c)] University of Corsica Pasquale Paoli
UMR CNRS 6134 Sciences for the environment
Technology information and communication project

[(a)] bpoggi@univ-corse.fr, [(b)]santucci@univ-corse.fr, [(c)]antoine-santoni@univ-corse.fr

ABSTRACT

The deployement of a Wireless Sensors Network appears as a strategical aspect in the environmental monitoring. A random deployement is efficient if the entities are mobile and able to take account of the positon of each others. If the WSN is static the deployment must be adapted at the monitoring area and must respect the intrinsic parameters of the WSN. To predict the best positionment of the WSN entities the simulation appears as an efficient way. This paper describes a framework based on the DEVS (Discrete Event System) formalism and GA (Genetic algorithms). This decision support tool for the deployment focuses on the improvement of two main characteristics of the network: the sensor coverage, and the connectivity between nodes.

Keywords: WSN, deployment, DEVS, Optimization, Simulation, decision support system

1. INTRODUCTION

Advances in Micro-Electro-Mechanical-Systems (MEMS) based sensor devices and miniaturization of processor and radio as sensor package have led to the emergence of Wireless Sensor Networks WSN) since the last decade. Nowadays the WSN appear like an efficent way for the environemental monitoring and the increasing of the different hardware platforms have contributed at several testebed in many research areas (Beutel and Römer 2009). It seems to be interesting to view the succes experiences in real world with WSN: animal habitat monitoring (Szewczyk and Mainwaring 2004), agriculture (Langendoen and Baggio 2006), volcano activities (Werner-Allen and Lorincz 2006), oceanography (Tateson and Roadknight 2005), health (Paek and Chintalapudi 2005), and wildifre (Antoine-Santoni, Santucci 2009). However, we can observe that these different papers don't integrate a realistic approach of the network deployement.

Indeed the deployment of Wireless Sensor Network can become one of the most important strategic aspects in the Quality of Service (QoS). The WSN can be identified by several characteristics: mobility or not of the nodes, routing or mac protocol, life of the node , connectivity, coverage, time of arrival of message, real time application. These characteristics are the intrinsic parameters of a WSN and it is important to determine their possible impacts on a deployment and inversely. In a static WSN, different of a mobile network, the entities of the network have the possibility to know the position of the neighboring nodes but they don’t have the possibility to move to enhance the link connectivity with the neighborhood.

It seems to be evident to say that the deployment of WSN is dependant of the area of deployement (AoD): an efficient WSN must to be different if the goal is to monitor an animal habitat monitoring or a wildfire monitoring The simulation appears like the best way to test and develop several strategies of WSN deployment to enhance the connectivity, the coverage and reduce the cost of a deployment with a diminution of the density of the netieis in the area. We can find different works on this research area (Younis and Akkaya 2008; Dhillon and Chakrabarty 2003; Efrat and Har-Peled 2005; Cheng Du 2005; Grandham and Dawande 2003; Antoine-Santoni and Santucci 2008) with a pertinent analysis by Otero and Kostanic (2009)

However all these works focus on a particular problem or specific parameters and they don’t propose a generic approach for the analysis of WSN deployment based on independent reflexion of domain and sometimes this limits the capacity to resolve some complex problems. Moreover, these approaches don’t define the way of development of decision support tool for the WSN deployement.

In this paper we intoruduce a generic open framework to simulate the WSN sensor deployement and optimize it by using Articificial Intelligence: specifically the evolutionary computation algorithms. The paper is organized as follows: in section 2 we present the bases of the framework, e.g. the DEVS formalism and the GAs approach. Section 3 deals with the description of our approach with the different models.

The results obtained when using the integration of genetic algorithms as optimization technique into a DEVSimPy modeling scheme is carefully described in section 4. Finally conclusions and perspectives are given in the last section.

2. DEVS FORMALISM AND GA

When people want to improve the performance of a studied system the modeling and simulation phases can be integrated with an optimization technique. This is what we call optimization via simulation (OvS). Optimization via Simulation is a structured approach to determine optimal settings for input parameters (i.e., the best system design), where optimality is measured by a function of output variables associated with a simulation model (Swisher and Hyden 2004; Garcia and Patek 2007). OvS problems can often be formulated as finding the minimum (or maximum) of a function (called in the OvS terminology objective function) having the vector x as parameters. Such a vector represents the set of decision variables according to the OvS terminology. To evaluate the objective function, we can only run simulation experiments at a particular value of x. Our framework is defined using an automatic integration of genetic algorithms optimization techniques into a discrete event modeling using the DEVS (Discrete EVent system Specification)

2.1. DEVS formlism

The DEVS formalism DEVS (Discrete Event System Specification) introduced by Professor Zeigler (1976) based on the definition of two types of components: atomic components which are the basic elements of the model behavioral and coupling components which correspond to a grouping of elements atomic behavior (description hierarchy).

The role of atomic components is to provide a local description of the dynamic behavior of the system studied. This type of component (atomic model) has state variables and ports of entry / exit, through which all interactions with the external environment occur. There are two types of external interaction: interaction of the component input (external events) that can be managed by external transition functions (named delta_ext) and output interactions managed by output functions (called lambda). The evolution of state variables of the component is dictated by the external transition function when interacting with the outside or by an internal transition function (named delta_int) when the component needs to evolve independently of its environment outside. Each state variable has a lifetime managed in a function of progress of time (called ta). The behavior of an atomic component is then obtained by the following algorithm:

1. The system is in a state s (initial state)
2. If no external event occurs, the component does not change state during a period determined by e = ta (s)
3. After this period, the component generates an output through lambda (s) and executes its internal transition function delta_int
4. If an external event occurs, the component changes state by executing delta_ext based on input values, the current state and the lifetime of this state.

An atomic component can not by itself account for the entire behavior of a complex system. Also, in order to have a phased approach to behavioral description, it is necessary to define another behavioral element: the coupling component. A coupling component (coupled model) to describe how many are interconnected elements (called sub-components) to form a new component. The specificity of coupling components is that they can be considered as basic elements in a coupling component of the highest level (Zeigler 1987). From the DEVS formalism briefly described above, an object-oriented simulation methodology has been implemented. The main idea is to automatically generate, from a given model, the corresponding simulation algorithms, allowing exploitation of the model. For this, it is necessary to associate each model a specific control structure. This control structure, called the simulator, a role for the management dynamics of the model on which it depends. The model simulation aims to generate output events, from input events from a given system. The simulation phase is then decomposed into two steps: the creation of the simulator and then use it in order to generate output events (outcomes). A simulator is represented by a graph type structure, called shaft simulation. This structure is described by a set of classes corresponding to different types of nodes constituting the tree simulation. For each model studied, the tree associated simulation is obtained by instantiation. This tree simulation has over operation of the model which it is attached. It is said that a simulator pilot the model associated with it. Each model is managed by a specific simulator, since any component of a model has an image and a single tree in the corresponding simulation. Communication between a model and its simulator is established through an exchange of messages. These messages are the events processed during the simulation process.

2.2. DEVSimPy

DEVSimPy software is a GUI for Python M & S (Capocchi and Santucci 2011) Automatic discrete event models described in the DEVS formalism. It was originally built to allow manipulation of graphical models PyDEVS to facilitate coupling. Indeed, the models are PyDEVS Python files using an API developed by researchers Bolduc and Vangheluwe (2001) from McGill University (Montreal) that creates DEVS models. These files contain a description of models and specification of couplings (links) between them. The only drawback, which does not depend PyDEVS, is the large number of errors caused by poor coupling due to a lack of attention during the connection ports between models. More models are, the more mistakes increase and pollute the thinking of developer around the model's behavior. DEVSimPy therefore avoids these errors by using a graphical

interface for creating and manipulating visual couplings.
The software DEVSimPy then evolved into a collaboration tool construction, simulation, storage and sharing libraries DEVS models from a design of the API PyDEVS.
DEVSimPy uses the graphics library and the wxPython API PyDEVS modeling and simulation. Of course, the core language which is the cornerstone tools DEVSimPy is giving birth to the Python language.

2.3. Library Genetic Algorithm

As explained in the introduction, the WSN deployment is a complex optimization problem. This one is described by several interrelated decision variables. Moreover WSN deployment depends on two aspects: the AoD characterisitics and the goals of the deployement. Artificial Intelligence is an efficient way to resolve some complex problems and we choose to integrate Genetic Algorithms in order to define an original approach.
Genetic Algorithms (GAs) are a bio-inspyred approach (Goldberg 1989; Holland 1992; Koza 1992; Mitchell 1996) based on an observation of adaptation in natural systems as it was described by Darwin (1836) on the origin of species. In GAs each population's individual represents a possible solution to the problem. The genetic code of an individual is represented by a list of bits. Over an iteration process as show in Figure 1, the population evolves toward improving solutions.

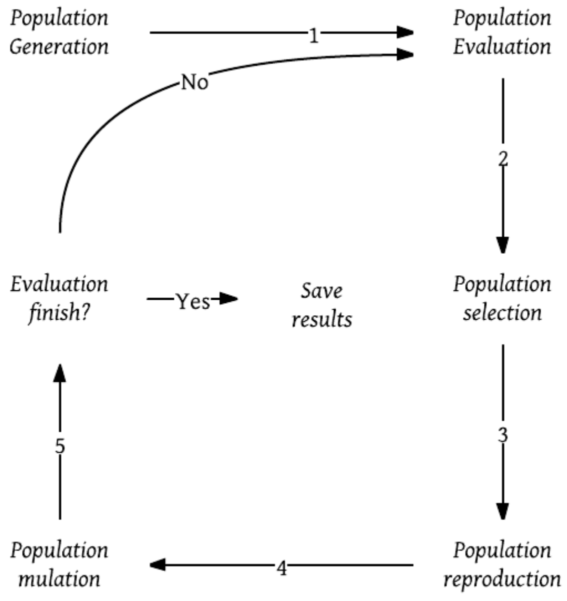


Figure 1 Genetic algorithms basic concept

The goal in the using of GAs is to provide a simulation and an optimization of WSN deployment in the same time. To reach this objective we choose to implement the different GAs functions as DEVS atomic models to allow building an interface between evaluation function and existing DEVS models as show in Figure 2.

When simulation begins the "population generator" model generates random solutions. Each created individual contains a list of node coordinates that represent the network topology. When all individuals are created they are send on the outpout port toward the "eval population" model.

Next the "population evaluator" model receives the population on it input port. The model can translate individual bits string representation if necessary. After this process the model broadcasts population over it output ports toward the decision support model(s) in order to produce simulation results. These results are stored until decision support simulation ends.The model ponderates the collected simulation results and computes a fitness for each solution. The fitness represents the response to the given problem like a score or an error value. When fitness computation ends, the population and individual fitness are send toward the selection model.

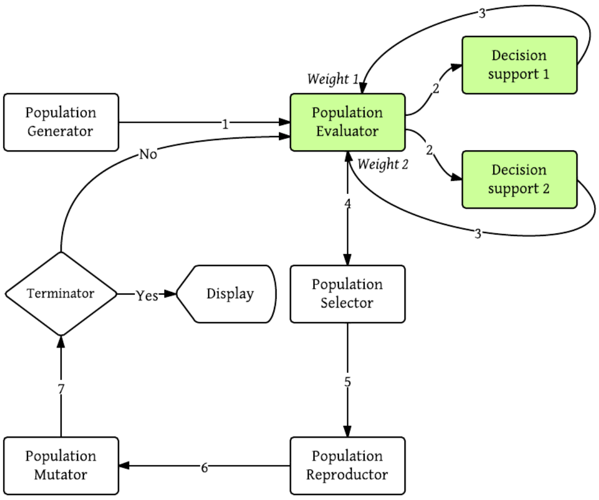


Figure 2 DEVS genetic algorithm concepts

The "population selector" model sorts the different individuals by their fitness and splits them into two groups:

1. The survivors
2. The condemned

These two groups are sent toward the "population reproductor" model.

By a crossover of genetic code the "population reproductor" will generate a new group of individuals from survivors. New children are created as visible on Figure 3

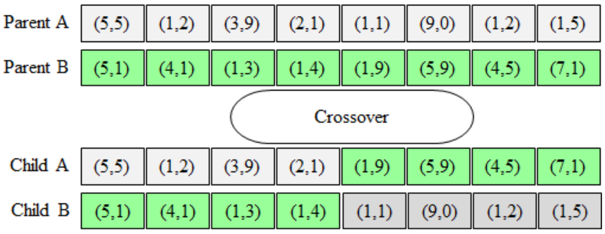

Parent A (5,5) (1,2) (3,9) (2,1) (1,1) (9,0) (1,2) (1,5)
Parent B (5,1) (4,1) (1,3) (1,4) (1,9) (5,9) (4,5) (7,1)
Crossover
Child A (5,5) (1,2) (3,9) (2,1) (1,9) (5,9) (4,5) (7,1)
Child B (5,1) (4,1) (1,3) (1,4) (1,1) (9,0) (1,2) (1,5)

Figure 3 Genetic Algorithm crossover

The model sends it new population toward the “population mutator’ model where the new solutions will be changed depending on mutation rate. This process avoids to stop into a local optimum.

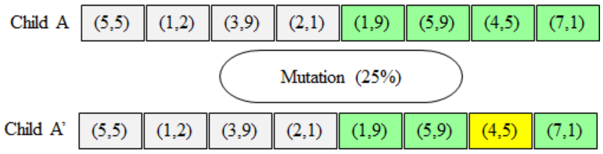

Child A (5,5) (1,2) (3,9) (2,1) (1,9) (5,9) (4,5) (7,1)
Mutation (25%)
Child A' (5,5) (1,2) (3,9) (2,1) (1,9) (5,9) (4,5) (7,1)

Figure 4 Genetic Algorithms mutation

During the optimization process user can vizualise the evolution into DEVSimPy using a specific pluggin. The user can drive the optimization process by defining a number of iterations to update the vizualisation interface.

The process loops until one or several fitness reach a user defined threshold or when a maximum number of iterations is exceeded.

3. WSN DEPLOYMENT MODELS

3.1. Connectivity model

WSN can be seen as a cooperative technology. Each sensor collects information and needs to relay them throught its neighbors towards the sink node and the gateway. The connecity between each other appears to be one of the main characteristic in the definition of Quality of Service (QoS) of a network. This QoS can be divided in two groups. The first group of values describes signal quality between each connected nodes named QoC (Quality Of Connectivity). The second group concerns numbers of neighbors for each node named QoN (Quantity of neighboor).

A DEVS model called “Connectivity” has been created to compute this type of simulation outputs.

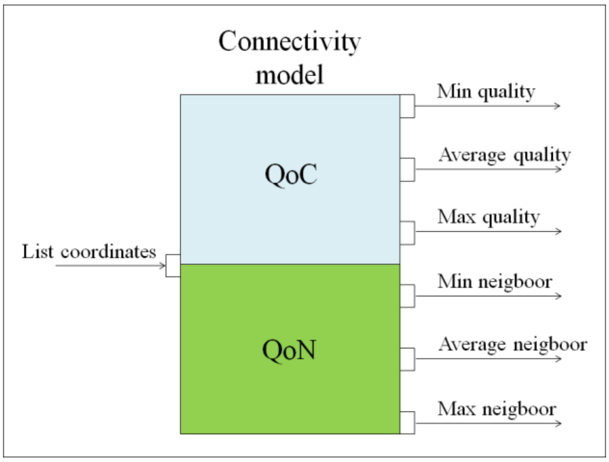


Figure 5 Connectivy model architecture

From a received list of coordinates during simulation on its single input port the connectivity model can generate different information using for the optimization of deployment process:

- Min quality : minimum quality signal
- Average quality : average quality signal
- Maximal quality : maximum quality signal
- Min neigboor : minimal number of neighbors
- Average neiboor : average number of neighbors
- Max neigboor : maximum of neighbors

QoC is estimated by using pysal (Python Spatial Analysis Library) and its module spatial weights. We define sensor communication range into the model. The Figure 6 represents for each link the estimatation model of the signal quality according to nodes distance. Different types of signal function can be used like: triangular, uniform, quadratic, epanechnikov, quartic, bisquare or gaussian.

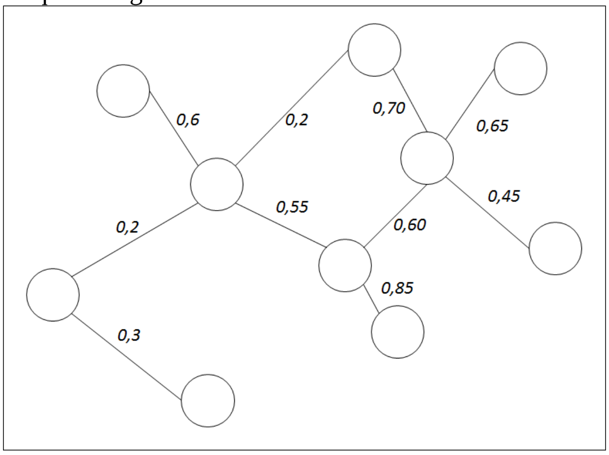


Figure 6 : Quality of links

The second group of characteristics concerns QoN. To limit nodes memory saturation and balance energy consumption over the network: the number of neighbor for each node must be the same. As shown in Figure 7 and Figure 8 the model computed the number of neighbors of each node according to a neigboorood level. This one is in relation with WSN application and specificities of communication protocols.

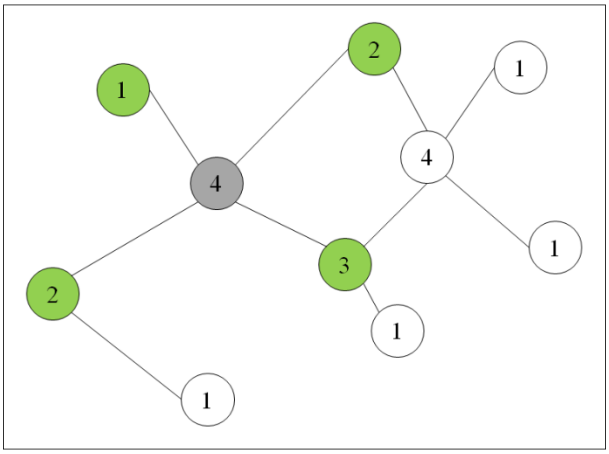


Figure 7 : Quantity of links with neighborhood level = 1

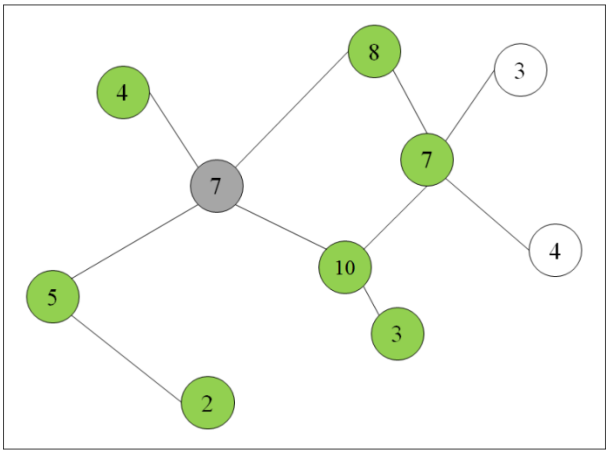


Figure 8 : Quantity of links with neighborhood level = 2

3.2. Coverage model

WSN goal is to collect data and information on a specific area. For example if the WSN application is to detect intrusion we need to cover all risks without exception.

In a WSN deployment each node localation influences other nodes locations.Therefore we need to put sensors at right places in order to maximize the area coverage and ensure a good distribution of equipements.

The coverage model provides two types of simulation output: converage informations and coverage repartition as shown in Figure 9.

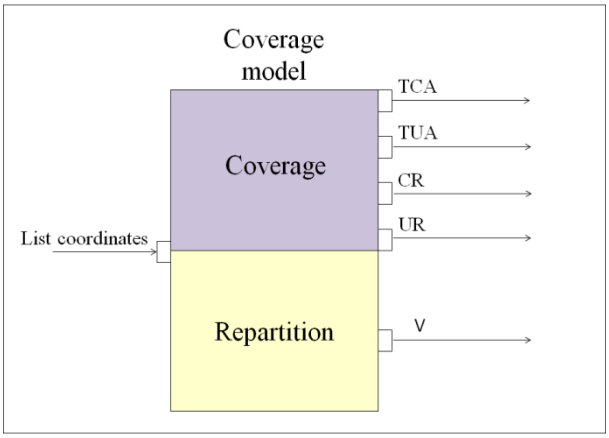


Figure 9 : Coverage model architecture

The coverage model provides different information computed with Shapely python library. The first information is TCA (Total Covered Area) givent by equation 1:

$$TCA \leftarrow \bigcup_{i=1}^{n} f(Si) \times g(Si) \cap TA \tag{1}$$

Where n is the number of sensor, f(Si) returns the sensor coverage range, g(Si) returns the sensor coverage attenuation depending on its location area or sub-area and TA referred total area. Then with this value the coverage model can easely compute the CR (Covered Rate) by equation 2:

$$CR \leftarrow TCA \times 100 \div TA \tag{2}$$

The second information is TUA (Total Uncover Area) given by equation 3:

$$TUA \leftarrow TA - TAC \tag{3}$$

With this value coverage model can compute the UR (Uncovered Rate) by equation 4:

$$UR \leftarrow TUA \times 100 \div TA \tag{4}$$

V is the variance of node number in each sub-area given by equation 5:

$$\sum_{i=1}^{n} (\alpha - \beta)^2 \tag{5}$$

Where n is the number of sub-area, α is the number of sensor into sub-area and β is the average number of sensor per sub-area.

As represented in Figure 10 we define five types of coverage (we can extend this number) capacity associated with an attenuation value:

- Full coverage (coverage = coverage)

- Good coverage (coverage = coverage x 0.80)
- Middle coverage (coverage = coverage x 0.50)
- Bad coverage (coverage = coverage x 0.20)
- Null coverage (coverage = 0)

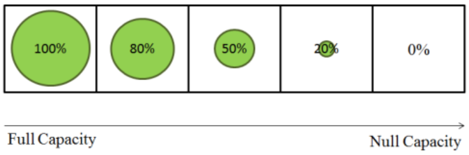


Figure 10 : Coverage Capacity

This model allows an easy representation of AoD. The AoD is exploding into several sub-areas with own values like signal attenuation as illustrated by Figure 11.

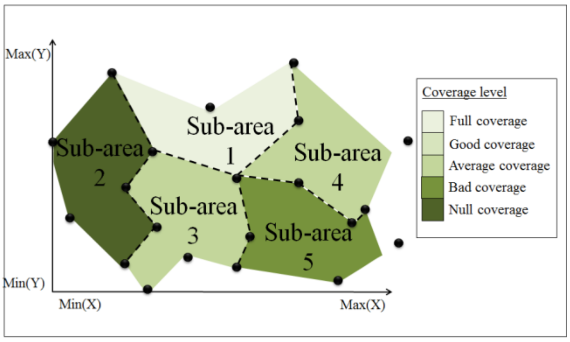


Figure 11 Area of Deployment

During simulation execution the model computes the different values using geometric operators such as intersection, union, difference as show in Figure 12.

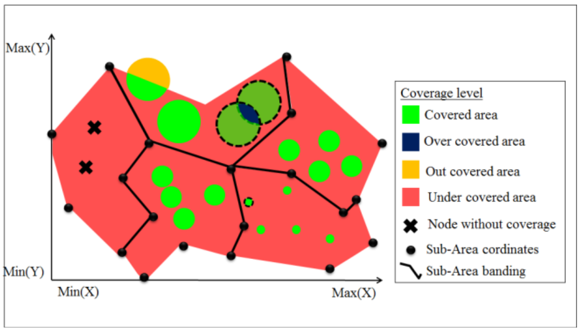


Figure 12 Covered area & uncovered area

4. RESULTS AND DISCUSSION

We have validated the previous DEVS models on three deployment optimization samples. In the first we only improve the connectivity. In the second we only improve the coverage and finally in the third we improve connectivity and coverage in a multicriteria application.

4.1. Connectivity

On this example we try to maximize the average number of neighbors for each deployed node. These nodes are deployed on AoD dimensioning 100 weights by 100 heights. The number of node is 30 with a transmission range of 10 meters.

If we observe Figure 13 we can see a random node distribution on the first iteration. Over the GAs evolution process we can see that deployment structure evolves toward height connectivity structure in witch all node are near. It is easely understandable: connectivity is inversely proportional to the distance between nodes.

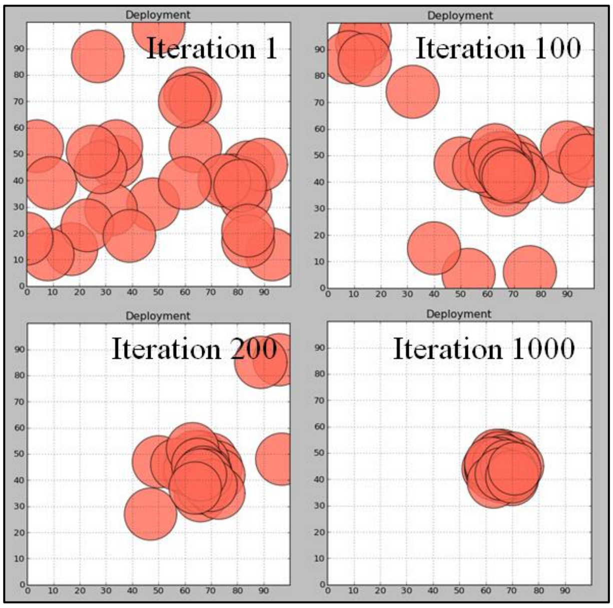


Figure 13 Connectivity optimization

4.2. Coverage

We try to maximize the deployment coverage illustrated by the Figure 15. The dimension of AoD is 100 meters weights by 100 meter height and sensor coverage range or detection range is 10 meters too. Our tool allows making distinction betweek sensing range and transmission range witch can be different.

The results show that coverage of 100 percent is really difficult to reach. This can be explaining by the difficulty to ajust cicrcle on space without blank areas. However we can note a improving of 23 percent of coverage area between the first random proposed deployment and the last optimized deployment. In the same time we can deduct that distance between nodes is proportional to the total coverage.

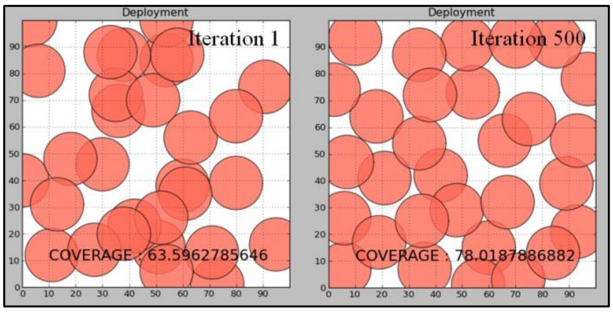


Figure 14 Coverage optimization

4.3. Qos Application

We apply these two concepts on a corsican area visible on Figure 15. After a manual and visual classification we isolate three differents sub-areas with specific coverage range atteunation as represented on Figure 16 Figure 16 and implemented into DEVSimPy as show in Figure 17:

1. Sub-area 1 : composed of rock presuming a good coverage
2. Sub-area 2 : composed of scrubland supposing a bad coverage
3. Sub-area 3 : composed of water imposing a Null coverage

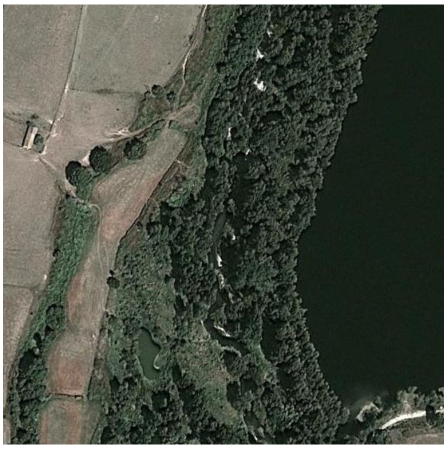


Figure 15 Area of Deployment (raster view)

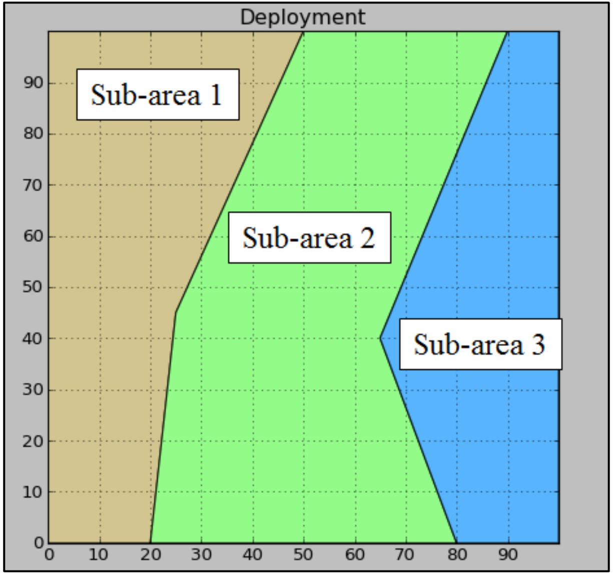


Figure 16 Area of Deployment (vector view)

We configure two groups of models: optimization models and decision support models through DEVSimPy interface. The solution's fitness are ponderate sums of coverage and connectivity outputs. The coefficients are 0.8 for coverage and 0.2 for connectivity. We consider this cofficients as criteria in our tool.

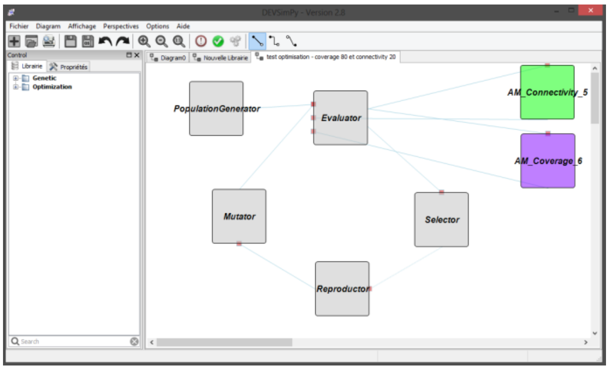


Figure 17 Deployment optimization

The Figure 18 represents the first deployment by out models. We can see a random and uniform node distribution between the different sub-areas. This configuration causes a poor AoD coverage. Moreover we can see a very low connectivity: the distance between each node is heigher than their coverage range. However two nodes are connected only and only if distance between them is lower than their range halved.

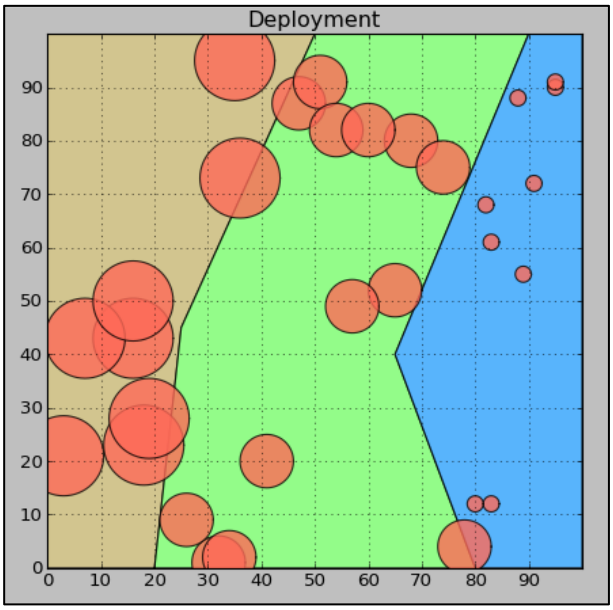


Figure 18 Generation 1

The Figure 20 represents the last deployment generation proposed by our tool. We can see the nodes move towards the sub-ara 1 where their coverage range is maximized as shown Figure 21. The results show the coverage maximization is satisfied. Moreover we can observe in the same time a height connectivy between the nodes result the connectivity criterion.

However some nodes stay in sub-area 2 where their coverage range is not maximized. This problem is characterstic for GAs. We are never sure to find the optimum.

As visible on the Figure 19, we can divide the curves evolutions into steps: between generations 0 and 60 the fitness are increasing quickly, then we observe the fitness inscreasing much more slowly.

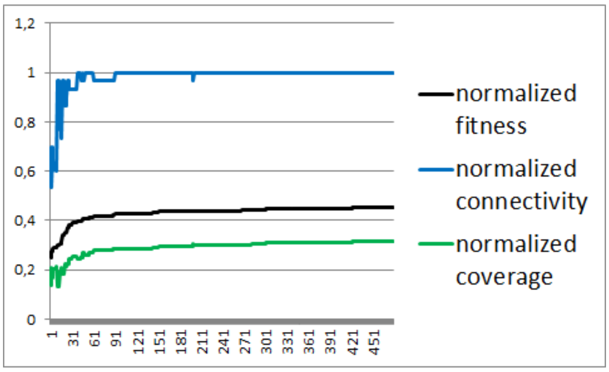


Figure 19 Normalized results

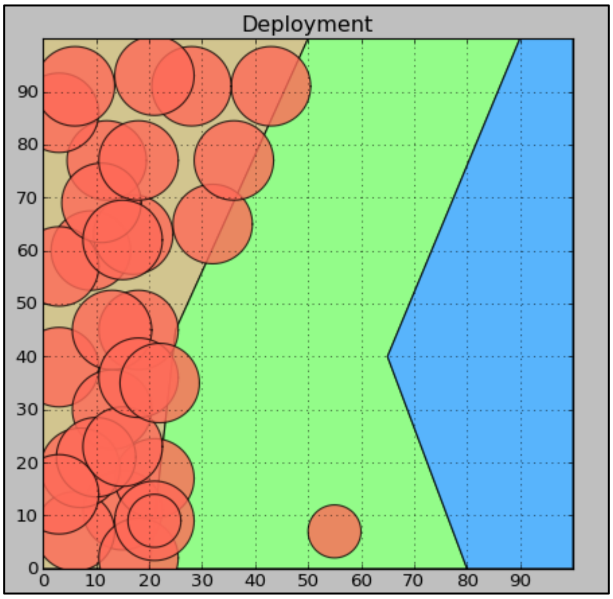


Figure 20 Generation 500

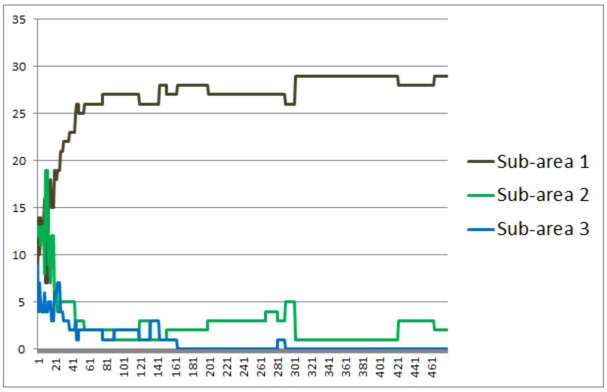


Figure 21 Node distribution

CONCLUSION AND FUTURE WORKS

In the paper we have introduced a generic DEVS framework using GAs to optimize WSN deployment according a development way targeting the definition of decision support tool.

Building on DEVS models, this framework allowed the simulation of WSN deployement according to two intrinsic parameters: connectivity and coverage. Our approach wants to be generic, not dependant of hardware platforms and limited by the domain. The first results are very interesting because they show the capacity to detect the best area for the deployment. The used example is our simulation is very simplistic however the objective is reached. The approach is able to propose the best deployement using an AI technical, AG. The GAs models implemented are generics and can be use for other optimization.

However this work is limited by the example and by the kind of AG. We propose to enhance this first approach by the three following axes:

- to test with a complex AoD, by example an insutrial site with more constrains
- to use DEVS parallel or Parallel Genetic Algorithms (PGAs) concepts to to reduce execution time because this aspect can become troublesome with the increasing of propulations
- to implement other metaheuristics algorithms like harmonic search or simulated annealing to compare the results and compute time.

REFERENCES

Antoine-Santoni, T., Santucci, J.F, De Gentili, E. Silvani, X., Morandini, F., 2009, Performance of a Protected Wireless Sensor Network in a Fire. *Analysis of Fire Spread and Data Transmission, Sensors*, 5878-5893

Antoine-Santoni, T., Santucci, J.F., De Gentili, E., Costa, B., 2008, Wildfire impact on deterministic deployment of a Wireless Sensor Network by a discrete event simulation, *Proceedings of the 14th IEEE Mediterranean Electrotechnical Conference (MELECON '08)*

Antoine-Santoni, T., Santucci, J.F., De Gentili, E., Costa, B., 2008, Wildfire impact on deterministic deployment of a Wireless Sensor Network by a discrete event simulation. *(MELECON '08)*

Beutel, J., Römer, K., Ringwald, M., Woehrle, M., 2009, Deployment Techniques for Sensor Networks Signals and Communication Technology, 219-248

Bolduc, J.S., Vangheluwe, H., 2001, the modeling and simulation package PythonDEVS for classical hierarchical DEVS, Technical report, MDSL-TR-2001-01, McGill University, Montreal, Canada

Capocchi, L., Santucci, J.F., Poggi, B., Nicolai, C., 2011, DEVSimPy: A collaborative Python software for Modeling and Simulation of DEVS systems, WETICE 2011, IEEE Computer Society

Cheng, X., Du, D.Z, Wang, L., Xu, B., 2007 Relay Sensor Placement in Wireless Sensor Networks, *IEEE Transactions on Computers*, 134-138.

Darwin, C., 1836, T*he origin of species*, New York, P.F. Collier

Dhillon, S.S., Chakrabarty, K., 2003 Sensor Placement for Effective Coverage and Surveillance in Distributed Sensor Networks, *Proceedings of IEEE Wireless Communications and Networking Conference*, New Orleans, LA, March.

Efrat, A., Har-Peled, S., Mitchell, J.S.B., 2005, Approximation Algorithm for Two Optimal Location Problems in Sensor Networks, *Proceedings of the 3rd International Conference on Broadband Communications, Networks and Systems*, Boston, Massachusetts.

Garcia, A., Patek, S. D., & Sinha, K. (2007). A decentralized approach to discrete optimization via simulation: Application to network flow. *Operations research*, *55*(4), 717-732.

Grandham, S.R., Dawande, M., Prakash, R., Venkatesan, S., 2003, Energy Efficient Schemes for Wireless Sensor Networks with Multiple Mobile Base Stations, *Proceedings of the IEEE Globecom*, San Francisco, CA.

Goldberg, D.E, 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc

Holland, J., 1992, *Adaptation in Natural and Artificial Systems*, Cambridge, MA: MIT Press

Koza, J., 1992, Genetic Programming: On the Programming of Computers by Means of Natural Selection, Cambridge, MA: MIT Press

Langendoen, K., Baggio, A., Visser, O., 2006, Murphy loves potatoes: experiences from a pilot sensor network deployment in precision agriculture, *Parallel and Distributed Processing Symposium. (IPDPS 2006) 20th International*.

Mitchell, M., 1996, *An introduction to genetic algorithms*, Cambridge, MA, USA, MIT Press

Otero, C.E, Kostanic, I., Otero, L.D, 2009, *Development of a Simulator for Stochastic Deployment of Wireless Sensor Networks JOURNAL OF NETWORKS, VOL. 4, NO. 8, OCTOBER 2009*.

Paek, J., Chintalapudi, K., Govindan, R., Caffrey, J., Masri, S., 2005, A wireless sensor network for structural health monitoring: Performance and experience, *the Proceedings of the. 2nd IEEE Workshop on Embedded Networked Sensors* (EmNets '05)

Swisher, J. R., Hyden, P. D., Jacobson, S. H., & Schruben, L. W. 2004. A survey of recent advances in discrete input parameter discrete-event simulation optimization. *IIE Transactions*, *36*(6), 591-600.

Szewczyk, R., Mainwaring, A., Polastre, J., Anderson, J., Culler, D., 2004, An analysis of a large scale habitat monitoring application, *Proceedings of the 2nd international conference on Embedded networked sensor systems* (SenSys '04), 214-226, New York, NY, USA

Tateson, J., Roadknight, C., Gonzalez, A., Fitz, S., Boyd, N., Vincent, C., Marshall, I., 2005, Real world issues in deploying a wireless sensor network for oceanography, *Proceedings of Workshop on Real-World Wireless Sensor Networks* (REALWSN '05)

Werner-Allen, G., Lorincz, K., Johnson, J., Lees, J., Welsh, M., 2006, Fidelity and yield in a volcano monitoring sensor network, *Proeedings. of 7th Symposium Operating Systems Design and Implementation*, 381–396, New York, NY, USA

Younis, M., Akkaya, K., 2008, Strategies and Techniques for Node Placement in Wireless Sensor Networks: A Survey, Ad Hoc Networks 6 (2008) 621-655.

Zeigler, B.P., 1976, Theory of Modeling and Simulation, Wiley Interscience, New York.

Zeigler, B.P., 1987, Hierarchical, modular discrete-event modelling in an object-oriented environment, SIMULATION, vol. 49, 1987, p. 219–230

Zennaro, M., Bagula, A., Gascon, D., Bielsa Noveleta, A., 2010, Planning and deploying long distance wireless sensor networks: the integration of simulation and experimentation, *Proceedings of the 9th international conference on Ad-hoc, mobile and wireless networks*, 191-204, Edmonton, AB, Canada

AUTHORS BIOGRAPHY

Bastien Poggi was born in Ajaccio, France in 1988. In 2011, he received a MSc in Computer Science from University of Corsica, Corte, France. Currently, he is prepared a PhD in the “Sciences for environnement” laboratory at University of Corsica. His main research focuse on DEVS (Discret EVent system), GAs (Genetics Algorithms) applied to WSN (Wireless Sensor Network).

Jean-François Santucci is Full Professor in Computer Science at the University of Corsica since 1996. His main research interests are Modeling and Simulation of complex systems. He has been author or co-author of more than 150 papers published in international journals or conference proceedings. Furthermore he has been the advisor or co-advisor of more than 20 PhD students and since 1998 he has been involved in the organization of ten international conferences. He is conducting newly interdisciplinary researches involving computer science, archaeology and anthropology: since 2006 he is working in interdisciplinary research topics: in the one hand he is performing researches in the archaeoastronomy field (investigating various aspects of cultural astronomy throughout Corsica and Algeria) and on the other hand he is applying computer science approaches such as GIS (Geographic Information Systems) or DEVS (Discrete EVent system Specification) to anthropology.

Thierry Antoine-Santoni is an associate professor in Computer Science at the University of Corisca since 2010. He maintained his doctoral thesis in 2007. His main topics of research are divided in two axes: modeling and simulation of complex systems and wireless sensor network. He developed different aspects and relations between using DEVS simulation and WSN testbeds for the monitoring of environmental and industrial phenomena.

HOW SMP SUITS THE NEEDS FOR THE EUROPEAN SPACE SECTOR

Vemund Reggestad[a], Anthony Walsh[b]

[a,b]ESA-ESOC, Darmstadt, Germany

[a]vemund.reggesta@esa.int, [b]anthony.walsh@esa.int

ABSTRACT

There is hardly any paper dealing with the topic of simulation and modelling for space applications without touching on the topic of model reuse. Simulation and modelling plays a key role during the entire lifecycle of a space project and the investments done in simulation models is a significant part of the overall cost of any space project. Due to this, it is natural that several ways have been invented to reduce the cost of simulation modelling by increasing model reuse. This paper will concentrate on how the ECSS-E-TM-40-07 Simulation Modelling Platform specification can be used as a fundament to build up an entire effective approach to simulation and modelling covering the entire space project

It will look at several areas like:

- Model development techniques covering the typical software development lifecycle, specification, design, implementation, testing.
- Model evolution from early concept studies (Phase A activities) until high fidelity models (Phase E activities).
- Model reuse between projects for models with similar requirements.
- Model design for reuse by applying reference architectures.
- Model exchange between organisations by building Library of Models.

Finally, by as well taking into account the importance of platform independency, the paper will show how the ECSS-E-TM-40-07 supports all the needs of the European Space Sector.

Keywords: Space, SMP

1. INTRODUCTION

Simulation techniques have been used with great success to support different aspects of a satellite mission lifecycle. However, this often involved the development of numerous simulators which were very different in terms of use-cases, size and scope. Often these simulators may have been implemented using different tools and approaches, and potentially running on different platforms under different operating systems.

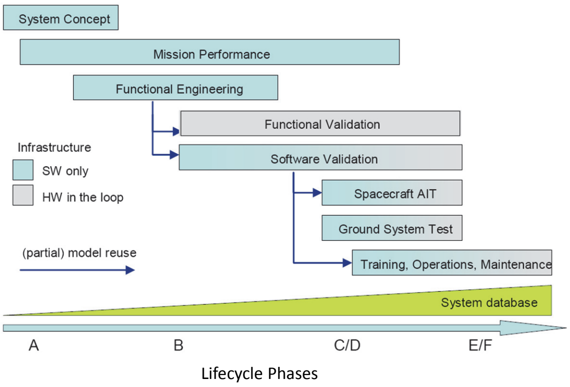


Figure 1: Simulation used across the mission lifecycle[1]

Building on these company-focused approaches, the use of simulation to support the system engineering activity has now also become part of ECSS – the ECSS-E-TM-10-21A Technical Memorandum (2010). ECSS-E-TM-10-21A identifies a set of simulation facilities (systems) deployed at various points (phases) throughout the lifecycle, each fulfilling a particular set of use-cases. In reality some of the individual identified facilities may in fact be combined into a single configurable multi-role system. This approach is far more logical and cost-effective compared to the development of a set of separate bespoke systems. However, it places a large emphasis on model re-use and simulation facility re-use throughout the lifecycle.

The ability to re-use models effectively and efficiently places a large emphasis on portable and configurable models, and the adoption of suitable standards (covering portability, model design, and simulator architectures). The re-use objective therefore has to be a planned part of the overall model or simulator development process. The fundamental aspects of reuse are shown in 0. This paper will describe each level in more detail as well as the use cases this reuse pyramid allows.

[1] Figure reproduced from ECSS-E-TM-10-21A

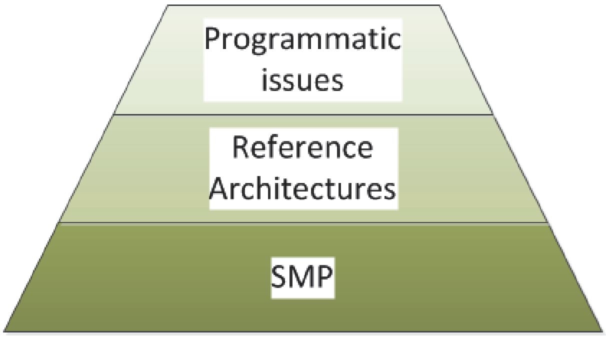


Figure 2: Reuse pyramid

2. THE SMP STANDARDS

In order to promote and support simulation model re-use, and facility re-use, the European Space Agency (ESA) has undertaken several key initiatives.

The Simulation Model Portability standard (SMP) was initiated by ESA in 1999. Work on the second version of the standard started in 2005 focuses on:

- model development and integration,
- and inter-model communication

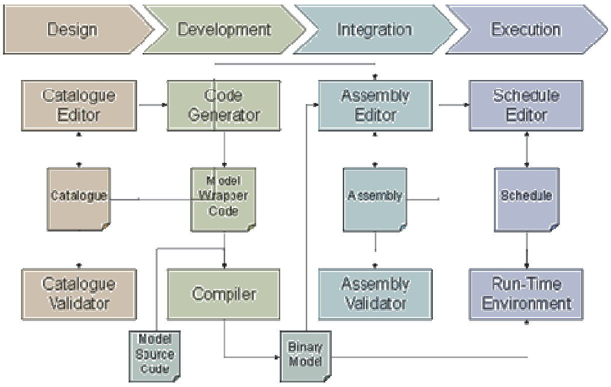


Figure 3: Model lifecycle and SMP2 artefacts

This process is most efficient when supported by suitable SMP2 tools, in particular allowing the user to create an SMP2-based design, and in general to handle the various SMP2 artifacts. This topic is covered in more detail in section 2.1 below.

Table 1: Development Environments and Run Time Environments

Development environments	Run Time environments
UMF SimVis	SIMSAT (ESA) Basiles (CNES) SimTG (Astrium) EuroSim (EuroSim Consortium)

In the past few years SMP2 has been revised into a Technical Memorandum within ECSS, ECSS-E-TM-40-07 (2005). Future follow-on activities are anticipated to transform this into an SMP standard within ECSS.

2.1. Modelling Environments and Tools

The SMP standard requires tooling support before it can be applied effectively. It also allows effective tools to be developed that aid the development process of models.

2.1.1. ESA Universal Modelling Framework (UMF)

UMF (Universal Modelling Framework) is an example of a tool enabling efficient development of models and integration of simulations with SMP (Fritzen et al. 2013).

A short overview of the steps are provided in 0:

- Requirements can be imported and mapped to design in UML.
- A UML based design approach is used to capture the models design in a platform independent way. From the UML design, SMP catalogues describing the models are exported.
- Based on the SMP catalogues, code generation and merging are supported to allow efficient implementation of the models behaviour.
- An extensive suit for testing of SMP models both at unit level and integration level are provided.
- Finally, the models can be packaged, and integrated into a ready to be started simulation. It can be distributed via the Library Of Models together with auto-generated documentation covering both user manuals and design documentation.

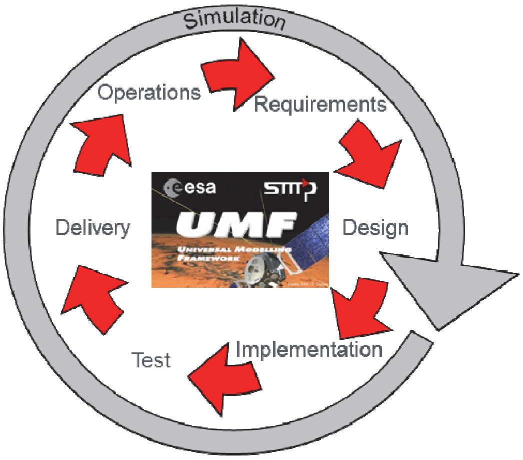


Figure 4: Model development cycle with UMF [9]

3. REFERENCE ARCHITECTURES

In addition to standards such as SMP2 a suitable reference architecture are required to effetely re-use model. Reference Architectures adds semantics for spacecraft system simulation which is not addressed by SMP-2 which is agnostic to the specific domain being simulated.

Within the space domain, two flavours of Reference Architectures exist today with different aims and use.

3.1. Intrusive Reference Architectures

ESA's European Space Operations Centre (ESOC) has created a so-called Reference Architecture (REFA) which is actively used at the core of the design of their SMP2-based Operational Simulators. Its use within ESOC simulators promotes consistency in design across the different mission satellite simulators, and further facilitates the re-use of SMP2 models.

REFA defines a set of standard interfaces between common satellite subsystems' models within the satellite simulator, and a set of base SMP2 models (REFA Interface Control Document, 2011). In addition to using REFA as-is, a simulator developer can also extend the REFA interfaces or models via inheritance to suit the design needs for a particular simulator.

3.2. Pure interface based Reference Architectures

SSRA (2010) and ISIS (2011) provides a basis for the Virtual System Model described in ECSS-E-TM-10-21A [1]. It defines an SMP2 reference architecture supporting the re-use of simulation models across various mission lifecycle phases and simulation facilities, or even across missions. ISIS strictly focus on interfaces to allow for model exchange between organizations and allow interoperability of models.

3.3. Overview of differences between reference architecture approaches

The following table show a short summary of the areas that may be covered by each of the types of Reference architectures:

Table 2: Standardization Area

Standardization area	IF based	Intrusive
Interfaces	X	X
Operability		X
Development approach		X
Common base classes		X
Approach for tracing/ debugging		X
Approach for installation and versioning		X

3.4. Current status in Europe

Several reference architectures have been developed in Europe during the last years to enable efficient reuse aiming at solving specific problems of specific organizations. This proliferation of reference architectures underline the fundamental need for standardization also in this field. However since most of the exiting solutions have been developed without taking the overall problem into account, it is currently problematic to achieve reuse between organizations using different architectures.

ESA is however currently initiating work to harmonize the various reference architectures used to ensure compatibility between Interface based and intrusive architectures in the various organizations.

4. PROGRAMMATIC ISSUES

4.1. Model exchange in practice

In order to efficiently reuse models, it is not sufficient to only overcome the technical issues of model integration. It is also required an easy way to in practice transferee models from Organization A to B, as well as a central place where the it is possible to get an overview of the available models.

To achieve this, a "Library of Models" (LoM) is needed. This is similar to common practice for release management via a Repository Manager (For example http://www.sonatype.org/nexus/).

Such a LoM must provide facilities to:

- Upload and download simulation models both in binary and source code format.
- Allow restricting access to models depending on license issues
- Provide protection for IPR issues.
- Provide standardized tags and attributes for models to allow for easy identification of suitable models.

5. MAJOR USE CASES FOR SMP

Following use-cases describes the different situations where SMP can be applied:

1) The SMP standard allow model exchange process between customer and supplier. Typically the supplier being a System integrator, i.e. an organization developing complete simulation solutions as part of the overall system to be developed. The customer either being the Spacecraft operators for operational simulators or other entities for Independent Design Verification.

2) The SMP standard allows outsourcing of simulation model developments, so that the System Integrator can concentrate on integration of models developed by domain experts. Such outsourcing clearly relay on a well-defined reference architecture as well.

3) The SMP standard allows model reuse by allowing system integrators to build a library of models suitable for reuse. The simulators can then be built by assembling already developed and validated models.

4) The SMP standard allows simulator end users (customers) to customize their simulation solutions by replacing a simulation model with its own custom version. Clearly such replacing imply a heavy revalidation of the overall system.

5) The SMP standard allow portability of simulations from one Run Time environment to another. This allows different organizations to harmonize internally on standard simulation environments, while still exchange simulation models with other organizations using other environments.

All of these use cases are summarized in 0.

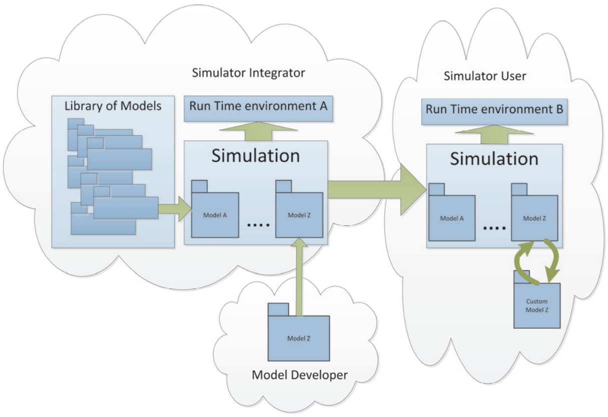


Figure 5: Summary of model exchange scenarios

6. CURRENT STATUS AND ISSUES

SMP is frequently used within the Space community, but a wide spread usage also outside the space domain is currently not taking place largely due to:

- It is a significant cost to upgrade existing tools to support the standard.
- Most larger organization has inhouse existing solutions that is suitable for their own development.
- Simulation technology is seen as "key competence" for several companies, hence there is no interested in standardizing it, since this would open it up for external competition.
- There is a need for standardization on reference architectures as well.

There are however resent signs that some of these issues may be resolved:

- Mathworks are evaluating to support SMP within Mathlab, hence removing the need for a SMP development or runtime environment to use the SMP standard.
- Organizations are realizing that it is extremely costly to maintain a state of the art internal simulation infrastructure, hence model portability raises on the agenda in several major European Companies.

7. SUMMARY AND CONCLUSIONS

The paper has described how SMP enables effective reuse of simulation and modelling in the entire Space sector. A review of the current status and issues focusing on the tree levels in the reuse pyramid has been done:

- Standardization (SMP)
- Reference Architectures
- Programmatic issues

This shows how SMP are used as a fundamental building block to achieve cost effective reuse for the need of the European Space Industry for the next decades to come.

REFERENCES

ECSS-E-TM-40-07 Volume 1A to 5A, 25-Jan-2011

Fritzen, P., Reggestad, V., Walsh, A, 2013. UMF – A Productive SMP2 Modelling and Development Tool Chain. Ellsiepen, EGOS 2013

ISIS Training Operation and Maintenance (TOMS) Interface Specification (ISIS), ISIS-SIM-IF-305-CNES_04 issue 4, 28/09/2011, CNES

REFA Interface Control Document Volume 1: Functional Interfaces, Reference: EGOS-SIM-REFA-ICD-1001 , Issue 1.6, 21-Apr-2011

SMP 2.0 Handbook, EGOS-SIM-GEN-TN-0099, Issue 1.2, 28-Oct-2005

System Modelling and Simulation, ECSS-ETM-10-21A, 16-April-2010

The Space Simulation Reference Architecture – Reference Architecture Specification Volumes 1-3, (SSRA.REP.001, SSRA.REP.002, SSRA.REP.003), Issue 1.0, 27-May-2010

Trends in European Space Simulation: Standards, Architectures and Tools across the Mission Lifecycle, Michael Irvine, Peter Fritzen, Peter Ellsiepen, RAST 2013

http://www.sonatype.org/nexus/

INTEGRATING AN AGENT-BASED MODEL OF MALARIA MOSQUITOES WITH A GEOGRAPHIC INFORMATION SYSTEM

S. M. Niaz Arifin[(a)], Rumana Reaz Arifin[(b,c)], Dilkushi de Alwis Pitts[(c)], Gregory R. Madey[(d)]

[(a)]Department of Computer Science and Engineering, University of Notre Dame
[(b)]Department of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame
[(c)]Center for Research Computing, University of Notre Dame
[(d)]Department of Computer Science and Engineering, University of Notre Dame

[(a)]sarifin@nd.edu, [(b,c)]rarifin@nd.edu, [(c)]dpitts@nd.edu, [(d)]gmadey@nd.edu

ABSTRACT

Agent-based models (ABMs) are used to model infectious diseases and disease-transmitting vectors. Malaria is a deadly infectious disease in humans, transmitted by *Anopheles* mosquito vectors. Although geographic information system (GIS) has been used before with ABMs, no ABM-based malaria study showed the usage of custom-built spatial outputs integrated within a modeling framework. In this paper, we show how to effectively integrate a malaria ABM with GIS-based, spatially derived parameters. For a specific study area, we process GIS data layers, create hypothetical scenarios, produce maps, and analyze biological insights. Results indicate that availability of resources and relative distances between them are crucial determinants for malaria transmission. The maps also reveal potential hotspots for the measured variables. We argue that such integrated approaches, which combine knowledge from entomological, epidemiological, simulation-based, and geo-spatial domains, are required for the identification of relationships between spatial variables, and may have important implications for malaria vector control.

Keywords: agent-based model, malaria, *Anopheles gambiae*, geographic information system

1. INTRODUCTION

Malaria is one of the oldest and deadliest infectious diseases in humans, transmitted by female mosquitoes of the genus *Anopheles*, which are regarded as the primary vector for transmission. Agent-based models (ABMs) can play important roles in malaria modeling, and answer interesting research questions. For example, ABMs can assist in selecting appropriate combinations of vector control interventions to interrupt malaria transmission, and in setting response timelines and expectations of impact.

Earlier, we developed a spatial ABM of the mosquito vector *Anopheles gambiae* for malaria epidemiology (Arifin, Davis, and Zhou 2011a; Arifin, Davis, and Zhou 2011b). Following a biological core model that describes the mosquito vector population dynamics, the ABM simulates the life-cycle of mosquito agents by tracking attributes of each individual mosquito.

A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. The idea of integrating GIS with ABMs is not new. Several studies, ranging across multiple domains, have shown such integration. For example, Brown *et al.* (2005) addressed the coupling of GIS-based data models with agent-based process models, and analyze different requirements for integrating ABM and GIS functionality. They illustrate the integration approach with four ABMs: urban land-use change, military mobile communications, dynamic landscape analysis and modeling system, and infrastructure simulations.

GIS has also been used in various epidemiological studies. For example, Gimnig, Hightower, and Hawley (2005) discussed the application of GIS to the study of mosquito ecology and mosquito-borne diseases, including malaria. Khormi and Kumar (2011) presented a review of mosquito-borne diseases, with examples of the use of spatial information technologies to visualize and analyze mosquito vector and epidemiological data.

However, no *model-based malaria study* has yet shown how to integrate an ABM with GIS, and thereby harness the full power of GIS, especially by utilizing custom-built spatial outputs. There is also a *vacuum of knowledge* in building robust integration frameworks that can guide the use of ecological, geo-spatial, environmental, and other types of features (related to malaria transmission) as model inputs, as opposed to simply use these features as cartographic outputs from the models (as done by most previous studies).

In this paper, we show how to effectively integrate a spatial ABM of malaria vector mosquitoes with a GIS. For a specific study area (Asembo, Kenya), we identify the relevant data layers, and collect, analyze, and prepare the data for the ABM. We rank different aquatic habitat types based on their characteristics. Then, we assign relative carrying capacities to the habitats, and build two hypothetical scenarios. Once the ABM is run with both scenarios, we analyze custom spatial variables (outputs of the ABM), which include adult abundance by location, cumulative biomass per aquatic habitat, cumulative number of females oviposited per aquatic habitat, and cumulative number of bloodmeals per house. Lastly, we produce GIS maps by overlaying the spatial variables on top of the relevant data layers, and analyze important biological insights as discovered from the maps.

The organization of this paper is as follows: Section 2 describes some relevant studies. Section 3 briefly describes the ABM, the study area, and the GIS-ABM Workflow. In Section 4, we describe, in details, the processing steps of the GIS data layers. Section 5 describes two hypothetical scenarios created for the ABM. Section 6 describes the assumptions of the simulations, and defines the four custom spatial variables produced as outputs of the ABM. Section 7 describes our results, and Section 8 concludes.

2. LITERATURE REVIEW

In this section, we discuss several malaria-related studies that use GIS, Global Positioning System (GPS), spatial statistical methods, geo-spatial features, *etc*. In general, GIS has been extensively used in epidemiological studies. In particular, for malaria as a disease, GIS has been used for measuring the distribution of mosquito species, their habitats, the control and management of the disease, *etc*. GIS and spatial statistical methods are regarded as important tools in epidemiology to identify areas with increased risk of diseases, and determine spatial association between disease and risk factors (Mmbando *et al*. 2011).

Mbogo *et al*. (2003) studied the seasonal dynamics and spatial distributions of *Anopheles* mosquitoes and *Plasmodium falciparum* parasites along the coast of Kenya. Using hand-held GPS, they recorded latitude and longitude data at each site, and produced the spatial distribution maps for three *Anopheles* species. Li, Bian, and Yan (2006) presented a spatially distributed mosquito habitat modeling approach, integrating a Bayesian modeling method with Ecological Niche Factor Analysis (ENFA) using GIS. They used data for seven environmental variables to represent the environmental conditions of larval habitats in the Kenya highlands. The *Malaria Atlas Project (MAP)* developed the science of malaria cartography by modeling the global spatial distribution of *P. falciparum* malaria endemicity (Hay and Snow 2006). Focusing on the spatial heterogeneity of malaria transmission intensity, they effectively produced and used maps as essential tools for malaria control (Hay *et al*. 2009).

Zhou *et al*. (2007) used GIS layers of larval habitats, land use type, human population distribution, house structure, and hydrologic schemes, overlaid with adult mosquito abundance, to investigate the impact of environmental heterogeneity and larval habitats on the spatial distribution of adult *Anopheles* mosquitoes in western Kenya. Mmbando *et al*. (2011) conducted a study of four cross-sectional malaria surveys in 14 villages located in highland, lowland, and urban areas of northeastern Tanzania during the rainy seasons. Their results show a significant spatial variation of *P. falciparum* infection in the region, identifying altitude, socio-economic status, high bednet coverage, and urbanization as important factors associated with the spatial variability in malaria. Ndenga *et al*. (2011) used a GPS unit to classify aquatic habitats within highland sites in western Kenya. They recorded the latitude, longitude, and altitude of the habitats, and classified them as natural swamp, cultivated swamp, river fringe, puddle, open drain or burrow pit, and showed that the productivity of malaria vectors from different habitat types are highly heterogeneous.

3. ABM, STUDY AREA, AND WORKFLOW

In this section, we describe the study area, the GIS-ABM Workflow, and the selected GIS data layers, which can be broadly classified into two categories: aquatic habitats and houses.

3.1. The Agent-based Model (ABM)

The agent-based model (ABM) was described earlier in (Zhou, Arifin, Gentile, Kurtz, Davis, and Wendelberger 2010). It is derived from a core entomological model of the dominant malaria vector species *An. gambiae*. The core model, essentially conceptual in nature, is governed by the biology underlying *An. gambiae,* and describes the vector population dynamics. The verification and validation processes for the ABM were described in (Arifin, Davis, and Zhou 2010a; Arifin, Davis, Kurtz, Gentile, and Zhou 2010b). For this study, we use a spatial extension of the ABM, which was described in detail in (Arifin, Davis, and Zhou 2011a; Arifin, Davis, and Zhou 2011b).

In the spatial ABM, each aquatic habitat is associated with a finite *carrying capacity* (*CC*), which is treated as a soft limit on the aquatic mosquito population that the aquatic habitat can sustain. The *combined carrying capacity* (*CCC)* for a given landscape (with one or more aquatic habitats) represents the sum of the *CC*s of all aquatic habitats.

3.2. The Study Area

For this study, a village cluster in Kenya’s Rarieda Division in Nyanza Province, known locally as *Asembo*, is chosen as the study area (see Figure 1). Asembo is located within a subsection of the Siaya and Bondo Districts in western Kenya. According to estimates from the 1989 Kenya Government census statistics, it covers an area of 200 km^2 and had a population of approximately 60,000 persons (Phillips-Howard *et al*. 2003). Asembo includes a study site of 15 villages (with an area of approximately 70 km^2 near Asembo Bay, and experiences intense, perennial malaria transmission (Nahlen, Clark, and Alnwick 2003).

The primary reason for selecting Asembo as our study area is the availability of data from the *Asembo Bay Cohort Project* (McElroy *et al*. 2001) and the *Asembo ITN project* (Phillips-Howard *et al*. 2003), which, in a series of 23 articles, report important public health findings from a successful trial of insecticide-treated bednets (ITNs) in western Kenya (Kazura 2003). Their research findings provide substantial evidence that high coverage of ITNs in the study area will result in significant health benefits for affected communities (Nahlen, Clark, and Alnwick 2003).

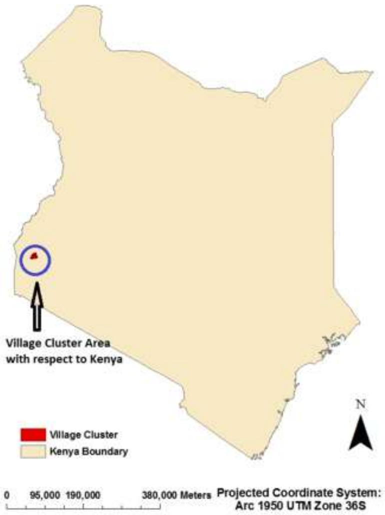


Figure 1: The Study Area of Village Clusters in Asembo, Kenya

3.3. The GIS-ABM Workflow

The GIS-ABM workflow is shown in Figure 2. The GIS module, using the ArcGIS software (ArcGIS Desktop 2011), produces, processes, and analyzes the relevant data layers, and converts them into plain-text ASCII format. The ASCII files are then converted into XML format by using a customized Java program (the Input Formatter), and fed as input to the spatial ABM. Once the ABM completes the simulations, the outputs are analyzed by using a custom-built Perl module (the Output Analyzer). Plots and other figures are then produced from the analyzed output. To perform the spatial analysis, we produce ASCII files from the analyzed output, and feed those into the GIS module. The GIS module then produces spatial maps with relevant information portrayed on top of the data layers.

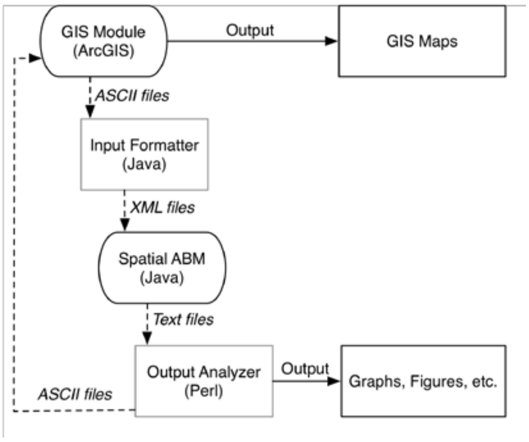


Figure 2: The GIS-ABM Workflow

4. GIS DATA LAYERS

The GIS data layers represent several thematic layers of the study area that are relevant to our spatial ABM. These layers fall into two categories: aquatic habitats and houses. The aquatic habitat types include two types of mosquito breeding sites (*type-1 breeding site* and *type-2 breeding site*), boreholes, pit latrines, and wetland. A type-2 breeding site is composed of a type-1 breeding site and several other data points (*e.g.*, compounds, boreholes *etc*.). Boreholes, also known as borrow pits, have great potentials as breeding sites in this area, and represent holes or pits made in the ground when local people use clay or soil for building houses, making pots, *etc*., thereby leaving depressions in the ground that easily get filled with rain water. Pit latrines are very common to households in the study area. The wetland represents a stretch of waterbody lying to the northwest corner of the study area. Human houses serve as bloodmeal locations for the mosquitoes, and include houses, huts, *etc*.

4.1. Processing the Data Layers with GIS

We start with the feature identification and extraction process for the whole of Kenya; then, we describe the *scale down* process to the study area of Asembo, followed by the selection of a subset of villages within Asembo, and finally, to the selection of a polygon within the village clusters, which is used as input to our spatial ABM.

We first identify and extract different water features (rivers, wetlands, and several water-points) and villages (including human houses) for all over Kenya, as shown in Figure 3. Different water features (rivers, wetlands, *etc*.) and villages are projected to the projection system Arc 1960 UTM Zone 36S. In Figure 3, the figure on the left shows different water features (rivers, wetlands and several water-points) all over Kenya, and the figure on the right shows villages for Kenya.

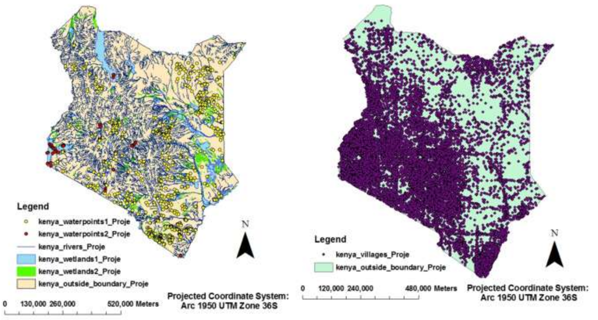


Figure 3: Water Sources and Villages Projections for Kenya

4.2. Selecting Aquatic Sites for the Study Area

We collect water features data for different types of aquatic sites. Each water source is assigned a unique ID ($ID_{data\text{-}feature}$).

Once we thoroughly examine the water sources' data layers, we encounter some overlapping problems.

To overcome these, we provide precedence ranking for the data layers, by sub-grouping water source layers based on their attributes, and creating new shapefiles. In the process, we combine similar types of water features in the same data layer.

Figure 4 shows the selected data layers for different water features. We use the *Select By Attributes* tool (with SQL query) to select features, and export the selected features to create new shapefiles. We also assign new IDs for each water feature type by using the *Field Calculator* tool to calculate value for the $ID_{data\text{-}feature}$ field.

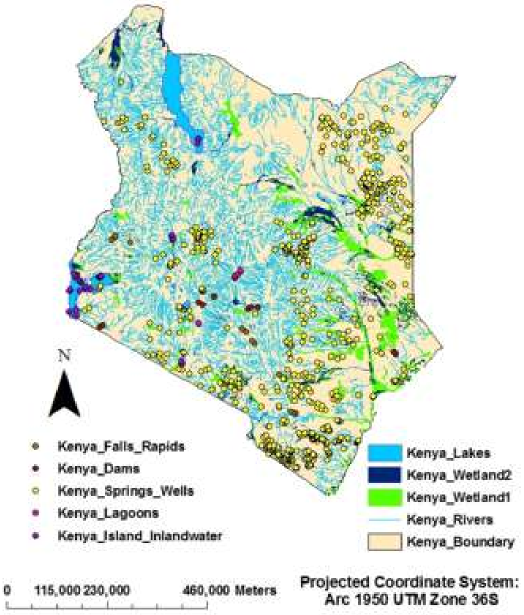


Figure 4: Selected Water Features for Kenya

4.3. Scaling Down to Village and Household Level

Since for the spatial ABM we need household level data that include water features available near the houses, we scale down the data to selected village cluster area from the entire Kenya boundary area.

We select a village cluster in Asembo, based on higher frequency of aquatic sites availability near households (than other clusters in Asembo). After analyzing the water features for all over Kenya, we also discover some wetlands and rivers features in the selected area. Figure 5 shows the selected village cluster area with houses and all water features located in the study area.

For reasons of performance and complexity (for example, large number of features), we further select a subset of villages from the village cluster area. The ABM, without explicit parallelization or multiple runs, can handle a landscape with maximum dimensions of *95 columns * 96 rows*. To reflect the available field data that points to limited flight ability and perceptual ranges of *Anopheles* mosquitoes, each cell in the landscape is set to *50 m * 50 m*, yielding a total area of ≈ 25 km^2. Hence, we further scale down the area, and select a 25 km^2 polygon, as outlined in magenta in Figure 5.

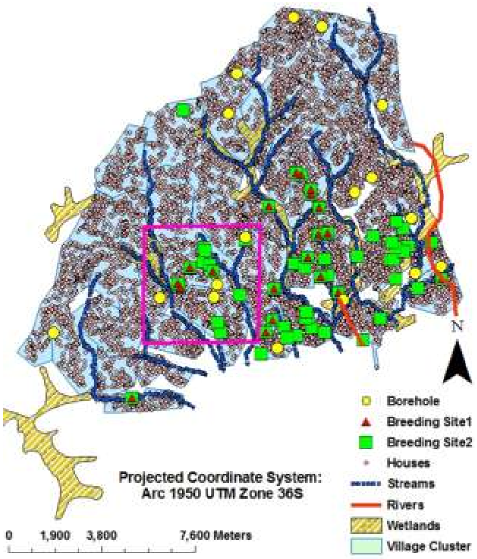


Figure 5: The Selected Polygon, Outlined in Magenta, within the Village Clusters in Asembo, Kenya

Next, we clip (crop) the aquatic habitats and houses within the outside boundary of the polygon. The clipped features include wetlands, streams, boreholes, breeding sites, and pit latrines. We eliminate the stream and river features, which, being moving (non-stagnant) water sources, are usually not considered as prospective breeding sites for *Anopheles* mosquitoes.

4.4. Conversion of Data Formats

Since the ABM needs data in the ASCII format, we first convert the selected layers to raster grid format. The cell size is set to *50 m * 50 m*, with the value field set to $ID_{data\text{-}feature}$. All point feature data layers for type-1 breeding sites, type-2 breeding sites, boreholes, pit latrines, and houses are converted using the *Point to Raster* tool. The data layer for pit latrines is created from the data layer for houses (since pit latrines are usually found inside household boundaries).

Due to the resolution (cell size), more than one feature type may fall in a single cell. In these cases, to calculate the number of features (of each type) in each cell, we set the *Cell Assignment Type* as *SUM*, since it sums the attributes of all points in the cell. Thus, it acquires the summation of the value fields (of $ID_{data\text{-}feature}$), and helps us to determine the number of features. Next, we set the extent for the conversion as the boundary coordinates of the polygon area shapefile (see Figure 5), and convert the raster files to ASCII format.

4.5. Generating the Study Area

Finally, we generate the study area for the ABM, which is shown in Figure 6: Map 1 shows the study area polygon for the ABM, outlined in magenta. The same polygon, within the village cluster area of Asembo, is shown in Map 2. In Map 3, the village cluster is marked in red circle within the map of Kenya. The figure clearly identifies the comparative scale down process of the area, as described previously in this section.

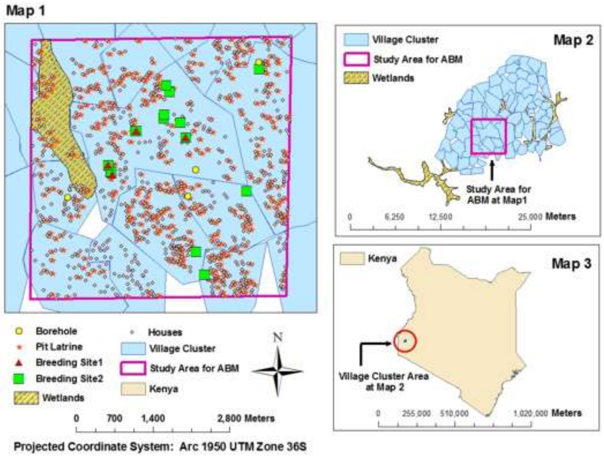


Figure 6: Study Area for the Spatial ABM

We reserve the use of shades of blue for all aquatic habitats, and square brown symbols for the houses. To aid in visualization, gridlines are also added to the map to every tenth point (starting from 1), using the Hawth's Tool (Hawth 2013), which we import into ArcGIS. The output map is shown in Figure 7.

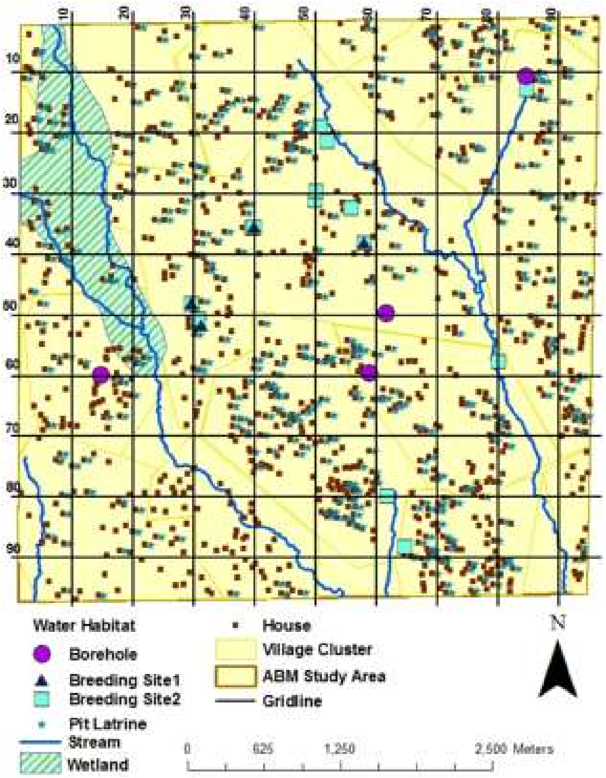


Figure 7: Study Area with Selected Data Layers and Gridlines for the Spatial ABM

4.6. Generating the Feature Counts for the ABM

From the GIS data layers described above (*i.e.*, houses and aquatic habitats), we generate the feature counts to use as inputs to our ABM. The feature counts, as shown in Table 1, appear as 1976 (982 aquatic habitats of different types, and 994 houses).

Table 1: GIS Feature Counts For The ABM

Feature Type	Feature Count
Type-1 breeding site	4
Type-2 breeding site	14
Borehole	4
Pit latrine	401
Wetland	559
House	994
Total	**1976**

5. CREATING SCENARIOS FOR THE ABM

To run simulations with the selected data layers, we create two hypothetical scenarios with different combined carrying capacities (*CCC*s, see Section 3.1).

We assign carrying capacities to the selected GIS layers that represent the aquatic habitats. However, since we cannot obtain absolute *CC* values for the habitats (due to the lack of habitat data), we assign relative *CC* values to the habitats based on the available spatial data, ensuring that the relative magnitudes of *CC*s are in accordance with: 1) the malaria vector productivity among distinct habitat types, and 2) the biological reality of the environment. For example, considering different cells in the spatial grid, a large breeding site cell would have higher *CC* than that of a wetland cell, although both cells represent the same surface area.

In terms of the magnitudes of the assigned *CC*s, we arbitrarily order the different aquatic habitat types in decreasing order of *CC* per cell: 1) type-1 breeding site, 2) type-2 breeding site, 3) borehole, 4) pit latrine, and 5) wetland. For wetland, which covers multiple cells in the northwest corner of the study area (see Figure 7), we assign the same *CC* value for each cell. In the future, when the data is available, the order, as well as the assigned *CC* values (to different aquatic habitat types), can be readily changed, and the ABM is ready to run with the newly assigned values.

To run the ABM with the selected data layers, we create two hypothetical scenarios with different *CCC*s by assigning relative *CC*s to the different aquatic habitat types, keeping the order of magnitudes intact. The *CCC*s for the scenarios appear as **21K** and **150K**, as shown in Tables 2 and 3.

Table 2: Dry Season Scenario **21K**

Feature Type	Feature Count	Assigned *CC*	Total
Type-1 breeding site	4	1000	4000
Type-2 breeding site	14	500	7000
Borehole	4	100	400
Pit latrine	401	10	4010
Wetland (each cell)	559	10	5590
		Total (*CCC*)	**21000**

Table 3: Rainy Season Scenario **150K**

Feature Type	**Feature Count**	**Assigned *CC***	**Total**
Type-1 breeding site	4	5000	20000
Type-2 breeding site	14	2000	28000
Borehole	4	1500	6000
Pit latrine	401	100	40100
Wetland (each cell)	559	100	55900
		Total (*CCC*)	**150000**

The two scenarios, depicted in Tables 2 and 3, may effectively represent two ecological settings with ***low*** (21K) and ***high*** (150K) potentials for mosquito populations, resembling the ***dry*** and ***rainy*** weather seasons, respectively, for the study area.

6. SIMULATIONS

We assume the following for the spatial ABM: the model starts with 1000 initial female adult mosquito agents (no male agents). All new agents (entering as eggs) are female. For each initial female agent, the state is set to *Gravid*, the age is set to 120 hours (for being in the *Gravid* state), and the agent is assigned to an aquatic habitat chosen at random. Since available field data points to limited flight ability and perceptual ranges of mosquitoes, the speed and range of movement (of mosquitoes) in our spatial ABM are controlled by special agent-level variables. Unlike other traditional malaria transmission models, we assume senescence (biological aging) of the mosquitoes, and the ABM implements age-specific mortality rates for the adult mosquitoes and the larvae (*i.e.*, the probability of death for mosquito agents increases with their age).

In order to seek for resources (aquatic habitats or houses), and hence to complete the gonotrophic cycles, the adult female mosquito agents move only while they are in *Bloodmeal Seeking* and *Gravid* states. At any point in the resource-seeking process, a mosquito's neighborhood is modeled as an eight-directional *Moore* neighborhood. The landscape is assumed to have a non-absorbing boundary, modeled topologically as 2D torus spaces). For details, see (Arifin, Davis, and Zhou 2011a; Arifin, Davis, and Zhou 2011b).

6.1. Spatial Variables

For the two scenarios (21K and 150K), the output of our spatial ABM includes four custom spatial variables:

1. *Adult abundance by location*: shows the distribution of the adult mosquitoes over the entire landscape at the end of the simulation.
2. *Cumulative biomass per aquatic habitat*: overlaid on top of the aquatic habitats, it represents the sum of biomass (eggs, larvae, and pupae) present in an aquatic habitat.
3. *Cumulative number of females oviposited per aquatic habitat*: also overlaid on top of the aquatic habitats, it represents the sum of the number of times female adults oviposited in the aquatic habitat.
4. *Cumulative number of bloodmeals per house*: overlayed on top of the houses, it represents the sum of the number of times female adults obtained bloodmeals in the house.

The last three spatial variables are sampled across all daily timesteps throughout the entire simulation. The output GIS maps, described in the next section, are produced by overlaying the above spatial variables on top of the relevant data layers. These variables allow us to analyze spatial correlations and find spatial patterns from the outputs of the ABM.

We use a special GIS map symbolizing technique known as ***graduated symbols***. Graduated symbols, used to compare quantitative values, vary in size according to the relative magnitudes of the values. In all output maps, we use graduated symbols as hollow circles, where the relative radii of the circles are determined by the output values generated by the ABM.

7. RESULTS

7.1. Mosquito Abundance (Non-spatial)

In the ABM, mosquito abundance depends, among other factors, on the carrying capacities of the aquatic habitats. Hence, not surprisingly, the 150K scenario yields much higher abundance than the 21K scenario, as shown in Figure 8. This also validates the abundance patterns usually observed in the *dry* and *rainy* seasons.

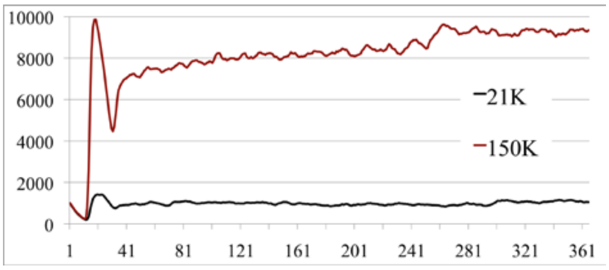


Figure 8: Mosquito Abundance (Non-spatial)

However, the output maps from all four spatial variables indicate that the *relative distances* between the aquatic habitats and the houses play a crucial role in determining the variables of interest, as shown in the following.

7.2. Adult Abundance by Location

Adult abundances by location are shown in Figures 9 and 10 for scenarios 21K and 150K, respectively. They indicate that higher abundances are associated with type-1 breeding sites, followed by type-2 breeding sites. Out of the four and 14 breeding sites (of type-1 and type-2, respectively), highest abundances are observed in locations where type-1 sites are in close proximity with type-2 sites, surrounded by human houses.

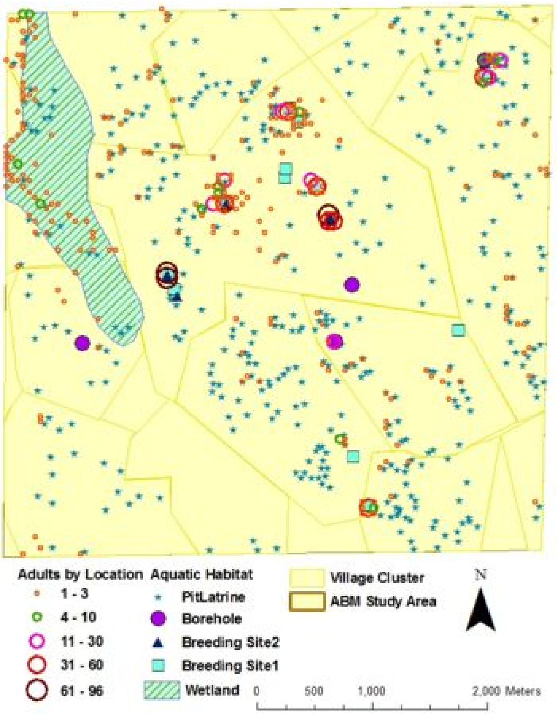


Figure 9: Adult Abundance by Location for 21K

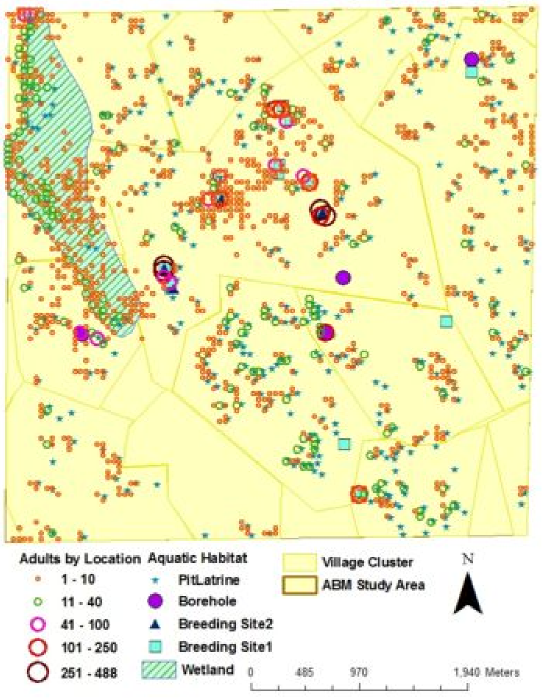


Figure 10: Adult Abundance by Location for 150K

We also observe very low (1-3 mosquitoes per cell) abundance in the wetland, which may be attributed to reduced human habitation around the wetland, and low carrying capacities associated with the wetland cells.

7.3. Cumulative Biomass and Females Oviposited

Figures 11 and 12 show the cumulative biomass *per aquatic habitat* for scenarios 21K and 150K, respectively.

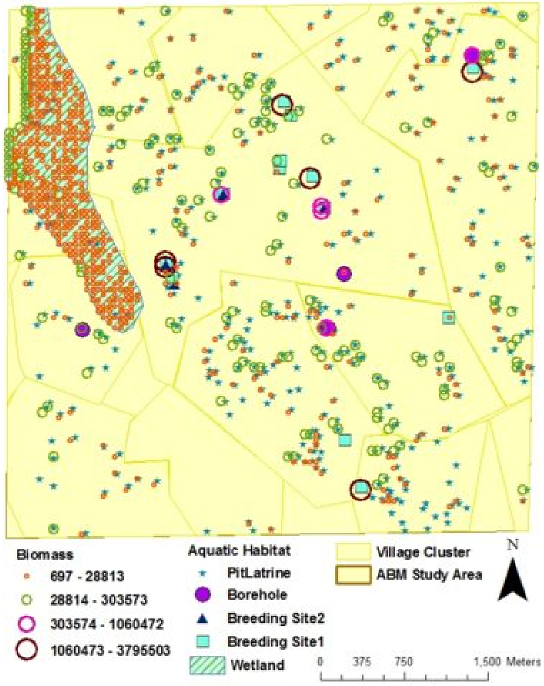


Figure 11: Cumulative Biomass for 21K

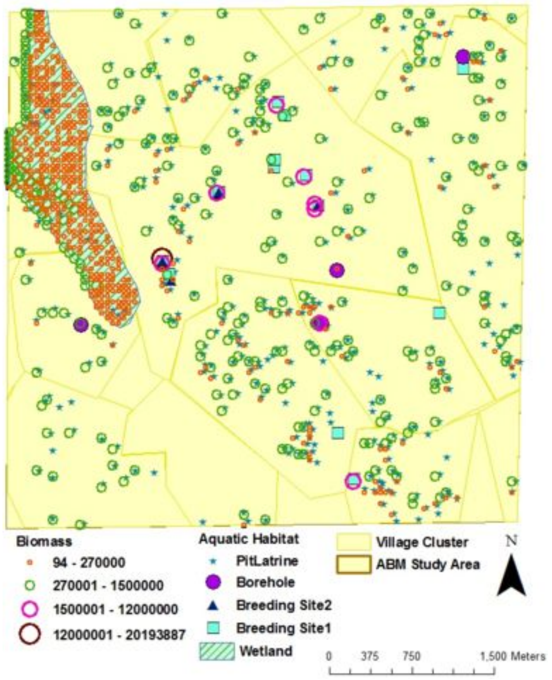


Figure 12: Cumulative Biomass for 150K

Figures 13 and 14 show the cumulative number of females oviposited *per aquatic habitat* for scenarios 21K and 150K, respectively.

Both of these metrics (Figures 11-14) show that higher abundances are associated with type-1 breeding sites, followed by type-2 sites, which are close to boreholes. However, an interesting insight reveals that two (out of 14) type-2 sites, suitable to yield high outputs (like other type-2 sites), yield only 0.07%-0.8% cumulative biomass, and only 0.005%-1% cumulative number of females oviposited, when compared to other type-2 sites.

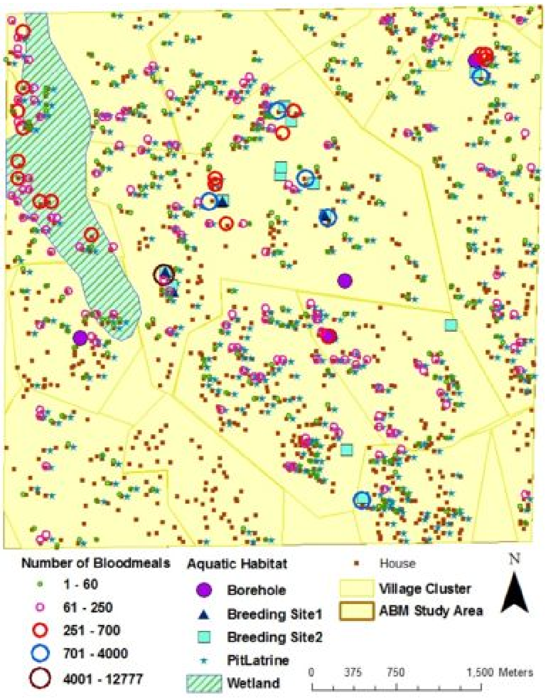


Figure 13: Cumulative Number of Females for 21K

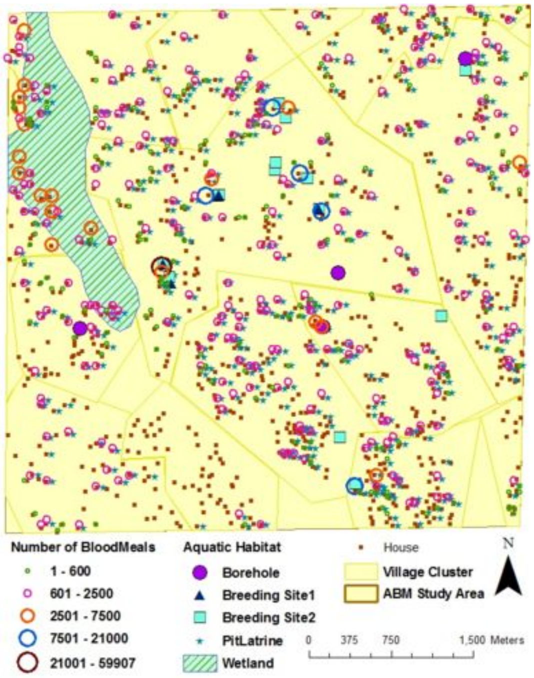


Figure 14: Cumulative Number of Females for 150K

Closer inspection of the corresponding output maps (Figures 11-14) reveals that the nearest human houses around these two type-2 breeding sites are situated much further than other type-2 sites. Since there are not enough houses in the close proximity, the female mosquitoes, ovipositing in these breeding sites, cannot find bloodmeals, and hence are forced to search longer distances. Since the mortality rate of mosquitoes increases with their age (recall that the ABM implements age-specific mortality rates that incorporate senescence, or biological aging), the additional delays in obtaining bloodmeals actually reduce abundance around these sites, causing much lower cumulative biomass and cumulative number of females oviposited.

7.4. Cumulative Number of Bloodmeals

Lastly, Figures 15 and 16 show the cumulative number of bloodmeals *per house* for scenarios 21K and 150K, respectively.

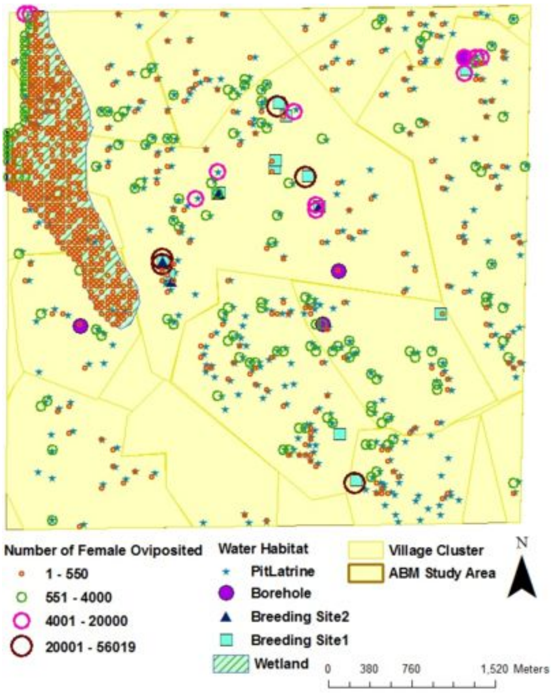


Figure 15: Cumulative Number of Bloodmeals for 21K

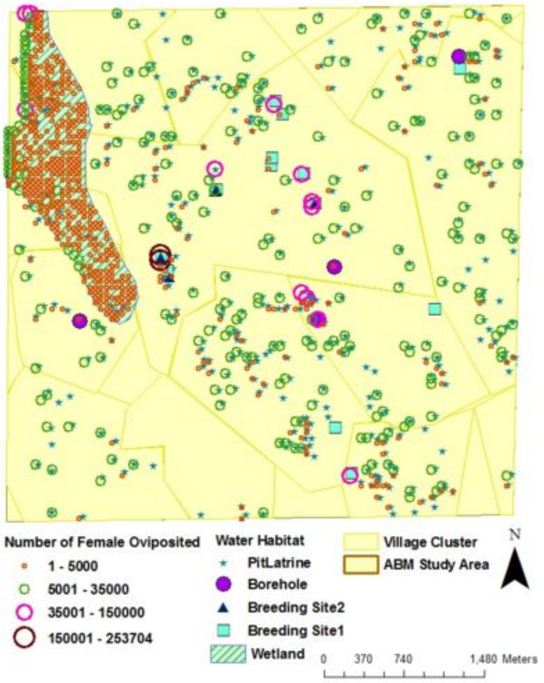


Figure 16: Cumulative Number of Bloodmeals for 150K

Figures 15 and 16 depict that higher values (number of bloodmeals) are associated with houses that have nearby type-1 and type-2 breeding sites, and moderate values are found in houses that have nearby aquatic habitats (of different types) with at least some carrying capacities. Interestingly, a large number of houses, located at the lower left quadrant of the study area, show no bloodmeals, due to the absence of aquatic habitats around them.

8. CONCLUSION

As our findings suggest, *availability* of the ecological resources, *i.e.*, the aquatic habitats and houses, and the *relative distances* between these distinct resource types, are two crucial determinants for the female mosquitoes to complete their gonotrophic cycles. From the viewpoint of mosquito agents, these resources directly define landscape features such as spatial heterogeneity, host availability, *etc*., the importance of which for vector control have been demonstrated by several studies. Reduced availability of either type of these spatial resources would prolong the gonotrophic cycle of the female mosquito, and potentially affect malaria transmission.

In our study, spatial analysis of the output variables generated by the ABM reveals important biological insights. The use of maps and spatial statistical methods allows readily identifying and displaying interesting spatial patterns, which, without the maps, are difficult to detect. The output maps also reveal potential ***hotspots*** with higher rates for the measured variables of interest.

The proposed robust integration framework also allows easy parameterization of the model. For example, the arbitrary order of the different aquatic habitat types, and the assigned *CC* per habitat, can be readily changed to suit new scenarios and/or new areas of study. This will allow the ABM to produce site-specific outputs (without the need of modifying the ABM itself). The simplicity in the scenario-based approach allows to feed in different scenarios to the ABM by using different *CCC*s for various aquatic habitat types, without requiring to change the data layers, features, *etc*., for future simulation runs.

Our results also indicate that disease-specific maps can play important roles in disease control activities, including monitoring the changes of malaria epidemiology, guiding resource allocation for malaria control, and identifying hotspots for further investigation. For example, the results highlight the importance of eliminating the aquatic habitats close to human habitations by means of environmental modifications and manipulations, supporting the arguments presented by several malaria control studies (*e.g.*, Fillinger and Lindsay 2011).

Although in this pilot study we handled a comparatively small study area of ≈ 25 km^2 (which transforms to a *95 columns * 96 rows* landscape for the spatial ABM), the methodology described here can be readily extended to include larger areas (*e.g.*, the whole Asembo area). For the new regions to be modeled, either real data can be used, or synthetic/predicted data can be interpolated from a few point regions (on which the described methodology is applied first).

We conclude that such integrated approaches, which combine knowledge from entomological, epidemiological, simulation-based, and geo-spatial domains, are required for the identification and analysis of relationships between various transmission variables, as demonstrated by our study. Eventually, such integration efforts may facilitate the *Integrated Vector Management (IVM)* agenda, promoted by the World Health Organization (WHO), to achieve improved efficacy, cost-effectiveness, ecological soundness, and sustainability of malaria vector control.

ACKNOWLEDGMENTS

Partial support for this work came from the Bill & Melinda Gates Foundation under the Vector Ecology and Control Network (VECNet) Project. We also wish to acknowledge helpful comments from Dr. Frank H. Collins and Dr. Neil F. Lobo at the University of Notre Dame.

REFERENCES

ArcGIS Desktop (Release 9.3) [Computer software]. (2011). Environmental Systems Research Institute (ESRI). Available from: http://www.esri.com/

Arifin, S. M. N., Davis, G. J., and Zhou, Y., 2010a. Verification & Validation by Docking: A Case Study of Agent-Based Models of *Anopheles gambiae*. *Summer Computer Simulation Conference (SCSC)*, 236-243, Jul. 2010, Ottawa, ON, Canada.

Arifin, S. M. N., Davis, G. J., and Zhou, Y., 2011a. A Spatial Agent-Based Model of Malaria: Model Verification and Effects of Spatial Heterogeneity. *International Journal of Agent Technologies and Systems*, 3(3):17–34.

Arifin, S. M. N., Davis, G. J., and Zhou, Y., 2011b. Modeling Space in an Agent-Based Model of Malaria: Comparison between Non-Spatial and Spatial Models. *Proceedings of the 2011 Workshop on Agent-Directed Simulation*, 92–99, Apr. 2011, Boston, Massachusetts.

Arifin, S. M. N., Davis, G. J., Kurtz, S. J., Gentile, J. E., and Zhou, Y., 2010b. Divide and Conquer: A Four-fold Docking Experience of Agent-based Models. *Winter Simulation Conference (WSC)*, 575 - 586, Dec. 2010, Baltimore, Maryland, USA.

Brown, D. G., Riolo, R., Robinson, D. T., North, M., and Rand, W., 2005. Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems*, 7:25-47.

Fillinger, U. and Lindsay, S., 2011. Larval source management for malaria control in Africa: myths and reality. *Malaria Journal*, 10:353.

Gimnig, J. E., Hightower, A. W., and Hawley, W. A., 2005. Application of geographic information systems to the study of the ecology of mosquitoes and mosquito-borne diseases. In: *Wageningen UR Frontis Series*, 9, 27-39.

Hawth's Analysis Tools for ArcGIS [Website]. (2013). Available from: http://www.spatialecology.com/

Hay, S. I. and Snow, R. W., 2006. The Malaria Atlas Project: Developing Global Maps of Malaria Risk. *PLoS Med*, 3(12).

Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., … Snow, R. W., 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med*, 6(3).

Kazura, J. W., ed. *The American Journal of Tropical Medicine and Hygiene*, (2003). 68(4 suppl).

Khormi, H. M. and Kumar, L. 2011. Examples of using spatial information technologies for mapping and modelling mosquito-borne diseases based on environmental, climatic and socio-economic factors and different spatial statistics, temporal risk indices and spatial analysis: A review. *Journal of Food, Agriculture & Environment*, 9(2):41-49.

Li, L., Bian, L., and Yan, G. 2006. An Integrated Bayesian Modelling Approach for Predicting Mosquito Larval Habitats. In: *UCGIS 2006 Summer Assembly*.

Mbogo, C. M., Mwangangi, J. M., Nzovu, J., Gu, W., Yan, G., Gunter, J. T., … Beier, J. C., 2003. Spatial and temporal heterogeneity of *Anopheles* mosquitoes and *Plasmodium falciparum* transmission along the Kenyan coast. *The American Journal of Tropical Medicine and Hygiene*, 68(6):734-742.

McElroy, P. D., ter Kuile, F. O., Hightower, A. W., Hawley, W. A., Phillips-Howard, P. A., Oloo, A. J., … Nahlen, B. L., 2001. All-cause mortality among young children in western Kenya. VI: the Asembo Bay Cohort Project. *The American Journal of Tropical Medicine and Hygiene*, 64:18-27.

Mmbando, B., Kamugisha, M., Lusingu, J., Francis, F., Ishengoma, D., Theander, T., ... Scheike, T., 2011. Spatial variation and socio-economic determinants of plasmodium falciparum infection in northeastern Tanzania. *Malaria Journal*, 10(1):145.

Nahlen, B. L., Clark, J. P., and Alnwick, D., 2003. Insecticide-treated bed nets. *The American Journal of Tropical Medicine and Hygiene*, 68:1-2.

Ndenga, B. A., Simbauni, J. A., Mbugi, J. P., Githeko, A. K., and Fillinger, U., 2011. Productivity of Malaria Vectors from Different Habitat Types in the Western Kenya Highlands. *PLoS ONE*, 6(4).

Phillips-Howard, P. A., Nahlen, B. L., Alaii, J. A., ter Kuile, F. O., Gimnig, J. E., Terlouw, D. J., ... Hawley, W. A., 2003. The efficacy of permethrin-treated bed nets on child mortality and morbidity in western Kenya I. Development of infrastructure and description of study site. *The American Journal of Tropical Medicine and Hygiene*, 68:3-9.

Vector Ecology and Control Network (VECNet) [Website]. (2013). Available from: http://www.vecnet.org/

Zhou, G., Munga, S., Minakawa, N., Githeko, A. K., and Yan, G., 2007. Spatial Relationship Between Adult Malaria Vector Abundance and Environmental Factors in Western Kenya Highlands. *The American Journal of Tropical Medicine and Hygiene*, 77(1):29-35.

Zhou, Y., Arifin, S. M. N., Gentile, J. E., Kurtz, S. J., Davis, G. J., and Wendelberger, B. A., 2010. An Agent-based Model of the *Anopheles gambiae* Mosquito Life Cycle. *Summer Computer Simulation Conference (SCSC)*, 201-208, Jul. 2010, Ottawa, ON, Canada.

AUTHORS BIOGRAPHY

S. M. Niaz Arifin is a PhD student in the Department of Computer Science and Engineering at the University of Notre Dame. His research interests include Data Warehousing, Agent-Based Modeling & Simulation, Geographic Information Systems, *etc*. He received his MS from the University of Texas at Dallas in 2006, and BS from Bangladesh University of Engineering and Technology (BUET) in 2004. He served as a software developer at Xcision Medical Systems, California and the Rails Online Database project at Sabre Holdings Corporation, Texas. **Rumana Reaz Arifin** is a PhD student in the Department of Civil & Environmental Engineering and Earth Sciences at the University of Notre Dame. Her research interests include Environmental Engineering and Geographic Information Science. She received her MSs in Civil Engineering (Notre Dame) and Geographic Information Science (University of Texas Dallas). She provides GIS support for Center for Research Computing (CRC) at Notre Dame. **Dilkushi de Alwis Pitts** is an Adjunct Research Assistant Professor at the University of Notre Dame. She earned a PhD in Environmental Engineering (specializing in Hydrology) at Cornell University, NY and a MS in Remote Sensing at the Rochester Institute of Technology, NY. She has nearly 17 years of experience assimilating spatial and attribute data at different spatial and temporal scales and up/down-scaling spatial data for integration into mathematical models and validation of model output. **Gregory R. Madey** is a Research Professor in the Department of Computer Science and Engineering at the University of Notre Dame. His research interests include Agent-Based Modeling & Simulation, Cyberinfrastructure, Open Source Software, Data Warehousing, Emergency Operations Management, Bioinformatics, Modeling Epidemiology, and Health Informatics. His recent research grants include a VECNet Cyberinfrastructure grant from the Bill & Melinda Gates Foundation, and Open Sourcing of the Civil Infrastructure Design grant from the US National Science Foundation.

AUTOMATIZATION OF A PROCESS OF AN INUNDATION AREA COMPUTATION

Pavel Veselý [(a)], Štěpán Kuchař [(b)], Jan Martinovič [(c)], Vít Vondrák [(d)]

[(a)] VŠB - Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[(b)] [(c)] [(d)] VŠB - Technical University of Ostrava
IT4Innovations
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[(a)] pavel.vesely@vsb.cz, [(b)] stepan.kuchar@vsb.cz, [(c)] jan.martinovic@vsb.cz, [(d)] vit.vondrak@vsb.cz

ABSTRACT

A FLOREON+ (FLOods REcognition on the Net) system for floods prediction and inundation areas computation has been developed to help with an operative disaster management and decision support. This article presents automated process of inundation areas computation based on open standards, unified process description and standardized internal component communication. It describes a case study and experimental use of business process description language such as BPEL in geoinformatic environment.

The system has been designed based on a service-oriented architecture (SOA). It contains a set of independent web services, orchestrated by business process interpreter application. External access to the process is clearly standardized implementing OGC WMS/WPS specifications, as well as communication between web services.

1. INTRODUCTION

FLOREON+ (FLOods REcognition on the Net) system (Vondrák, Martinovič, Kožusznik, Štolfa, Kozubek, Kubíček, Vondrák and Unucka 2008; Martinovič, Štolfa, Kožusnik, Unucka and Vodnrák 2008; Unucka, Martinovič, Vondrák, and Rapant 2009) for floods prediction and inundation areas computation has been developed to help with an operative disaster management and decision support. Main goal of our work is to develop versatile methodology for automation of FLOREON+ processes. It has to include system design, usage of language for business process description and execution, communication standards and selection of basic technology platform (Enterprise Service Bus or a similar framework for the process execution). Independency of system's components must be preserved to achieve maximum flexibility of the design and make the components reusable for different processes.

2. RELATED WORK

One of the first works, that shows GIS as several independent cooperating services, is stated in (Alameh, 2003). Implementation of SOA in GIS is described by several projects. They are mostly created as an implementation of INSPIRE architecture, such as in (Friis-Christensen, Bernard, Kanellopoulos, Nogueras-Iso, Peedell, Schade and Thorne, C. 2006). It shows SOA in a prototype of GIS for forest fire assessment. The prototype demonstrates solution of real problems using rapid development of a distributed application, which is facilitated by Spatial Data Infrastructure (SDI) as a basis for distributed service oriented geoprocessing.

An article by Andreja Jonoski (Jonoski, 2012) describes current trends in the field of hydroinformatics. Standardization of communication between web applications is a basic challenge. Therefore several common used formats exist for web services communications and spatial data exchange, especially Open Geospatial Consortium (OGC) formats (WMS, WFS, GML) and its derivates, such as WaterML. The next topic of the current trends is Spatial Data Infrastructure (SDI) data sources. The most important SDIs are National Spatial Data Infrastructure (NSDI) in USA, INSPIRE in EU or United Nation's UNSDI. Using standards enable sharing datasources and making distributed calculations. The article provides an example of a web based application DIANE-CM. It benefits from cooperation of independent web services and spatial data sources for flood risk management.

Using geospatial services in environmental systems describes (Granell, Díaz, Gould, 2010) using SOA applied to alpine runoff models. The application includes usage of geospatial services facilitating discovery, access, processing and visualization of geospatial data in a distributed manner.

Communication between geospatial services must be standardized and unified. Project MEDSI (Rocha, Cestnik and Oliveira 2005) shows example of implementation of OGC standards in a crisis management GIS. The project verified implementation of a prototype based on OGC standards. The implementation was deployed on several software platforms, including open source tools UMN MapServer, GeoServer, as well as commercial Geomedia Web Server and ESRI ArcIMS. The prototype was completed by catalogue services in accordance with a standard Catalogue Service for Web (CSV). It allows searching datasources by its metainformation.

Next work (Castronova, Goodall, Elag, 2012) advances an idea of interoperability between multiple independent spatial web services by presenting a design for a web service, which is build in accordance to the OGC Web Processing Service (WPS) protocol. The WPS is used as an interface between hydromodeler environment based on Open Modelling Interface (OpenMI) standard and client-side workflow.

One of the big challenges in environmental GIS using SDI is integration of multiple geodata sets and takeover geodata from different datasources. (Foerster, Lehto, Sarjakoski, Sarjakoski, Stoter, 2009) presents a process for geodata generalization and schema transformation in a web service's architecture to achieve interoperability between different geodata.

The final step to make our computation process fully automated is an orchestration of web services. Research project Orchestration Services for GeoWeb (Prager, Klímek, Růžička, 2009) describes standards in a field of service's cooperation, such as Business Process Execution Language (BPEL), Ontology Web Language for Services (OWL-S) or XML Process Definition Language (XPDL). It shows practical implementation of BPEL using Oracle BPEL process management and cooperation of geoservices in a workflow on Oracle's Business Process Management platform.

Calling geoservices in a workflow we have to take into account asynchronous communication (Zhao, Peisheng, Di, Liping, Yu, Geong, 2012). Asynchronous workflow can be distinguished at calling asynchronous web services in a workflow or calling a workflow as an asynchronous service. The article (Zhao, Peisheng, Di, Liping, Yu, Geong, 2012) describes several case studies. They have some common features: calling asynchronous services, using OGC standards (WPS, WCS, GML) and using WS-BPEL to describe workflow itself.

Our work shows an application of standards and ESB technologies in a real-world geoinformation system. The work links geoinformation services (in accordance to OGC standards) to a workflow in an open-source ESB platform OpenESB. We concentrate on practical usage of ESB technologies such as BPEL in hydrologic simulations. As opposed to the projects we mentioned before we add ESB and workflow technologies to GIS. Our application shows an importance of the transformation of geodata, when the data are moved between the hydrological models. The transformation to common standards is very important to achieve independence of the web services, which gains an access to the models.

3. METHODOLOGY

The first step is to describe the process of flood prediction using its graphical representation in Business Process Model and Notation (BPMN) (Object Management Group 2008) (see Figure 1). This human-readable form shows an overview of execution flows and data transmission between components of the system.

In order to execute actions within business processes with web services, the process has to be described using some process execution language as Web Services Business Process Execution Language (WSBPEL) (OASIS 2007) or similar language such as Microsoft's Workflow Foundation XAML (Microsoft MSDN 2013). Transformation from BPML to the execution language depends on the used. It is described later in this paper.

Each web service involved in the process must be described by Web Services Description Language (WSDL) (W3C, 2001). This is an XML-based interface description language that is used for describing the functionality offered by a web service. The WSDL document for a web service of the process is used in combination with XML schema description of a Web Processing Service (WPS) standard.

The most important goal is to standardize external interfaces of the process and internal service communication. To ensure independence of each component inside the process and its integration with other processes, common communication standards have to be used. Several specialized standards for exchange of hydrologic data or communication with web services exist (Vitolo, Buytaert, Reuser 2012). Hydrologic data can be stored in a common geoformat GML/KML or its specialized derivates WaterML and UncertML. Standards OGC WPS, WMS or WFS (Open GIS Consortium 1998) can be used to assemble a query to a geoinformation web service, including a spatial query. Web Processing Service (WPS) standard is used to execute a spatial function on a web service. Caller of the process can request a visualized map and set its parameters using the Web Map Service (WMS) standard.

4. SOFTWARE ENVIROMENT

Open Enterprise Service Bus (OpenESB) tool have been selected as a platform for creating workflow and executing BPEL (OpenESB, 2012). This tool has several advantages:

- Open source software with strong community support without any additional costs,
- Comfortable and powerful user interface based on NetBeans IDE tool for rapid development of services and workflows,

- Because the platform is Java-based, the solution can be deployed to different operating systems without changes.
- Solutions created on OpenESB platform are reliable, strong and scalable, appropriate to create crisis management applications.

Final application services were deployed on Oracle GlassFish server. It is a reference implementation of Java EE platform sponsored by Oracle Corporation. The OpenESB is strongly integrated with GlassFish and deployment of the services is completely automated.

5. BUILDING THE PROCESS

The process goes through several tasks during its execution:

- Data transformation.
- Rainfall-Runoff simulation.
- Hydrodynamic simulation.
- Data transfer.

Figure 1 shows positions of each task in the process. Some of them, such as data transformation, are executed repeatedly in different places and with different input/output arguments.

Before each step, data should be transformed into a format that is used by the processing component/model. Data in the process has spatial information, so data transformation is not changing only the format of the data inside storage, but a conversion of spatial coordinates may be needed, such as conversion between different spatial projection systems. Results are converted to some of the standardized or commonly used formats (GML, Shapefile) or into a geodatabase storage using a geointerface, such as PostGIS.

Rainfall-Runoff simulation component involves four rainfall-runoff models: HEC-HMS (HEC-USACE 2010), HYDROG (HySoft 2010), MikeSHE (DHI 2011a) and our own in-house model called Math1D. All of them can be accessed using one web service. Caller determines the model to perform the simulation by a specialized parameters set in the input of the process. Different models need different data input formats and so the data conversion has to be executed before passing inputs into the selected model.

Results of the Math1D model can be statistically evaluated by modelling the uncertainty of its input parameters. The Monte-Carlo simulation method is used for estimating possible river discharge volumes based on the uncertainty of precipitation and meteorology forecast and provides several confidence intervals that can support the decisions in the operational disaster management (Kuchař, Kocyan, Praks, Litschmannová, Martinovič, Vondrák 2012).

The hydrodynamic modelling phase is executed after the rainfall-runoff results are acquired from the models in the first step. HEC-RAS (HEC-USACE 2010) and MIKE 11 (DHI 2011b) hydrodynamic models are used in the FLOREON+ system. Selection of the computation model is made during the process execution based on the parameter set at the start of the process. This parameter also governs the data format for data conversion before execution of the selected model.

The main idea of data transfer is not to transmit all the data between web services, but send and retrieve only links to the data stored in a place that is accessible by both sides of the web service call. This kind of transitions is fully accepted by the OGC WPS standard.

Final data are converted to OGC GML format and stored in PostgreSQL database through PostGIS interface or visualized according to OGC WMS request and sent back to the caller.

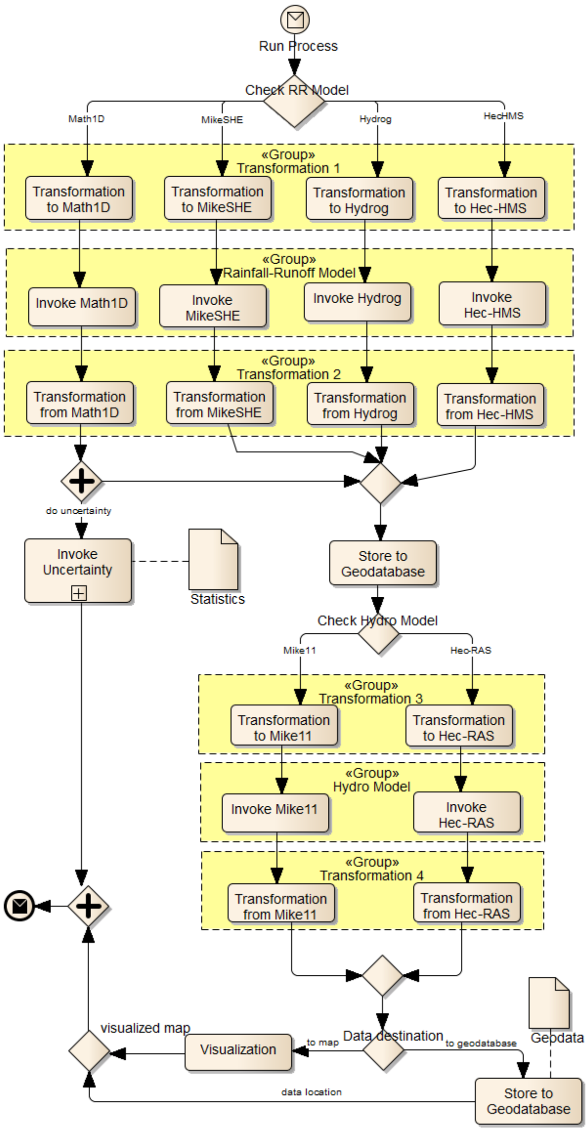


Figure 1: Process description using BPMN notation

6. CREATING WSDL FILES

Because the process will be published as a web service, it needs a Web Services Description Language (WSDL) document. The service will be standardized in accordance with OGC WPS standard. The format of the communication with the service is described in a schema descriptor (XSD) file on a web page http://schemas.opengis.net. This schema is set as a type in definition of the WSDL document. The WSDL itself was created using OpenESB's NetBeans IDE. WSDLs of the services of rainfall-runoff and hydrological models describe input parameters of the models and a format of their responses.

7. A USE OF THE WPS STANDARD

The WPS standard has been used, because it has two main advantages:

- It is supported by OGC, the leading international industry consortium in a field of geo-web industry standardization.
- It has universal design with ability to carry information with different types and structures.

In a BPEL process description we use "assign" BPEL element to map WPS XML nodes and attributes to model service's properties. The WPS contains "Input" nodes with identifier and data. A BPEL "predicate" node that consists XPath functions were created in BPEL process. A predicate applies a condition to a node that can have multiple values. Values of selected nodes are assigned to web service's parameters. Also some transformations from string type to number were done during assignment.

WPS request can contain an envelope with coordinates of an area of interest. This envelope is used to draw a final map with floodlakes.

8. PROCESS DESCRIPTION IN BPEL

Creating BPEL document on the basis of BPMN diagram depends on OpenESB tools for designing of the process. This is how BPMN objects correspond with OpenESB notation:

- BPMN start and end event are equal to receive and reply activity,
- BPMN activities are mostly created as invoke web service activity,
- BPMN connections are created as assign activity,
- BPMN gateways are equal to their appropriate structured activities.

We created BPEL document in NetBeans IDE. Than we added actions into the process, started from receive activity, web services invocation and assign activities, up to reply action. Partner links where created from receive and reply action using main service WSDL. It represents request/response of the web service of the process. Next partner links where created from each invocation activity to appropriate web services with service's WSDL file. Assign activity where set to connect input and output parameters of the web services and/or the receive/reply actions. We created simplified version of the process as a sample of the OpenESB BPEL diagram and for testing purposes. This simplified process shows Figure 2. A partner link to a WSDL of the process is on the left side of the figure. Next partner links on the right side shows connection to WSDL of the model services.

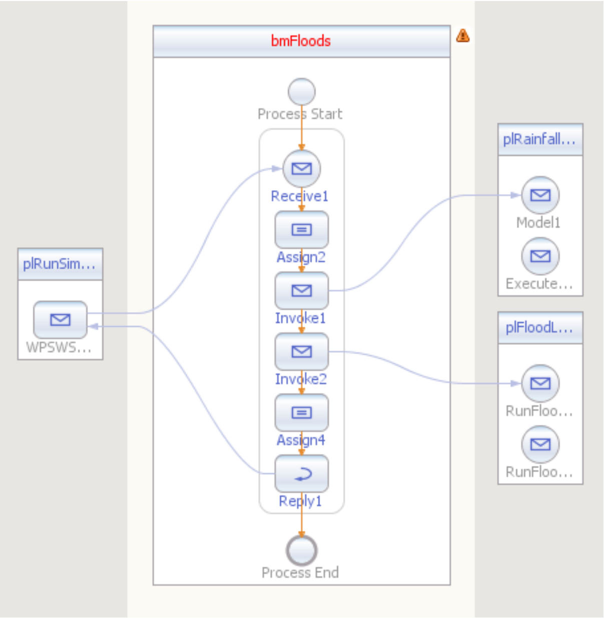


Figure 2: Simplified BPEL process

9. CREATING SERVICE ASSEMBLY

The service assembly (calling Composite Application in OpenESB speak) is a group of service units gathered together to create single application. The service assembly includes metadata for "wiring" the service units together (associating service providers and consumers), as well as wiring service units to external services. This provides a simple mechanism for performing composite application assembly using services (Java Community Process, 2013). We used NetBeans IDE provided with OpenESB to create composite application. Figure 3 shows diagram of connections from input request ports (HTTP and SOAP) to the BPEL process and from the process to external services of rainfall-runoff and hydrological models.

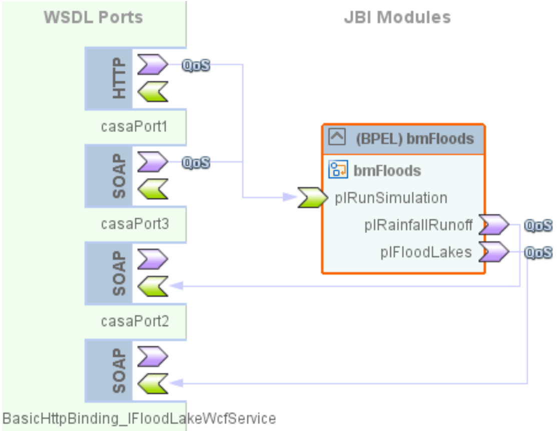


Figure 3: Diagram of composite application

10. EXPERIMENTAL RUN OF THE APPLICATION

We run application several times using a WPS request with basic parameters passed to the process. All parameters where passed using common WPS node "Input". Parameter's names where set in a subnode "Identifier" and values in a subnode "DataLiteral". First parameter "modelId" commands the process to use specific model. In our first test case it was HEC-HMS model for rainfall-runoff and hydrodynamics. Second test case we do the same with Math1D model. Parameter "basinSchemaId" sets preset schematization of the river, out tests used river Olše. "startTime" and "endTime" parameters set time period for data computation. Parameters from input request were transformed to rainfall-runoff and hydrodynamic web service's form during an execution of BPEL commands.

Table 1 shows computing times of rainfall-runoff service for different time periods. It shows that the computing time is very weakly dependent on the selected time period. Computing times in rainfall-runoff are much shorter than computing in hydrodynamic services. They are shown in table 2. Selected time period in a hydrodymanic significantly rises computing time up to several hours.

Table 1: Rainfall-runoff services run times

Rainfall-runoff services		
Time period	HEC-HMS	Math1D
48 hours	29.4 s	20 s
72 hours	29.7 s	21 s
96 hours	30.2 s	23.4 s

Table 2: Hydrodynamic services run times

Hydrodynamic services		
Time period	HEC-HMS	Math1D
6 hours	1597 s	1684.1 s
12 hours	2338.5 s	2651.1 s
24 hours	4325.2 s	4654 s

The whole process will be executed automatically in a selected time interval or casually by user's demands. Data sources of the models are refreshed by more precise values in a six hours interval. The same interval was set for automatic process execution. This execution can be made for a long time period 24 hour or even 48 hour. The time periods for casual execution of the process which will be run by user's request shouldn't be longer than 6 hours.

11. CONCLUSION

This paper describes automation of a computation of flood lakes. The process is built on enterprise service bus platform using BPEL to achieve flexibility and reliability. User or caller of the process can choose computational rainfall-runoff model and hydrodynamic model just by setting correct input parameters. Thereafter each step in the process is done automatically. The process is posted as a web service in accordance with OGC WPS standard. It allows integration of the service to complex geographic information systems in a further work. Each web service created by this work is reusable and can be accessed from other processes, which will be created in next work.

12. FUTURE WORK

We want to create deeply standardized and universal system. Each web service, which is applied in the system, can be in accordance with OGC standards. Data transformation wasn't fully tested and it can't be presented in the paper. In next work we will finish transformation services to make each part of the system fully independent.

Input data, which are used inside the models for computation, will be in a future automatically extracted from different data sources. This extraction can start the process of inundation areas computation or it can be a part of the process.

ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

REFERENCES

Alameh, N., 2003, Chaining geographic information Web services, *IEEE Internet Computing*, vol. 7 (5): 22-29.

Castronova A. M., Goodall J. L., Elag M. M., 2012. Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard, *Environmental Modelling & Software*, pp. 72-83, Volume 41, March 2013.

DHI, 2011. *MIKE SHE - integrated catchment modelling*. Available from: http://www.dhisoftware.com/Products/WaterResources/MIKESHE.aspx [10th May 2013].

DHI, 2011. *MIKE 11 - river modelling unlimited*. Available from: http://www.dhisoftware.com/Products/WaterResources/MIKE11.aspx [10th May 2013].

Friis-Christensen, A., Bernard, L., Kanellopoulos, I., Nogueras-Iso, J., Peedell, S., Schade, S., & Thorne, C., 2006. Building service oriented applications on top of a spatial data infrastructure– a forest fire assessment example. *AGILE-Shaping the future of Geographic Information Science in Europe*, 2006, Visegrád, Hungary.

Foerster T., Lehto L., Sarjakoski T., Sarjakoski L. T., Stoter L., 2012, Map generalization and schema transformation of geospatial data combined in a Web Service context, *Computers, Environment and Urban Systems*, pp. 79-88, Volume 34, Issue 1, January 2010.

Granell C., Díaz L., Gould M., 2010, Service-oriented applications for environmental models: Reusable geospatial services, *Environmental Modelling & Software*, pp. 182-198, February 2010.

HEC-USACE, 2010. *Hydrologic Engineering Center – US Army Corps of Engineers*. Available from: http://www.hec.usace.army.mil/ [10th May 2013] .

HySoft, 2010. *HySoft.* Available from: http://www.hysoft.cz/ [10th May 2013].

Java Community Process, *JSR-000208 Java Business Integration*, 2013, Available from: http://jcp.org/aboutJava/communityprocess/final/jsr208/index.html [9. June 2013].

Jonoski, A., 2012, *Hydroinformatics and Decision Support: Current Technological Trends and Future Prospects*. BALWOIS 2012.

Kuchař, Š., Kocyan, T., Praks, P., Litschmannová, M., Martinovič, J., Vondrák, V., 2012. Simulation of Uncertainty in Rainfall-Runoff Models and their Statistical Evaluation in the Floreon System. *The 11th International Conference on Modeling and Applied Simulation 2012*, pp.128-133. 19-21 September, Vienna, Austria.

Martinovič, J., Štolfa, S., Kožusnik, J., Unucka, J., and Vodnrák, I., 2008, FLOREON - the system for an emergent flood prediction, *ECEC-FUBUTEC-EUROMEDIA*, Porto, Portugal.

Microsoft, MSDN, 2013. *XAML Overview (WPF).* Available from: http://msdn.microsoft.com/en-us/library/ms752059.aspx [10th May 2013].

Object Management Group, 2008. *Business Process Modeling Notation*, V1.1. Available from: http://www.omg.org/bpmn/Documents/BPMN_1-1_Specification.pdf [10th May 2013].

Open GIS Consortium, 1998. *OGC® Standards and Specifications.* Available from: http://www.opengis.org/public/abstract.html [10th May 2013].

OpenESB, 2012. *The Open Enterprise Service Bus.* Available from: http://www.open-esb.net [5th June 2013].

OASIS, 2007. *Web Services Business Process Execution Language Version 2.0*. Available from: http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html [10th May 2013].

Prager, M., Klímek, F., Růžička, J.. 2009, *GeoWeb Services Orchestration Based on BPEL or BPMN*, GIS Ostrava 2009.

Rocha, A., Cestnik, B., Oliveira, M. A., 2005. *Interoperable geographic information services to support crisis management*. Berlin Heidelberg: Springer.

Unucka, J., Martinovic, J., Vondrak, I., Rapant, P., & Brebbia, C. A., 2009. Overview of the complex and modular system FLOREON+ for hydrologic and environmental modeling. *In Proceedings of the 5th International Conference on River Basin Management*, Malta, 2009. pp. 207-216. WIT Press.

Vitolo, C., Buytaert, W., Reuser, D. E., 2012. Hydrological Models as Web Services: An Implementation using OGC Standards. *HIC- 10th International Conference on Hydroinformatics*, Hamburg, Germany.

Vondrák, I., Martinovič, J., Kožusznik, J., Štolfa, S., Kozubek, T., Kubíček, P., Vondrák, V. , Unucka, J., 2008. A description of a highly modular system for the emergent flood prediction, *Computer Information Systems and Industrial Management Applications, 2008, CISIM '08.* 7th.

W3C, 2001, Web Services Description Language (WSDL) 1.1. Available from: http://www.w3.org/TR/wsdl [4th June 2013].

Zhao, P., Di, L., & Yu, G., 2012. Building asynchronous geospatial processing workflows with web services. *Computers & Geosciences*, 39, 34-41.

TECHNICAL AND ECONOMIC VERIFICATION OF THE CONVENIENCE IN REENGINEERING A PRODUCTION LINE USING SIMULATION TECHNIQUES

Domenico Falcone[(a)], Antonio Forcina[(b)], GianPaolo Di Bona[(c)], Vincenzo Duraccio[(d)],
Alessandro Silvestri[(e)], Cristina Cerbaso[(f)]

[(a) (b) (c) (d) (e) (f)] Department of Civil and Industrial Engineering
University of Cassino and Lazio Meridionale
03043 Cassino (FR) – Italy
tel. 0039-776-299635 - Fax 0039-776-310812

[(a)] falcone@unicas.it, [(b)]a.forcina@unicas.it, [(c)]dibona@unicas.it, [(d)]duraccio@unicas.it,
[(e)]silvestr@unicas.it, [(f)]c.cerbaso@unicas.it

ABSTRACT

The work shows some proposals for achieving the production flow optimization of an engineering company production lines, operating in the automotive sector. Moving from the analysis of the actual line production efficiency, either by simulative techniques, either by technical-economic analyses, some improvement actions have been proposed and validated. The results obtained, referring to production capacity and equipment state, have point out that the proposed solution permits both a productive flow optimisation and a productiveness increase.

Keywords: modelling, validation, productiveness, bottle neck

1. CASE STUDY

The company is a leading global supplier of bearings, seals, mechatronics, lubrication systems and services which include technical support, maintenance and reliability services, engineering consulting and training.

It is a global company, established in Europe, North and Latin America, Asia and Africa. Today, it is represented in more than 130 countries. The company has more than 100 manufacturing sites and also sales companies supported by about 15,000 distributor locations; a widely used e-business marketplace and an efficient global distribution system.

The company works mainly through three business areas: Strategic Industries and Regional Sales and Service, servicing industrial original equipment manufacturers and aftermarket customers respectively, and Automotive, servicing automotive producers and aftermarket customers. It operates in around 40 customer segments, whereof examples include cars and light trucks, wind energy, railway, machine tool, medical, food and beverage and paper industries.

Technical development, quality and marketing have been strongly in focus since the beginning. The Group's efforts in research and development have resulted in numerous innovations, forming bases for new standards, products and solutions in the bearing world. Due to a reorganization of the establishments in Europe, some plants have started producing a new type of ball bearing, such as the studied one.

The bearing actual production capacity was much lower than the theoretical. It has been necessary, therefore, the identification of production process improvements, to be adopted quickly and with no waste of efforts. The use of simulation techniques allowed obtaining the above goals.

2. DESCRIPTION OF THE LINE

Inside the considered plant, the production is organized by production channels, ie small units including all operations, machines and resources required for bearing production, starting from raw materials or semi-finished products to obtain the finished product.

The whole manufacturing process consists of six main phases: Moulding, Turning, Heat treatment, Facing, Grinding, Assembly.

The object of the study, channel 9, on which it carries out grinding, lapping and assembly, consists of two branches developing longitudinally and parallel to each other to meet in assembly (Figure 1).

On the left line grinding machines and monitoring devices for inner ring are located, while, on the right line, the machines for outer ring processing are placed.

At the top of the line, all the rings are subjected to a 100% dimensional control through head-line control devices, in order to verify the actual match of measurement and processing schedule.

On the inner ring the following operations are performed:

- *facing*,
- *face grinding* (material is removed from the faces of the ring),

- *throat and flange grinding*,
- *hole grinding*,
- *throat lapping* (carried out to achieve high surface finish).

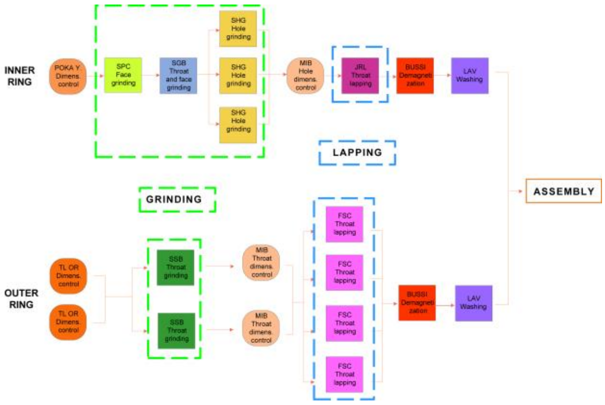


Figure 1: Channel 9: grinding and lapping

For the latter operation, they don't use grinding wheels, but shaped and with a decreasing particle size stones.

On the outer ring, instead, throat grinding and lapping are performed. Following lapping, all the rings pass inside demagnetizers with the function of removing dirt and metal particles residues from the rings.

Subsequently, they pass through special washing machines, in order to eliminate any further processing residues, harmful for the bearing.

IR and OR rings, previously machined on the grinding lines, are placed in the assembly line to be attached with other components characterizing the ended bearing (cages, spheres, shield, etc.).

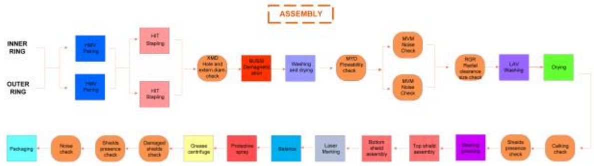


Figure 2: Channel 9: assembly

The operations on the branch assembly are the following (Figure 2):

- *pairing*, by which inner ring, outer ring and spheres are matched using machines able to realize radial clearance customer requirements;
- *stapling*, ie metal or plastic cages assembly;
- *hole and outer diameter size check*;
- *flowability, noise and radial clearance size check*;
- *washing and drying operations*;
- *bearing greasing*;
- *marking*;
- *amount of inserted grease check*;
- *sprinkling with a protective spray*;
- *packaging*.

3. LINE EFFICIENCY ANALYSIS

Observations in order to assess the efficiency of the line have been carried out. In particular, it has been observed the trend of the grinding and lapping branches production flow.

The theoretical production of the line is 2100 pieces/hour.

In practice, this value is not reachable, even considering the ideal time machine.

There are two machines, indeed, SHG and Denison, with maximum theoretical production capacity, respectively, of 1831 and 1800 pcs/h, because, at higher speeds, they would produce rings with sizes outside required tolerances and, therefore, by discard.

Besides, it has been noted as OR branch was faster than IR. It has been found, in fact, that the sum of the working average time on IR was greater than OR of more than 2 seconds, both considering measured cycle time both the ideal ones.

The problem is represented by hold grinding machine (SHG), that is the bottleneck of the line.

The presence of these differences on processing time has important issues about the whole process upstream, with birth of queues and stops along the line.

These are difficult to dispose of because of a substantial problem involving continuous accumulation of OR before assembly line and IR before SHG machine.

Our analysis has permitted to esteem the real line productive capacity and efficiency.

Table 1: Production capacity and line efficiency

	Channel 9
Gross production (expected)	**2100 units/hour**
Actual mean production (standard conditions)	**1220 units/hour**
Productive loss	**-880 units/hour**
Efficiency (%)	**58,1 %**

4. PROCESS FLOW OPTIMIZATION

Analyzing the data, it has come to define measures to increase production capacity.

Since the fully automated line, it's not very useful to focus on the operators. However, the work team experience is essential both to solve occurring problems and for operations speed carrying out such as tool change or machinery cleaning.

Nor it is possible to think of reversing some processes, which must necessarily be carried out according to a defined sequence.

The possible proposed actions, therefore, are the following:

1. IR interoperational buffer introduction

Introduction at the end of IR branch, of a buffer containing inner rings ground and lapped on another line working the same type of ring and in excess.

This operation would compensate for the deficiency due to the slow processing of this part of the line.

2. Adding a machine SHG

An alternative is the addition of a fourth machine SHG, the bottle neck of the line.

Assuming the new machine cycle time the average of the analogous SHG cycle times, this operation would increase the daily production capacity of about 2400 pieces and would reduce the production speed difference between OR and IR branches, although still OR processing faster.

3. SHG time machine reduction

It's the most practical alternative, since it would change less the line.

Reducing to 6 seconds the cycle time for all three SHG, measured time of one of the three machines, there would be a production capacity increase.

To achieve that, it should understand the causes that make:

- the three machines processing times variables between them;
- cycle times higher than theoretical time machine.

As regards the first point, it's important that SHG machines are slightly different from each other, in fact one of them is older than the others.

Furthermore, even if the initial setting is the same for the three machines, there are parameters only adjustable manually by the operator.

The operator also intervenes changing working parameters when he detects problems such as high amounts of waste.

The hole diameter size, moreover, are slightly different between the pieces, always within the tolerance range (10 µm).

Regarding the second point, the reasons may be different.

- the material quality is very important; it was noticed, in fact, that thermally treated rings within the plant are qualitatively better than external suppliers rings. Therefore they can be processed at higher speeds without causing many rejects or blocks machine.
- Sometimes the SHG are purposely slowed down by operators, for example when following assembly line is blocked.

Therefore, the operation that could be made is to redesign three machines processing cycle, optimizing it, ie to re-examine the values of the process parameters, the type of tool and the combination of the two, in order to obtain a reliable processing as much as possible. The three machines are similar, but not identical, then the set process parameters, which are the same, could be optimized only for one of them.

Secondly, should be adopted actions, such as a machine internal control system, aimed at minimizing subjectivity of operations.

5. VERIFICATION THROUGH A SIMULATION MODEL

To value the technical and economic suitability of the proposed interventions, it has been designed, using a dedicated software, a process model.

The fullness and the truthfulness of the simulation model have been tested in different conditions and on long simulation time periods.

The productiveness data has been obtained referring to a continuum operative period of 48 hours (a total production of 76,665 units – 1600 pcs/h), to guarantee the simulation stabilisation and reliability.

The results achieved by the model were, in terms of production trend, fully comparable with the ideal ones.

Time analysis shows how the bottle neck of the line is represented by SHG, which can produce a maximum of 1674 pcs/h. This value is slightly higher than the one obtained from the model, since the latter considers set-up time, inevitably present.

The simulation model results have been compared to real data, validating the designed simulative analysis.

The problems, coming out from the simulation model, are the same as those noticed during the process observation. Particularly evident are the following aspects:

- material accumulation on OR branch, after a short time. The OR branch is much faster than IR one, so the OR buffer, placed before coupling machine HMV, and dedicated machines buffers fill chain up to create a partial block of the branch, which restarts each time a piece from IR branch comes (Fig. 3).

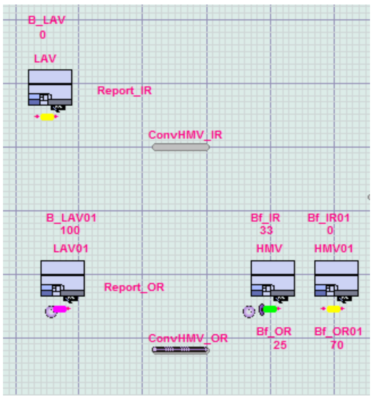


Figure 1 – Accumulation in the OR buffer placed before coupling machine HMV

- sliding bearing on the assembly line goes perfectly, buffers are empty.
- concerning to IR branch, the problem is on SHG (Fig. 4): in correspondence of these machines, the line has a large quantity of pieces into dedicated buffers and on conveyor belts, because SHG are unable to dispose of incoming pieces slowing down the entire process.

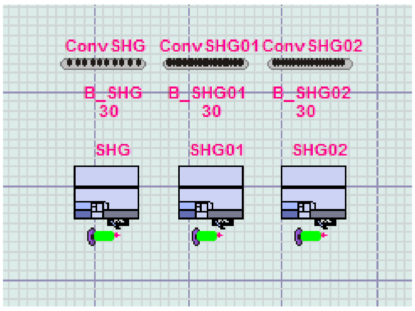


Figure 2 – Accumulation at the SHG

This feature is also known by the machines work data obtainable by the simulation model and shown below. The SHG (Fig. 5), in fact, works almost all the time, having always available workpieces.

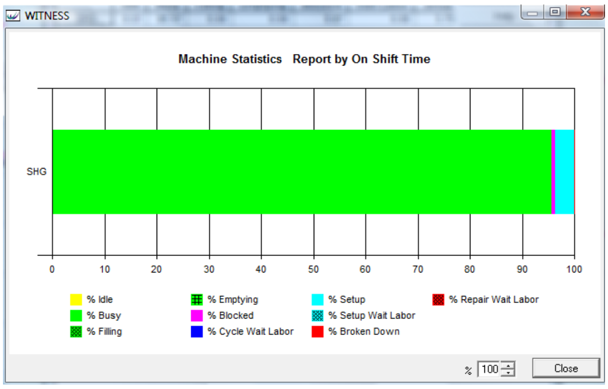


Figure 3 - SHG field data after 24 hours of production

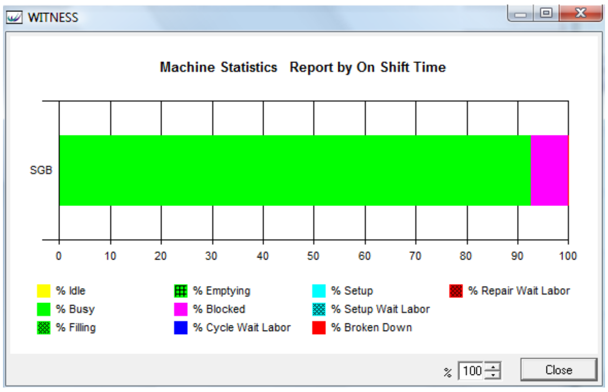


Figure 4 - SGB field data after 24 hours of production

Also the SGB (Fig. 6) runs continuously, although it is blocked for a certain period of time since, otherwise, being faster than SHG, it would produce an excessive number of pieces.

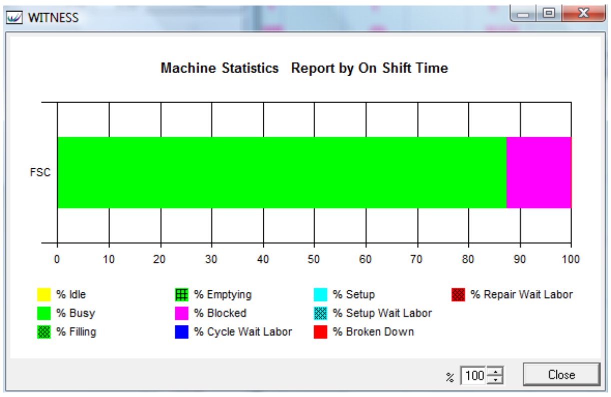


Figure 5 – First FSC field data after 24 hours of production

The FSC percentage blocking (Fig. 7), instead, is linked to the higher processing speed of the branch OR than IR, so it must be stopped for a period to adapt to the production on the other branch.

It's also noted from work data that the second MVM (Fig. 8), on the assembly line, is, for most of the time, on hold of work, so underutilized.

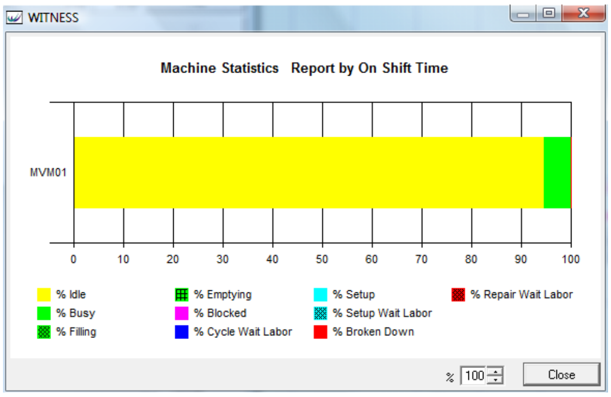


Figure 6 - MVM2 field data after 24 hours of production

The simulation model also confirms proposed changes validity: all three alternatives would lead to increase production capacity.

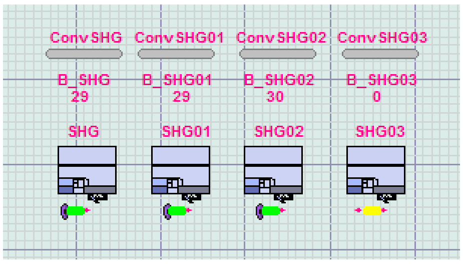


Figure. 7 – Addition of a forth SHG: situation after 24 hours on IR grinding branch

For example, figure 9 shows how a fourth SHG allows to have less accumulation in the buffers and on the conveyor belts before SHG, even after 24 hours, showing how, unlike before, these machines are able to dispose of the production.

6. TECHNICAL AND ECONOMIC ANALYSIS OF THE PROPOSED MODEL

The simulation time relative to the proposed model refers to 48 hours of continuum line activity. The data coming out from the changed model, shows a great increase both in hourly production and machines saturation.

1. Inserting a buffer of 1000 pieces per day, would increase the daily production capacity of about 1000 pieces, actually 745 whereas the actual production and model production differ by a certain percentage. This happens because there are other influential parameters, such as blocks machine, for example for faults, and waste.
2. The addition of a fourth machine SHG with a cycle time assumed as the average of the similar SHG cycle times, increases the theoretical daily production capacity of about 3200 pieces (actual 2400 pieces) and reduces the production speed difference between OR and IR branches, while remaining OR machining faster.
3. The last alternative, finally, would carry an actual production increase of about 1950 pieces.

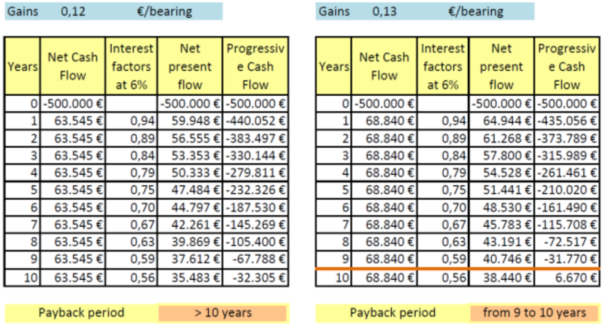
Subsequently, the new model has been economically validated.

1. The interoperational buffer introduction would increase slightly the efficiency but, at the same time, would cost just as little, ie as the cost of the buffer, which is assumed of 1000 €.
2. Relatively to the second solution, it is considered:

- The cost of the machine of € 500,000,
- The payback period of 10 years,
- The selling price of the bearing of 1 €,
- 220 working days per year, with an increase in annual production amounted to 529,540 pieces;

and it is assumed that the additional produced quantity is actually sold.

Under these conditions, assuming different values of gain from the sale of a bearing and interest rate of 6%, with *discounted payback* technique of valuation of investments, it will get different payback times (Fig. 10 a,b).



Gains 0,12 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Net present flow	Progressive Cash Flow
0	-500.000 €		-500.000 €	-500.000 €
1	63.545 €	0,94	59.948 €	-440.052 €
2	63.545 €	0,89	56.555 €	-383.497 €
3	63.545 €	0,84	53.353 €	-330.144 €
4	63.545 €	0,79	50.333 €	-279.811 €
5	63.545 €	0,75	47.484 €	-232.326 €
6	63.545 €	0,70	44.797 €	-187.530 €
7	63.545 €	0,67	42.261 €	-145.269 €
8	63.545 €	0,63	39.869 €	-105.400 €
9	63.545 €	0,59	37.612 €	-67.788 €
10	63.545 €	0,56	35.483 €	-32.305 €

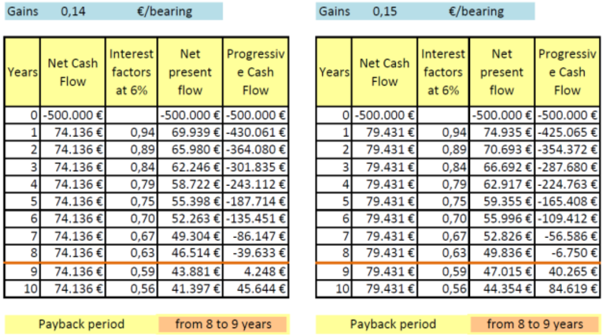
Payback period > 10 years

Gains 0,13 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Net present flow	Progressive Cash Flow
0	-500.000 €		-500.000 €	-500.000 €
1	68.840 €	0,94	64.944 €	-435.056 €
2	68.840 €	0,89	61.268 €	-373.789 €
3	68.840 €	0,84	57.800 €	-315.989 €
4	68.840 €	0,79	54.528 €	-261.461 €
5	68.840 €	0,75	51.441 €	-210.020 €
6	68.840 €	0,70	48.530 €	-161.490 €
7	68.840 €	0,67	45.783 €	-115.708 €
8	68.840 €	0,63	43.191 €	-72.517 €
9	68.840 €	0,59	40.746 €	-31.770 €
10	68.840 €	0,56	38.440 €	6.670 €

Payback period from 9 to 10 years

Figure 8 a) - Payback discounted method



Gains 0,14 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Net present flow	Progressive Cash Flow
0	-500.000 €		-500.000 €	-500.000 €
1	74.136 €	0,94	69.939 €	-430.061 €
2	74.136 €	0,89	65.980 €	-364.080 €
3	74.136 €	0,84	62.246 €	-301.835 €
4	74.136 €	0,79	58.722 €	-243.112 €
5	74.136 €	0,75	55.398 €	-187.714 €
6	74.136 €	0,70	52.263 €	-135.451 €
7	74.136 €	0,67	49.304 €	-86.147 €
8	74.136 €	0,63	46.514 €	-39.633 €
9	74.136 €	0,59	43.881 €	4.248 €
10	74.136 €	0,56	41.397 €	45.644 €

Payback period from 8 to 9 years

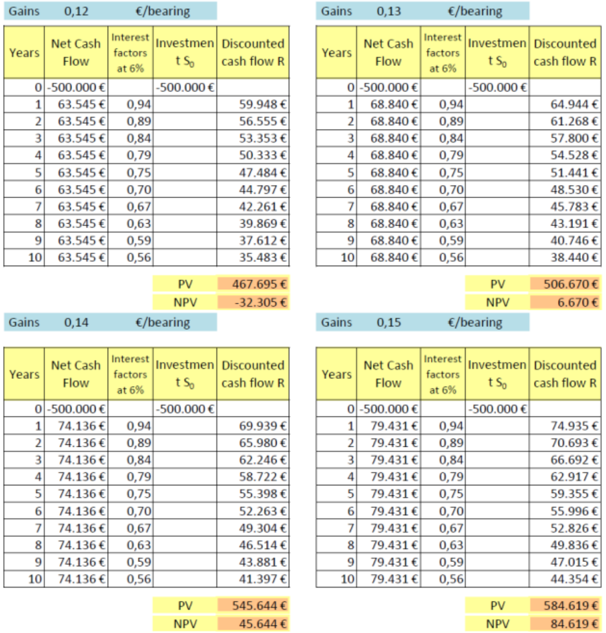
Gains 0,15 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Net present flow	Progressive Cash Flow
0	-500.000 €		-500.000 €	-500.000 €
1	79.431 €	0,94	74.935 €	-425.065 €
2	79.431 €	0,89	70.693 €	-354.372 €
3	79.431 €	0,84	66.692 €	-287.680 €
4	79.431 €	0,79	62.917 €	-224.763 €
5	79.431 €	0,75	59.355 €	-165.408 €
6	79.431 €	0,70	55.996 €	-109.412 €
7	79.431 €	0,67	52.826 €	-56.586 €
8	79.431 €	0,63	49.836 €	-6.750 €
9	79.431 €	0,59	47.015 €	40.265 €
10	79.431 €	0,56	44.354 €	84.619 €

Payback period from 8 to 9 years

Figure 90 b) - Payback discounted method

Increasing even one euro cent profit from the sale of a bearing, it can see as the recovery time decreases and simultaneously the total gain increases – NPV Method (fig. 11).



Gains 0,12 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Investment S_0	Discounted cash flow R
0	-500.000 €		-500.000 €	
1	63.545 €	0,94		59.948 €
2	63.545 €	0,89		56.555 €
3	63.545 €	0,84		53.353 €
4	63.545 €	0,79		50.333 €
5	63.545 €	0,75		47.484 €
6	63.545 €	0,70		44.797 €
7	63.545 €	0,67		42.261 €
8	63.545 €	0,63		39.869 €
9	63.545 €	0,59		37.612 €
10	63.545 €	0,56		35.483 €
			PV	467.695 €
			NPV	-32.305 €

Gains 0,13 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Investment S_0	Discounted cash flow R
0	-500.000 €		-500.000 €	
1	68.840 €	0,94		64.944 €
2	68.840 €	0,89		61.268 €
3	68.840 €	0,84		57.800 €
4	68.840 €	0,79		54.528 €
5	68.840 €	0,75		51.441 €
6	68.840 €	0,70		48.530 €
7	68.840 €	0,67		45.783 €
8	68.840 €	0,63		43.191 €
9	68.840 €	0,59		40.746 €
10	68.840 €	0,56		38.440 €
			PV	506.670 €
			NPV	6.670 €

Gains 0,14 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Investment S_0	Discounted cash flow R
0	-500.000 €		-500.000 €	
1	74.136 €	0,94		69.939 €
2	74.136 €	0,89		65.980 €
3	74.136 €	0,84		62.246 €
4	74.136 €	0,79		58.722 €
5	74.136 €	0,75		55.398 €
6	74.136 €	0,70		52.263 €
7	74.136 €	0,67		49.304 €
8	74.136 €	0,63		46.514 €
9	74.136 €	0,59		43.881 €
10	74.136 €	0,56		41.397 €
			PV	545.644 €
			NPV	45.644 €

Gains 0,15 €/bearing

Years	Net Cash Flow	Interest factors at 6%	Investment S_0	Discounted cash flow R
0	-500.000 €		-500.000 €	
1	79.431 €	0,94		74.935 €
2	79.431 €	0,89		70.693 €
3	79.431 €	0,84		66.692 €
4	79.431 €	0,79		62.917 €
5	79.431 €	0,75		59.355 €
6	79.431 €	0,70		55.996 €
7	79.431 €	0,67		52.826 €
8	79.431 €	0,63		49.836 €
9	79.431 €	0,59		47.015 €
10	79.431 €	0,56		44.354 €
			PV	584.619 €
			NPV	84.619 €

Figure 11 - Net present value method

The costs associated with the production capacity increase by performing a reduction of SHG times to measured time for one of the three machines (6 seconds), are not very high. Annually, there is an increase of cost of about € 2260 due to increased consumption of grinding wheels compared to an increase of gain from bearings sale of approximately €

51,300 (assuming earnings per bearing of € 0.12). This is true assuming the tool wear does not increase by reducing to these values the cycle time and, therefore, the diamond coating interval should be not reduced.

7. CONCLUSIONS

In this paper a production process re-engineering methodology is proposed.

The planning process has been supported by techno-economic analysis, using simulation techniques. The developed procedure has enabled the company to achieve significant benefits.

The advantages obtained thanks to the application of simulation techniques to the bearing production process, include:

- increase in the productive efficiency;
- increase in the machines saturation;
- reduction in the bottle necks.

The first solution is characterized by a very low realization costs, 1000 €, compared to a production capacity increase by 2.5% (about 750 pcs/day), but it requires continuous availability of components from another production line. The second alternative is the most expensive to implement, due to the machine cost (500,000 €), but it allows the production capacity increase higher, equal to 8.2% (about 2400 pcs/day).

The third option would be certainly more convenient, since it does not involve changes to the production channel and, at the same time, it would lead to an efficiency increase slightly less than adding a fourth SHG machine (6.6%, about 1950 pcs/day). The cost would be definitely lower, about 2250 €, but it should study how actually to be able to implement this solution.

REFERENCES:

Grimaldi M., Cricelli L., Rogo F., 2013. A methodology to assess value creation in communities of innovation. *Journal of Intellectual Capital* 13(3): 305 -330.

Naylor T.H., Bality J.L., Burdick D.S., Chu K., 1968. *Computers Simulation Techniques*. New York: John W & Sons.

Morgan, B.J.T., 1984. *Elementary Simulation.* London: Chapman 6 Hall.

Martinoli B., 1988. *Guida alla simulazione: metodi, linguaggi e modelli di ricerca operativa per simulare processi ed eventi con esempi relativi ad applicazioni aziendali*. Franco Angeli Editore.

Bratley P., Fox B.L., Schrage L.E., 1983. *A Guide to Simulation,* New York: Springer.

Falcone D., Duraccio V., Silvestri A., Di Bona G., 2006. Improvement of performances of an optical system for defectiveness survey in a company of the automotive field. *Proc. Summer Simulation Multiconference*. 2006, Calgary, Canada.

Falcone D., De Felice F., Di Bona G., Silvestri A., 2003. Improvement of the moulding process in an automotive company through the employment of Robust Design. *Proc. Modelling and Simulation*. 2003, Palm Spring, CA.

Beniaafar S., 1992. *Intelligent Simulation for Flexible Manufacturing System: an Integrated Approach, Computer & Industrial Engineering*. Vol.22 n°3.

Caron F., 1992. Introduzione ai linguaggi e modelli di simulazione. *Simulare per decidere meglio*. Convegno IRI. 6-7 febbraio 1992, Milano.

Silvestri A., Falcone D., Di Bona G., Duraccio, V. Forcina A., 2011. Modeling and Simulation of an assembly line: a new approach for assignment and optimization of activities of operator. *Proc. The 10th International Conference on Modeling and Applied Simulation*. Rome, Italy.

Cricelli L., Grimaldi M., Levialdi N., 2009. Modelling the competition of an HNO vs. an MVNO in the mobile telecommunication industry, *International Journal of Technology, Policy and Management* 9(3): 277 - 295.

Duraccio V., Forcina A., Silvestri A., Di Bona G., 2013. Productive Line Reengineering Through Simulation Techniques. *32nd IASTED International Conference on Modelling, Identification and Control.* 11-13 febbraio 2013, Innsbruck.

Duraccio V., Falcone D., Silvestri A., Di Bona G., 2006. Project of an Agv Transport System through Simulation Techniques. *Proc. International Workshop on Modeling & Applied Simulation*. 2008, Campora S.Giovanni Italy.

Colorni, A., 1984. *Ricerca Operativa*. Milano CLUP

Averil L. M., 1991. *Simulation Modeling Analalysis*. II edit, Mc Graw-Hill.

Falcone D., Di Bona G., Forcina A., Silvestri A., Pacitto A., 2010. Study and modelling of very flexible lines through simulation. *Proc. Emerging Applications in Industry and Academia Symposium*. 2010, Orlando, Florida.

Falcone D., Duraccio V., Di Bona G., Silvestri A., 2005. Technical and economical analysis of the layout of a palletization plant through simulation techniques. *Proc. Summer Simulation Multiconference*. 2005, Philadelphia USA.

Tocker, K.D., 1963. *The Art of Simulation,* English London: University Press.

De Felice F., Silvestri A., Di Bona G., Forcina A., Petrillo A., 2008. The Project Management through performance indexes in an automotive company, International Project Management Association. *Proc. 22nd World Congress IPMA.* 2008, Rome, Italy.

De Felice F., Petrillo A., 2012. Productivity analysis through simulation technique to optimize an automated assembly line. *Proceedings of the IASTED International Conference Applied Simulation and Modelling 2012,* June 25 - 27, 2012, Napoli, Italy.

WITNESS MANUAL USER, Release 9.10, AT&T ISTEL.

EXPLORING UNKOWN NETWORKS USING A COOPERATIVE MAS-BASED APPROACH

Pedro Simeão Carvalho, Rosaldo J. F. Rossetti, Ana Paula Rocha, Eugénio C. Oliveira

Informatics Engineering Department, Artificial Intelligence and Computer Science Lab
Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, S/N, 4200-465 Porto, Portugal

{pedro.simeao, rossetti, arocha, eco}@fe.up.pt

ABSTRACT

This paper reports on a novel method to explore and map an entirely unknown network using a cooperative Multi-Agent System (MAS) to extract knowledge or information from nodes and connections. We consider the likely presence of obstacles, eventually making the network disconnected. The MAS architecture is applicable to a vast range of scenarios. Our main goal is to discover the entire network as quickly as possible, characterizing its nodes' meta-structures. In this paper, we propose a novel method that relies on agents that can communicate to each other through simple messages, ensuring that there is no resource sharing. The proposed method is compared to other two non-cooperative methods through simulation, in order to establish a basis for comparison. Preliminary results show that our cooperative approach produces better results than the other two implemented and guarantees that the entire network is explored at the end.

Keywords: network exploration, multi-agent systems, cooperation in MAS.

1. INTRODUCTION

The motivation for this work is the need to completely explore an *a priori* unknown network, defined as an abstract set of nodes connected to each other through edges. There is also the need to consider the presence of "dark nodes" on the network, which are nodes that can be neither analyzed by agents nor even crossed by them while moving over the network (e.g. physical or abstract obstacles, unreadable nodes, etc.)

This problem specification considers no specific application domain and is applicable to a vast range of scenarios, provided they can be represented as a network. This approach might be used, for instance, in space exploration vehicles or robots, data block processing, navigation systems, social networks, and generic network discovery. Thus, our main goal is to discover the entire network as quickly as possible, characterizing its nodes' meta-structures and/or its connections.

In the past years, the widely adoption of Multi-Agent Systems (MASs) has increased in problem solving across many areas, such as informatics, intelligent systems and even the industrial sector. The MAS concept allows the use of distributed computation that can be either virtual or physical. These agents are autonomous, share an environment through communication and interactions, and make decisions according to the situation (Parker 2003). This approach is useful when processing a considerable amount of data such as in large networks, because this exploration and processing can be divided into small pieces and performed by individual agents. We believe that this idea can be used when exploring networks as well, so they can increase the overall performance (Tan 1993).

Although there are many search algorithms used in graphs or networks (Knuth 1997, Knuth 1998), such as Breadth-First Search (BFS) (Bader 2006, Yoo 2005), Depth-First Search (DFS), Dijkstra and Kruskal, they are all thought to search for values on networks, not to explore them. These algorithms were not thought to be used either for very large and highly connected networks, for the presence of obstacles or for being used by multiple agents simultaneously. Moreover, some algorithms do not consider the distance between nodes or even a cost to travel through them. Thus, they cannot be used in our scenario. There are still other algorithms based on the Ant Colony approach (Weyns 2007, Claes 2011, Colorni 1991, Dorigo 1992) that might be interesting to this scenario but they do not fully meet our constraints.

We propose a novel method using a MAS architecture to explore a network, where agents can only communicate between them using simple messages, so as to share information and to accomplish the entire discovery. This method guarantees almost full isolation of the agents and no resource sharing. These agents do not need to negotiate, so they are naturally fault safe considering the overall process.

To prove the usefulness of this method, we had developed a simulation with two other methods as a basis for comparison. The application domain for this simulation was the discovery of a new planet by space vehicles of different types. We implemented three types of agents, each of them with a different method (random, non-cooperative and cooperative). The comparison was made using the full exploration mean time for each agent type.

The remainder of this paper is structured as follows. Section 2 discusses on the proposed approach, where we present the problem formalization and our solution method. In Section 3, preliminary results are presented and analyzed. Related work on network exploration are presented in Section 4, whereas in Section 5 we draw conclusions about the proposed method.

2. PROPOSTED APPROACH

The basis for the cooperative discovery approach proposed in this paper is the use of altruistic agents that work together to discover all the network.

The core for this agents is the A* algorithm, whose premises rely on agents featuring two sets of node identifiers. These sets contain the nodes that the agent wishes to explore (SN$_{wish}$) and those it has already explored (SN$_{explored}$). At the beginning, both SN$_{wish}$ and SN$_{explored}$ are empty. This is the only information that an agent needs to store in its memory. Each agent has memory and only keeps knowledge of its experience and history; thus, despite what it actually knows, an agent is unaware of the remaining network and of what other agents know. Therefore, each agent is independent from other agents, which guarantees that there is no memory/resource sharing and no breaking points all over its activity.

Since an agent is isolated, as mentioned before, it can only send and receive messages to and from other agents. These messages can be directed to one agent in particular, or be broadcast across the network to all other agents. Thus, there must exist a messaging service to provide this requirement. There are only three types of messages that the agents can send:

1. Inform all other agents that it is visiting one node (Message$_{inform}$); this is a broadcast message;
2. Ask all agents what is the content of their SN$_{wish}$ set (Message$_{ask\text{-}help}$); this is a broadcast message;
3. Reply to a specific agent that previously has sent Message$_{ask\text{-}help}$ the content of the SN$_{wish}$ set (Message$_{response\text{-}help}$).

These messages will be used on specific situations and will be explained with the algorithm. Each agent has its own mail box.

The visibility of each agent on each node is limited to the node itself and to the node's adjacencies, which means that the sole information it can have are nodes' identifier and whether all connected nodes are dark nodes or not. That means that an agent can "see" which connected nodes can be explored prior to moving to them. This is particularly useful, e.g. when an agent is exploring a map and there is a wall on his path.

Figure 1 shows the state diagram for this method, representing an algorithm overview. More specifically, at each step, an agent does the following (considering that the agent is inside a node):

1. Read mail box and process all messages (explained ahead);
2. Extracts the needed node information;
3. Sends a broadcast message to all other agents (Message$_{inform}$) saying that the node being analyzed is already explored, so other agents can remove this node from theirs SN$_{wish}$, if it exists, and add it to SN$_{explored}$;
4. Adds that node identifier to SN$_{explored}$;
5. From the neighbor nodes, chooses the ones that are not dark nodes, and adds them to SN$_{wish}$, if not present in SN$_{explored}$;
6. If SN$_{wish}$ is not empty, chooses the nearest node from SN$_{wish}$ (if there are more than one, use random choice) and goes to it (then, return to step 2);
7. If SN$_{wish}$ is empty, sends a broadcast message to all other agents (Message$_{ask\text{-}help}$), asking them to send their SN$_{wish}$ contents, so that it can have new nodes to explore.

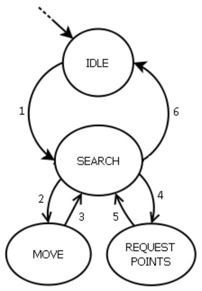


Figure 1: Agent State Diagram

There is no defined algorithm to move across nodes; one can use any heuristics, depending on the problem. When moving between nodes, it executes steps 2, 3 and 4. Thus, the agent maintains its goal and builds on the work. Step 6 can be performed using a heuristic utility that tries to maximize any goal. However, this is optional.

The mail box reading process uses the following rules, according to the message received:

- If Message$_{inform}$: reads the node identifier from the message. Removes that node from SN$_{wish}$, if exists. Add that node to SN$_{explored}$;
- If Message$_{ask\text{-}help}$: sends the content of Message$_{wish}$ to the agent that sent this message;
- If Message$_{response\text{-}help}$: copy the received nodes identifiers to SN$_{wish}$.

When SN$_{wish}$ is empty and there is no response for Message$_{ask\text{-}help}$ (Message$_{response\text{-}help}$ messages), the work is done. That means there is no known nodes to further explore, and the job is complete.

This method guarantees that the entire network is discovered, because each agent contributes, at each step, to this process. Despite these agents are isolated and do not explicitly share information about their knowledge, they can guarantee, as a whole, that the network is fully explored. At an earlier stage, each agent tries to explore all the nodes that are in their SN_{wish} set; later on, when their job queue is empty, they try to collaborate with other agents and do their work.

Each agent is independent and acts by itself, so there is no critical point in the whole process (i.e., a coordinator base) - the exploration is just cooperative, and there is no need to have a centralized coordinator. That means there are no breaking points in the exploration, which makes them naturally fault safe and act as reactive agents. In other words, if one agent fails, the full exploration is not compromised. According to Stan Franklin and Art Graesser, these agents are autonomous, goal-oriented, temporally continuous and communicative (Graesser 1996).

3. PRELIMINARY RESULTS AND DISCUSSION

To test the proposed method, we developed a simulated case study, using the *REPAST*[1] simulation framework. The chosen scenario was the exploration of an unknown planet by some space rovers (space exploration vehicles/robots).

3.1. Simulation setup

The planet surface was represented in a 2D perspective, using the torus concept. The surface had many obstacles, representing the "dark nodes". Each node (position), represented by the coordinate pair (X, Y), was connected to the closest eight near cells (degree equals 8) using the Moore's neighborhood concept. There was a starting point, called "base", where all the robots were at the beginning of the simulation, and 250 obstacle points (representing walls or rocks that the robot cannot cross), distributed in different forms, in order to provide a complete experience. Figure 2 depicts the map (size was 50x50) used for this experiment.

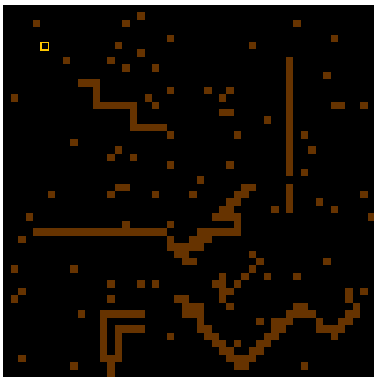


Figure 2: map with obstacles (brown points) and one base (yellow square)

In order to establish a basis for comparison of the proposed method, we developed three types of agents: random agents (A_R), non-cooperative (selfish) agents (A_{NC}) and cooperative (altruistic) agents (A_C).

A_R agents, at each iteration, choose the next node randomly from the connected nodes. They have no knowledge or memory. This is the simplest agent.

A_{NC} agents use pre-determined blocks to explore (areas), previously calculated, and they follow it, avoiding obstacles when found. Although the division of this scenario (network) in blocks is based on the map size, this division can be done by other heuristic that do not need that information. Moreover, the block-division algorithm does not know the map content, either. The reason for choosing this algorithm was its simplicity to be implemented; nonetheless, it could be any other algorithm as long as it is non-cooperative-based.

The third type of agent, A_C, uses the methodology presented in this paper, whose performance we want to compare with the former two.

At each simulation step ('tick' in *REPAST* framework), the order of execution of each agent was random. In order to compare all agents, we measured the number of explored positions on the map over time (ticks) for each agent type.

We did seven tests, with 10, 15, 20, 25, 30, 35 and 40 agents of each type. For statistical significance, we run each experiment 160 times; the average of the ending time was calculated, and the less (lower values) the better.

3.2. Results and Discussion

In some of our tests, our A_R (random) got results over 7000 ticks in all rounds, so it will not be considered in the analysis. This happens because this map size is too big for this number of agents, so they need more time to explore everything.

Table 1 and Table 2 show the results obtained in the experiments.

Table 1: Execution ticks averages

Nr. of Agents	A_R	A_{NC}	A_C
10	N/A	878.98	484.16
15	N/A	665.64	408.56
20	5733.76	537.66	412.18
25	4132.44	440.61	413.36
30	3948.95	479.74	371.99
35	3306.01	425.40	388.91
40	3038.31	425.30	386.26

In Table 1 we present the ticks average of complete exploration for all the algorithms. As expected, A_R gets better as there are more agents to explore this network. However, those results are much higher than those corresponding to A_{NC} and A_C. A_C got in all tests always a better result than A_{NC}.

[1] http://repast.sourceforge.net

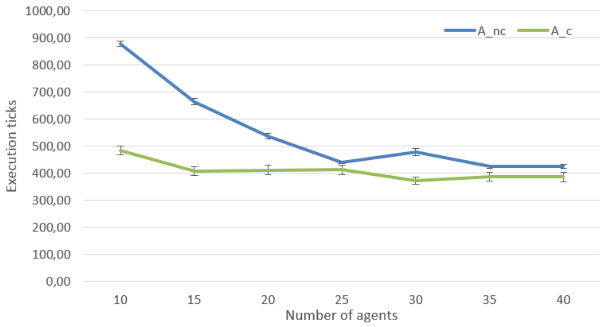


Figure 3: Execution ticks average for A_{NC} (blue line) and A_C (green line), with confidence range level of 95%.

Table 1 also shows that the performance obtained in the A_C method has a low variance. In other words, the difference between the best and worst value for A_{NC} was 453.68 ticks and for A_C was just 112.6 ticks.

Table 2 shows the precision of the mean obtained in Table 1, with a confidence level of 95%, calculated with equation 1 for experimental results (JCGM 2008, Carvalho, et al. 2012, Simões 2008). Figure 3 shows the execution tick average for A_{NC} and A_C, for all experiments, with precision range indicated on Table 2.

$$\sigma_{m(95\%)} = 2\sigma_m = 2\frac{\sigma}{\sqrt{N}} \approx 2\sqrt{\frac{\sum(x_i-\bar{x})^2}{n(n-1)}} \qquad (1)$$

Table 2: Standard Deviation of the Mean (confidence level of 95%) in ticks and corresponding percentages.

Nr. of Agents	A_R	A_{NC}	A_C
10	N/A	9.92 (1.1%)	15.97 (3.3%)
15	N/A	12.61 (1.9%)	16.76 (4.1%)
20	604.01 (10.5%)	10.01 (1.9%)	18.07 (4.4%)
25	304.80 (7.4%)	5.23 (1.2%)	17.87 (4.3%)
30	382.79 (9.7%)	13.07 (2.7%)	13.55 (3.6%)
35	289.04 (8.7%)	5.61 (1.3%)	15.96 (4.1%)
40	364.30 (12.0%)	7.12 (1.7%)	17.09 (4.4%)

As shown in Table 2, the confidence level for the mean of the A_C method is around 4%, which represents a good precision value. Applying the range of $\bar{x} \pm \sigma_{m(95\%)}$ for all results, we can prove that A_C is still always better (lower) than A_{NC}.

Figure 4 and Figure 5 show two simulation executions, indicating the number of explored positions *vs.* tick time, for each type of agent. We can see that the evolution of the methods are different. A_C proves to be linear through time, dealing well with the obstacles found.

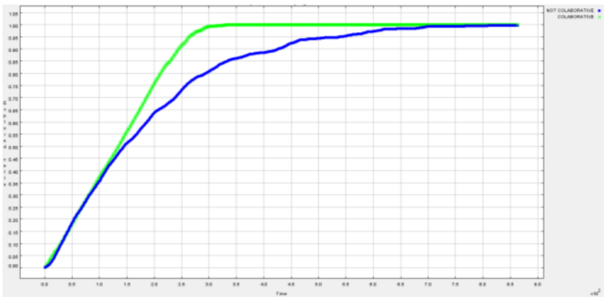


Figure 4: Execution example with 10 agents. Explored nodes *vs.* ticks. Blue line - A_{NC}; green line - A_C.

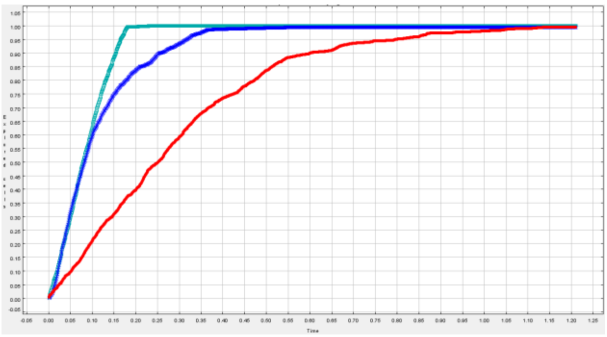


Figure 5: Execution example for 20 agents. Explored nodes *vs.* ticks. Red line - A_R; blue line - A_{NC}; green line - A_C.

These results shows that our proposed method is always better than the other two used for comparison. This happens in result for the existence of dark nodes inside the network, which oblige agents to cope with them. In these experiments, our network has a lot of combinations of dark nodes (horizontal, vertical and diagonal walls, "S-shapped" walls and isolated obstacles). Although we can use some simple and reactive obstacle avoidance algorithm such as Bug1 or Bug2 (Ribeiro 2005, Stepanov 1990, V. a. Lumelsky 1990, Choset 2005), they still have to cope with them, so agents will have to overcome these obstacles. In our experiment, A_{NC} had predefined areas to explore; sometimes, agents needed to abandon their "working area" to avoid some obstacles, which proves to be a high cost to the overall exploration process.

Figure 6 shows the states evolution of our A_C in a simulation with 40 agents, as defined in Figure 1. In the initial phase there are more agents exploring than moving. In the middle of the execution, search and moving states are equal and then, when the network starts to be fully explored (approximately 90% at 100 ticks), all agents start to move and request points. That means only approximately 10% of the network is still not explored and it consumes almost the same time to explore it as in the first 90%. This difference is a result of the traveling time of agents to explore some missing points across the network. Figure 7 can be used to check the evolution of this experience.

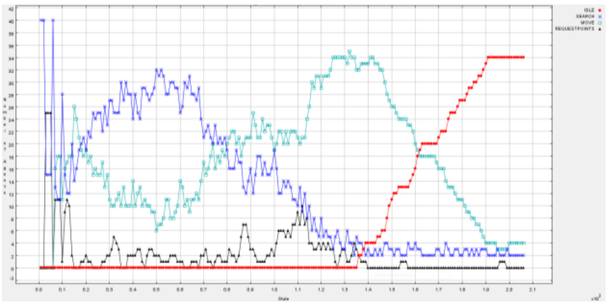


Figure 6: Collaborative agents' state for an execution example with 40 agents. IDLE - red line; SEARCH - blue line; MOVE - green line; REQUEST POINTS - dark line.

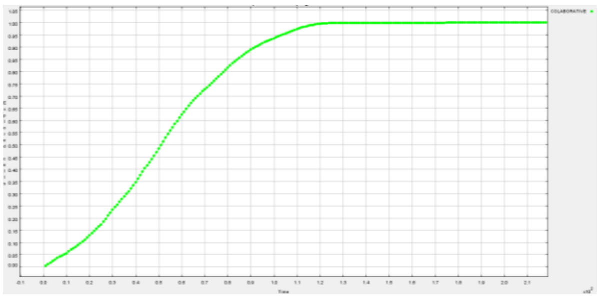


Figure 7: Execution time with 40 A_C (green line)

The fact that agents A_C do not need to negotiate between them has the consequence that there is no need for a synchronizing phase. Even without sharing information about the network, all agents proved effective when coping (on their own only) with unpredictable obstacles during exploration.

4. RELATED WORK

Network exploration and network search in computer science and mathematics is not a recent topic. There are many algorithms that try to optimize the searching mechanism on a network, to find the best path between two nodes or even to explore its metadata (Leiserson 2001). There are also other approaches based on the Ant Colony concept that use the environment as a communication medium, to perform a cooperative exploration. Furthermore, some authors have used parallel processing in their algorithms, as well as MAS architectures to perform their tasks using different approaches such as decentralized search (Zhang 2005).

However, to the best of the authors' knowledge there are no such methods or algorithms that meet all constraints for the type of networks here presented. Consequently, this work was not based on any previously published works.

5. CONCLUSIONS

In this paper, we presented a novel method that performs a cooperative exploration of an unknown network with obstacles using a MAS-based approach, in order to explore it as quickly as possible. Our concept relied on isolated and altruist agents that communicate through simple messages in order to exchange some information.

We performed a basis for comparison with other two non-cooperative methods through simulation. Our method always outperformed the non-cooperative ones whenever dark nodes were present on the network. The precision of our results was about 4%, for a confidence level of 95%. Thus, results show that our agents have a higher performance when there are a few agents running the exploration algorithm for a given network, compared with the other methods. That means cooperative agents can produce better results when dealing with large networks. This is also true even though a dark node is an articulation point in the network, for instance, when the associated graph representing the network will be disconnected.

As our agents will explore the whole bunch of nodes within range, we only need to ensure that the entering points are normally distributed over the network to guarantee total coverage. Moreover, our agents are kept isolated, with no resource sharing, and there is no need of negotiation, so there is no fault point in the whole process. Also, they do not try to foresee the future, so they are very reactive according to the environment. Their behavior relies on the information that they have at that very moment.

Further improvements may include, for instance, developing better communication and path calculation algorithms, to optimize and minimize the length of the path traveled by the agent when it finds a dark node or needs to travel between nodes. Thus, the future goal is to minimize the wasting time in any journey and use it for discovery purposes.

Depending on the scenarios, other improvements or modifications can be made, such as the existence of a "refresh" time of the nodes or the need for a specific agent to analyze a nodes' set. Our method is flexible in many aspects in order to adapt to existing constraints.

The knowledge of each agent is relatively exclusive to its execution but, as a whole, agents learn and behave in an implicit way. This method guarantees that the entire network is explored to the end, despite the network size or characteristics. Thus, this method can be applied either to extract nodes metadata or to analyze their connections using a cooperative MAS approach.

ACKNOWLEDGMENTS

The authors are indebted to José Pedro Silva for helpful discussions about some of the ideas presented in this work at its earlier stages.

REFERENCES

Bader, D.A., Madduri, K., 2006. Designing multithreaded algorithms for breadth-first search and st-connectivity on the Cray MTA-2. *Proceedings of the 35th International Conference of Parallel Processing (ICPP 2006)*, pp. 523--530, August, Columbus (Ohio, U.S.A).

Carvalho, P.S., Sousa, A.S., Paiva, J., Ferreira, A. 2012. *Ensino Experimental das Ciências. Um guia para* *professores do ensino secundário. Física e Química.* Porto: Editorial U.Porto.

Choset, H., Lynch, K.M., Hutchinson, S., Kantor, G., Burgard, W., Kavraki, L.E., Thrun, S., 2005. *Principles of robot motion: theory, algorithms, and implementations.* Cambridge: The MIT press.

Claes, R., Holvoet, T., Weyns, D., 2011. A decentralized approach for anticipatory vehicle routing using delegate multiagent systems. *IEEE Transactions on Intelligent Transportation Systems.* 12 (2), 364--373.

Colorni, A., Dorigo, M., Maniezzo, V., 1991. Distributed optimization by ant colonies. *Proceedings of the First European Conference on Artificial* Life, pp. 134--142 (Paris, France).

Dorigo, M., 1992. *Optimization, learning and natural algorithms.* Thesis (PhD). Politecnico di Milano (Italy).

Graesser, A., Franklin S., 1996. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. Proceedings of the 3rd *International Workshop on Agent Theories, Architectures, and Languages*, pp. 21-35, Springer-Verlag London (UK).

JCGM, 2008. *Evaluation of measurement data - Guide to the expression of uncertainty in measurement.* Sèvres: Bureau Internationale des Poids et Mésures.

Knuth, D.E., 1997. *The Art Of Computer Programming.* 3. Vol. 1. Boston: Addison-Wesley.

Knuth, D.E., 1998. *The Art of Computer Programming.* Vol. 3. Harlow: Addison-Wesley.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C., 2001. *Introduction to Algorithms.* Cambridge: The MIT press.

Lumelsky, V.J., Skewis, T., 1990. Incorporating range sensing in the robot navigation function. *IEEE Transactions on Systems, Man and Cybernetics,* 20 (5), 1058-1069.

Parker, D.C., Manson, S., Janssen, M.A., Hoffmann, M.J., Deadman, P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers,* 93 (2): 314--337.

Ribeiro, M.I., 2005. Obstacle avoidance. *Instituto de Sistemas e Robótica, Instituto Superior Técnico.*

Simões, J.A.M., Castanho, M.A.R.B., Lampreia, I.M.S., Santos, F.J.V., Castro, C.A.N., Norberto, M.F., Pamplona, M.T., Mira, L., Meireles, M.M., 2008. *Guia do Laboratório de Química e Bioquímica.* Lisboa: Lidel.

Lumelsky V.J., Stepanov, A., 1987. Path-planning srategies for a point mobile automaton amidst unknown obstacles of arbitrary shape. *Algorithmica,* 2, 403-430, Springer.

Tan, M., 1993. Multi-agent reinforcement learning: Independent vs. cooperative agents. *Proceedings of the 10th international conference on machine learning,* vol. 337, pp. 330-337, Amherst (Massachusetts, U.S.A.).

Weyns, D., Holvoet, T., Helleboogh, A., 2007. Anticipatory vehicle routing using delegate multi-agent systems. *Proceedings of the Intelligent Transportation Systems Conference (ITSC 2007. IEEE),* pp. 87--93. September. 30-October 3, Seattle (Washington, U.S.A.)

Yoo, A., Chow, E., Henderson, K., McLendon, W., Hendrickson, B., Catalyurek, U., 2005. A scalable distributed parallel breadth-first search algorithm on BlueGene/L. *Proceedings of the ACM/IEEE Conference on* Superconductiong, pp. 25--25, Noverber 12-18, Seattle (Washington, U.S.A.).

Zhang, J., Ackerman, M.S., 2005. Searching for expertise in social networks: a simulation of potential strategies. *Proceedings of the Iinternational ACM SIGGROUP Conference on Supporting group work (ACM 2005),* pp. 71--80, Sanibel Island (Florida, U.S.A.).

AUTHORS BIOGRAPHY

Pedro Simeão Carvalho is graduated in Informatics and Computing Engineering where he took his master degree; he is now working at software developer TLANTIC SI enterprise, in mobile systems department.

Rosaldo J. F. Rossetti is an assistant professor at Faculty of Engineering, University of Porto, and member of the Artificial Intelligence and Computer Science Laboratory.

Ana Paula Rocha is an assistant professor at Faculty of Engineering, University of Porto, and member of the Artificial Intelligence and Computer Science Laboratory.

Eugénio C. Oliveira is a full professor at Faculty of Engineering, University of Porto. Prof. Oliveira is the head of the Artificial Intelligence and Computer Science Laboratory, and his research interests are mainly in the field of Multi-Agent Systems and their applications

USING NATURAL INTERFACES TO INTERACT WITH A VIRTUAL CONTROL DESK OF A NUCLEAR POWER PLANT

Maurício A. C. Aghina[(a)], Antônio Carlos A. Mól[(b)], Carlos Alexandre F. Jorge[(c)], Paulo Victor R. Carvalho[(d)] , Ana Paula Legey[(e)] , Gerson G. Cunha[(f)], Luiz Landau[(g)]

(a), (b),(c),(d) Instituto de Engenharia Nuclear /Comissão Nacional de Energia Nuclear, , RJ, Brazil
(b)Instituto Nacional de Ciência e Tecnologia de Reatores Nucleares Inovadores/CNPq, Brazil
(b), (e) Universidade Gama Filho, Brazil
(f), (g) Universidade Federal do Rio de Janeiro, Brazil

[(a)] mag@ien.gov.br, [(b)] mol@ien.gov.br, [(c)] calexandre@ien.gov.br , [(d)] paulov@ien.gov.br , [(e)] analegey@hotmail.com,
[(f)] gerson@ufrj.br [(g)] landau@lamce.copp.ufrj.br

ABSTRACT

This paper reports results achieved in a development for a virtual control desk, interfacing with a nuclear power plant's control system. This virtual control desk was developed aiming to combine the dynamics simulation of a nuclear power plant operation, with high fidelity control desk's visual appearance. Natural interfacing techniques where used to interact with this virtual control desk, as spoken command recognition, head tracking and body tracking. The combination of such interfacing techniques could improve user interfacing, through exploring each technique's advantages for specific tasks. For instance, spoken command recognition is used for switching among different frame views of the virtual control desk, while head tracking techniques are used for specific frame exploration to access indicators and controls. Body tracking is similar of head tracking, but the actuation of the controls of the virtual control desk are made by computational viewing of hands, instead of mouse.

Keywords: Non-conventional interfaces, Virtual environments, Virtual control desk, Nuclear plants, Speech recognition, Face tracking, Body tracking.

1. INTRODUCTION

Nuclear power plants (NPP) involve high safety requirements in their operation, thus operators must keep them into normal operational conditions, or act appropriately and fast to bring them back to normal conditions in the occurrence of any abnormal ones. Operators must run very efficient training, to prepare facing postulated incidents or accidents. These training used to be carried out through the use of full-scope control desk models, which resembled real ones with high fidelity relatively to visual appearance. This favored training in that users could see all variable indicators, actuator controls and alarm indication in the same positioning as they would do in real control desks.

Virtual reality (VR) techniques can improve user interfacing with NPP simulators, – or simulators of any other industrial plants –, since virtual control desks (VCD) resemble much the real ones in which they are based. Therefore, the VCD approach combined with the computer-based simulators bring both the online dynamics simulation capability and the high fidelity visual appearance, favoring user training and ergonomics evaluation.

Such a VCD has been developed at *Instituto de Engenharia Nuclear* (Nuclear Engineering Institute – IEN), a research and development (R&D) center belonging to *Comissão Nacional de Energia Nuclear* (Brazilian Commission of Nuclear Energy – CNEN), (Aghina et al., 2008). An existing NPP computer-based simulator was integrated with this developed VCD through computer networking, either local or through the Internet.

New interacting modes, like natural interfaces, were included for a friendlier user interfacing, to enable user interaction in front of projection screens, for example, free from computer keyboard and mouse. These new interacting modes comprise an automatic speech recognition (ASR) system, head and body tracking systems that will be discussed in the following sections.

Three systems of computational viewing devices were tested, two head tracking systems, with and without visual markers, and Kinect, from Microsoft, for body tracking. This later device can acquire 3D data of the coordinates of the human body and then interact with the VCD. This interaction can control the viewing of the VCD and controls the positioning and actuation of display cursor, by hands. We observed that a combination of these interacting modes served better user interfacing, than using one or another alone. A comparative analysis of the two head tracking systems, and the body tracking is also performed.

2. RELATED RESEARCH

A previous R&D was carried out at IEN for the development of a computer-based NPP simulator, in cooperation with the Korea Atomic Energy Research Institute (KAERI) and with the International Atomic

Energy Agency (AIEA). This cooperation resulted in a new laboratory at IEN in 2003, named *Laboratório de Interfaces Homem-Sistema* (Human-Systems Interface Laboratory – LABIHS), (Carvalho and Obadia, 2002; Santos et al., 2008).

The developed NPP simulator was coupled with a synoptic windows-based interface that communicated with the simulator through computer networking. This whole system has been used through these years at IEN for operator training and to support ergonomics evaluation. The later led to modification proposals in the form information is presented for users, so as to improve it from the ergonomics point of view, and consequently improve operational safety (Carvalho et al., 2008; Santos et al., 2008; Oliveira et al., 2007). Figure 1 shows a view of a reduced-scale model of the real control desk in that the current VCD was based. Figure 2 shows a view of the computer-based NPP simulator room, where it is possible to notice the use of multiple computer screens, to minimize the need for switching among many frame views; even so, it may still be a difficult task.

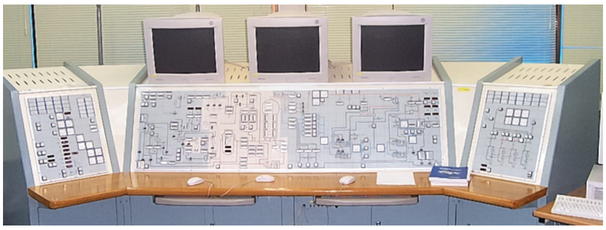


Figure 1: Reduced-scale full-scope model of the real control desk.

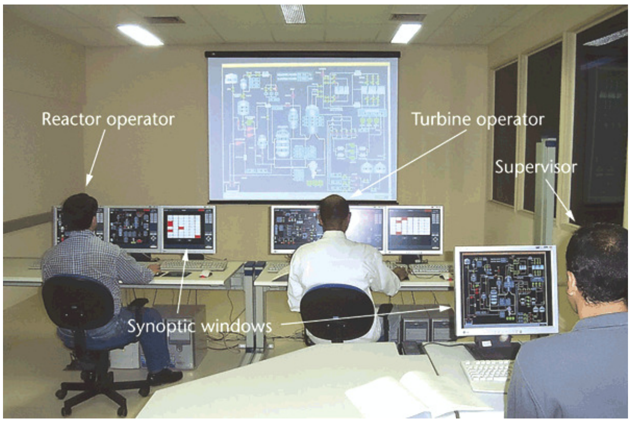


Figure 2: Computer-based NPP simulator room.

Relatively to VR approaches in designing and evaluating control desks or rooms, there are other R&D groups running similar works directed also towards the nuclear field, as can be verified by the following references (Drøivoldsmo and Louka, 2002; Nystad and Strand, 2006; Markidis and Rizwan-uddin, 2006; Hanes and Naser, 2006). The VR approaches enable the evaluation and decision making about location of variable indicators, actuator controls and alarms displays, to serve a better operator acting to run normal NPP operations and to mitigate any abnormal conditions, before the construction o real control desks. This enables also the modification of existing ones, to improve their design from the ergonomics perspective.

An overview of all the R&D developed at IEN, from the beginning with first VCD results up to the present with the most recent results is described in (Aghina et al., 2008).

3. THE VIRTUAL CONTROL DESK

The VCD was developed from the beginning to consist in an interactive interface with users, what led to the choice of OpenGL graphics pack for C/C++ languages. The VCD design was fully based on the real control desk shown in Figure 1, considering all variable indicator and actuator control types, as well as alarm indicators. These interface types were created as different classes, each on replicated as needed. Photos were used as textures for all these interfaces and for the front VCD front panel, with the schematic connections shown. Figure 3 shows a complete view of the VCD; compare it with Figure 1. Figure 4 shows a close perspective view of the VCD.

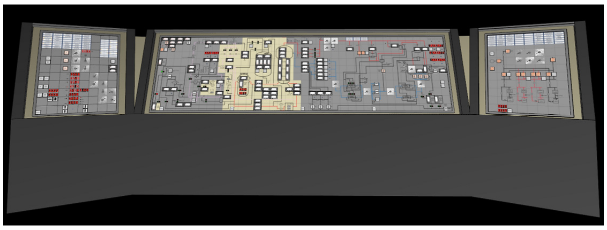


Figure 3: VCD’s complete view.

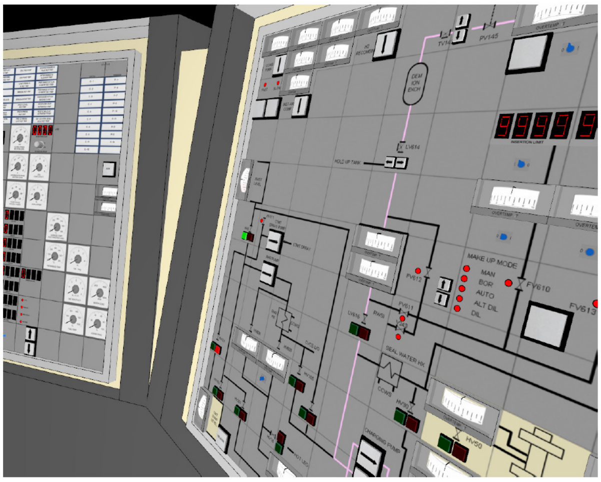


Figure 4: A close perspective view of the VCD.

4. NATURAL INTERFACES

Friendlier man system interfaces were developed to improve users’ interaction with the system, so they could interact in front of a projection screen, computer screen, or any other display, free from keyboard and mouse. This thus would enable more natural interaction forms, consequently improving user immersion. In a first stage, an ASR system was developed as the friendly interface, through which users could interact with the VCD, by spoken commands as: “left”, “right”, “up”, “down”, “zoom in”, “zoom out”, and so on. But the R&D staff soon became to try other friendly interaction modes, besides this. In particular, two head tracking techniques and body tracking were implemented and tested, and either one of them combined with the ASR system.

4.1. Speech Recognition

The ASR made use of well-known techniques for small-vocabulary isolated-word recognition system, as using Cepstral analysis (Rabiner and Schafer, 1978; Oliveira, M. V.; Moreira, D. M.; Carvalho, P. V. R.; 2007. Construção de interfaces para salas de controle avançadas de plantas industriais, *Ação Ergonômica,* Vol. 3, 8-13.
Oppenheim and Schafer, 1989) for parameter extraction, and neural networks (NN), (Haykin, 1999) for pattern recognition. The system was first implemented offline, and then upgraded to full online system, including online NN training, as detailed in the following. The speech recognition is currently performed in the following steps: (i) speech detection; (ii) end-point detection; (iii) word segmentation; (iv) parameter extraction; (v) pattern recognition. All steps are detailed.

In the first step above (i), speech sound is automatically detected through a specified threshold, adjusted experimentally, to identify speech above background noise, and start recording.

Then, in the second step (ii), once recorded with a fixed (sufficient large) time window, – also adjusted experimentally –, end-point detection is performed to isolate the spoken word itself from the background noise, following a simple approach using short-time energy (Pinto *et al.*, 1995).

In the third step (iii), the (already isolated) spoken word is segmented in an approximate range of 30 ms, where speech can be considered stationary (Rabiner and Schafer, 1978). Segmentation is performed with Hamming window (Rabiner and Schafer, 1978), with a fifty-percent superposition to compensate for attenuation in the segments' ends (Lima *et al.*, 2000; Diniz *et al.*, 1999).

Parameter extraction (iv) is performed in a simple and readily implementable form as the Cepstral coefficients obtained from the Fourier transform analysis (Lima *et al.*, 2000; Diniz *et al.*, 1999). The speech signal $s(n)$ can be considered as a convolution between an excitation signal $u(n)$ with the human vocal tract $h(n)$, as shown by Equation 1. The vocal tract is a time-varying system modeled by the variable filter $h(n)$; this is the reason because the speech signal must be segmented in a short-time range before being processed by the Fourier analysis. Then, speech is analysed in both time and frequency, leading to the spectrogram, as shown in Figure 5. Cepstral analysis performs deconvolution between excitation and vocal tract response, according to Equation 2. The later one can be used for pattern recognition. Twelve Cepstral coefficients belong to the vocal tract (Rabiner and Schafer, 1978) were extracted, discarding the zero-index one (Deller *et al.*, 1993), along fifty segments, resulting in six hundred parameters per word for pattern recognition.

$$s(n) = u(n) * h(n) \tag{1}$$

$$c_S = IDFT\{\log|DFT[s(n)]|\} = IDFT\{\log|S(k)|\} = \tag{2}$$

$$= IDFT\{\log|U(k)||H(k)|\} = IDFT\{\log|U(k)| + \log|H(k)|\} = c_U + c_H$$

where:

- IDFT: inverse discrete Fourier transform
- $S(k)$: DFT of the speech signal
- $U(k)$: DFT of the excitation
- $H(k)$: DFT of the vocal tract
- c_S: Cepstrum of the speech signal
- c_U: Cepstrum of the excitation
- c_H: Cepstrum of the vocal tract

The six hundred-dimensional data form a random vector that, with all realizations comprised by the repetitions for all the commands to be recognized, form a data set that is used for training a NN for the pattern recognition stage. In a former implementation, two parallel NN of different topologies were used for a voting system (Jorge *et al.*, 2010).

Currently, only one feed-forward NN (FFNN) trained with a more robust backpropagation (BP) -based training algorithm is used instead. The training algorithms implemented perform adaptation of learning rates (Jacobs, 1988; Cichocki and Unbehauen, 2003), what leads to faster convergence in flat regions in the search space, while slower convergence around minima. The effect of this approach is a more robust convergence in regions quite different from parabolic-like minima, as steep gutter-like ones, for example. This originated a class of backpropagation algorithms known as resilient backpropagation (RPROP), (Riedmiller and Braun, 1992), more robust for convergence.

The current implementation is based in part on the Silva and Almeida algorithm (Silva and Almeida, 1990; Cichocki and Unbehauen, 1993), according to Equation 3. Global adaptive learning rate was also implemented (Cichocki and Unbehauen, 2003), similar to the local one but using only one learning rate for the whole NN. The implemented code enables some choices for users, as: (i) online or batch training (Haykin, 1999); (ii) possible use of moment (Haykin, 1999); (iii) global or local adaptive learning rate (Cichocki and Unbehauen, 2003).

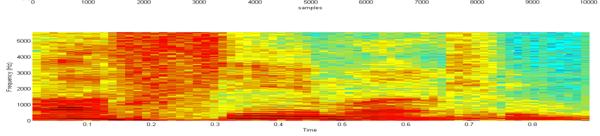


Figure 5: The time-domain spoken command "acima" (Portuguese word for "up") in the upper part; and its corresponding spectrogram in the lower part, where it is possible to notice the time-varying nature of the speech spectrum.

4.2. Head Tracking

Two approaches for head tracking were implemented: with and without visual markers. But in both cases, the purpose is to enable a more natural interaction between users and system. Head tracking turned out to be an interesting possibility after some experimentation with the ASR system. A user had to talk commands to do anything during operation of the VCD, since moving its view up, down, left or right to zooming it in or out, or switching view among different VCD's modules (see Figure 3, where it is possible to see three main VCD's modules). Imagine that user had to have a close view of some detail in the left module, for example; he or she would have to switch view to that module, then adjust the zoom to a specific indicator, every action controlled by voice commands; it might be a difficult task to speak repeatedly "left", "right", and so on, until focusing in the desired indicator. Thus, the R&D staff became searching for other types of interaction modes. In fact, both the ASR and the head tracking systems can be combined depending upon the task to be executed. With the head tracking approach, the above mentioned task (seeing a detail information in the left VCD module) could be executed by moving user's head to the left, and the image would then turn to that side; and to look in more detail an indicator, he or she needed to approach head towards the screen, and the projected image would zoom in. Head tracking is based on a six-degree of freedom (6 DoF) information, three for displacements relatively to the three Cartesian axes, and other three for rotation angles relatively also to the same three axes, as illustrated by Figure 6. This work makes use of only three degrees of freedom that are, according to Figure 6: (i) yaw, (ii) pitch and (iii) forward-back, the later to enable zooming in and out the VCD.

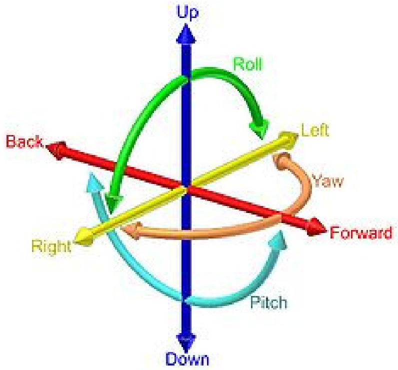


Figure 6: 6 DoF for head tracking.

4.2.1. Head Tracking with Visual Markers

This interaction mode makes use of visual markers attached to user's head, with an infrared (IR) sensor, – Trackir5, supplied by Natural Point (www.naturalpoint.com) –, enabling head pose estimation. Three reflective markers are fixed at user's head. The pose is estimated based on projective geometry computation, by a freeware library, – OptiTrack –, supplied by the same company. Tests showed this approach results in a good accuracy in head pose estimation. Figure 7 shows a view of this interaction mode in operation.

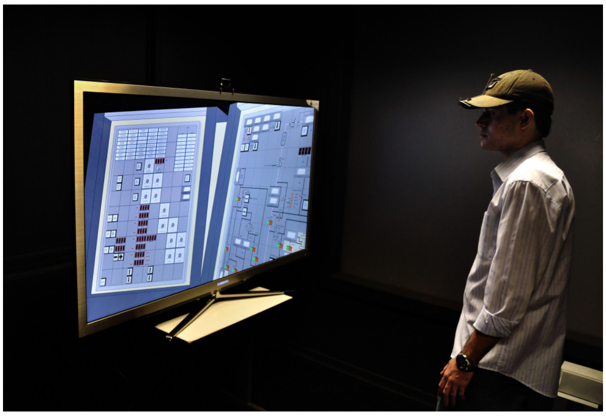


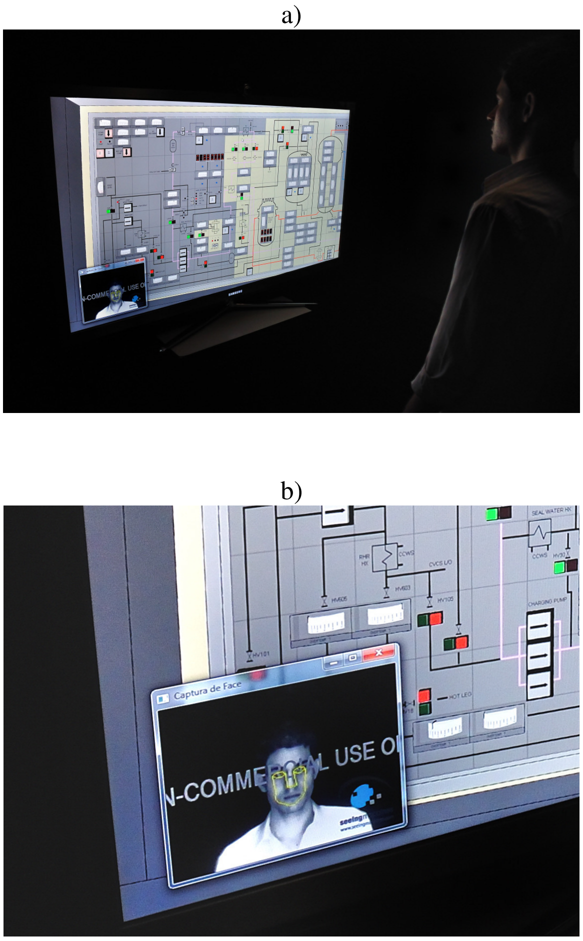
Figure 7: The head tracking interaction mode with visual markers.

4.2.2. Markerless Head Tracking

Another approach was also implemented and tested, markerless head tracking system. The code used performs head pose estimation based on tracking some points detected in users face, with six degrees of freedom, and is a proprietary library named FaceAPI, supplied by Seeing Machines (www.seeingmachines.com). The source code is not available, and the company does not supply any details about the tracking methodology used. Thus, it is used as an executable called by our application. It operates with either webcams or IR cameras.The disadvantage of this approach, after performing some tests, is that it does not has good accuracy as the other approach with visual markers, specially when user turns his or her head to the sides; at twenty degrees to both sides, the estimation of head angle sometimes oscillates, making the VCD image on screen to shake for both sides. The advantage is that interaction is more natural, since there is no need of using markers on user's head.

Another disadvantage is that this system can not be used with head-up display. Since it is trained to detect faces, any other device in one's head makes face undetectable by the code. Other details such as glasses or beard can also cause problems in face detection.

Besides this, tests showed that using the head tracking system (be it with or without markers) as the only interaction mode would also cause problems, similarly as pointed out in the first paragraph of section 4.2. In that case, executing all commands by voice would be a difficult task, but in the present one too, as explained in the following. Imagine the same situation mentioned in the first paragraph of section 4.2, repeated here for convenience: say a user had to have a close view of some detail in the left module, for example; he or she would have to switch view to that module, then adjust the zoom to a specific indicator, every action controlled by head movements. It might be a difficult task and cause discomfort for user to move head to the left sufficiently to switch to the left VCD module's view, and then approaching his or her head to the projection or computer screen for zooming in, until focusing in the desired indicator. That is the reason for using both interaction modes, as explained in Section 4.2. Figure 8 shows a screen shot of the markerless head tracking system in operation; Figure 8a shows a whole view of this system, while Figure 8b shows a detailed view of the points tracked in the user's face.


a)

b)

Figure 8: a) The markerless head tracking system; b) A detailed view of the points tracked in the user's face.

4.3. Body Tracking

The Kinect sensor was used for a full computational viewing of the operator body, to operate the VCD without any kind of mechanical interface with the operator.

The Kinect sensor was originally intended to be a device that recognizes the user's movements to the MicroSoft Xbox 360 (www.xbox.com), which allows him to control games through gestures and voice commands. Its main hardware components are an RGB camera, a depth sensor that consists of an infrared light emitter and a camera to get this emitting light, multi-array microphones, an engine tilt, and three-axis accelerometer.

The Kinect depth sensor (wikipedia.org/wiki/Kinect) consists of an infrared light source, an emitter projecting a pattern of dots that are read back by a monochrome infrared camera, which is called structured light (wikipedia.org/wiki/Structured light). The sensor detects segments reflected by the pattern dots and converts the images into a depth map, so that the image has, besides the x, y coordinates, the z axis distance of the objects in the scene image to the Kinect sensor.

In this project, we used the open source OpenNI library that contains a middleware (NITE), which, through its algorithm, can recognize the shape of the user's body and provides the x, y, z body coordinates, such as: head, shoulders, chest, hands, etc. so that through them you can make an interaction with VCD. It also presents the image using the Kinect OpenGL graphical library format, which is the same used in the VCD.

To capture the coordinates three softwares were used:

1) OpenNI

The OpenNI is a software design framework (OpenNI.Org) focused on interoperability of natural interaction devices.

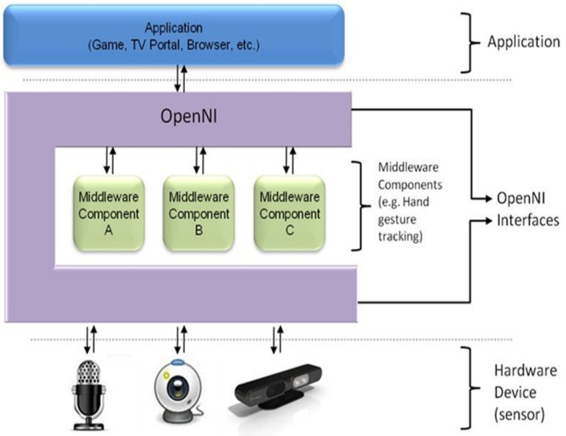


Figure 9: OpenNI Arquiteture

2) Kinect Driver sensor for Windows

The idea of OpenNI is to support various types of natural interaction devices. This specific Microsoft Kinect Sensor driver for Windows must be installed (github.com/avin2/SensorKinect).

3) NITE

This software is a middleware made to operate together with OpenNI. It was developed by Prime Sense (www.primesense.com/nite), which is the company that makes the Kinect hardware. It analyzes the image with depth information generated by Kinect and generates the x, y, z point coordinates of the operator's body.

We used an application of NITE, "players", which generates 14 coordinates x, y, z of the skeleton of the operator, as shown in Figure 10. The advantage of using this application is that the generated image is made using the OpenGL visualization library, which is the same used in VCD, facilitating the integration of the two programs. The final program displays two windows and the skeleton of the VCD, as can be seen in figure 11.

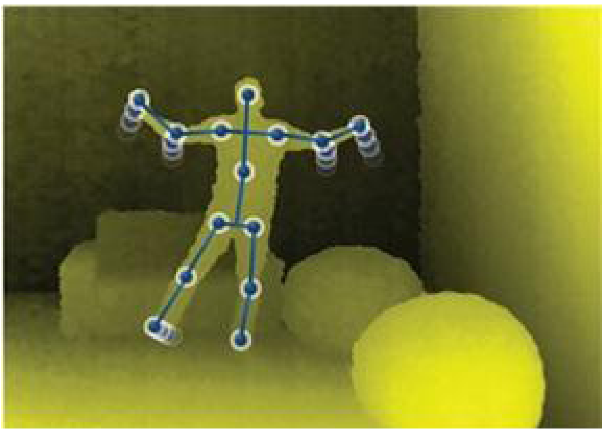


Figure 10: Joints of the skeleton generated by the "players"

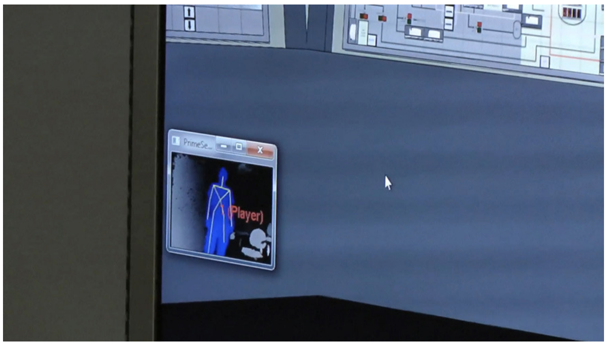


Figure 11: Image of skeleton integrated with VCD

The desired controls for natural interaction with VCD are:

- VCD rotation on its vertical axis.
- Zoom to Observe Details of the VCD.
- Control Actuation of VCD.
- Show Predetermined Regions of Interest of the VCD.

4.3.1. VCD rotation on its axis vertical

The NITE model treats the head as if it were a unique skeleton joint as is impossible to make head tracking, with a unique coordinate.

To face the problem of getting the head tracking for VCD rotation we used the body rotation angle of the shoulders position. The arc tangent between the coordinates of the shoulders was used, where:

θ = atg ((Δ shoulders z) / (Δ shoulders x)) (3)

This approach is practical because it is independent of the distance from the sensor to the operator.

4.3.2. Zoom to observe details of the VCD

The Zoom function was used with the z coordinate of the chest, making the visualization of VCD turn larger or smaller, with the approach or departure of the user's body.

4.3.3. Control actuation of VCD

The control actuation of the VCD is made by the buttons of the desk, that are the way to pass the user information to program operation of the simulator, such as: turn pumps on or off, deploy control rods in a deliberate shutdown of the reactor, etc.

With Kinect, the cursor movement is done by moving the right hand. With OpenNI NITE middleware, using the application "Players", it is possible to extract the centroid x, y position of the right hand in the space depth map.

With the mouse Windows API, it is possible to match the hand's position in 640X480 points resolution (depth map) to place the cursor on any size of viewed VCD screen.

For the actuation of the control buttons of the VCD, we used a methodology to recognize the opening and closing of the right hand to click the cursor.

NITE provides the position of the centroid of the hand. The space around the centroid of the hand was analyzed in the depth map. This area constitutes a square of approximately 20 cm sides that corresponds to a 50x50 dot matrix in the dept map.

With access to this matrix, we can analyze the difference of the Z coordinate of each point, relative to the Z coordinate of the centroid point of the hand, as seen in Figures 12 and 13. If the absolute value of each difference is less than 20 cm, the result is included in a summation. As the area of the open hand on the matrix is greater than that of the closed hand, the summation of the differences is bigger. This process is repeated for each generation of VCD frame. If the summation is 40% larger or smaller than the previous one, it is determined whether the hand is open or closed and thus the click is triggered or not.

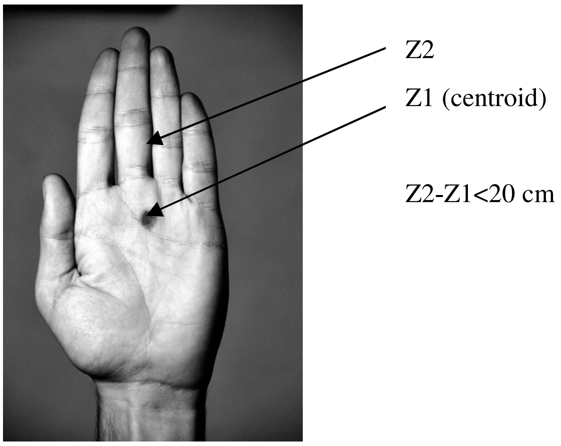


Figure 12: Open hand Centroid

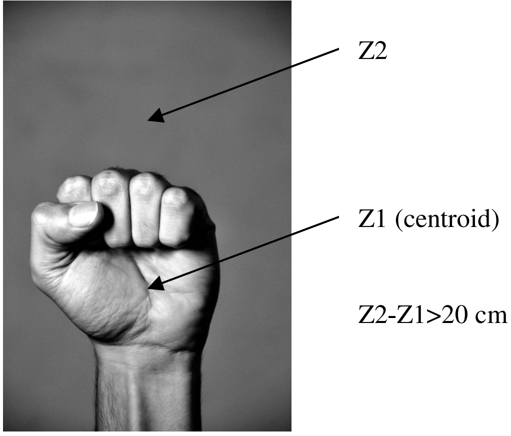


Figure 13: close hand Centroid

4.3.4. Predetermined Regions of Interest of the VCD

The VCD has five predetermined regions of interest to facilitate rapid observation of the VCD by the operator. To operate Kinect we used the X axis of the left hand movement. The image of the VCD is calculated continuously, generating multiple frames per second. In each VCD frame the current position is analyzed and compared to the previous left hand position, calculating the centroids differences in relation to the x-axis. A negative difference decreases the region of interest and a positive increases it, making a sequential selection of the regions of interest.

5. RESULTS

5.1. Comparative Analysis between marker-based and markerless head tracking systemsis Length of the Paper

This section shows a comparative analysis between the two head tracking methods used in this work, with and without markers. The user's head angles for the left and right sides were measured for both methods, approximately around three positions: (i) frontal, with user looking directly to the screen; (ii) rotated to the left, approximately at –20 degrees; and (iii) rotated to the right, approximately at +20 degrees. Table 1 shows results for the marker-based method (using TrackIR5), while Table 22 shows similar results for the markerless method (using FaceAPI). Each line in both Tables is in fact an average among ten measurements. One can verify the higher standard deviation (Std. Dev.) for the later method, what means a higher pose estimation instability for angles around 20 degrees, to the left or to the right sides. Thus it follows some advantages and disadvantages of both methods in Table 3.

Table 1: Marker-based tracking performance

	Frontal		Rotated to the left		Rotated to the right	
	Average (Degrees)	Std. Dev. (Degrees)	Average (Degrees)	Std. Dev. (Degrees)	Average (Degrees)	Std. Dev. (Degrees)
1	0,508	0,200	-26,72	0,190	24,369	0,097
2	1,306	0,177	-22,15	0,134	27,679	0,139
3	0,256	0,237	-20,22	0,151	27,696	0,125
4	-0,758	0,214	-25,20	0,156	28,001	0,129
5	0,566	0,291	-21,48	0,078	26,597	0,151
6	-1,179	0,263	-17,06	0,115	26,966	0,172
7	-0,153	0,183	-18,28	0,135	28,669	0.127
8	0,015	0,155	-19,02	0.087	24,890	0,159
9	-0,744	0,140	-22,31	0,083	26,795	0,115
10	0,657	0,211	-22,03	0,170	20,284	0,180

Table 2: Markerless tracking performance

	Frontal		Rotated to the left		Rotated to the right	
	Average (Degrees)	Std. Dev. (Degrees)	Average (Degrees)	Std. Dev. (Degrees)	Average (Degrees)	Std. Dev. (Degrees)
1	-0,566	0,231	-16,24	0,576	14,218	1,927
2	-0,432	0,153	-15,05	1,396	14,307	1,509
3	-0,625	0,122	-16,90	1,286	18,249	1,889
4	-0,512	0,161	-16,20	0,963	15,518	1,462
5	-0,072	0,119	-14,59	1,589	13,266	4,203
6	-0,619	0,155	-13,22	0,586	16,255	0,607
7	-0,336	0,406	-20,97	0,343	17,923	1,18
8	-1,937	0,174	-17,44	0,531	13,549	0,353
9	1,117	0,3425	-15,28	1,475	12,742	0,552
10	-1,218	0,164	-16,38	0,861	14,417	0,648

Table 3: Comparative analysis between the marker-based and the markerless tracking methods relatively to some usage parameters

	Marker-based method	Markerless method
Camera type	Proprietary IR camera	General purpose camera (webcam)
Camera characteristics	Operates with low illumination	Higher tracking errors for low sensitivity cameras and low frame rates (requires minimum of 30 fps)
Pose estimation	Robust pose estimation even for high rotation angles	Poorer pose estimation for high rotation angles
Tracking errors	Losses tracking if there are reflexive objects in scene	Do not loose tracking when new face appears in scene

INITIALIZATION AND OPERATION	Immediate tracking from the beginning	Requires from 5 to 6 s from initial to current tracking
SOFTWARE AND LIMITATIONS	Free software, can be used for commercial purpose by $150.00 (including camera)	Free software demo for non-commercial applications, full software with paid license
COUPLING WITH HEAD-UP DISPLAY	Can be coupled with head-up display	Can not be coupled with head-up display

5.2. Interaction based on the ASR system

Interaction based on the ASR system was analyzed by some figures of merit, as explained in the sequel. As already mentioned, the current ASR system's implementation makes use of FFNN. First figure of merit is the word error rate (WER), defined as:

$$\mathrm{WER} = (S + I + D)/N \qquad (4)$$

where:

- *N*: total number of words to be recognized
- *S*: number of replaced words
- *I*: number of inserted words
- *D*: number of deleted words

The three later are errors. WER should be ideally zeroed. Another figure of merit is the word recognition rate (WRR), defined as:

$$\mathrm{WRR} = (1 - \mathrm{WER}) \qquad (5)$$

which should be ideally one. We show these later results, in the sequel.

Another important figure of merit is the real time factor (RTF), defined as:

$$\mathrm{RTF} = TP/TA \qquad (6)$$

where:

- *TA*: the input word duration
- *TP*: corresponding processing duration.

This should ideally be as low as possible.

Experiments were carried out by using cross-validation (Haykin, 1999), where the data set was split into a number of subsets, some of them comprising the training subset, and the remaining, the test subset. Learning was repeated, each time using a different subset as the test one. Table 4 shows the scheme adopted. The whole set was split into four subsets, three of them comprising the training subset at each learning repetition. Bold characters indicate the test subsets. The last two columns show the figures of merit's results for each individual experiment run.

Table 4: Cross-validation scheme adopted

	Subsets				WRR (%)	RTF
Experiment 1	**25 %**	25 %	25 %	25 %	94.5	0.026
Experiment 2	25 %	**25 %**	25 %	25 %	98.2	0.023
Experiment 3	25 %	25 %	**25 %**	25 %	97.1	0.023
Experiment 4	25 %	25 %	25 %	**25 %**	93.5	0.025

The resulting average WRR among the four learning repetitions was: WRR = 95.8 %. The resulting average RTF was: RTF = 0.024.

5.3. Body Tracking

Tests were made taking the screen cursor in a lower region of the screen and placing it on a upper button and triggering it by closing the hand. This procedure was done with two users to test diversity operation between users.

Thirty samples were conducted of two users in two sets of tests, separated by a space of time, so that the sample would not be users addicted (Table 5).

Table 5: Press Button Tax error

30 presses (unit 1)				30 presses (unit 2)		
	Right presses	Wrong presses	error (%)	Right presses	Wrong presses	error (%)
user 1	26	4	13,3	25	5	16,6
user 2	24	6	20	27	3	10

6. CONCLUSIONS

The developed VCD seems to be a very good alternative relatively to the synoptic windows-based approach, by aggregating both the high fidelity visual appearance with the corresponding real control desk, and the computer-based PWR NPP simulation system's functionalities.

Relatively to the interaction modes, it seems that a combination of the ASR with either one of the head tracking systems would result in a more natural users' interaction with the VCD. Each interaction mode is used for the tasks it serves best, with ASR performing better for switching among different VCD views, – what one could call macro movements –, and with head tracking performing better for small movements within a particular VCD module view, – what one could call micro movements.

The intention to use the Kinect Sensor, was integrate all the above mentioned interfaces in one device. Making use of the body to control the VCD. Changing regions of interest using the left hand, their microphones for voice commands and the right hand to control the screen cursor and your click, making the user interface with the VCD more natural as possible. With the use of the various possibilities of interfaces, the intention of this work was offer to the user several options for their interaction with the VCD, and so enable him to make a choice for their preferred use. MCV developed, with its low cost of construction, its similarity to the original control board and its natural interfaces, which help the user to have interaction more friendly, show the originality of this work in the nuclear field.

7. REFERENCES

Aghina, M. A. C.; Mól, A. C. A.; Jorge, C. A. F.; Pereira, C. M. N. A.; Varela, T. F. B.; Cunha G. G.; Landau, L.; 2008. Virtual control desks for nuclear power plant simulation: improving operator training, *IEEE Computer Graphics and Applications*, Vol. 28, No. 4, 6-9.

Aghina, M. A. C.; Mól, A. C. A.; Jorge, C. A. F.; Espírito Santo, A. C.; Freitas, V. G. G.; Lapa, C. M. F.; Landau, L.; Cunha, G. G.; 2011. Virtual control desks for nuclear power plants, In: Tsekov, P.; *Nuclear Power – Control, Reliability and Human Factors*, InTech: Rijeka, Croatia, 393-406.

Carvalho, P. V. R.; Obadia, I. J.; 2002. Projeto e implementação do laboratório de Interfaces Homem Sistema do Instituto de Engenharia Nuclear, *Revista Brasileira de Pesquisa e Desenvolvimento*, Vol. 4, No. 2, 226-231.

Carvalho, P. V. R.; Santos, I. J. A. L.; Gomes, J. O.; Borges, M. R. S.; Guerlain, S.; 2008. Human factors approach for evaluation and redesign of human-system interfaces of a nuclear power plant simulator, *Displays*, Vol. 29, No. 3, 273-284.

Cichocki, A.; Unbehauen, R.; 1993. *Neural Networks for Optimization and Signal Processing*, John Wiley & Sons.

Deller Jr., J. R.; Proakis, J. G.; Hansen, J. H.; 1993. *Discrete-Time Processing of Speech Signals*, MacMillan, New York.

Diniz, S.; Thomé, A. G.; Santos, S. C. B.; Silva, D. G.; 1999. Automatic speech recognition: a comparative evaluation between neural networks and hidden markov models, *International Conference on Computation Intelligence for Modeling, Control and Automation (CIMCA 99)*, Vienna, Austria, February 17 – 19.

Drøivoldsmo, A.; Louka, M. N.; 2002. Virtual reality tools for testing control room concepts, In: Liptak, B. G.; *Instrument Engineer's Handbook: Process Software and Digital Networks*, third ed., CRC Press.

Foley, J. D.; Wallace, V. L.; Chan, P.; 1998. *Human Computer Interaction*, Prentice-Hall.

Hanes, L. F.; Naser, J.; 2006. Use of 2.5D and 3D technology to evaluate control room upgrades, *The American Nuclear Society Winter Meeting & Nuclear Technology Expo*, Albuquerque, NM, 12 to 16 November.

Haykin, S.; 1999. *Neural Networks – A comprehensive foundation*. Upper Saddle River: Prentice-Hall.

IAEA TECDOC 995; 1998. Selection, specification, design and use of various nuclear power plant training simulators.

ICRP Publication 60; 1991. ICRP Publication 60 – Recommendations of the International Commission on Radiological Protection.

Ishii, H.; 2008. The Non-conventional User Interface and Its Evolution, *Communications of the ACM (CACM) – Special Issue "Organic User Interfaces"*, Vol. 51, No. 6, 32-36.

Jacobs, R. A.; 1988. Increased rates of convergence through learning rate adaptation, *Neural Networks*, Vol. 1, 295-307.

Jorge, C. A. F.; Mól, A. C. A.; Pereira, C. M. N. A.; Aghina, M. A. C.; Nomiya, D. V.; 2010a. Human-system interface based on speech recognition: application to a virtual nuclear power plant control desk, *Progress in Nuclear Energy*, Vol. 52, No. 4, 379-386.

Jorge, C. A. F.; Mól, A. C. A; Couto, P. M.; Pereira, C. M. N. A.; 2010b. "Nuclear plants and emergency virtual simulations based on a low-cost engine reuse", In: P. V. Tsvetkov (Ed.); *Nuclear Power*, InTech: Rijeka, Croatia, 367-388.

Kim, Y. H.; Park, W. M.; 2004. Use of simulation technology for prediction of radiation dose in nuclear power plant, *Lecture Notes in Computer Science*, Vol. 3314, 413-418.

Lima, A. A.; Francisco, M. S.; Lima Netto, S.; Resende Jr., F. G. V.; 2000. Análise Comparativa de Sistemas de Reconhecimento de Voz, *Simpósio Brasileiro de Telecomunicações*, Gramado, Rio Grande do Sul, Brazil, p. 001-004.

Markidis, S.; Rizwan-uddin; 2006. A virtual control room with an embedded, interactive nuclear reactor simulator, *The American Nuclear Society Winter Meeting & Nuclear Technology Expo*, Albuquerque, NM, 12 to 16 November.

MSDN, Microsoft Developer Network; http://msdn.microsoft.com (Most recent access in April 2012).

NaturalPoint a; TRACKIR, by Natural Point, Inc.: http://www.naturalpoint.com/trackir/ (Most recent access in April 2012).

NaturalPoint b; OptiTrack, by Natural Point, Inc.: http://www.naturalpoint.com/optitrack/ (Most recent access in April 2012).

Nitendo; Wii system, by Nitendo of America, Inc.: http://www.nitendo.com/wii/ (Most recent access in April 2012).

Nystad, E.; Strand, S.; 2006. Using virtual reality technology to include field operators in simulation and training, *27th Annual Canadian Nuclear Society Conference and 30th CNS/CNA Student Conference*, Toronto, Canada, 11 to 14 June.

Oliveira, M. V.; Moreira, D. M.; Carvalho, P. V. R.; 2007. Construção de interfaces para salas de controle avançadas de plantas industriais, *Ação Ergonômica*, Vol. 3, 8-13.

Oppenheim, A. V.; Schafer, R. W.; 1989. *Discrete-Time Signal Processing*. Englewood Cliffs: Prentice-Hall.

Pinto, R. G.; Pinto, H. L.; Calôba, L. P.; 1995. Using neural networks for automatic speaker recognition: a practical approach, *38th IEEE Midwest Symposium on Circuits and Systems*, Rio de Janeiro.

Rabiner, L. R.; Schafer, R. W.; 1978. *Digital Processing of Speech Signals*. London: Prentice-Hall.

Riedmiller, M.; Braun, H.; 1992. RPROP – a fast adaptive learning algorithm, *Seventh International Symposium on Computer and Information Sciences (*ISCIS VII*)*, Antalya, Turkey.

Santos, I. J. A. L.; Teixeira, D. V.; Ferraz, F. T.; Carvalho, P. V. R.; 2008. The use of a simulator to include human factors issues in the interface design of a nuclear power plant control room, *Journal of Loss Prevention in the Process Industries*, Vol. 21, No. 3, 227-238.

SeeingMachines; FaceAPI, by Seeing Machines: http://www.seeingmachines.com/product/faceapi/ (Most recent access in April 2012).

Silva, F. M.; Almeida, L. B.; 1990. Speeding up backpropagation, In: Eckmiller, R. (Ed.); *Advances of Neural Computers*. Elsevier Science Publishers, p. 151-158.

SIMULATION MODEL TO ANALYZE TRANSPORT, HANDLING, TEMPORARY STORAGE AND SORTING ISSUES: A VALUABLE WAY TO SUPPORT LAYOUT, SYSTEM DEFINITION AND CONFIGURATION AND SCHEDULING DECISION

Sergio Amedeo Gallo[(a)], Riccardo Melloni[(b)], Teresa Murino[(d)]

[(a, b)]Department of Engineering Enzo Ferrari, DIEF (ex Department of Mechanical and Civil Engineering, DIMeC),
University of Modena and Reggio Emilia, ITALY
[(e)] Department of Chemical, Material, and Industrial Production, University of Naples – Engineering Faculty– Federico II
Napoli, ITALY

[(a, b)](sgallo, riccardo.melloni)@unimore.it, [(c)]teresa.murino@unina.it

ABSTRACT

The following paper describes the use of simulation models to analyze a common production scenario in manufacturing plants: many assembly lines producing specific families of items in a large variety of versions; following handling and transport system, evaluated to be effective, costless, with an adequate capacity to be not the bottleneck of the whole system, and integrated with a storage/sorting system to decoupling the assembly phase and the following ones as completion/test and expedition phase, included recovery and reform of the expedition lot.

To analyze a system like this, a simulation model has been developed with a flexible AGVs transport system, a Transit Point Warehouse acted by AS/RS, used also as sorting media. The size of the warehouse incoming bay, number of warehouse sub allocation zones or aisles, the capacity of all queues, all control logics for AGV's have been evaluated in a pre-modeling phase, and then evaluated by the model.

Keywords: AS/RS, Simulation Models, Handling and Transport Systems, AGV Systems, DSS.

1. INTRODUCTION

This paper focus on the evaluation among different solutions and layout configuration of the transfer and handling in a definite and already existing assembly system, and with the definition of many process parameters, with the aim to reach a better control on information about transferring process, an, moreover, an improved attitude to support productive contest evolution, that tends to an increase of levels of flexibility and mix and volumes variation.

Moreover, all operating logics have been evaluated, too.

Some results and related considerations have been outlined in previous papers, Gallo S. A., Melloni R. (2007), adopting as analysis tool a simulation model with a AGVs, diffuse and multi allocated storage system.

The original system we started from, is a production and assembly plant of engines, produced on demand, and in a large amount of versions. The part of the production/assembly system we considered start with some assembly lines each one configured to assembly just defined typologies of item families. Items travel in the system among distinct shops bringing on information, in the form of attributes, to define the sequence to follow and many process parameter.

Any item family is characterize by size, weight, relevant features, but, depending on customers specifications, national laws, final finishes or service to act, the number of item exploits to hundreds.

Furthermore, items have to be tested and finished in two other shops or areas, first one is at the end of the plant area for safety and noise problems, separated from the main area by a wall, instead, the finishing area is at the end of the assembling lines, back in the flow.

Each of the phases is decoupled with the others: we consider as time horizon of reference, a day, that is divided in single or double shift depending on the specific area of the process. In this time slot, labor rates for all systems, have to be equal, and the throughput must be those required by the Production Plan, PP.

The issue and proposal of the analysis is the definition of a new layout, new logics and new handling, storage, sorting systems. Furthermore, we had to respect some constrains, to define all configuration parameters, and maximize the system performance in terms of efficiency, cost, flexibility: problems of sizing of the specific parts of the systems, of defining the number and the capacity of vehicles, of dimensioning buffering areas and related means, emerge.

The issue addressed concerns the analysis of a possible alternative solution for the internal distribution of the engines from the assembly phase, through testing and finishing departments, up to the shipments area. The original distribution was carried out by a rigid transport system, a conveyor. This system performed its task, but it offered a high level of stiffness for production and distribution, as well as structural rigidity, preventing internal enhancement of the layout, such as the easy feeding of finishing materials.

We looked for a collecting, transport, storage and sorting system that present an high level of continuity, efficiency improvement, traceability and tracking of items. It should develop a system that allow continuous distribution and an increase in the current efficiency.

The previous research had focused on the overall analysis of the system, the quantification of flows of items, on the identification of the logical operation of the system, the definition of a model representing the as-is system in the real operating configuration. Successively, this was followed by a pre-analysis of the system, its constraints, and its need for the aspects related to the handling, to the interoperational storage, with the aim of the identification of one or more types of systems adequate to the replacement of the conveyor; after this preliminary analysis, we propose some interesting solution, a system that provided for an intermittent transport of groups of engines, loaded on racks, distributed as dynamic buffers at all operational phases, handled by automatic guided vehicles AGV (Towing Load).

In this work, as a replacement system, instead of the main collecting conveyor, it was considered an AGVs system, but with the presence of an automated warehouse, AS/RS, to decoupling the stages, that is the solution we refer to in the present work.

We can observe the representation of the structure of the model made in the simulation. You can see that the AS/RS stands along the entire area covered by the test brakes and the finishing department.

The need of the assessment of alternative systems lies in the difficulty that the existing system has to support the evolutionary scenario over time: the requests for supply of engines have been transformed by requests for large quantities of the same type and version to small lots with an high customization level. This made production planning and resource scheduling much more complex, as the line balancing, material flows management, production planning, and the system is increasingly required a high level of reconfigurability in terms of flexibility and elasticity.

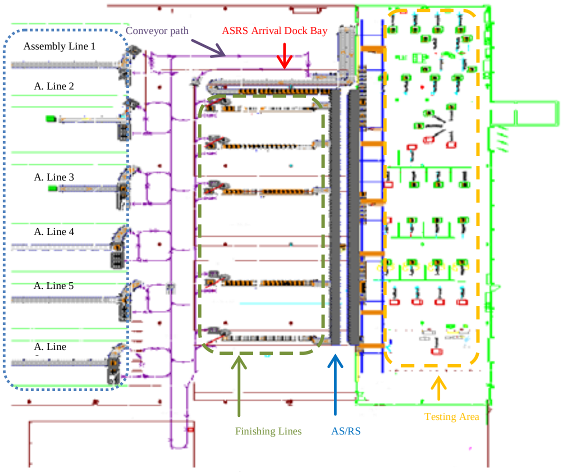


Figure 1: Snapshot of System layout in the present configuration.

Additionally, the main conveyor represents a fixed installation which limits the viability of the area and induced limitations to the realization of an efficient power supply system of the component parts, such as the feeding to the finishing lines.

The main conveyor also acts as a decoupling/storage system between the assembly stage and the other of testing and finishing, due to the different capacity which they present. This aspect, due to the lack of identification systems of the specific engine at a specific position of the main conveyor, makes engines traceability impossible, engines that, sometimes, for the space limitations may not be downloaded at their destinations, leaving many engines turning on the conveyor for many times, with break of sequence and need of reconstructing lots.

2. LITERATURE REVIEW

An important place in this work it will play the discrete event simulation, used to verify the goodness of the solutions, and to support decisions about system configurations, to evaluate obtained outputs, used in the comparison of the solutions.

One very interesting development of the simulation is to support operational simulation as a planning and control of short-term production and logistic systems, with the creation of simulation models detailed and updated as much as possible, that, in integration with enterprise information systems make possible to simulate in real time, in parallel to the real system, and evaluate the different decision alternatives.

The areas to which the operational simulation can benefit are design, scheduling, capacity planning and control. Cho, S. (2005).

Ceric and Hlupic (1993) present an approach to use simulation in evaluating different system configurations between the various alternatives. The conceptual model is made with active cyclic diagrams. The simulation result in a high level of complexity due to the wideness and dynamics of the system. They are useful when the real system to compare with, does not exist, the validation is made of independent type verification and validation based on face validity (expert consultation) and not on statistical analysis to measure and error checking.

Simulation models can be used to improve the performance of production systems, and in this case, the process of improving the performance tend to the evaluation of the system in terms of interactions and interdependencies of the elements of the system. An example is presented by Alan and Pritsker (1997), as regards the analysis of the performance of existing systems. The authors use simulation to analyze the performance improvement process, and focus on criticality of the system, interactions and dynamics.

Another example of performance evaluation of the system is presented by Ueno, Sotojima Takeda (1991). As in Alan and Pritsker (1997), simulation is used as a tool and a DSS to support the redesign of the production process, to identify dynamically bottlenecks or machines with the lower production rate. The aim of the simulation is to define a new configuration of the production line with the minimum cost.

The use of simulation for strategic decision making, presented by Kumar and Nottestad (2006), was aimed to redesign production system in order to improve the productive capacity. It has been developed a Discrete Events Simulation system, DES to configure a line semi-automated production as part of process to improve the existing production process capacity. The simulation is validated with Design Of Experiments (DOE), used to interpret the results and information on specific parameters. Surfaces of response were produced to evaluate system behavior depending by control factors, such as average cycle time machines, buffer capacity, MTTF of machines, number of parallel machines and the size of the lots.

An tactical level example of simulation models is provided by Watson, Medeiros and Sadowski (1997). The authors analyze the release schedule of orders for a make-to-order production. For these models MRP logic, with infinite capacity and lead time defined on historical data and past experience, is adopted. Often, schedules are obtained that are often unfeasible.

One particularly accurate use of simulation models to schedule flow shop systems has been proposed by Vaidyanathan, Miller and Park (1998). The system is divided in scheduling program, used to generate the daily schedule, and the simulation model, that uses the obtained scheduling to simulate the system and improve it.

To analyze job shop systems we find many approaches. The approach proposed by Selladurai, Aravindan, Ponnambalam and Gunasekaran (1995) tries to elaborate scheduling using the simulation directly, limited to rules of dispatching.

Systems have been created that allow allows to decompose NP-hard problems into sub problems, with the aim of solve problems of scheduling multi - objective with setup time variables.

Examples have been proposed by Yang and Chang (1998), based on a Pareto analysis in the field of production systems, for multi-objective optimization approach. The proposed methodology is compared with the traditional approach and several heuristic rules for dispatching based on numerical examples.

More recently, we can cite some study that use simulation models as cost effective means to evaluate checking and configuring the plant layout, considering both the process line features and the material handling features and carefully integrating them together toward best efficiency in Kolkka, Rajagopalan, and Suksi (2013). The study of configurations covers the layout, but also deals with the individual lines in-feeding and out-feeding systems with appropriate storage areas and system, along with the way the product is packed and handled before being shipped to the customer. The findings of study are checked by real-time simulation to arrive at the best capacity of the material handling equipment, like cranes, AGVs, transfer conveyors, building area, types of storage, cycle time.

In Malmborg (2001), is presented a model for configuring storage racks in an AS/RS systems with multi-shuttle machines. The models goes on to extend consolidate rules for sizing storage racks based on defined performance levels of system utilization. The models forecast the relative proportion of different types of order picking cycles used in a system, with the use of stochastic, analytical model of system interleaving.

In Jane and Laih (2005), in a synchronized zone order picking system, all the zones process the same order simultaneously. This paper develops a heuristic algorithm to balance the workload among all pickers so that the utilization of the order picking system is improved and to reduce the time needed for fulfilling each requested order. With a similarity measurement, a natural cluster model, which is a relaxation of the well-studied NP-hard homogeneous cluster model, is constructed.

In Perry, Ronald, Hoover, Stewart, Freeman and David (1983), is showed the use of a simulation model as a design aid for an AS/RS, where the flow of bins from storage locations to work stations via conveyor and return to original storage locations, is controlled by heuristic acted on a computer. The initial model goes into a cost effective system for achieving desired performance goals through judicious use of a detailed, stochastic simulation model.

3. SYSTEM LAYOUT DESCRIPTION

Our analysis refers to a portion of the production, handling of engines produced between departments.

The current market needs tend more and more to the creation of small and customized lots. This means to produce engines of defined types, but with levels of "adaptation" and personalization which greatly vary in relation to the use of the engine, in relation to the market, that affect legal emission levels, etc., which means that product flows very vary, subject to change that define different product codes, and thus different sequences between the stations that perform the process.

3.1. Assembly lines

The assembly area shows six parallel assembly lines distinct by family and version of product. They work in a single daily shift.

In the original system, at the end of assembly phase, the operator retrieves the motor by an hoists and places it on the main conveyor, to be moved to its next phase, the testing phase. The original main conveyor is a trays conveyor.

The engines, in next phases of the process, will be again repositioned on the conveyor to be transported in the direction of the finishing department.

The speed of the conveyor is on about 1.8 m/s, to safely allow placement and removal of engines from the operators. Conveyor conformation is continuous, with a referable complex "U" shape, and allows the completion of the entire ride in about 4.5 h.

It works both as an handling mean, but also is used as a dynamic warehouse being composed of 450 trays. It was placed between the heads of the final part of the assembly lines, and the finishing ones, but another part served the testing area. It performed all its duties, however, problems growth related to viability, material handling, tracking and traceability of the items.

3.2. Testing room, testing department

The testing area, on the right of layout, consists of six areas in which there are dynamometric brakes grouped and allocated appropriately in relation to the tests characteristics, to the capacity requirements they can perform. This phase is bottleneck for the process, and works in a double daily shift.

Each group of brakes is provided with buffering rollers, sized in relation to ensure the feeding of brakes, on movement reliability and speed, space constrains, and to decouple this phase. On each roller conveyor, operators positioned motors to be tested, unloading them, via hoists, from the main conveyor. Once the test is complete, engines were repositioned on the conveyor.

3.3. Finishing Lines and Applications

Continuing along the path of the conveyor, the tested engines were taken, by the operators of the finishing department, to be positioned above rollers, used as buffering docks, placed to the side of the specific lines. Here, operators perform a first composition of the batch of shipment, using visual recognition and information with the engines.

Originally, finishing department had five lines, and works in a single daily shift.

3.4. Shipping Area

In the left lower side of the layout there is the shipping area that follow the Finishing phase. Handlings are made by forklifts. Load Unit, LU, stored to be shipped, are reassembled to definitely constitute lots of shipment, if this activity is not done before. Identical motors of an order lot are scheduled in repeated sequences defined to achieve a good balance of the single assembly lines, and to balance the load in the referable time unit (shift - hour), feeding next steps based on available capacity offered by these. All this causes a break in the FIFO sequence and integrity of the lots, in such a condition impossible to recover.

4. DEVELOPMENT OF THE ANALYSIS

The original system of handling effectively responds to the characteristics and requirements as:

1. Handling of LU between assembly and testing area, and between testing and finishing.
2. Handling continuous LU.
3. Possibility of an intermediate buffering to decoupling all phases because of the different timing of the departments.

The solution we consider, since the preliminary assessment, and confirmed by the results of the simulations, showed better performances compared to the initial situation with regard to the opportunity of having the tracking and tracing of engines, to support the ability to trace more easily shipping batch within the system; to the level of the occupation of the soil for the storage of working or finished engines; to the opportunity of having more finishing lines with better correspondence between the line and type engine; to the increasing of the capacity to configure specific finishing on more than one finishing lines corresponding to different outlets of the warehouse; to the ability to perform the secondary material feeding to the lines, especially finishing ones, more efficiently; to the reduction of transfer times between each stage; and, finally, to the reduction of throughput limitation due to the handling system.

To meet the identified needs, an AS/RS system has been considered. After the analysis of the path the original main conveyor, after the analysis of all flows and the area of interest in the plant, considering the limitation of space, a possible location of the warehouse is represented by area along the wall in front of the finishing department toward the testing department. This positioning, furthermore, would allow a direct distribution of the LU at all finishing and testing areas, and avoid the use of further transport systems to distribute engines to the various brakes, such as the AGV themselves, conveyors, reducing their number, capacity and with a limited need of human supervision, with an affordable investment and exercise cost, and, in addition, coping with the scarcity of space.

The system that appeared most appropriate and convenient, in terms of quick response and continuity of supply, in terms of flexibility and versatility, low operating costs and the absence of limitation to the practicability of the area, as distribution and handling system with a very high level in automation, was represented by AGVs.

Different systems of transport of Unit Loads of engines from the assembly lines are critical, as already seen in previous works, because they are rigid with fixe occupation of the area, as conveyors, because of danger problems and psychological impact, as with overhead chain conveyor. The forklifts handling of single items or LU by carried by human, imply high use of labor and related costs, subjective management of movement, or in the best case, supported by systems of flow analysis to be integrated to carriers. Instead, the system based on the use of AGV should meet the automation needs in the creation of LU and their delivery to the loading bay on the AS/RS, a relevant cost of investment, but a lower operating cost, big reliability, tracking and traceability of item flows, a level of continuity of flows to feeding all parts of the system, very easy to module.

The first issue to face with, is the one of the dimensioning of the LU, as number of items to group together, and as physical dimension of the LU, considering the different sizes of the items to collect. This issue affect both the handling interface and the relative automation level, both the opportunity of standardization of the LU, both the space consuming, the number of travel missions.

In the logistic principles, it is very important have an standard type of LU for all systems that participate in the definition of the logistic system and for the entire chain to use standard equipment along all the system, to define warehouse loading bays, and avoiding the need for adaptation in the future and a structural reconfiguration of the store.

Observing the size of the engines, Table 1, may be defined a basic type of handling pallet or rack, on which to mount each engine, that could be able to support the movement until the end of the finishing phase, where it will be separated and recovered, of dimensions 800 x 600 mm, which allows to support any size increase.

Table 1: Size and Weight features of items

Features of Engines								
Type	Assembly destination	1D (l)	2D (L) [mm]	3D (h)		Weight [kg]		Volume $[dm^3]$
9LD	Gruppo 9	559	633	600		110		212,308
11LD_522	Gruppo 11	500	656	558		153		183,024
11LD_626	Gruppo 11	484	770	586		170		218,390
LDW502	VETTURETTE	407	404	472		60		77,610
LGW523	VETTURETTE	406	420	420				71,618
LGW627	SARMAS	406	412	449				75,105
LDW702	SARMAS	412	421	515,5		66		89,415
LDW1003	SARMAS	412	513	515,5		85		108,954
LDW1204	SARMAS	440	593	515,5		96		134,504
LDW1204/T	SARMAS	480	593	556,5		101		158,402
LDW1404	SARMAS	412	596	515,5		98		126,582
LDW903	SARMAS	412	513	515,5				108,954
RD2	RD	559	633	600		110		212,308
MD2	MD	464	485	498		63		112,070
MD3	MD	500	666	558		153		185,814
Analysis Outlines								
Mean		456,867	553,867	525		105,417		143,341
MAX		559	770	600		170		218,390
MIN		406	404	420		60		71,618

Initially, it was considered, a trip for each item. This would have implied a number of trips, at least multiple by ***2n*** of the number of engines (a trip for each phase, with n phases, plus the return), and a number of vehicles too excessive (a high way traffic area..), so we considered to manage items movements in groups. Six is the maximum number of grouped items, based on size and weight of the LU, and on available space. But the real size for any LU was defined in attributes whose value is specific for each item. Engines are supported in wheeled racks.

Is possible consider the transfer of the engines from the assembly lines to this support mean, in an automated way, by defining a further conveyor section, transversal to assembly lines, with a size verified trough the simulation. The same is made when AGV have to download items on Arrival Dock Bays, ADB at the AS/RS.

Vehicles used for handling through the assembly departments and storage could have been used to distribute engines from the storage area to the testing department, using the space, freed by the removal of the conveyor, however really small. A second hypothesis for same aim, was to use a portion of the main conveyor.

The final hypothesis, the one we chosen, analyzed, modeled, has been to use the AS/RS as a engines sorter directly to the area of use.

This decision was based on the following considerations:

- the need to limit the number of LGV missions to timely supply items to specific brakes,
- considerations about the spaces of the testing area, too narrow,
- considerations about the presence at the test area of the bridge cranes, useful for transferring of engine on the brakes, in a very simple way,
- considerations about the capacity level for the Storage & Retrieval Machine (SRM), and about the opportunities raising from AS/RS availability to use this as a sorting mean, also.

This opportunity has been possible thanks to the size of the warehouse itself and its physical location, i.e. parallel to the areas of testing. With the substitution of the conveyor, and the freed space, it's possible place the structure of the warehouse in a way that can be used both as a storage system but both as a sort system, also: engines are brought to the areas of destination by gravity roller conveyors on different levels (input and output from the tests) and placed on trolleys or roller using the pre-existing system overhead crane. Once back in the AS/RS, from testing brakes, via different conveyor, the engines are repositioned within the AS/RS structure.

If carefully designed the warehouse allows us to reconstruct the lots of shipping, allowing the finishing process to only play the task required to them: to finish engines and place them in the appropriate structures of shipping (pallets or crates).

The distribution towards the finishing area is done with the same methodology of the distribution in testing. Moreover, it is possible to eliminate the phase times of both downloading the item from the conveyor to buffering rollers, and both from buffering rollers to finishing lines, directly connecting the downloading zone of AS/RS, with the finishing lines. To avoid that operators remain blocked among finishing lines, we supposed to have downloading bays at an high level of the AS/RS, that, with descending conveyors, could feed finishing lines. This create a passage used by operators to access the workstation, and allows to use this space to recharge materials kits to finishing lines, without crossing AGVs paths.

When the finishing phase is ended, the handling of lots of shipping will be carried out by the same AGVs, since the route is in part superimposed to the return path to the AS/RS. Vehicles, after each delivery mission, has to query for the next mission.

The extent of the AS/RS Arrival Dock Bay has been evaluated after many simulations, to absorb the peaks of engines, massively delivered at certain times, to free the head of the assembly lines. The delivered engines have to wait to be uploaded into the warehouse from the S&RM, that can be just one because just one aisle can exist, and whose capability is lower than the sum of all assembly rates.

4.1. Distribution between assembly and storage

The work of analysis, the logical definition, the modeling and programming of the logic, begins with the study of the first operation performed by the main conveyor, the engines transport from the assembly lines to testing area. To define the characteristics of the truck to use to design the LU to be handled, to evaluate the optimal amount of engines in the LU, and consequently how many vehicles use, we made many preliminary simulations carried out with the simulation software.

The first step was to emulate the production of engines made during the reporting period, in terms of quantity for each assembly line, with their defined characteristics such as type, version, reference times for the various stages, destinations, etc.., defined in attributes of representative engine load, to determine the need for transportation of AGV. We have used the data collected previously collected in tables sorted by line.

The model read production plan data supported on spread sheets in .csv format. In the data sheets is defined the Assembly Plan, AP, of the line of relevance. In this way is very easy processing many different data sets, and evaluate the model and logic sensitivity to the typology variation of items, their distribution among the versions.

Observe and describe what is collected in them:

- Date: scheduled work for that day.
- Customer: customer name.
- Type of engine: type of engine in production.
- Line Phase Time (distribution parameters): time required for the assembly of the engine.
- Version K: is the engine code, identifies type, customer, and lot release testing of the engine.
- Inspection Time (distribution parameters): time required to stay on dynamometers for testing.
- Finishing Time (distribution parameters): time required for finishing the motor.
- Type of test: possible types of tests to be performed.
- Q: amount of motors assembled in sequence, or the Order Quantity. Does not represent the amount of engines of a lot.
- Sorting Codes: codes to determine stations, phases and lines that have to cross the engine.

Table 2: Extract of Assembly Plan for engine family.

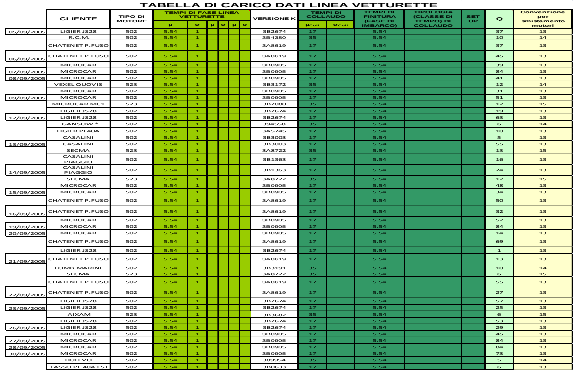


TABELLA DI CARICO DATI LINEA VETTURETTE

	CLIENTE	TIPO DI MOTORE	TEMPI DI FASE LINEA VETTURETTE						VERSIONE K	TEMPI DI COLLAUDO		TEMPI DI FINITURA (FASE DI IMBARCO)	TIPOLOGIA (CLASSE DI TEMPO) DI COLLAUDO	SET UP	Q	Convenzione per smistamento motori
			μ	σ	μ	σ	μ	σ		μColl	σColl					
05/09/2005	LIGIER JS28	502	5.54	1					3B2674	17		5.54			37	13
	R.C.M.	502	5.54	1					3B4380	35		5.54			10	14
	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			37	13
06/09/2005	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			45	13
	MICROCAR	502	5.54	1					3B0905	17		5.54			39	13
07/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			84	13
08/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			41	13
	VEXEL QUOVIS	523	5.54	1					3B3172	35		5.54			12	14
	MICROCAR	502	5.54	1					3B0905	17		5.54			31	13
09/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			51	13
	MICROCAR MC1	523	5.54	1					3B2080	35		5.54			12	15
	LIGIER JS28	502	5.54	1					3B2674	17		5.54			19	13
12/09/2005	LIGIER JS28	502	5.54	1					3B2674	17		5.54			63	13
	GANSOW *	502	5.54	1					394558	35		5.54			6	14
	LIGIER PF40A	502	5.54	1					3A5745	17		5.54			10	13
	CASALINI	502	5.54	1					3B3003	17		5.54			5	13
13/09/2005	CASALINI	502	5.54	1					3B3003	17		5.54			55	13
	SECMA	523	5.54	1					3A8722	35		5.54			13	15
	CASALINI PIAGGIO	502	5.54	1					3B1363	17		5.54			16	13
14/09/2005	CASALINI PIAGGIO	502	5.54	1					3B1363	17		5.54			24	13
	SECMA	523	5.54	1					3A8722	35		5.54			12	15
	MICROCAR	502	5.54	1					3B0905	17		5.54			48	13
15/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			34	13
	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			50	13
16/09/2005	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			32	13
	MICROCAR	502	5.54	1					3B0905	17		5.54			52	13
19/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			84	13
20/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			14	13
	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			69	13
	LIGIER JS28	502	5.54	1					3B2674	17		5.54			1	13
21/09/2005	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			13	13
	LOMB.MARINE	502	5.54	1					3B3191	35		5.54			10	14
	SECMA	523	5.54	1					3A8722	35		5.54			6	15
	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			55	13
22/09/2005	CHATENET P.FUSO	502	5.54	1					3A8619	17		5.54			27	13
	LIGIER JS28	502	5.54	1					3B2674	17		5.54			57	13
23/09/2005	LIGIER JS28	502	5.54	1					3B2674	17		5.54			25	13
	AIXAM	523	5.54	1					3B3682	35		5.54			6	15
	LIGIER JS28	502	5.54	1					3B2674	17		5.54			53	13
26/09/2005	LIGIER JS28	502	5.54	1					3B2674	17		5.54			29	13
	MICROCAR	502	5.54	1					3B0905	17		5.54			45	13
27/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			84	13
28/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			84	13
30/09/2005	MICROCAR	502	5.54	1					3B0905	17		5.54			73	13
	DULEVO	502	5.54	1					3B9954	35		5.54			5	14
	TASSO PF 40A EST	502	5.54	1					3B0633	17		5.54			6	13

The correct attributes permit to generate the correct types of engines, to create the correct sorting, AGVs request, testing typology, and make possible the correct grouping in the LU at the end of the line.

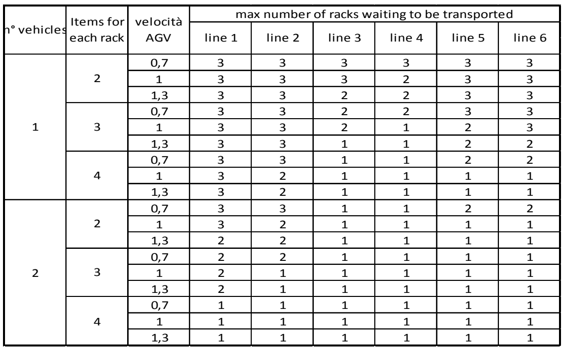
Furthermore, those items with higher production rate must have of the highest levels of priority in the request of AGVs, and this last consideration would imply the block of AGVs to serve just more produced items, putting in infinite queue the entire list of the other missions of the trolley, including those related to the track feeding where necessary. This has been a very interesting issue in the definition of the control rules of missions of AGVs. In fact, the logic to place an AGVs request based on the filling level of the rack/LU at any station has been one of the solution to manage fluently AGVs missions.

Another requirement is the reduction of the number of trips, and then the question of how many engines send for each mission.

To thoroughly have some determinations for many logic and configuration choices, many initial simulations were ran concerning the modeling of the process of movement between engine assembly and loading bay to the warehouse, only, in order to define the size of LU and their consequent numbers, and the space required for the accumulation of LUs at assembly lines, parameterized to the number and speed of available vehicles (1, 2) in each test simulation.

We summarize obtained results in the next table showing not only the importance of the number of motors to be sent to the warehouse, but also the intrinsic link between the speed of AGVs, and the speed of the used Storage & Retrieval Machine, S&RM.

Table 3: Rack Quantity required at each line end for each AGVs speed.



h° vehicles	Items for each rack	velocità AGV	max number of racks waiting to be transported					
			line 1	line 2	line 3	line 4	line 5	line 6
1	2	0,7	3	3	3	3	3	3
		1	3	3	3	2	3	3
		1,3	3	3	2	2	3	3
	3	0,7	3	3	2	2	3	3
		1	3	3	2	1	2	3
		1,3	3	3	1	1	2	2
	4	0,7	3	3	1	1	2	2
		1	3	2	1	1	1	1
		1,3	3	2	1	1	1	1
2	2	0,7	3	3	1	1	2	2
		1	3	2	1	1	1	1
		1,3	2	2	1	1	1	1
	3	0,7	2	2	1	1	1	1
		1	2	1	1	1	1	1
		1,3	2	1	1	1	1	1
	4	0,7	1	1	1	1	1	1
		1	1	1	1	1	1	1
		1,3	1	1	1	1	1	1

We can see that the solution of a single AGV, is acceptable with the exception just in the case of groups of four items. The limitation arise from the need to place in line 1, the one dedicated to the most intensive production of engine n°1, more than 3 racks, given that peak quantity of them.

It is noted, instead, the easiness of response, to the assembly lines, with two vehicles. This will be the solution that will be accepted, also verified when considering handling requests from the finishing shop, when AGVs will also play new missions, thus increasing the whole number of them.

Validated the solution on two vehicles, we started to evaluate the amount of motors to transport that will form the LU. As we can see in the table, the best responses occur by increasing the amount of motors transported per trip. Looking just at the missions required to serve assembly line, the reduction of the grouping value, lower the peak at the ASRS Arrival

Dock Bay, ADB, since the S&RM has more time to store the engine, and consequently reduces the maximum length of the conveyor used as ADB. On the other side, such definition of the LU, more continuous and fluid, however, affect missions number, and or the vehicles number, both much larger.

Under these opposing effects, it is much more preferable promptly to dispose and route assembled engines at the end of assembly lines, that cannot accumulate to much, due to space in that area, and size a greater loading bay, so there is a tendency towards larger LU.

To transport any of known engines in groups of three, the rack structure we evaluate be larger than 1800 x 800 mm instead, to transport items of any size in groups of four, will be of 1200 x 1600 mm. The structure dimension for the transportation of four of the motors is largely acceptable, and is, then, validated, the choice to move more motors is discarded for the excessive size of the structures suitable to their handling. Obviously, these handling facilities never should miss at the end of assembly lines, otherwise you would have halted all production, which is unacceptable. To assure their availability where required, during the simulation, a cyclic control of the availability of LU racks within their bays, is performed. This task of feeding racks, is the first in the priority list of the AGV. The maximum acceptable number for each zone is equal to three because, otherwise, it would be necessary too much space just to host structures.

The evaluation of response capacity of the S&RM at the ADS Bay of warehouse has been particularly complex in the quantification of the racks and LU size.

If the S&RM should work exclusively for storing engines of the LU coming from the assembly, with normal levels of performance, it would be able to cope with, just with an adequate dimensioning of the loading bay: despite the speed of examined S&RM, anyway, it results a peak which determines the need for the incoming roller of a length that is enough to constitute a decoupling buffer between assembly/transport and storage. In particular, the maximum accumulation will be achieved in the hours close to break time, when assembly line, that work on a single shift, have produced all the items of the day, testing brakes are not able to equal the ratio, and when the whole system requires the maximum services of S&RM. To overcome this problem, it should be considered the possibility of offset of few hours shifts of shops, anticipating or delaying the time at which the stock will have difficulty.

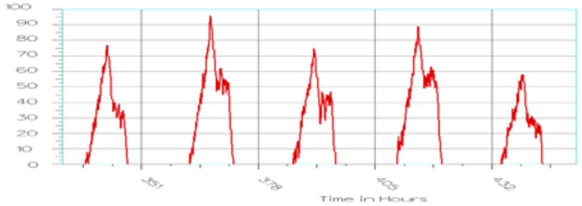


Figure 2: Cyclicality of waiting items amount Waiting at ADS Bay.

4.2. Distribution between storage and testing shop

Once in the AS/RS, items must wait until it is free the next phase, the testing one, to be moved again.

No changes have been done with respect to the initial structure, except for the replacing of the main conveyor with new rollers available at the benches of dynamometric brakes, capable of supporting a couple of motors, just those to be mounted on brake testing as soon as the previous one is released at the end of testing. Such substitution reduce the buffer protection of the resource bottleneck to a minimum, comforted by the performances of the AS/RS used also as sorter.

Initially, it was thought to use as a Downloading Bays toward brakes, from the AS/RS, a portion of the main conveyor that occupied or the entire length of the longer side of the testing shop, but we considered the space limitation and the need that it should would have to be a closed loop, not suitable to sort appropriately items to their destinations.

In a second step, it was decided to use the AGV to delivery engines to each dynamometric brakes, which made the request, an hypothesis also discarded because of limited space and the resulting chaotic traffic, and the potential lack of responsiveness offered by a discontinuous handling system. It appeared interesting, instead, the idea of using the AS/RS as sorting mean, assigning it sorting tasks to the testing department.

The distribution is found to be simple because it was sufficient to use the rollers at the side of the wall corresponding to the AS/RS, where previously there was the conveyor, and using the overhead cranes to handle items toward serving trolleys in the vicinity of brakes, shorter than previous and original ones.

The engines, after test, had to return inside the AS/RS, also placed on a roller conveyor along the same wall of the AS/RS, as those direct to the testing phase. The solution was then to position the rollers on two levels, with appropriate gradients to have a movement flexibility handling by operators, always comfortable.

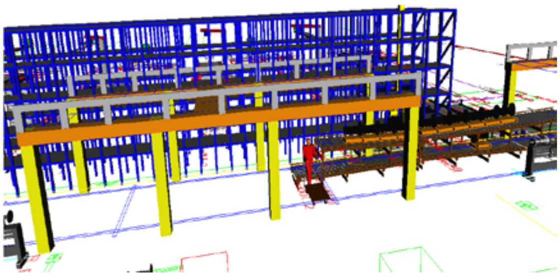


Figure 3: A snapshot from the testing area toward the AS/RS and the rollers at overlapped levels.

The length of the rollers are proportional to the needs: capacity and space are balanced to both the AS/RS maximum delay, and both the delay of the bridge crane in the sorting stage of the testing.

Motors, to access the testing shop have to be requested and ordered directly from the brakes buffers, as soon as they have ran a test. The order arrives at the AS/RS, that interrogates its inventory and observe if there are engines for the request, if there are, they are taken and routed on the conveyor outlet, vice versa is there aren't, no mission is done, but the request remains stored in the logic infinite queue, the order lists.

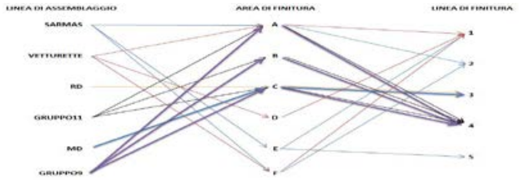


Figure 4: From To Chart in the observed configuration.

The assignment of the zone of the AS/RS, whose allocation rule is a Class Based Storage Policy, is made on the base of the testing destination of each engine, and, consequently, is made as soon as the engine is taken from S&RM from ADS Bay. Warehouse shelving is divided into dedicated zones.

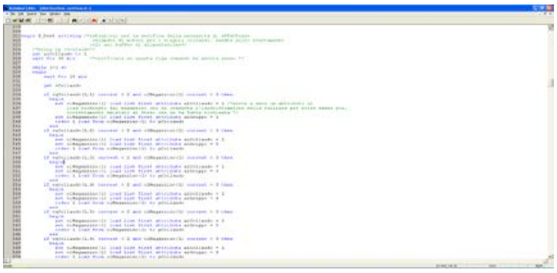


Figure 5: A snapshot of the logic code to advance items toward testing destinations.

Each buffering area manage an internal fictitious queue (OrderList), containing any of the useful attribute values, as the engine typology and version, the Order Line that generates that item, the entry time in the buffer, that when questioned from the brake banks, will answer on the availability of engines defined features. For the distribution was considered the FIFO logic.

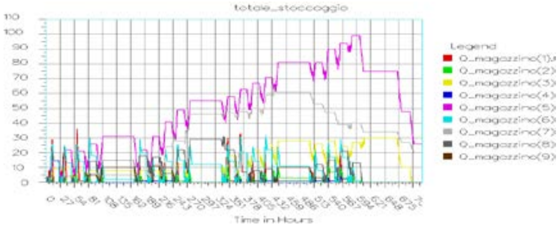


Figure 6: A snapshot of the variability of levels of engines in AS/RS waiting to be moved to test brakes.

The available capacity offered by the testing brakes, considering that is a high time consuming phase, has been highlighted in preliminary simulations. This phase is the only one that requires a daily double shift to accomplish the Assembly Plan.

In the graph it is represented the stock level for each area of the AS/RS, each one devoted to feed brakes for specific test typologies, and shows the difference reported by two zones of warehouse:

First, zone 7 is not too much critical for the low number of engines produced, but should be assessed more accurately if production level should reach higher values. In this case, is possible evaluate the opportunity to re configure, with appropriate setup of 4.5 hours, alternative brakes. Second, is critical, instead, the curve number 5, the zone devoted to store items with the higher intensive flow, which tends to rise, and when it happens, is possible use a jolly brake, in the same area, at the expense of the engines of other typologies, that are usually served by that brake.

4.3. Distribution between testing and finishing

The engines, then, travel from the exit of the tests in the direction of the corresponding finishing lines. As previously occurred, even in this case is the S&RM that reads and defines the exact destination that should have the motor. For the destination pre-assigned attibutes are read. The engine is, then, taken from the testing brakes and placed on the rollers to be routed through AS/RS to the corresponding finishing lines.

Since the opportunity of using the AS/RS as sorter too, instead of buffering conveyor, it is thought to achieve better results routing engines on as many rollers, as the finishing lines, adjacent to the wall of the warehouse, to be taken by operators with the use of gantry cranes, avoiding the engine search and by providing a more robust and defined routing.

Moreover, we decided to use the AS/RS as a sorter to automatically deliver directly to the finishing lines, enabling operators to always be in the possession of the motors to work avoiding the transshipment between the two rollers.

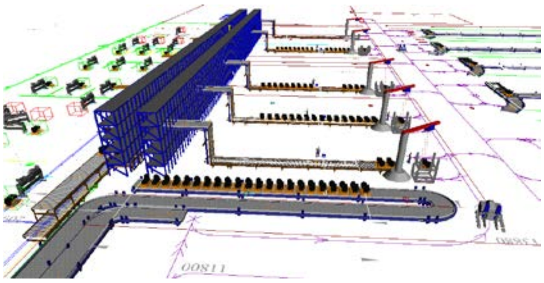


Figure 7: A snapshot of the finishing area with the engine distribution from the top.

Furthermore, the solution we propose, avoids to close operators inside finishing lines, preventing them to have escaping routes or easiness of walking, and makes possible using this corridor to the supply finishing materials or kits. The items comes out from the AS/RS not at the ground level, but at a higher aerial level, at about five/six meters, and brought to fair quote by from automatic descenders.

To compensate for the excess distribution that would reach to the finishing lines, it was considered a double criteria control logic based both on the AS/RS residual storage capacity, and both on the free space on the service rollers at the finishing lines.

It has been defined a logic activated by the engines after the testing phase that evaluate the amount of motors present in the finishing line of destination, and, in case of availability of space, place an handling order to the S&RM that move the item in appropriate roller, as soon as possible, otherwise ask for a moving mission to the bays of wait, close to the dedicated opening areas, at destination descenders, one for each outlet.

From the graph it is possible to note the necessity of engine storage capacity, especially in the early hours of the morning, in correspondence of the activity of testing, that starts two hours before assemblies, while this possibility decreases in the course of the day.

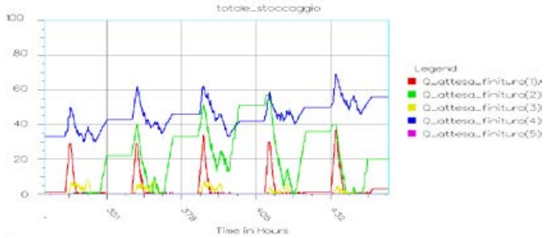


Figure 8: Trends of the engines number waiting to be moved to finishing Area.

It should be noted that the lines of destination, as configured in the real situation, would lead to an excessive load handling for S&RM, therefore was more appropriate, without losing adherence to the real system, reallocate finishing lines in correspondence of the testing areas of the same items, but on the opposite side with respect of the AS/RS.

We arrived at the end of the process, the engines must now be placed in transport units to be shipped. In original situation, movements of these units were done by trans pallet driven by operators. It is natural consider of using AGVs for this last transportation. The motors are placed at the end of the finish using hoists or cranes in pallet or boxes suitable for shipping and then are picked up by the AGV toward the expedition area.

We provided to use another item attribute to be used to define an exact quantity of engines that it was appropriate to recompose lots of shipment. This information, anyway was not available at the time of our work, and we considered the blind hypothesis a generic quantity of five engines per lot.

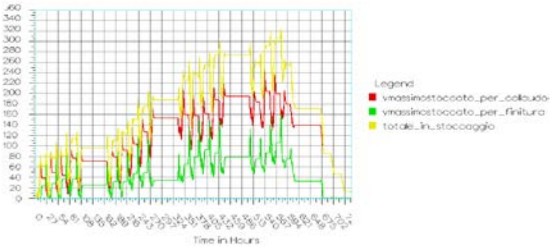


Figure 9: Amount of whole items at testing area (red), finishing area (green), and in all the system (yellow).

Engine attributes, read at the initial phase of the simulation, presents not only the destination of finishing information, but also identify the type of contract and customer. The shipment logic can keep an item in the AS/RS before complete the finishing phase and the shipment too, case for shipment to finishing, till the last item of the shipping lot is not already tested: any engine taken from the testing will check if is the last of its shipment order, that ask for the release to the finishing area, from the AS/RS of all other items to be shipped, allowing operators to have motors in the correct order.

In the following lines an extract of lines of code responsible for the activation of the logic model for the selection and identification of lots and handling.

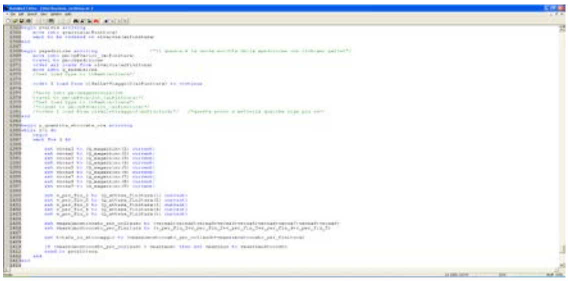


Figure 10: Snapshot of the logic to activate selection and identification of lots relative handling

CONCLUSIONS

The analysis of a new handling and transport system that could replace the conveyor has been completed according to the solutions, that in the preliminary analysis, resulted most adequate among those remaining solutions not yet verified, through simulation models.

The models and analysis of the produced results have supported both the definition and validation of the general structure, of features to consider, the definition of configuration parameters to be used, such as speed and acceleration of AGV and S&RMs, number of vehicles, items number to be aggregated in LU, management rules of the system, layout definition, locations, buffer sizing, warehouse sizing, and so on.

A comparison of the tested system with those previously assessed, not presented here, can show the important differences, relating to occupation of space, economic investment, interoperational buffer level, reconstitution of lots control. The current solution shows prevalence in space occupation, lower incidence of human activity due to the increased level of automation, a greater rigidity in accompanying the flow variations, with a trend towards greater initial investment cost.

The initial obtained results led to the definition of an AS/RS, used not only for its most obvious function, but also the possibility to be used to distribute motors to the finishing lines and to the areas of testing, allowing to reduce the levels of motors in interoperational buffer, obtaining a greater availability of space within the dimensions already departments themselves. This, in the perspective of improvement and increase of production, is transformed into the possibility of developing its departments through the introduction of new machinery or new lines, or, as seems appropriate, in the definition of finishing lines in greater numbers than at present, each dedicated to specific codes, perhaps replicated more than one times in case of request of capacity increase for a specific code in the finishing activities, and to achieve more flexibility.

The use of the simulation tool has allowed us to light on aspects of detail in the definition of some characteristics of the assumed system: the number of vehicles needed, interoperational buffer sizing, performance parameters of LGV and S&RM, level of aggregation of the LU, up to highlight the limits of efficiency of the model configurations assumed, and suggest the characteristics of structurally different solutions, as in the case of the use of two S&RMs.

The proposed system and related models have been defined to have the highest possible degree of flexibility, resilience and reconfigurability (also made with many configuration value supported on data sheet , on reading files, external to the application, easy to reconfigure, as well as in the definition of the structural characteristics of the system.

Under these considerations, it could be nice continue to evaluate results with different configuration, and different production data, both in terms of quantity and mix.

The model can support any item information to enforce item recognition and traceability and tracking, but it isn't based on very intensive automation level. Order lots integrity in guaranteed as FIFO policies accordingly with lean production principles, also if it is applied in a smooth and tolerant way: because the randomness of the system and the simulation model too, sometimes one previously introduced item can be over passed by another one of same family, version and client order.

AS/RS has been concept as a transit point and sorting system, combining the storage feature with the skill of manage the sorting need of items in an automated way.

But not all the system is been thought adopting automated means, just where considered more effective.

The control logic to manage final assembly area fulfilment, and the consequent need to be freed, and the coupled logic to pick items as the last process phases are going in shortage represents a satisfying mix of push pull logic: the main activation impulse is a production plan already processed based on Due Date respect, but the advancing logic of prepared lots of items, especially the transport system is phased on a hierarchical control logic that mixes buffers area limitation constrains and shortages for buffers feeding lasts of processing phases, or bottleneck machines, as the testing machines that represent the ones to be operative and working because they have the longest processing time.

REFERENCES

Alan A., Pritsker B., 1997. Modeling in performance-Enhancing processes. *Operation Research*, Vol.45, No.6, pp 797-804.

Ceric V., Hlupic V., 1993. Modelling a Solid-Waste Processing system by discrete event simulation. *The Journal of the Operational Research Society*, Vol.44, No.2, pp 107-114.

Cho S., 2005. A distributed time-driven simulation method for enabling real-time manufacturing shop floor control. *Computers & Industrial Engeneering* , 49, 572-590.

Kumar S., Nottestad D. A., 2006. Capacity design: an application using discrete-event simulation and designed experiments. *IIE Transactions*, 38, 729–736.

Kolkka, T., Rajagopalan, J., Suksi, J. 2013. Benefits of advanced configuration and simulation study. *Iron and Steel Technology*, 10 (6) , pp. 65-72.

Jane, C., Laih, Y.-W., 2005. A clustering algorithm for item assignment in a synchronized zone order picking system. *European Journal of Operational Research*, 166 (2) , pp. 489-496.

Gallo S. A., Melloni R., 2007. Valutazione della scelta di una soluzione alternativa di sistema per il trasporto motori con tecniche simulative: il caso Lombardini motori. XXXIV Convegno Nazionale Animp-Oice-Uami.

Malmborg, C.J., 2001. Estimating cycle type distributions in multi-shuttle automated storage and retrieval systems. *International Journal of Industrial Engineering: Theory Applications and Practice*, 8 (2) , pp. 150-158.

Perry A., Ronald F., Hoover, Stewart V., Freeman, David R., 1983. Design of an AS/RS using simulation modeling. *Proceedings of the International Conference on Automation in Warehousing*, pp. 57-63 4.

Selladurai V., Aravindan P., Ponnambalam S.G., Gunasekaran A., 1995. Dynamic simulation of job shop scheduling for optimal performance. *International Journal of Operations & Production Management*, Vol. 15 No. 17. pp. 106-120.

Ueno N., Sotojima S., Takeda J., 1991. Simulation-Based Approach to Design a Multi-Stage Flow-Shop in Steel Works. *IEEE*.

Vaidyanathan B. S., Miller D. M., Park Y. H., 1998. Application Of Discrete Event Simulation In Production Scheduling. *Proceedings of the 1998 Winter Simulation Conference*. D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

Watson E. F., Medeiros D. J., Sadowski R. P., 1997. A Simulation-Based Backward Planning Approach For Order-Release. *Proceedings of the 1997 Winter Simulation Conference*.

Yang J.; Chang T. S.; 1998. Multiobjective scheduling for IC sort and test with a simulation testbed. *IEEE transactions on semiconductor Manufacturing*, vol. 11, no2, pp.181-231 (19), pp. 304-315

A PIPE SPOOL FABRICATION SIMULATION MODEL

Labban, R.[(a)], Abourizk, S.[(b)], Haddad, Z.[(c)], El-Sersy, A.[(d)]

[(a)]University of Alberta, 3-015 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA
[(b)]University of Alberta, 5-080 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA
[(c)] Consolidated Contractors Group, 62B Kifissias Avenue, Maroussi, 15125, Athens, GREECE
[(d)] Consolidated Contractors Group, 62B Kifissias Avenue, Maroussi, 15125, Athens, GREECE

[(a)]labban@ualberta.ca, [(b)]abourizk@ualberta.ca, [(c)] zuhair@ccc.gr [(d)] asersy@ccc.gr

ABSTRACT

When dealing with larger and more complex construction operations, which are more difficult to manage using traditional project management tools, computer simulation methods have shown to be effective in designing and analyzing construction processes, regardless of the complexity or size. A simulation model can be built to describe the construction activities of a scope of work ranging from large, complex industrial projects to a simple room of a small building. Using simulation, engineers can test out different construction scenarios, estimate resource utilization and find bottlenecks, and forecast time and cost requirements without having to go to site. This paper describes the pipe spool fabrication model built at Consolidated Contractors Group (CCC).

Keywords: pipe spool fabrication, discrete event simulation

1. INTRODUCTION

On a large industrial project, pipe spool fabrication is a major component of the construction operation. It is also a relatively short term, complex construction process often riddled with uncertainty due to the intrinsic unique nature of its outputs and the numerous factors affecting its activities. As such, it is important for all stakeholders to have a good grasp of the performance of pipe fabrication shops and their ability to meet the site pipe installation schedules. The ability of computer based modeling and simulation to model resource and activity interactions, queuing, and uncertainties renders it a good fit for modeling the pipe spool fabrication process.

2. PIPE SPOOL FABRICATION MODEL

2.1. Background

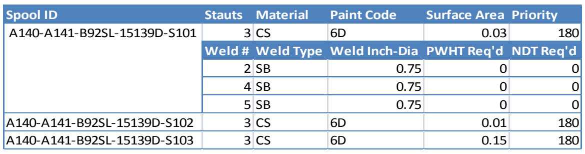
Construction contractors on large industrial projects often build one or more project specific pipe fabrication shops to handle the pipe spool fabrication scope. These shops are built to handle a specific set of pipe fabrication activities including cut and bevel, fit-up, welding, QC inspection, post weld heat treatment, non-destructive testing, and painting. Each of these activities is repeatedly performed by a specific type of crew on pipe spools. Each time it is performed, a crew is utilized for a certain duration and the result is specific progress of a pipe spool along its path to completion. With the large number of spools and their diverse characteristics and resource requirements, forecasting pipe spool fabrication activity completion and optimizing resource allocation and utilization becomes a complex task well suited to computer modeling and simulation.

2.2. Simulator Design and Development

The simulator was developed to aid stakeholders in arriving at answers to the issues stated above. The first step was the abstraction of the real world situation into a simulation model representing the operations of a pipe spool fabrication shop, including detailing the product and process definitions for all the main activities. In order to understand the nature of how pipe spool fabrication activities were performed on construction sites, extended visits to multiple mega industrial projects were conducted to observe and document the above mentioned set of activities. Benchmarking for every activity was conducted via numerous observations of the activity being performed on different spools of varying characteristics. Both crew composition information and productivity figures were collected. In this paper we will not deal with the analysis of the observed productivity data; this matter will be dealt with at a different time. Instead, for this paper, we will assume the resulting productivity norms deduced from the observations as our activity productivity norms for the tasks. The simulator was developed as a discrete event simulation model with spools as the main entity. For the welding tasks, welds are the entities - where spools are split into their constituent welds - in order to process welds individually and collect their artificial history.

2.3. Product Definitions

Product definition for spools to be processed by the simulator is a straightforward process where only those spool characteristics required for simulating the fabrication activities were specified for each spool. It is organized into a two-level hierarchy for spools and their relevant joints. Data for spools include spool ID, current spool status, line class, material type, paint code, surface area, and spool specific priority information. Data for joints include weld type, inch-dia, post weld heat treatment (PWHT) requirement and non-destructive testing (NDT) requirements.



Spool ID	Stauts	Material	Paint Code	Surface Area	Priority
A140-A141-B92SL-15139D-S101	3	CS	6D	0.03	180
	Weld #	Weld Type	Weld Inch-Dia	PWHT Req'd	NDT Req'd
	2	SB	0.75	0	0
	4	SB	0.75	0	0
	5	SB	0.75	0	0
A140-A141-B92SL-15139D-S102	3	CS	6D	0.01	180
A140-A141-B92SL-15139D-S103	3	CS	6D	0.15	180

Figure 1: Sample Spool-Weld Hierarchy

2.4. Process Definitions

For the process definitions we needed to define the activities and flow required to fabricate the different spools. For each activity, the type of resource (crew) required and its relevant productivity had to be identified. For the purpose of this paper, we will only cover the main activities of the pipe spool fabrication process, and not include logistical activities such as crane and trailer operations.

2.4.1. Pipe Spool Fabrication Activity Flows

The following three figures are snapshots of the different activities represented in the DES model. For each activity, the required crews and time to perform the activity is decided based on spool characteristics. Figures 2a, 2b, and 2c, below, depict the DES flow of spools through the "Cut," "Bevel," and "Fit-up" activities.

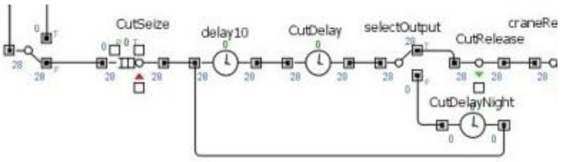


Figure 2a: Cut

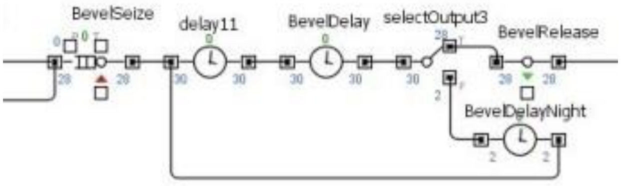


Figure 2b: Bevel

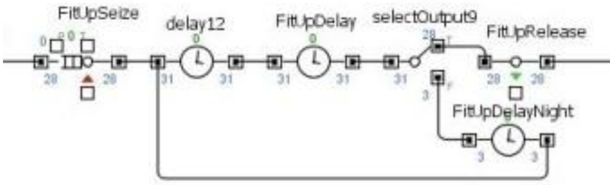


Figure 2c: Fit-up

Figures 3a, 3b, 3c, and 3d depict the DES flow of spools through the welding process. Based on spool configuration, spools are routed either to manual welding stations, or to automatic welding machines. For welding, each spool entity is split into its welding entities, based on the number of shop welds required. Based on spool and weld characteristics, (1) the appropriate number of welders is assigned to each weld, and, accordingly, (2) the weld duration is derived. Splitting the spool entity into weld entities allows us to process welds independently and collect their respective artificial history individually. Once all welds are processed, the weld entities are batched into a spool entity again, which then undergoes "QC Release." In QC Release a certain fraction of spools, based on norms derived from site observations, is sent back to repair due to defects in the welding process.

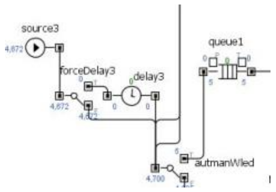


Figure 3a: Routing to Manual or Auto-Welding

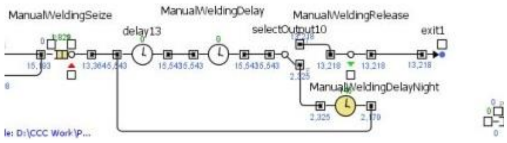


Figure 3b: Welding

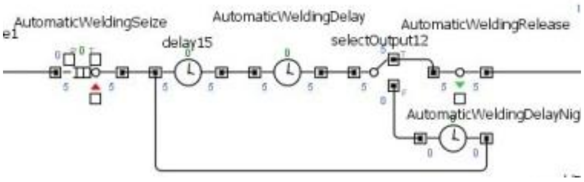


Figure 3c: Auto-Welding

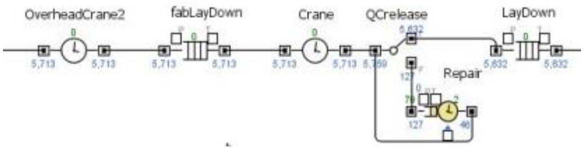


Figure 3d: QC Release

Figure 4 shows the DES flow of spools through the PWHT, NDT and painting activities. Not all spools require PWHT, and not all spools require NDT. The flow and logic control of the model automatically detect this from the spool information and associated tasks are initiated accordingly.

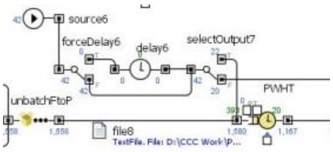


Figure 4a: PWHT

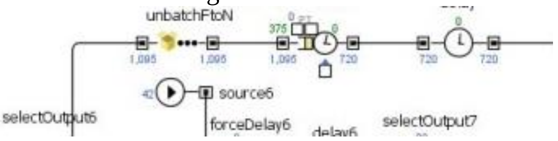


Figure 4b: NDT

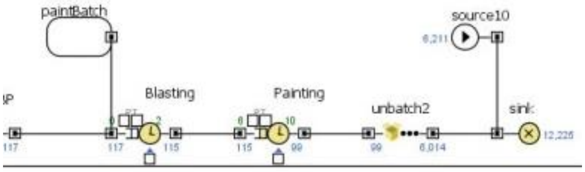


Figure 4c: Painting

2.4.2. Resource Definitions

Each of the pipe fabrication activities is associated with a resource type. Each resource type is typically a crew composed of a group of workers required to perform a specific task. Following is a table showing typical crew compositions on a large industrial construction project. Notice that certain worker types are shared amongst the various crew types.

Crew Type	Worker Type 1	Worker Type 2	Worker Type 3	Worker Type 4	Worker Type 5	Worker Type 6	Worker Type 7	Worker Type 8	Worker Type 9
Cut	1	2	1	2					
Bevel		1							
Fit-up		2	2		1	4			
Weld Size 1			1				1		
Weld Size 2			2				2		
PWHT			1					2	
Painting	1		3						4

Figure 5: Typical Crew Compositions

A table showing typical worker availability over time on an industrial project is shown below. Workers available make up the required crews (resources) for the activities which are then captured to simulate the performance of a task on a spool or weld.

Worker	Month							
Type	1	2	3	4	5	6	7	8
1	21	23	25	25	25	25	25	25
2	85	90	95	95	100	100	100	100
3	80	82	85	86	104	104	104	104
5	37	38	40	41	42	42	42	42
6	90	95	100	100	110	110	110	110
7	92	99	112	129	135	135	135	135

Figure 6: Typical Worker Availability Over Time

2.5. Model Structure

All the above pieces come together as in the structure shown in the figure below.

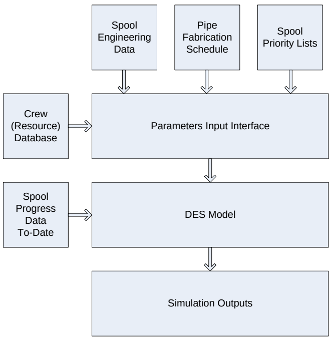
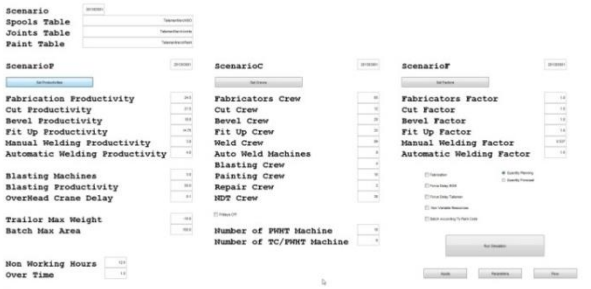


Figure 7: The Overall Model

2.5.1. Parameters Input Interface

This module allows the user to define the different parameters of the simulator and run the model. “Spool Engineering Data,” “Pipe Fabrication Schedule,” “Spool Priority Lists” and “Crew Database” feed into the “Parameters Input Interface” module. “Spool Engineering Data” provides engineering characteristics of the spools; “Pipe Fabrication Schedule” and “Spool Priority Lists” provide, respectively, the activity schedule information related to pipe fabrication and the priority lists for spool fabrication requirements produced by the engineers; “Crew Database” provides resource information, namely the number of crews available over time of each crew type. Through the “Parameters Input Interface” the user can change the location of the feed data, assumed productivities for the different tasks, working hours, and fix the number of crews at a constant level throughout the simulation duration instead of reading them from the relevant feed.



Scenario
Spools Table
Joints Table
Paint Table

ScenarioP
Fabrication Productivity
Cut Productivity
Bevel Productivity
Fit Up Productivity
Manual Welding Productivity
Automatic Welding Productivity
Blasting Machines
Blasting Productivity
OverHead Crane Delay
Trailor Max Weight
Batch Max Area
Non Working Hours
Over Time

ScenarioC
Fabricators Crew
Cut Crew
Bevel Crew
Fit Up Crew
Weld Crew
Auto Weld Machines
Blasting Crew
Painting Crew
Repair Crew
NDT Crew
Number of PWHT Machine
Number of TC/PWHT Machine

ScenarioF
Fabricators Factor
Cut Factor
Bevel Factor
Fit Up Factor
Manual Welding Factor
Automatic Welding Factor

Figure 8: Parameters Input Interface

2.5.2. Discrete Event Simulation Model

The discrete event simulation model (DES) carries all the DES flow and logic required for running the simulation model. It includes all the tasks along with their corresponding parameters, resource pool requirements, and duration formulas.

Figure 9: DES Components

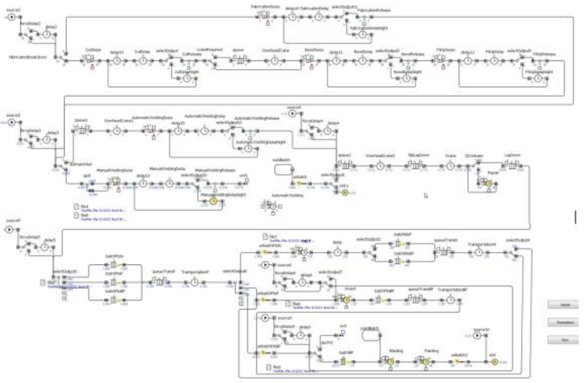
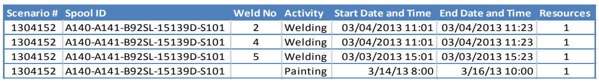


Figure 10: DES Flow

It also receives from the "Spool Progress Data To-Date" the latest spool statuses snapshot which allows it to insert each spool into the DES model at the correct stage through the flow. This enables users to run simulations on existing site data taking into account the current situation of the project.

2.5.3. Simulation Outputs

The simulator produces, as its main output, a comprehensive set of data comprised of the artificial history of the simulated pipe fabrication operations. The result set (as in the figure below) contains a record of the activities performed on the corresponding entities (spools or welds) utilizing the required resources. For each entity/activity/resource occurrence, the data contains a start date and time, an end date and time, and a number of resources utilized for the duration.



Scenario #	Spool ID	Weld No	Activity	Start Date and Time	End Date and Time	Resources
1304152	A140-A141-B92SL-15139D-S101	2	Welding	03/04/2013 11:01	03/04/2013 11:23	1
1304152	A140-A141-B92SL-15139D-S101	4	Welding	03/04/2013 11:01	03/04/2013 11:23	1
1304152	A140-A141-B92SL-15139D-S101	5	Welding	03/03/2013 15:01	03/03/2013 15:23	1
1304152	A140-A141-B92SL-15139D-S101		Painting	3/14/13 8:00	3/16/13 10:00	1

Figure 11: Sample Simulation Outputs

2.6. Verification, Validation, & Accreditation

Credibility of a model which is expected to help manage large industrial construction projects is of utmost importance in order for stakeholders to accept and adopt the model.

In order to verify the model, both unit tests on each of the tasks within the model, and an overall system test were run. Outputs after the tests were compared with expected results based on predetermined inputs and ensured the model and its components were correctly implemented.

Validation of the model was done in two steps. First, the model flow and logic were compared and confirmed against a conceptual model design based on workflows and information collected from actual pipe fabrication operations on large industrial projects. Then, the model was run with historical data from multiple projects and its outputs compared to historical results to ensure the model was behaving as per its design purposes.

Initial informal accreditation was done through a test implementation of the model at a large industrial construction project. Stakeholders gave positive feedback and are employing the simulator to forecast pipe spool fabrication progress based on pre-determined resource availability. Further accreditation of the model within the corporate environment is planned through more implementations on upcoming industrial projects.

3. DISCUSSION

3.1. Potential Benefits

The pipe spool fabrication model described in this paper provides the stakeholders and end users with a tool which allows them to proactively perform low level resource planning on pipe fabrication activities on large industrial projects. The model is able to run in a predictive mode where no progress on site has been made or in a management mode where progress information for spools is already available and forecasting can be made using existing progress data as a starting point for the model to run. Retrospective running of the model is also possible for change impact assessment or lessons-learned analysis. Simulation model run results include a complete simulated history for each activity/spool(weld)/resource combination with resource and time requirements to be performed. Simulation run results can be analyzed to answer performance questions including, but not limited to: Which of my resources along the flow is acting as a bottleneck in the operation? Which of my resources is under employed? What is the expected resource histogram for the pipe spool fabrication operation to finish on time? With the available resources, how much time will the pipe fabrication operations need to finish? Will my fabrication activities meet my schedule milestones? Will the fabrication activities finish on time for a certain priority sub-area?

3.2. Future Work

The pipe spool fabrication model described in this paper is a first step at aiding in the management of pipe spool fabrication operations. Further development and enhancements to this model include:

1. Integration of a material constraint module. The current model assumes drawings and material to be ready and available for pipe spool fabrication to commence. As observed on projects, material availability is sometimes an issue for pipe spool fabrication shops. A simulation model which manages material availability in accordance with spool engineering data and material delivery schedules will potentially add value to the inputs of the current pipe spool fabrication model.
2. The flow of the current model ends once spools are painted (and transported to laydown) and are waiting to be installed on site. A further foreseen development is the

addition of a pipe installation simulation model which covers the pipe spool construction process until the spools are installed in place on site and given final release.

4. CONCLUSION

This paper presented a special purpose discrete event simulation model for managing pipe spool fabrication operations in pipe fabrication shops on industrial projects. The simulation model helps stakeholders manage their activities and perform low level resource planning for all shop pipe spool fabrication activities. The main benefits of the model are (1) predictive analysis of fabrication resource requirements, and (2) managing operations and forecasting resource and time requirements during project execution. Future work includes adding enhancements to the model including a material constraint module and a pipe installation module.

ACKNOWLEDGMENTS

The pipe fabrication model described in this paper was developed and implemented at Consolidated Contractors Group (CCC) to aid in estimating, planning and managing pipe fabrication operations on large industrial construction projects.

REFERENCES

AbouRizk, S., 2010. Role of simulation in construction engineering and management. *ASCE Journal of Construction Engineering and Management* 136(10):1140-1153.

Anylogic, N.d. *The big book of Anylogic*. Anylogic. Available from: http://www.anylogic.com/the-big-book-of-anylogic [accessed 14 April 2013].

Barrie, D.S., Paulson, B.C., 1992. *Professional Construction Management, Including C.M., Design Construct, and General Contracting*, 3rd ed. New York: McGraw-Hill, Inc.

Department of Defense (DoD) Modeling and Simulation Coordination Office (M&SCO), N.d. *Verification, validation, & accreditation (VV&A) recommended practices guide (RPG).* Available from: http://msco.mil/VVA_RPG.html [accessed 14 July 2013].

Hu, D., Mohamed, Y., 2011. Effects of pipe spool sequences in industrial construction processes. *Proceedings of CSCE 2011 Construction Speciality Conference, Canadian Society of Civil Engineers,* pp. CN144-1 – CN144-10. June 14-17, 2011, Ottawa, ON.

Pritsker, 1986. *Introduction to Simulation and SLAM II*, 2nd ed. New York, NY, and West Lafayette, IN: Wiley and Pritsker Associates.

Sadeghi, N., Fayek, A. R., 2008. A framework for simulating industrial construction processes. *Proceedings of 2008 Winter Simulation Conference,* pp. 2396-2401. December 7-10, 2008, Miami, FL.

Song, L., Wang, P., AbouRizk, S., 2006. A virtual shop modeling system for industrial fabrication shops. *Simulation Modeling Practice and Theory* 14(5), 649–662.

AUTHORS BIOGRAPHY

RAMZI LABBAN is a PhD candidate in Construction Engineering and Management at the Department of Civil and Environmental Engineering, University of Alberta. He is also Manager of Computer Modeling and Simulation at Consolidated Contractors Company (CCC). His email address is labban@ualberta.ca.

SIMAAN M. ABOURIZK holds an NSERC Senior Industrial Research Chair in Construction Engineering and Management at the Department of Civil and Environmental Engineering, University of Alberta, where he is a Professor in the Hole School of Construction. He received the ASCE Peurifoy Construction Research Award in 2008. His email address is abourizk@ualberta.ca and his web page is http://www.irc.construction.ualberta.ca.

ZUHAIR HADDAD is vice president, corporate affairs, and chief information officer for CCC. He is in charge of mapping out and implementing IT and communication strategies and manages the information systems, communication, and E-procurement departments. He also oversees the plant department, managing CCC's extensive fleet of construction equipment. His work experience includes several construction projects in Saudi Arabia. He has interests in developing 3D-based construction controls software. He holds a civil engineering degree from Cal State-Chico and a master's in construction management from Stanford University.

AMR EL-SERSY holds a Ph.D. degree in Engineering and Construction Management from the University of California at Berkeley. He has over 20 years of experience. Currently, he is the Group Manager Learning and Innovation at Consolidated Contractors Company (CCC). He is responsible for the strategy and implementation of corporate-wide Knowledge Management program as well as many innovative ideas and R&D initiatives. Prior to this assignment, he was deputy General Manager for the EPC oil & gas business unit of CCC. Earlier, he was the Manager of Engineering and Construction systems group where his responsibilities included overseeing the implementation of different engineering, procurement and construction IT systems. He led the development of several corporate initiatives such as the Enterprise Risk Management program. Before joining CCC, he held several project management positions with international general contractors and construction management consulting firms.

A MODEL BASED DECISION SUPPORT SYSTEM FOR LOGISTICS MANAGEMENT

V. Boschian[c], M.P. Fanti[(b)], G. Iacobellis[(b)], G. Georgoulas[(c)], C. Stylios[(c)], W. Ukovich[(a)]

[(a)] University of Trieste, Via Valerio 10, 34127 Trieste, Italy
[(b)] Polytechnic of Bari, Via Orabona 4, 70125 Bari, Italy
[(c)] Technological Educational Institute of Epirus, GR-47100 Arta, Greece

[(a)]{valentina.boschian, walter.ukovich}@di3.units.it ,[(b)] {fanti, iacobellis}@deemail.poliba.it,
[(c)] georgoul@teleinfom.teiep.com, stylios@teiep.gr

ABSTRACT

This paper deals with the specification of a Decision Support System (DSS) that has to manage the flow of goods and the business transactions between a port and a dry port. The paper investigates the case of the broader area of Trieste port and specifies the DSS that manages the import flows of freights between dry port and seaport. An integrated approach is designed for the tactical level decision strategy based on simulation optimization, where metaheuristic algorithms are applied.

Keywords: Decision Support Systems, Discrete Event Simulation, Optimization, Metaheuristic algorithms

1. INTRODUCTION

The current concept of a dry port directly connected with a seaport (through a city center) is an essential matter since it opens a new series of problems to be faced. With the establishment of a dry port directly connected with a seaport, new operational problems arise, since the logistics operations between the two terminals must be coordinated and synchronized. Moreover, the informative systems of the two terminals must be integrated to manage all information regarding the operations of both the two terminals. For this problem, strategic, tactical and operational, (Giani et al., 2004), problems arise since it is necessary to define the better allocation of the various services between the two terminals at different scales. The term ‘dry port’ was introduced by the Economic Commission for Europe in 2001 to denote an inland terminal directly linked to a maritime port. Subsequently, Roso et al. (2008) formalized this term to denote not only a terminal linked to another one but also a terminal where some typical services of a seaport are moved to provide more available space and to require less service time at the port area. At the international literature there are several papers that analyze intermodal terminals and in particular container terminals (Stahlbock et al., 2008).

In addition, the connection between a seaport and a dry port is not yet investigated in depth by the research community. In fact, a few papers deal with the analysis of the impact of new road and railway networks on the logistic system in an intermodal container environment at strategic level via simulation models (Parola and Sciomachen, 2005). Moreover there are no studies on how and what services have to be moved to the dry port.

In such context, Discrete Event Simulation (DES) models are widely used to describe decision making and operational processes. In Ramstedt and Woxenius (2006) a thorough literature analysis about the simulation of the decision-making process within a transport chain is presented and in Gambardella et al., (1998), a resource allocation problem in an intermodal container terminal is simulated. Duinkerken et al. (2006) present a comparison among three transportation systems for the overland transport of containers between container terminals and a simulation model was developed to assist in this respect. The model is applied to a realistic scenario for the Maasvlakte situation in the near future. In (Ottjes et al., 2006), a generic simulation model structure for the design and evaluation of multi-terminal systems for container handling is proposed.

While several research activities have been realized, and are on-going, related to the intermodal transportation, it has to be outlined that no specific activities have been performed in the direction of applying modeling, simulation techniques and new technologies to address the dry-port challenges. Therefore this paper proposes and studies the impact of an integrated Information and Communication Technology (ICT) solution to manage the problem of the connection of a port and a dry port area. In particular, the proposed results are focusing on the management of the logistic operations at the tactical level.

The aim of this paper is to contribute in the specification of a model based Decision Support System (DSS) to integrate logistics management and decision support for intermodal port and dry port facilities. In particular, the case study that considers the port of Trieste (Italy) and the dry port of Fernetti is analyzed. This case study is examined in the SAIL project, sponsored by the EU 7th Framework Programme, Marie Curie IAPP.

The paper is organized as follows: Section 2 describes briefly the case study, Section 3 presents the methodology to specify the proposed DSS, Section 4 reports the discussion about the application of the DSS and Section 5 concludes the paper.

2. THE CASE STUDY

2.1. The port and the inland terminal of Trieste

The logistics system in the Friuli Venezia Giulia region (Italy) is particularly significant both for its geographical location, as the meeting point of the trans-European Corridor V and the Adriatic Corridor, and for its concentration of ports and land, sea and railway transport networks. A requirement analysis identified two different configurations for the specific test case: one for the containers and one for the trucks. In the port of Trieste, the traffic of trucks directed to Turkey through a Roll-on/roll-off Traffic (Ro-Ro) service represents a consolidated traffic. The containers have large areas to be warehoused in the port. On the contrary, a limited space is dedicated to the truck parking area. Hence, the study of the optimization of movements of trucks between the port area and the dry port area is crucial and needs the application of suitable management strategies.

Before going into the details of the freight flow, this section aims at defining the main involved actors, users and stakeholders. We are mainly focusing on the connections between the port of Trieste and the Intermodal Terminal of Fernetti that operates also as a dry port area of the Trieste port. In particular, the port of Trieste is a free port and the Port Authority has the role of controlling, coordinating and managing the port operations.

In the analysis of the case study, we consider the flow of goods that are managed by the following actors:

- the freight forwarder: it is represented by a company operating as intermediary, taking care of authorization procedures and documents. The freight forwarder has a key role for the organization of the flow of goods and information;
- the final customer: there are several customers involved and the flow of goods always begins with an order of the customer;
- the shipping agent: it operates as intermediary, taking care of authorization and booking procedures. The shipping agent also has a key role in the organization of the flow of goods and information;
- the terminal operator: it provides a full range of additional services including container freight station, warehousing and storage, survey, container repair and maintenance and dedicated areas. Goods transported in containers are unloaded in the terminal area ;
- customs: The Custom Agencies of Fernetti and Trieste. Customs is the authority responsible for collecting and safeguarding customs duties and for checking the flow of goods. The customs clearance procedures can be performed at the origin of the flow, in Fernetti or in the Port of Trieste.

We present two different ways of managing the typical interactions of cargo with the authorities and the infrastructure operators in the export phase: a) the current situation, called case "as is", b) the new solution, called case "to be". In this paper, for the sake of brevity we do not describe the import phase.

2.2. Description of the export phase: case AS IS

The export flow of goods considered for this case study is divided in the following phases:

1. *New order from the customer*: if the proprietary of the goods decides to perform all the customs clearance procedures in the plant, the domiciled procedure authorized by the Customs Authority is carried out. Otherwise, the customs clearance will be performed in another point of the transportation flow. It must be noted that for containers, at this stage, booking of a place on ship can be performed, while in the case study such anticipated booking of a place on ship for trucks is not accepted by the shipping agents.
2. *Choice of transportation mode*: the goods are ready to be sent. There are two different possibilities: by road (i.e. goods boarded as a complete truck or trailer) or by railway (i.e. goods boarded as a complete truck).
3. *Arrival of trucks*: the flow is organized in two different ways depending on the type of means of transport to be boarded. More precisely, a complete truck (with the trailer and the cab) has to stop in the dry port area before entering the port; a trailer has the possibility to choose either to go directly to the port or to pass through the dry port area (typically these represent the 30% of the total flow of trailers). When a truck arrives at the Fernetti intermodal terminal, its arrival is registered.
4. *Inside the Fernetti terminal*: if goods are already cleared, then there is a truck parking area, where truck is waiting to be called for the boarding in the port of Trieste. Otherwise, there is a dedicated area where the truck will be moved and the customs clearance procedures will take place.
5. *Customs clearance procedures*: the bill of loading and the "cargo manifest" are transferred to the customs in order to transfer information about all the transported goods. Then the bill regarding all the customs duties is prepared: it contains a customs code, the origin of the goods, its value and profit. Successively, the customs duties are paid.

6. *Booking*: when a truck arrives at Fernetti, the truck driver books a place on a Ro-Ro vessels and gives to the shipping agents all the needed documents. When the ship arrives at the port and it is ready to be loaded, the truck driver receives a communication through a variable message panel. At this point, the truck driver receives back a certificate that enables him to go in the port.
7. *Transportation phase to the port*: the truck driver leaves the Fernetti terminal and goes to the port of Trieste. In order to avoid too many delays, each truck driver has an hour and a half to reach the port.
8. *Security checks*: goods arriving at the port may have to be checked by the customs. If the freight has to be checked, the truck driver has to move the truck to a special area for the security check operations, made by the Customs Agency and by a Customs Anti-Fraud Service (Italian acronym is SVAD), to take place.
9. *Boarding*: at this point of the flow the trucks or the trailers are ready to be loaded on ship.

Looking at the intermodal terminal of Fernetti, there are two different areas: A) an area for the trucks that have to perform customs clearance operations which has 252 numbered parking places and there is a ticket to be paid of 10€ for 72 hours. For each ship about 15-20 trucks perform the customs clearance in this area. B) a second area where the trucks are waiting to be called in the port: it has 120 non numerated places and the ticket is paid by the shipping agent. This is the zone that has a function of a real dry port area.

Of the total number of trucks arriving at the port of Trieste, the 30% of them pass through Fernetti. In particular, there is a local regulation imposing that all the complete trucks have to stop at Fernetti. Therefore, the remaining truck trailers that have to be loaded on the Ro-Ro ship arrive directly at the port by motorway or by railway. There are 3 trains per day arriving at the port and the traffic arriving by the railway represents the 15% of the total Ro-Ro traffic.

It is very important for the shipping agent to monitor the freight once it has left the Fernetti terminal to reach the port. The shipping agent calls the truck driver to come in the port one hour and half before the loading while the travelling time is about half an hour. Nevertheless, there is a percentage of 5% of trucks that do not arrive on time and therefore the customs have to check the freight and often because of this extra imposed delay the truck cannot be loaded. Currently the freight is not continuously monitored and that causes several problems to the planning of the loading of the ship.

Normally a Ro-Ro ship transports 238 units, and one third of them are complete trucks. The volume of the traffic in 2010 was of about 105.000 loaded units (37.000 complete trucks and 68.000 truck trailers), divided in 15 ships per week. In normal conditions the waiting time before loading is about 25/30 hours but in case of congestion it can even reach 100 hours.

2.3. Description of the export phase: case TO BE

The following description aims to highlight the changes that are foreseen in our “to be” solution after the introduction of the DSS. The following description concerns only the changes of the new scenario:

1. *New order from the customer.*
2. *Transportation planning*: through the booking service provided by the ICT platform, customers will be able to buy services from suppliers (e.g., freight forwarders). The supplier will be able to book a place on a vessel and to start the delivery processes.
3. *Arrival of trucks*: if a ship is ready and there is enough available space, then the DSS communicates it to the truck driver before entering in Fernetti through the ICT platform and the Variable Message System signals. Therefore, the truck driver can go directly to the port or, proceed, as planned at the beginning, through Fernetti. A truck that arrives to Fernetti is automatically registered and communicated through the ICT system to the shipping agent and to the terminal operator.
4. *Inside the Fernetti terminal.*
5. *Customs clearance procedures*: the customs receive in advance in electronic format the data needed for the customs clearing operations. Truck will be assigned an on board computer: in such a way the transport will be easily recognized/detected and the clearance procedures can proceed faster. In such a case all the information can be elaborated in advance and the on board system guarantees that the transport will be completed without delays. Hence, the local Customs Office will eventually have to undertake a reduced number of checks, being sure that what was found/registered in previous control has not changed during the transport.
6. *Booking*: through the ICT platforms it is possible to book a place in the ship for the truck using the booking service application. Then the DSS plans the bookings and through the ICT integrated tools sends back the information at the shipping agency, which calls the truck driver as soon as the ship is ready to be loaded.
7. *Transportation phase to the port*: when the truck driver leaves the Fernetti terminal, the ICT platform checks if the ticket has been paid and communicates to the shipping agent the exit time. Once the truck arrives at the port, the entrance time is communicated to the shipping agent.
8. *Security checks*: by monitoring the truck by GPS system, the ICT platform is able to

communicate to the customs if there is a need of security check. In particular, if there are not anomalies or if a delay is registered but the truck has respected the foreseen path, security checks may be avoided; otherwise the truck will be checked. Thus, the ICT platform helps to reduce the number of unnecessary inspections and to improve the targeted checks.

9. *Boarding*.

Summing up, the DSS leads to an optimized utilization of resources, both human and machines, that are involved in the process of management of goods.

3. MODEL BASED DECISION SUPPORT SYSTEM

Modern DSS spans a wide range of technologies (Turban et al. 2010). In this paper, we consider a model-based DSS to tackle at the tactical level of the logistics processes. Model based DSSs give the possibility for the user to manipulate model parameters, to examine the sensitivity of the results or to conduct what-if analysis.

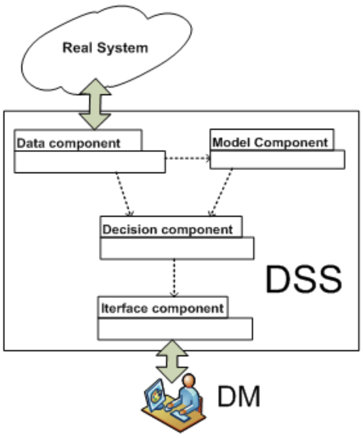


Fig.1: The general DSS Configuration

Figure 1 sketches the overall structure of the DSS: it receives data from the real system, elaborates them and finally suggests decisions to the Decision Maker (DM). Then the DM can use the DSS to change and set the parameters of the real system. Several categories of DSSs exist but almost all of them share common characteristics. DSS include three main components: the data component, the model component and the interface component (Turban et al., 2010).

The data component usually consists of a database and a database management system (DBMS). The used data can be internal, if they come from organization's internal procedures and sources such as products and services prices, recourse and budget allocation data, payroll cost, cost-per-product etc. (Turban et al., 2010). External data can be related with competition market share, government regulations and may come from various resources such as market research firm, government agencies, the web etc. In some cases, the DSS can have its own built database or it may use other organizational databases either by connecting directly with them or by using data available from reports (Turban et al., 2010).

The model component mainly includes mathematical models describing the operations of the organization in various levels and the type of functions used according to the operation that they have to support.

The decision component is in charge to suggest and support the DM during his decision process. It can merge information coming from the data component and the model component in order to propose solutions to the DM through the interface model.

The interface component is responsible for the communication and interaction of the system with the decision makers. Regardless of the quality and quantity of the available data, the precision of the model in describing the organizations procedures and even the hardware capabilities, the interface of the system must ensure that the decision maker will be able to take advance of the system capabilities.

3.1. The Discrete Event Simulation

Simulation is a descriptive tool that can be used for both prediction and exploration of the behavior of a specific system. A complex simulation embedded in a model-driven DSS can help a decision maker plan activities, anticipate the effects of specific resource allocations and assess the consequences of actions and events (Power et al. 2007). "What if analysis" performed via a simulation tool can offer extra confidence to the DM than just the simple presentation of numbers in a tabular format. Simulation enables the evaluation of the behavior of a particular configuration or policy by considering the dynamics of the system.

The starting point of the simulation model developed for the DSS is the description of the system by Unified Modeling Language (UML) (Boschian et al., 2011, Fanti et al., 2012). Successively the UML model is translated into a simulation model, whose dynamics depend on the interaction of discrete events, such as demands, departures and arrivals of transporters at facilities, acquisitions and releases of resources by vehicles, blockages of operations. In particular, we implement the model described in the UML framework in the Arena environment (Kelton et al., 2009), a discrete event simulation software particularly suitable for dealing with large-scale and modular systems. The following steps can be performed in order to specify the simulation:

1. the Arena modules are derived from the UML activity diagram elements, by establishing a kind of mapping between each Arena module and the UML graphical element of the activity diagrams;
2. the simulation parameters are included in the Arena environment, i.e., the activity times, the process probabilities, the resource capacities,

and the average input rates are assigned. Nevertheless, these specifications can be modified in every simulation cycle and enable the choice of the scenarios for the case study implementation and management;

3. the simulation run of the experiments is singled out. The performance indices are determined and evaluated with suitable statistics functions.

In order to analyze the system behavior, a set of performance indices are selected:

- the system throughput, i.e., the average number of containers delivered per time unit by the inland terminal;
- the lead time (LT), i.e., the average time elapsed from two particular activities in the system;
- the average percentage utilization of the resources.

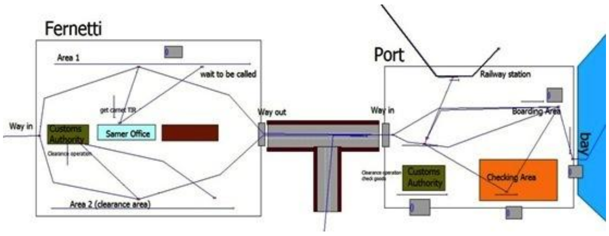


Fig.2: The snapshot of the Arena Model.

Figure 2 shows a snapshot of the model implemented in ARENA environment depicting the main components of the system:

- the Fernetti inland terminal is described by two areas: Area 2 for units that have to do the Customs operation and Area 1;
- the transport system is the stretch of highway connecting the Fernetti area to the Port;
- the port area, including Railway, the Customs Authority office, the checking area, the boarding zone and the bay.

3.2. Harmony Search Optimization

The Metaheuristics are methods that guide other procedures (heuristic or truncated exact methods) to enable them to overcome the trap of local optimality. Although these methods (tabu search, simulated annealing, etc.) are generally designed for combinatorial optimization in the deterministic context (Zapfel et al., 2010) and may not guarantee the convergence, they have been quite successful when applied to simulation optimization. Among the various metaheuristic approaches Harmony Search (HS) is quite new method that has been applied successfully in various areas (Geem 2010).

HS is a metaheuristic that mimics in a way the musicians improvising process (Geem et al., 2001). It was originally developed for integer variables but since then variants for both integer and binary variables have been proposed. The method uses a Harmony Memory (HM) that stores the best so far candidate solutions which form a pool for creating new candidate solutions.

The set of operators include:

A) Random selection: a new value is chosen randomly out of a candidate set with a probability (1-HMCR). B) Memory consideration: one value is chosen out of the HM set, with a probability equal to the harmony memory considering rate (HMCR). C) Pitch adjustment: a value that has been selected in the previous step of memory consideration is further changed into neighboring values with a probability equal to the pitch adjusting rate (PAR).

If the newly generated vector is better than the worst vector in HM with respect to the objective function, the former takes the place of the latter.

As it has been identified by many studies, in a simulation optimization framework, for each candidate solution we need to use at least some predefined number of replications in order to reduce the effect of the simulation noise (Schmidt et al., 2006). As it was described above, HS is searching the solution space based on the information stored in the HM. It is therefore important to update the HM in a consistent manner in the presence of (simulation) noise.

Therefore after each iteration of the algorithm, which involves a predefined number of replications for the new candidate solution (in the case of integer variables a lookup table for already visited solution can save some execution time at the expense of increasing the memory requirements) an Optimal Computing Budget allocation (OCBA) step is involved to assign the number of replications to a set, including the new position as well as the harmonies stored in HM in order to exclude the worst of them from the HM.

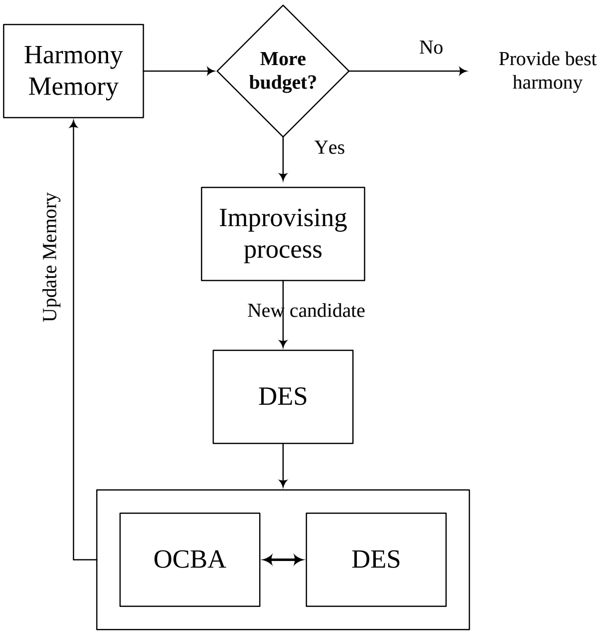


Fig.3: Schematic representation of the integration of DES+HS+OCBA for DSS.

OCBA was recently proposed (Chen et al., 2010) as a procedure to optimally allocate a predefined number of trials/replications in order to maximize the probability of selecting the best system/design. OCBA tries to maximize the probability of Correct Selection P (CS) (the probability of actually selecting the best among the k designs). More specifically the Approximate Probability of Correct Selection is a lower bound of the P (CS).

The aforementioned technologies are integrated in order to provide the DM with a tool to make (near) optimal decisions in a reasonable amount of time. Figure 3 depicts the architecture of the system.

4. CONCLUSIONS

This paper specifies a DSS for the management of complex logistic system. In particular, we examine the case of the Trieste port with the terminal of Fernetti (Italy) where we model, simulate and optimize the vehicle flows between these two coupled ports. There is proposed an integrated system to optimize the operation for the export phase. The integrated Decision Support System is based on a Discrete Event Simulation Module combined with Optimal Computing Budget Allocation (OCBA) that are further optimized with the Metaheuristic approach of Harmony Search (HA). The proposed framework is going to be further tested and it seems to be suitable for many other applications areas.

ACKNOWLEDGEMENT

This work was supported by the E.U. FP7–PEOPLE–IAPP–2009, Grant Agreement No. 251589 (SAIL).

REFERENCES

Boschian, V., Dotoli, M., Fanti, M. P., Iacobellis, G., and Ukovich, W. (2011). A Metamodeling Approach to the Management of Intermodal Transportation Networks. *IEEE Transactions on Automation Science and Engineering*, 8(3), pp. 457-469.

Chen, C.H., and Lee, L.H. (2010). *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. World Scientific Publishing Co.

Duinkerken M.B., Dekker R., Kurstjens S.T.G.L., Ottjes J.A., Dellaert N.P. (2006). Comparing Transportation systems for inter-terminal transport at the Maasvlakte container terminals. *OR Spectrum* 28,469–493.

Fanti, M. P., Iacobellis, G., Georgoulas, G., Stylios, C., and Ukovich, W. (2012). A Decision Support System For Intermodal Transportation Networks Management. In Proceedings *The 24th European Modeling & Simulation Symposium*, September, 19-21, Vienna, Austria.

Fu, M.C., Chen, C.H., and Shi, L. (2008). Some topics in simulation optimization. In Proceedings of *Winter Simulation Conf, IEEE*, pp 27-38, Piscataway, NJ.

Gambardella L.M., Rizzoli A.E., Zaffalon M. (1998). Simulation and planning of an intermodal container terminal. *Simulation*, 71, 107-116.

Geem, Z. W. (2010). *Recent advances in harmony search algorithm*. Springer Verlag.

Geem, Z.W., Kim, J.-H. and, Loganathan, G.V. (2001). A new heuristic optimization algorithm: harmony search, *Simulation* 76 (2), 60– 68.

Giani, G., Laporte, G., and Musmanno, R. (2004). *Introduction to Logistics Systems Planning and Control*. Willey.

Kelton, W. D., Sadowski, R. P., and Swets, N. B. (2009). *Simulation With Arena,5th ed.* MA: McGraw-Hill, Boston.

Ottjes J.A., Veeke H.P.M., Duinkerken M.B., Rijsenbrij J.C., Lodewijks G. (2006). Simulation of a multiterminal system for container handling. *OR Spectrum* 28, 447–468.

Parola F., Sciomachen A. (2005). Intermodal container flows in a port system network: Analysis of possible growths via simulation models, *International Journal of Production Economics*, 97, 75-88.

Power, D. J., and Sharda, R. (2007). Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*, 43(3), 1044-1061.

Ramstedt L., Woxenius J. (2006). Modelling Approaches to Operational Decision-Making in Freight Transport Chains", *Proc. 18th NOFOMA Conference*, Oslo, 7-8 June 2006, Norway.

Roso V., Woxenius J., Lumsden K. (2008). The dry port concept: connecting container seaports with the hinterland. *Journal of Transport Geography*.

Schmidt, C., Branke, J., and Chick, S. (2006). Integrating techniques from statistical ranking into evolutionary algorithms. *In Applications of Evolutionary Computation*, Volume 3907 of LNCS, pp. 753–762.

Stahlbock R., Voß S. (2008). Operations research at container terminals: a literature update. *OR Spectrum*, 30, 1-52.

Turban, E., Shardaand, R., and Delen, D. (2010). *Decision Support and Business Intelligence Systems.* Prentice Hall.

Zapfel, G., Braune, R, and Bogl, M. (2010). *Metaheuristic Search Concepts to Production and Logistics. A Tutorial with Applications*. Springer.

APPOINTMENT SCHEDULING FOR A COMPUTED TOMOGRAPHY FACILITY FOR DIFFERENT PATIENT CLASSES USING SIMULATION

F. Boenzi[(a)], G. Mummolo[(a)], J.E. Rooda[(b)]

[(a)]DM3 - Facoltà di Ingegneria del Politecnico di Bari (Italy)
[(b)]Department of Mechanical Engineering - Eindhoven University of Technology (The Netherlands)

[(a)]boenzi@poliba.it, [(a)]mummolo@poliba.it, [(b)]j.e.rooda@tue.nl

ABSTRACT

In the present paper, a discrete event simulation model of a CT facility within a hospital is presented. The examination facility has to serve different patient classes with different priorities. At the strategic level, outpatient daily access is filtered out by means of the adopted appointment schedule (AS) scheme, whereas, at the tactical level, the decision about which patient to examine is taken by the established priority rule. Being the two levels inter-related, a comprehensive model of the examination process can help analyzing the different patient flows from a global point of view, taking into consideration equipment utilization and patient service performance, both in terms of waiting time at the facility and appointment interval for outpatients. Despite the model has been developed for a specific case-study, it is flexible and different data and settings could be easily implemented. Furthermore, some general considerations are drawn.

Keywords: simulation, health care, outpatient scheduling, CT examinations

1. INTRODUCTION

Complex diagnostic services, as in the case of CT (Computer Tomography) scans or MRI (Magnetic Resonance Imaging), require expensive equipment and very specialized human resources, making their full utilization an unavoidable necessity. As a order of magnitude, the cost of an MRI machine ranges between 1 and 2 million euros, depending on the magnetic field intensity, along with huge costs for building and preparing the space it will occupy. The cost of CT equipment is similar, essentially depending on the number of slices the machine is capable of producing for image computation. Generally, in hospitals, these types of resources are utilized for serving at least both classes of patients: inpatients (patients at the hospital wards) and outpatients, so that “customers” compete for accessing them in short periods of time. The idea behind is that making the resource shared is beneficial for reducing its idle time and achieving better utilization. In outpatient clinics, managers have looked to the popular policy of “Open Access” (“do today’s demand today”) as a solution for avoiding wasted capacity due to no-shows. Alternative booking techniques, based on short booking window and on the optimal policy from a Markov Decision Process, can perform even better in terms of smoothing out the demand and reducing peak work-load, as illustrated by (Patrick 2012). In hospitals, being only outpatient access planned in advance, in the Appointment Schedule (AS) definition phase, i.e. defining the number of service slots per time session and number of appointments at the beginning of each slot, the scheduler has to take into account allowance for the random (internal) demand component. In daily operations, priority rules are usually implemented (i.e. decision about which class of patient should be served first when both, random and planned demand, are present at the facility). When the diagnostic service facility is also open to patients from the Emergency Department (ED), they generally must be served as soon as possible, possibly on their arrival, unless the resource is already busy. Similarly, priority of outpatients over inpatients is justified by the simple consideration that, in any case, inpatients have to wait in their wards, whereas, for outpatients, excessive waiting time determines overall negative service perception, as pointed out by Sickinger and Kolisch (2009). Occasionally, outpatient prolonged waiting time could lead to over timing the appointment admittance time span and to service denial or incurring additional costs. On the other hand, it should also be noted that too long inpatient waiting time, due to examination postponing, could lead to un-necessary longer stays in hospitals and increasing cost for bed occupancy (Green et al. 2006).

2. LITERATURE SURVEY

The first research papers on outpatient appointment schedule date back to the ‘50s (Bailey and Welch 1952). Since then, many authors have dealt with this subject in many different settings; an overview can be found in Cayirli and Veral (2003). More specifically, as regards to a CT scan facility, open to the above mentioned patient flows, Kolisch and Sickinger (2008) propose a mathematical model, involving a Markov Decision Process. In their model, it’s possible to distinguish two levels: an upper level which we could consider as “strategic”, regarding the outpatient Appointment

Schedule (AS) and a lower level, which we could regard as “tactical”, involving decisions to undertake on the run at a discrete number of time-points. These decision levels were yet characterized by (Green et al. 2006), who proposed a finite-horizon dynamic control model for the two capacity management tasks (appointment scheduling and real-time capacity allocation), highlighting their interrelation, and applied it to an MRI hospital facility. At the decision level, the choice at the beginning of each time-slot is about serving waiting scheduled outpatients and/or inpatients, stated that if an emergency patient were generated in the previous time-slot, he must always be served. Stochasticity in the system is introduced by diverse probabilities associated to the different classes of patients: no-shows for outpatients and random arrival of an inpatient and/or an ED patient in the previous slot (the limitation to one is among the model assumptions). Instead, service time, equal to the slot time, is assumed to be deterministic. The underlying Markov Decision Process is determined by the established decision policy, which should aim at maximizing an expected total reward function. In general the latter consists of the sum, over the entire appointment time span, of a linear combination of served inpatients’ and outpatients’ reward, waiting costs and penalty costs for service denial at the end of the shift (ED patients excluded). The optimal policy can be found by the “backward induction algorithm” (Puterman 2005). However, since the authors observe that “the acceptance for computer-based decision rules in medical environments is low”, they investigate the performance of “simple decision rules which can be applied manually”. The examined rules are LCA (Linear Capacity Allocation) introduced by (Green et al. 2006), FCFS (First-Come-First-Served) and Random, which are compared in combinations of three different scenarios (generated varying the problem parameters) and three different AS schemes from literature (2BEG, Block and Threshold). Even though the LCA rule performs better, the authors underline the importance of the fairness of the rule for the reduction of the perceived waiting time by the patients. FCFS, contrary to LCA, is a very simple and fair decision rule regarding its inter-class selection behavior, so that it can be considered as a “fair heuristic”.

In successive work (Sickinger and Kolisch 2009), the authors focus their attention on the “strategic” level, on the basis of the results obtained optimally solving the associated stochastic dynamic program. They carry out an empirical study on a two CT scans examination service with 8 slot available, under three increasing system utilization levels (number of scheduled patients equal to 4,8,12 respectively). They compare the values of the objective function resulting from a proposed Generalized Bailey-Welch (GBW) schedule, a Neighborhood Search (NS) heuristic and the optimal scheduled (obtained by full enumeration). The authors find that the GBW rule and the NS heuristic generate optimal or near-optimal solutions for low and medium utilization, whereas, for high utilization, the GBW rule, contrary to the NS heuristic, does not provide optimal solutions any more. They also analyze the impact of the function parameters on the results (adopting NS schedule as reference) and highlight the gap with GBW. In particular, in cost structures characterized by relevant penalties in case of denial of service for outpatients, the GBW rule becomes the best choice. Anyway, it’s observable that, in case of high utilization, nothing assures that all the outpatients will be served within the service time period and for this reason the authors themselves recommend the calculation of a scheduled optimal number.

The problem arising from the co-existence of random urgent patients aside the scheduled ones is often faced by leaving some slots “open” to accommodate urgency. Taking this into account and treating the position of a couple of open slots as a decision variable, Klassen and Rohleder (1996) carry out a full factorial ANOVA analysis on a simulation model (in SIMAN IV simulation language) of a family medicine clinic. The authors adopt 10 known pre-defined AS rules as second decision variable and take into account two environment factors involving probabilistic considerations (3 possible mean values and 5 levels of percentages of clients with “low” standard deviation of lognormal service time distributions). They analyze the system performance in terms of WIT (sum of expected total clients’ waiting time and expected total server idle time costs) and other secondary measures. In their model there is not a decision process and a decision policy (tactical level) because of the presence of only two patient classes (scheduled and urgent calls) and of the assumption that the clinic could accept at most two urgent patients per session (number equal to the open slots). According to their findings, simple rules like 2BEG (Bailey’s rule, with 2 clients in the first slot) and 4BEG (4 clients at the beginning) perform worse than rules which take into account client’s classification into two possible service time variance groups (low and high), when assigning them to the slots. In particular, the LVBEG rule (low variance clients at the beginning) proves to be the best rule, also for its equanimity in balancing clients’ time and server time. However, its practical implementation requires the availability, at the clinic, of recorded information about clients’ past service times and more attention by the receptionist. The simplest rule FCFA (first-call-first-appointment) is, in all the examined cases, in the group of the best rules and should be preferred for its simplicity. In successive work (Rohleder and Klassen 2002), the authors modify and expand their model, addressing the issues of rolling-appointment horizon and variable demand load, using simulation. They carry out a full factorial analysis considering six demand patterns, six overloading rules and three rule delay periods. Results are summarized in a matrix that outlines good managerial choices for each scenario.

Kaandorop and Koole (2007) consider the problem of optimal outpatient appointment schedule, in which only this class of patients is present. The objective function to be minimized is a linear combination of mean waiting time, mean doctor's idle time and mean tardiness (time exceeding the given session period). They prove that the proposed local search algorithm converges to the global optimum under the problem stated conditions; moreover, they highlight that "for certain parameters value the Bailey-Welch rule is indeed optimal". It should be noted that service time is assumed to be exponentially distributed, which is quite uncommon in healthcare services.

As seen in literature, several cases of appointment service systems, accounting or not for additional random demand, exist, which makes very difficult drawing out general rules, easily understandable by healthcare operators. Simulation in this field of study is deemed to be a very flexible tool, especially for the capability of including particular singular features of real-case systems, typical of the healthcare sector. In the present paper, a discrete event simulation model of one CT examination server within a urban hospital is considered.

The remainder of the work is organized as follows: in Section 3, the process of CT examination demand generation is illustrated, in Section 4 the model is presented, in Section 5 simulation results are illustrated and commented, finally Section 6 presents the conclusions.

3. CT PROCESS DESCRIPTION

The CT scan is located at the radiology department of a large community hospital in Basilicata region (Italy); a schematic layout of the department is reported in Figure 1. In the blank rooms, other type of equipment (traditional x-ray technique machines and ultra-sound scanners) or technical rooms are present, and, on the left part of the figure, the proximity with the ED department, located on the same floor, is depicted.

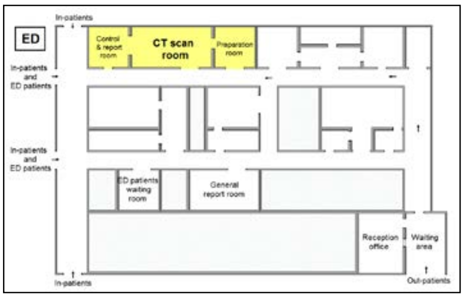


Figure 1: Radiology Department Layout

The diagnostic examination facility has to serve different patient classes with different priorities; specifically, in descending priority order: a) ED patients, b) urgent in-patients, c) outpatients and d) non-urgent in-patients; each of them, except for scheduled patients, is characterized by a random arrival process. These patients flows are coexisting during the reception opening time (from 8 a.m. to 8 p.m., from Monday to Friday), whereas access for urgent cases of type a) and b) is possible at any time. On their arrival, patients c) and d) are checked-in and their data put in the Radiology department Information system (RIS) by reception staff, whereas for patients a) and b), the examination requests are generally sent in electronic form, which makes check-in possible also when the reception is closed. The list of waiting patients builds up the work-list for the technologist in the CT control room. Once registered, outpatients, stationing in a dedicated area, wait for call and generally reach the examination room autonomously, following signposts; sometimes, they are accompanied by relatives or by department attendants. In-patients and ED patients are always accompanied by attendants or by one of the technologists themselves, because generally they are on wheelchairs or wheel-beds. A relevant difference between outpatients and other types of patients, in addition to diverse priority, is that, for CT scan examination which require intravenous administration of a contrast medium by a nurse, patients have to pass first through a preparation room, whereas hospitalized and ED patients generally can access the CT room directly. In the remainder of the paper, the preparation phase for outpatients has not been taken into account for total process time quantification (as if it were part of waiting time) because the two processes (examination and preparation), taking place in two different rooms, can be regarded as independent and parallel processes (i.e. a patient can undergo a CT scan, while the next one is being prepared). The two activities can still, in rare cases, overlap. This happens when preparation time takes too long (especially with elderly persons) and, meanwhile, the CT room has turned free or in the case when, at the end of the examination, a patient pleads indisposition and has to be monitored by the nurse, who, therefore, can't take care of the next one, causing delays.

As regards the process of examination generation, for inpatients (urgent and not), examination requests are generated by doctors in the various hospital departments and ED patients requests are generated, when needed, after their arrival at the ED. For out-patients, the requests are generated by family or speciality doctors. After that, possible points of access to health-care services are by phone-call to a unified regional call-centre or by taking the paper request to a "CUP" (unified booking centre) office. In any case, information about the next available appointment slot in a health-care structure able to dispense the requested service are shared in real-time. Requests are added so forth and build up waiting lists. For outpatients, a random generation process is adopted at the origin of the demand, whereas their daily access to the examination process is filtered out through the AS scheme established by the department director. Inpatient and ED patient access requests can be considered random generated and flow to the daily work-list; anyway, at the "tactical" level, the decision about which patient to examine is set by the mentioned priority rule. The described process and its integration with the examination process are illustrated in Figure 2, in which the time-span within the two vertical dashed lines represents the time elapsed from the input of the request into the "CUP" system to the appointment day.

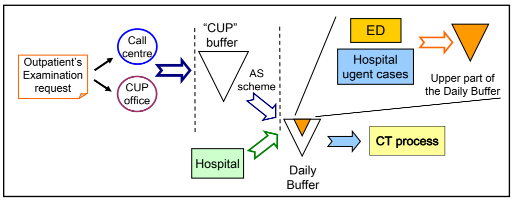


Figure 2: Examination Generation process and buffering

The aim of the present paper is analyzing the performance of the system altogether in terms of waiting time for the different patient classes and machine utilization by means of simulation. Differently from existing literature, which generally focuses on patient waiting time during the appointment session, for outpatients it's very important to know also the "long" waiting time (usually measured in weeks) to the appointment day. Moreover, according to recent regional legislation, outpatients have to be differentiated at the origin, in order to account for particular urgent cases and their waiting time must comply with specified limits. Three classes of outpatients have been characterized depending on this time limit: appointment within 10 days, denoted as "B"; within 30 days, denoted as "D"; without particular time constraint, denoted as "N". Correspondingly, three different booking lists have been set up at "CUP" and at the radiology department. Outpatient waiting time in the department on the appointment day could be considered a secondary measure of performance, whereas it remains very important for emergency cases. Of course, the two aspects are strongly inter-related because, at the strategic level (AS), scheduling a bigger number of outpatients can shorten their waiting queue and "long" waiting time, but, on the other hand, can congest the system, increase unacceptably waiting time for the random low-priority component and eventually lead to lateness of the examination session (overtime), post-poning or cancellation of scheduled patients. The simulation model aims at offering a global view of the system performance when the different patient flows are coexisting.

4. MODELLING

4.1. Patient data

The employed data come from two different sources: RIS data reporting the CT examinations performed at the radiology department for each patient class in the period October 2012 – May 2013 and data of examination requests (classified as outpatient N, D or B) arrived at the "CUP" office over the same period. For the codification of the various types of examinations into the model, a national codification system, set up in (VV.AA. 2006), has been partially used. The latter system, also for accounting purposes, consider a single body-part CT scan as a coded examination. In reality, the majority part of patients undergo some typical CT sequences (as in the case of chest-abdomen, brain-chest-abdomen or "total-body" scan); for these examinations, aggregating the sequence into a whole has led to create appropriate additional new codes. RIS data for outpatients have not been used, since their access to the service is regulated by means of scheduling. For inpatients and ED patients RIS data have been filtered with the check-in time, excluding the examinations not within the reception opening days and hours of the day. The resulting examination mix is reported in Figure 3. Average inter-arrival time and throughput values of examination requests (outpatients) and of patients (ED and inpatients) are reported in Table 1 (averaged over the net reception worked hours in the observed period).

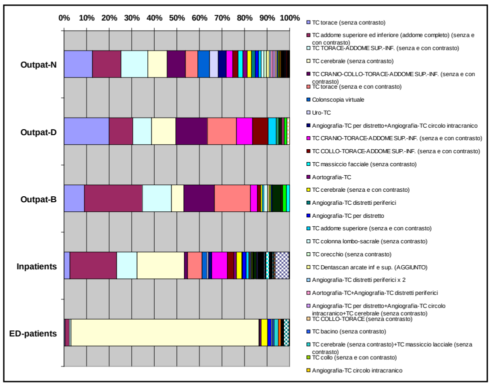


Figure 3: Patient Examination Mix

Inpatients and urgent inpatients have not been differentiated with regard to the examination mix (i.e. any type of examination could be requested with urgency); their relative proportions in current data are 78.8% and 21.2% respectively.

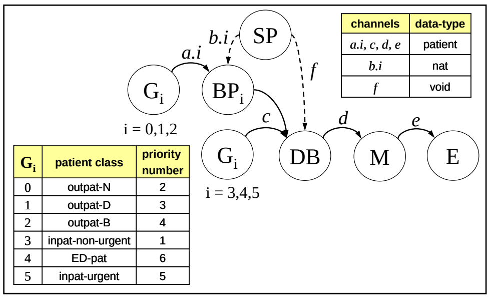
Table 1: Patient Flow Data

	Outp-N	Outp-D	Outp-B
t_a (min/pat)	105.67	1276.55	798.99
Thrput (pat/h)	0.57	0.047	0.075
	Inpat	**ED-pat**	**Urg-Inpat**
t_a (min/pat)	93.80	175.45	348.15
Thrput (pat/h)	0.64	0.34	0.17

4.2. Discrete-event simulation model

The model is a stochastic discrete-event simulation model built by means of the process algebra language Chi 1.0 (Hofkamp and Rooda 2007). With Chi, a symbolic representation of a system is translated into a model, consisting of parallel processes, which communicate, in a synchronous way, one with each other via channels. Data exchanged on channels can represent physical entities (e.g. patients) or information contents (e.g. signals, data, etc.). Among the principal advantages of the language are its capability of preserving formalism and it's ease of comprehension and transparency even to non-expert people. Process based language Chi has been used in manufacturing modeling as well as in the health-care sector. For example in (Jansen et al. 2012), an aggregate model of an MRI department is developed employing effective process time (EPT) concepts.

In the present paper, instead of aggregating examination data, their differentiation is maintained on the basis of the examination types because of the necessity, for future model developments, of characterizing some particular types, for which separated outpatient booking lists are currently adopted in the department (in addition to the three mentioned waiting lists). In Figure 4, the Chi model of the CT diagnostic service is depicted, along with two tables: one summarizing the correspondence between generator processes and classes of patients and priority rank; the other reporting channels and exchanged types of data. The arrowed lines represent the patient flows and the dashed lines represent exchanged data.



channels	data-type
a.i, c, d, e	patient
b.i	nat
f	void

G_i	patient class	priority number
0	outpat-N	2
1	outpat-D	3
2	outpat-B	4
3	inpat-non-urgent	1
4	ED-pat	6
5	inpat-urgent	5

Figure 4: Model and Processes

The model consists of six patient generator processes G_i, three buffer processes BP_i, the scheduling process *SP*, the daily buffer process *DB*, the examination process *M* and the exit process *E*.

The generators G_i $(i=0,1,2)$ represent the arrival processes of examination requests arriving at "CUP", for the three types of waiting queues (N, D and B). Examination requests (patients in the model) are then put in queue in the buffer processes BP_i, as happens at the booking office. Patient data-type contains information regarding the code of the requested examination, chosen randomly on the basis of the found mix (see previous section) and the assigned priority. The arrival process is modeled as a homogeneous Poisson process (HPP), with mean inter-arrival time according to Table 1; this assumption of the model comes from the hypothesis that the served population remains constant and there is not a seasonal component. The generators G_i $(i=3,4,5)$ represent generators of the random component of the demand, represented by inpatients, ED patients and urgent inpatients. For each of these generators also, a HPP is assumed. This hypothesis is indeed an approximation, because variability in the course of the day and eventually from a day of the week to another day occurs. Generators of inpatients and urgent inpatients utilize the same examination mix, but have different mean inter-arrival times, respecting the proportions found in current data.

Process *SP* represents the scheduling process for authorizing outpatient buffers BP_i to release fixed numbers of patients to process *DB* in the course of the day, according to the established AS. This information is communicated to each BP_i via channels *b.i* of integer numbers. The probability of no-shows is not included in the current model and therefore the scheduled number always corresponds to the number of released patients, unless the buffer becomes empty. *SP* assumes also the function of a cyclic clock in the model, because it takes care of the passing time at disposal to complete the daily schedule. At the end of the day, a signal is sent via channel *f* to process *DB*. Generators G_i $(i=3,4,5)$ are not "filtered" by a scheduling process, but are directly linked to the *DB* process.

Process *DB* simulates the daily buffer of patients to be examined each day, sorted according to their priority number. This buffer is indirectly related to the *SP* process, since its filling up follows a cyclic behavior, on the basis of the AS; in addition, it is also subject to "disturbances" due to the unplanned arrival (according to HPPs) of the other types of patients. At the end of the day, a signal is received by *SP* via channel *f* and the remaining numbers of patients are monitored, in order to calculate average values. The most critical event which could happen is that in this buffer there are still outpatients; this means that the random arrival of patients with greater priority has prevailed, impeding the completion of the daily schedule. According to the hospital staff, this event is rarely possible, but, in any case, scheduled outpatients must be examined that day. The same could happen for inpatients, who are normally examined only during the reception opening time. However, to a limited extent, this is not a serious problem, considering that inpatients are in wards at the hospital. Patients remaining in the daily work-list are normally examined in overtime and this doesn't have consequences on the next appointment sessions. Therefore, in the model, buffer *DB* is emptied at the end of each day.

Process *M* represents the examination process, in terms of CT room occupation time and is modeled as a time delay for the patient. For some examination codes, collection of empirical data has led to determine maximum likelihood estimated parameters of *gamma* probability distributions, with acceptable results of goodness-of-fit tests at the significance level 0.05 (Boenzi et al. 2012). For other codes, for which commonly used PDFs don't fit satisfactorily, empirical distributions are implemented. For all the other CT examination codes with a scarce number of observations, process mean values μ are assumed on the basis of technologists' esteems. *Gamma* PDFs are then adopted, assuming a worst-case approach with regard to the highest variability for process time. Among the observed examination types, the maximum found coefficient of variation $c=\sigma/\mu=\sqrt{(1/\alpha)}=0.55$ is selected and it is employed to calculate a common shape parameter $\alpha=3.3$. Then, different scale parameters β are calculated as μ/α.

In the model, at last, patients are sent to the exit process *E*, in which indicators regarding patient flow (data deriving from time-stamps at each stage) are calculated.

5. SIMULATION RESULTS

All the following simulation results are calculated as average values of five independent simulation runs, time-terminated at 500.000 minutes, time limit which corresponds to approximately 3 years of continuous reception time operation. In all the simulations, the system starts in empty conditions, i.e. preceding patient queues are disregarded.

5.1. Current system

The AS currently adopted at the radiology department is depicted in Figure 5.

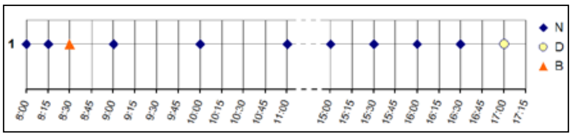


Figure 5: Current AS for CT examinations

The daily schedule (Monday – Friday) follows the reported scheme, except for Wednesday, when only the first part of the schedule (till the dashed lines) is adopted, because in the afternoon a special health monitor program is in place. Consequently, in the simulation model, a generic cycle of the scheduling process *SP* (see Section 4.2) can be either 720 minutes long for a full day (8 – 20) and 420 minutes for half a day (8 – 15). A complete weekly cycle comprises four full days and one half-day, i.e. 3,300 minutes in total. It can be observed that, in the first part of the morning shift, an approach similar to the Bailey-Welch rule is adopted, placing three patients, with only 15 minutes time-elapse, in the first operative hour. This can account for possible no-shows (not modeled), but inevitably increases patient waiting time. It can also be observed that, in the morning, the examination slot duration is set to one hour, whereas in the afternoon shift it is reduced to half an hour. This scheduling decision is due to the assumption that the random demand, especially by ED patients, is more intense in the morning. The model presented in this paper, however, doesn't take this aspect into consideration. Outpatient admission is distinguished substantially between two blocks and the last appointment in the morning is at 11 a.m. The clearance time between the two blocks is devised, according to the department staff, just in order to process the main part of inpatient examinations. Comparing the weekly accepted number of outpatients with the observed average weekly demand (Table 2), it's possible to observe that, with the current schedule, some extra capacity is employed. This is clearly an effective strategy to reduce waiting queues.

Table 2: Outpatient Examination Demand

Average weekly demand (over 35 weeks)		**Current weekly schedule**	**Extra-capacity**
Outp-N	31.23	41	1.31
Outp-D	2.59	4	1.54
Outp-B	4.13	5	1.21

Simulation results are reported in Tables 3 and 4. The Appointment Interval (A.I.) represents the time-span for obtaining an appointment at the facility (measured in effective days, including Saturday and Sunday) and waiting time (w-time) values, instead, refer to waiting at the department. Utilization of the system is 0.498.

Table 3: Outpatient Flow Performance

	Outp-N	**Outp-D**	**Outp-B**
Avg Thrput (pat/h)	0.572	0.048	0.077
Avg A.I. (days)	**0.31**	**1.79**	**2.96**
Min A.I. (days)	0.0001	0.0094	0.0138
Max A.I. (days)	2.06	10.39	15.61
Avg w-time (min)	**9.50**	**8.08**	**12.01**
Min w-time (min)	0	0	0
Max w-time (min)	182.11	93.97	114.10

Table 4: Hospital Patient Flow Performance

	Inpat	**ED-pat**	**Urg-Inpat**
Avg Thrput (pat/h)	0.63	0.34	0.17
Avg w-time (min)	**17.51**	**6.64**	**7.14**
Min w-time (min)	0	0	0
Max w-time (min)	273.75	116.33	116.14

Similar results are obtained assuming that, in the model, the examination process start is systematically 40 minutes delayed, from 8 a.m. to 8.40 a.m. This is a pessimistic but realistic assumption, because, occasionally, due to organizational reasons and limitation of personnel resources, it could happen that the CT facility is not fully operational at the start of the shift. Waiting time results are reported in Table 5, in which an increase for low-priority patients can be observed and the minimum waiting time for B-type outpatients is ten minutes, as expected. The above assumption will be held also in the following.

Table 5: Patient Flow Performance at the department with delayed CT room availability

	Outp-N	**Outp-D**	**Outp-B**
Avg w-time (min)	**29.29**	**7.91**	**15.28**
Min w-time (min)	0	0	**10**
Max w-time (min)	219.02	77.36	101.78
	Inpat	**ED-pat**	**Urg-Inpat**
Avg w-time (min)	**28.05**	**7.60**	**8.71**
Min w-time (min)	0	0	0
Max w-time (min)	297.88	112.24	137.17

Since the system starts in empty conditions and an extra-capacity is put at disposal, outpatients are characterized by very brief average appointment time-spans, also with regard to maximum values. In these conditions, if the examination demand remained the same, waiting lists would be progressively emptied and the strategic objectives, regarding reduced appointment time-span, met. Over-sizing of the current AS is also testified by Table 6, reporting the time average buffer size and the average number, at the end of a generic cycle, of additional patients which could have entered the system if the buffer could have released the requested number.

Table 6: Outpatient Buffers and additional potential patients

	Outp-N	Outp-D	Outp-B
Avg Buffer size (patients)	1.83	0.93	2.17
Avg additional patients (patients/cycle)	1.95	0.28	0.17

5.2. New schedule analysis

In this and in the following sections, some scenario hypothesis are formulated and simulation results are illustrated and compared. The first hypothesis to be investigated assumes an increment of outpatients exactly equal to the current weekly scheduled number. Results are reported in Table 7. Utilization of the CT room (0.557) is only marginally improved, but the average A.I. for outpatients of type B (and, much worse, its maximum value) is above the acceptable limit and tends to increase in time, i.e. the system doesn't reach a stationary condition. This trend is also confirmed by the buffer time-history reported in Figure 6 (reporting the results of five simulation runs) and is due to the variability of the arrival process.

Table 7: Outpatient Flow Performance under the hypothesized increment and the current AS

	Outp-N	Outp-D	Outp-B
Avg Thrput (pat/h)	0.737	0.0701	0.0874
Avg A.I. (days)	**4.08**	**28.79**	**21.78**
Min A.I. (days)	0.0042	2.20	0.77
Max A.I. (days)	**15.35**	**57.98**	**56.47**
Avg w-time (min)	30.09	9.41	14.92
Min w-time (min)	0	0	**10**
Max w-time (min)	219.71	111.16	93.09

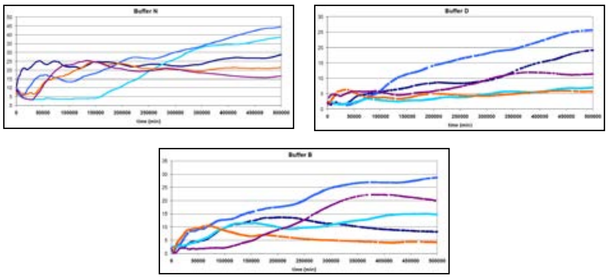


Figure 6: Outpatient Buffers time-history

In order to accommodate the increased examination demand, a new AS, reported in Figure 7, is proposed and tested. On its basis, the weekly admitted outpatient numbers increase according to Table 8.

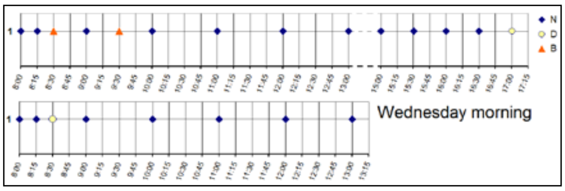


Figure 7: New Proposed AS

Table 8: Outpatient Examination Demand Increment

Hypothesized outpatient weekly demand		New weekly schedule	Extra-capacity
Outp-N	41	51	1.24
Outp-D	4	5	1.25
Outp-B	5	8	1.6

Simulation results are reported in Table 9, showing that offering extra-capacity assures achieving the stated appointment-time objectives, also with regard to maximum waiting time.

Table 9: Outpatient Flow Performance under the hypothesized increment and the new AS

	Outp-N	Outp-D	Outp-B
Avg Thrput (pat/h)	0.741	0.0747	0.0912
Avg A.I. (days)	**0.28**	**2.56**	**1.05**
Min A.I. (days)	0.000084	0.00926	0.00159
Max A.I. (days)	1.68	11.44	4.84
Avg w-time (min)	26.40	9.47	15.11
Min w-time (min)	0	0	0
Max w-time (min)	242.22	105.37	122.76

5.3. Hypothesized scenarios

Maintaining the illustrated hypothesis of increased outpatient demand and employing the newly proposed AS, additional "what-if" scenarios take into consideration the eventuality of rising up of the random urgent patient component (ED patients and urgent inpatients), who are assigned greater priority ranks with respect to outpatients. The scenarios comprise the following: a_1) increasing ED demand 50%; a_2) increasing ED demand 100% (doubling the current figure); b) changing the percentage of urgent inpatients from the current figure (21.2%) to three different levels: b_1) 50%, b_2) 63.3% and b_3) 75%, constant in time. The second level has been calculated in such a way that the summed throughputs of current ED patients and urgent inpatients is equal to the summed throughputs of current urgent inpatients and doubled ED patients, i.e. the urgent throughput is the same for a_2) and b_2). Scenarios a_1) and a_2), even if clearly over-estimated, could be the consequence, for example, of the closure of one or more neighboring EDs. As regards to scenarios b), they come from the consideration that, realistically, inpatient whole throughput can't increase, because it is linked to the hospital bed capacity. Instead, its urgent component could increase, considering that hospitalized patients, as society, are an aging population and that employing the form of urgent examination request could be increasingly utilized by doctors at wards, in order to shorten their dismissal. As also pointed out in Section 4.2, the major impact that urgent patients can have on the appointment schedule is not succeeding in examining all the planned outpatients. In order to monitor this, the remaining number of patients in the daily work-list, in the course of a simulation run, is summed up and average values over the total number of daily cycles are calculated. Therefore the average value can also be regarded as the probability of finding a certain type of patient remaining in the work-list, at the end of a generic daily schedule. Comparative results are reported in Figure 8. Urgent patients do not represent a serious concern because access for them is granted at any time: therefore they can always be present in the work-list. Instead, non-urgent inpatients should be preferentially completed during the reception opening time. It can be observed that in scenario a_2), due to their low priority and the increased number of high priority patients, inpatients have to wait and therefore it's more likely to find any of them in list at the end of a day. In scenarios b), this effect is mitigated, also compared to initial system conditions, because quantitatively their presence is reduced. For outpatients, the probability of over-timing is, in all the examined cases, very low. It can be observed that the impact on over-timing in scenario b_2) is more severe, compared to scenario a_2), even if the urgent throughput is the same. This can be explained considering the different examination mix of inpatients and ED patients and the greater variability for the first ones, with longer time examinations. In general, the impact of increasing urgent inpatient percentage is greater than increasing ED throughput, even if in the first case there is not a net increment of patients.

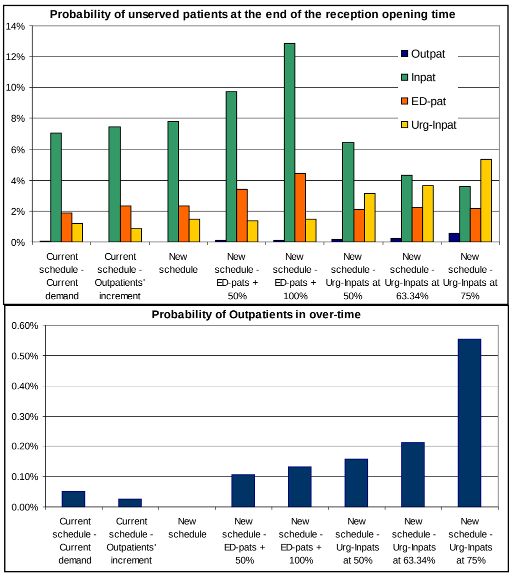
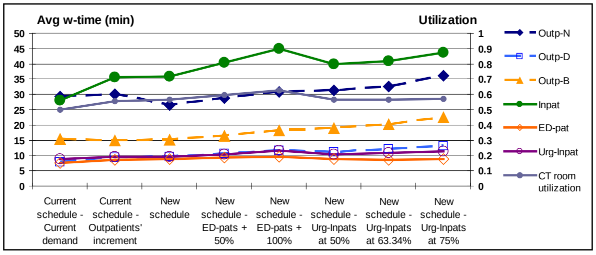


Figure 8: Average Number of patients remaining in the Daily Work-list per cycle

In Figure 9, average waiting time values at the department, along with CT room utilization, are summarized. It can be noticed that, in general, waiting time is acceptable, even though, especially for outpatients of type N, it could be improved changing the AS and taking into account, in simulation, the possibility of no-shows. Inpatients are, in all the examined cases, the most penalized service users and the impact of their lowest priority is particularly evident on the maximum waiting time (not reported). In turn, this could eventually lead to the decision of postponing them to the next daily cycle, incurring costs for additional hospitalization. In all cases, utilization is low, ranging from around 0.5 to 0.62.

Figure 9: Average Waiting Time and Utilization



6. CONCLUSIONS

In the present paper a discrete-event simulation model of a hospital CT facility has been presented. The aim of the model is giving a global view of the problem of coexisting admitted classes of patients, in terms of local performance (waiting time at the department for the different patient classes and CT equipment utilization) and long-term performance, represented by the average appointment interval for outpatients. The last is determined by the AS policy and, as illustrated by means of a case-study, the two aspects are inter-related. Current situation and some hypothesized scenarios, employing a different AS, have been illustrated and qualitatively compared. Even though the obtained simulation results refer to a particular case-study, two general recommendations can be drawn: 1) in schedule planning, setting up extra-capacity with respect to the current average external examination demand, instead of offering a capacity strictly equal to it, prevents the making up of appointment queues, because of the random nature of the process. Therefore, external demand should be periodically monitored by the "CUP" staff and eventually determine the AS redesign.

2) It should be avoided a too rigid application of the priority rule for non-urgent inpatients (lowest priority) in order to prevent examination postponing. Therefore an alert system for excessive waiting time should be implemented, permitting in some cases overriding the rule at the expense of outpatients. This could in turn cause the increase of average outpatient waiting time, but, as illustrated, over-timing is a rare event.

Future work comprises model validation, for which data collecting is in progress, and finding strategies for AS improvement.

REFERENCES

Bailey, N.T.J., Welch, J.D., 1952. Appointment systems in hospital outpatient departments. *Lancet* 259: 1105–1108.

Boenzi, F., Mummolo, G., Rooda, J.E., 2012. Analysis of a Diagnostic Radiology Department with different patient flows using different data sources. *Proceedings of the International Workshop on Applied Modelling and Simulation*, 114–124. September 24-25, 2012, Rome, Italy.

Cayirli, T., Veral, E., 2003. Outpatient scheduling in health care: a review of literature. *Prod Oper Manag* 12: 519–549.

Green, L.V., Savin, S.V., Wang, B., 2006. Managing patient service in a diagnostic medical facility. *Oper Res* 54(1): 11–25.

Hofkamp, A.T., Rooda, J.E., 2007. C*hi 1.0 Reference Manual*. Eindhoven University of Technology. Available from: http://seweb.se.wtb.tue.nl/sewiki/_media/chi/chi10_refman1689.pdf [Accessed 15 July 2013].

Jansen, F. J. A., Etman, L. F. P., Rooda, J. E., Adan, I. J. B. F., 2012. Aggregate simulation modeling of an MRI department using effective process times. *Proceedings of the 2012 Winter Simulation Conference*, 919–930. December 9-12, 2012, Berlin, Germany.

Kaandorop, G., Koole, G., 2007. Optimal outpatient appointment scheduling. *Health Care Manag Sci* 10: 217–229.

Klassen, K.J., Rohleder, T.R., 1996. Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management* 14: 83–101.

Kolisch, R., Sickinger, S., 2008. Providing radiology health care services to stochastic demand of different customer classes. *OR Spectrum* 30(2): 375–395.

Patrick, J., 2012. A Markov decision model for determining optimal outpatient scheduling. *Health Care Manag Sci* 15: 91–102.

Puterman, M.L., 2005. *Markov decision processes: discrete stochastic dynamic programming*. New York. Wiley.

Rohleder, T., Klassen, K.J., 2002. Rolling Horizon Appointment Scheduling: A Simulation Study. *Health Care Manag Sci* 5: 201–209.

Sickinger, S., Kolisch, R., 2009. The performance of a generalized Bailey–Welch rule for outpatient appointment scheduling under inpatient and emergency demand. *Health Care Manag Sci* 12: 408–419.

VV.AA., 2006. *Metodologia di determinazione dei volumi di attività e della produttività dei medici radiologi - Nomenclatore SIRM-SNR delle prestazioni radiologiche.* OMICRON Editrice Genova. Available from: http://www.sirm.org/index.php/documenti/doc_download/35-nomenclatore-e-calcolo-volumi-di-attivita [Accessed 15 July 2013].

SIMULATION MODELS TO SUPPORT GALB HEURISTIC OPTIMIZATION ALGORITHMS BASED ON RESOURCE BALANCING BASED ON MULTI OBJECTIVE PERFORMANCE INDEX

Sergio Amedeo Gallo[a], Giovanni Davoli[b], Andrea Govoni[c], Riccardo Melloni[d]

[a, b, c, d]Department of Engineering “Enzo Ferrari” (ex DIMeC),University of Modena and Reggio Emilia, ITALY

[a]sgallo@unimore.it, [b]giovanni.davoli@unimore.it, [c]andrea.govoni@unimore.it, [d]riccardo.melloni@unimore.it

ABSTRACT

The following paper face with an approach to analyse a multi model manual assembly line, and the following heuristic algorithm to optimize the scheduling of tasks to the available stations, respecting of a set of restrictions, as task/station obligation, and aiming to optimize a multi objective function, based on time, balancing utilization rates, and line balancing costs elements.

This problem can be considered as belonging to the wide area of GALB Problems.

Some strategy about resource scheduling opportunities has been considered.

The original referable configuration of the considered system is an assembly line with six stations, and with six operators.

In the present step, the solution we experimented, shows a redundant (doubled) number of stations, to be a more flexible solution compared to previous solutions, with a layout too much specific and referred to the particular tasks and constrains profile, based on real data.

Keywords: workstation design, work measurement, ergonomics, decision support system

1. INTRODUCTION

As in previous works, in following lines we describe main models and system features.

We face with a manual assembly lines.

Items advancement on line is done on a accumulating conveyor system, so **line** is **paced**, but **not synchronized**. Just a single accumulation among stations, is allowed, in the original system. This opportunity makes the flexibility level higher, but we have to consider no inter operational buffer.

Assembly line process a very large variety of items, defined in families, 6 in the original Assembly Plan, (**AP**), that differ for size, features, optional, lot size. Spurred by the increasing market competition, and customers’ requirements, are commonly accepted very low quantity for single order. Tasks assignment to stations must respect efficiency targets, as the maximization of utilization rate, balancing concerns, assignment constrains restrictions of specific tasks to specific stations where special machines are available. This constrains are commons for all items of the mix.

Assembly plan data, **AP** Data, and system configuration arise from those of a real assembly line. Task times are stochastically distributed based on real observations. The influence of stochasticity is not the focus in the present scheduling problem, because no cost for off line completion can be considered, and because tasks that do not respect line cycle time, cause just the tack time increase of the single item, but not of the mean of performed Tack Time for the whole lot.

No incompletion costs and operators moving cost are considered, because both negligible compared to operating cost, both because of the short distance to cover moreover, no costs related to operators training, or both changes of the tasks to operate, has been considered, because, we are facing with a Multi Model Assembly Line, and, furthermore, operators are supported with displays showing instructions for tasks to operate.

The same tasks of different items, has operating times, that can vary for each item, for the operation declared in the same way. All tasks in the whole annual assembly mix in AP are represented on the precedence diagram, uniform for all items. To assembly a model not all the 34 operations are needed for a specific item, depending on features and optional. Each item has a defined number of operation which ID number increase as the assembly process goes on.

The performance parameters are the production rate, to be maximized, that means to reduce tack time, and, at the same time, optimize the internal balancing of both tasks for stations, both labour level among operators. Based on these criteria, a multi objective function with the aim to minimize the whole lot assembly cost, calculated on the effective tack time and on the current scheduled resources and their balancing level, has been defined.

The results demonstrate the capability of the proposed algorithm of dealing with the multi objective nature of the re-balancing problem. Solutions with advantages both in tack time reduction, and both on balancing improvement are obtained.

The heuristic algorithm is based, at each step on a logic trying to improve the previous balancing configuration.

We built a virtual model of the assembly line in a simulation environment, to test and measure performances of the heuristic algorithms, but, also all the algorithm code has been implemented in the same software platform, so simulation has been used not only as a verifying tool, but also as a solving or solution finder, and as task and resource scheduler.

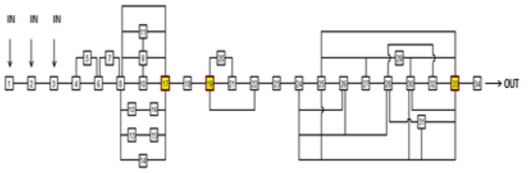


Fig. 1: combinatorial diagram of sequences

The definition of the scheduling in the AP is oriented to the lean production philosophy: first production order has to be produced first, too. Anyway, at least, one week, is the planning time horizon an order can be shifted inside the scheduling window, without affect due dates. So, it is possible over pass the rule to route assembly orders as they were placed in the order list, to get better system performances.

Precedent models have been developed. The first one represents the “as is” configuration of the firm for strategies and configurations, is a basic model to be compared with our improved one. It was also used to verify and validate the virtual representation comparing to available real data. Also any implicit scheduling and assigning rule has been checked and verified.

The following models, last ones before the present one, Gallo and al. (2012), is in many parts similar to the present. So it seems appropriate to recall some feature and logic.

They were based on task assigning rules, attempting to fulfil stations adding tasks to, in sequence, till total station time doesn’t overpass Reference Tack Time, **RTT**, calculated for each Order Line, **OL**, and item. Moreover, additional control code was devoted to check if any constrained task is joined, and in this case, provide to verify the station where to assign that task, if not the current one, and to calculate all parameters for intermediate stations.

Additional rules evaluated if all already scheduled operators were idle, and, if not, the algorithm attempted to re allocate the idle ones to more suitable stations. Again, all new values of parameters were calculated and compared to the old ones, as the objective cost function. Just the best allocation survived and recorded.

Moreover, if some station was undercharged, a routine incremented recursively the tack time, till all those stations became empties, so that released operators could be re - assigned to over charged station.

The constraint position of a special chamfer machine, allocated on a defined station, and the assignment constrains of other operations, to stations where other equipment is available, dramatically limited opportunities to gain better performances based on balancing the line with a mixed sequence of selected groups of items extracted from the AP, conveniently defined for quantity and for typology, and seemed, not feasible (balancing on scheduling).

Finally, in the last previous models, to improve performance, based on residual efficiency edges, i.e., on the complementary values of the SUCs, was defined a double line to assemble coupled mixed items, contemporary, each one on one of two lines. This kind of configuration was based on some aspect of constrains position, and on the specific AP data.

The opportunity we found to achieve better performances, grouping single units of different available orders, was of **coupling tasks to be assembled**, on two **parallel lines**, fed in counterflow. A single operator should be assigned to corresponding stations, one on the first line and one on the parallel line, and they should have to complete their operations, alternatively, on both stations of two lines. An improved performance was gained, and was possible operate more assembling strategies.

At this time, a very relevant observation raised in our mind: the parallel coupled line can be a general and versatile solution when applied to more general data sets and different constrains position and configuration?

NO, of course!

In fact, the previous proposed solution to balance the line, was too much based on the specific constrains and available data. Profiles of Stations Utilization Rates, SUCs resulted of similar shape for the large part of available items, differing, often, just for the scale of the RTT. Also, the need to find good matchable items, close enough in the AP, is not sure to face with, also considering that the amount to produce, or the correspondent assembly time, should be quite correspondent. On the opposite, experimenting some heuristic rule to assign to same operators more than one station charge, seems more likely hopeful, especially watching at some item, with residual edges for resources SUCs, no further more improvable because of line configuration, and constrains accomplishment.

So the gauntlet to build a more general, flexible, versatile system configuration, with the associate heuristic logic strategies, was picked up, and challenge started.

1.1. Present models configuration

The strategies to distribute tasks to stations are similar to the previous models.

Also in the present one, task assignment to station has to respect efficiency concerns, as the maximization and the balancing of utilization rate, but, first, specific task assignment restrictions, because of the need of special machines, available at defined station.

The initial heuristic strategy, **config_1,** try to assign tasks in sequence to stations, till the RTT is fulfilled, or till total station time doesn’t overpass calculated referable task time. An additional control logic to verify if a specific task is one constrained to be assigned to a specific station, and in case, the consequent logic to point to the correct station, and to calculate all parameters for intermediate stations, is present, also.

In **config_2**, it is present a logic to evaluate if there is some unused resource, and in this case, the heuristics reassign to the more overcharged stations to help the more suitable stations already assigned.

Again, in **config_3**, it is present a logic to evaluate if, after the first assignment, some undercharged station/operator results. The charging level is defined as a the percentage ratio of the current RTT. A routine defines a RTT recursive increment till, all the undercharged station become empties, so freed operators could be reassigned to over charged stations, as in the previous mentioned case, **config_4**, and **config_5**.

Simulation models can read AP data form a CSV file, with any useful attribute to be used to characterize the specific OL, and the configuration: time distribution parameters, item definition, order quantity, order date, etc.. In this way, is very easy to change configuration.

Any time a new strategy is applied, all performance parameters, as station/operators utilization coefficients, UCs, Direct Assembly Cost, are calculated, stored and compared to best performing configuration emerged at the previous step. Just the better, for each item, is the one considered for the final solution.

But, considering that in the previous model, in the case of the single line, at the end of the cascade of logic steps, nothing more to achieve a better performance was possible, based on work assignment balancing strategy, the current new strategy was considered. The present idea is to increase the number of the stations, with the aim of enforce balancing opportunities based on the resource/work balance.

We decide to create two stations where before one, to have a more relaxed assigning opportunity. Also the constrains position at specific stations has been doubled, to keep proportion with the original system:

- line configuration, the number of available stations, the equal number of assigned operators has been doubled to 12.
- Any equipment available at a constrained i-th station, has been located at 2 * i-th station.
- The tacks assignment logics in the heuristic is the same.
- At the end of the logic cascade, when no further opportunity to achieve a better balancing based on works contents of stations, an adding logic starts to calculates the maximum value of the station time, in other words, the line tack time, and searches to find stations which work load have a sum equal to the current line tack time for each order.

After any assignment strategy was tried, after the reallocation of idle or freed operators to the currently overcharged stations, in the respect of the constrains, it is possible assign operators more than one stations.

To support this opportunity, a **U shape line** has been considered. A display on the top of any station shows which is the current item, which is the mounting cycle and parts, and if and where to go. Not all of initial 12 stations and operators, will be scheduled as final optimized configuration. For each item or OL just one configuration is the best, with specific number of stations, and operators, too.

Models can be applied to a wide variety of systems, with different number of stations, and different constrains positions, just integrating new additional constrains rules based on new configuration values. Data have been those arising from the real system.

Achieved results showed good improvements compared with initial solutions, and any time a new strategy was applied.

2. LITERATURE REVIEW

An assembly line is a flow-oriented production system, where the operative location units performing work, referred to as stations, are sequentially aligned. Work pieces move on transportation systems as a conveyor.

Their configuration and planning is relevant both as a optimization problem both because they are systems at medium intensive capital.

Assembly Line Balancing Problem (**ALBP**) means the assignment of tasks to stations and operators on a line, whereas the items are produced at pre-specified production rate. Configuration planning covers both all tasks allocation and both decisions related to equipping and aligning the productive units for a given production process, including setting the system capacity (cycle time, number of stations, station equipment) as well as assigning the work content to productive units (task assignment, sequence of operations).

Since the times of Henry Ford and the model-T, customer requirements, and consequently, production systems, have changed in a way to increase dramatically customization of their products. The high level of automation of assembly systems and the fixed movement system make the (re)-configuration of an assembly line critical.

In literature, there is a wide variety of algorithms to solve ALBP, any one facing a partial part of the problem, or oriented to a particular system or configuration.

Many of them consider the problem too much statically, just under a one point of view.

But the increasing need to face continuous changes in customer's requirements, as product design, restyling and lot quantity needed, enforced with high customization and reduction of time-to-market, push to test dynamic versions of ALBP solution procedures.

Those modifications imply a very high flexibility level for the line.

ALBP consists of assigning tasks to stations in such a way that (Salveson, 1955):

- each task is assigned to one and only one station;
- the sum of performance task times assigned to each station does not exceed the cycle time;

- the precedence relationships among the tasks are satisfied;
- some performance measures are optimized.

Most procedures consider the types **I and II ALBP**, based on minimization of the number of stations, given a desired cycle time or minimization of the cycle time, given a desired number of stations, respectively.

Because of the simplifying assumptions of this basic problem, this problem was labelled simple assembly line balancing (**SALB**) in the universally accepted review of Baybars (1986). Subsequent works attempted to extend the problem by integrating practice relevant aspects, like U-shaped lines, parallel stations or processing alternatives (Becker and Scholl, 2006), referred to as general assembly line balancing (GALB).

Scholl (1995), and Pierreval et al. (2003) proposed a very large and comprehensive reviews of the approaches developed to solve the problem.

Ghosh and Gagnon (1989) defined a taxonomy to classify ALBP solution procedures under two key aspects, mix or variety of items produced on a single line and the nature of performance task times: single model lines or multi/mixed model lines manufacturing more items in batches or simultaneously; deterministic ALBPs, in with performance task times constant, or stochastic ALBPs, with stochastic task times distributed according to a specific distribution function.

ALBP can be solved to optimize both time - and cost, as reported in Amen (2000, 2001) and Erel and Sarin (1998), which concern the deterministic and stochastic versions of the problem, respectively.

Moodie and Young (1965), Raouf and Tsui (1982), Suresh and Sahu (1994), Suresh et al. (1996) have proposed time-oriented algorithms, improving procedures developed for the single-model deterministic problem, with the aim of minimize stations number and the over time to complete the work off the cycle time.

In any case, relevant incompletion costs often occur in stochastic assembly lines.

A multi objective cost function often is needed.

Two cases, both described in literature:

- the whole line is stopped till the over work is completed (Silverman and Carter, 1986);
- incomplete products get completed off-line.

Kottas and Lau (1973, 1981) proposed heuristic procedures to minimize both the total labour cost and the expected incompletion cost. Extensions of the Kottas and Lau's (1973) method were developed by Vrat and Virani (1976), Shtub (1984).

Sarin et al. (1999) proposed, not so general as Kottas and Lau's (1973), a branch and bound heuristic to minimize the total labour cost and the total expected incompletion cost with good results.

Erel and Sarin (1998) noticed the difficulty of methods in literature to model real conditions, and suggested that newer works should be oriented at useful studies, with impact on real-life assembly lines.

Rekiek (2000) observed that differences among ALBP and real-life statements were the multi-objective nature of the problem, no so considered in literature.

Some studies deal with the re-balancing problem of an existing line, as Sculli (1979, 1984) and, Van Oyen et al. (2001) considered the re-balancing of an existing line, under fluctuations of operator output rates or equipment failures, in short-term problem. The proposed solution to avoid temporary imbalance on the line has been the dynamic work sharing.

Rekiek et al. (2002) demonstrated that the integration between heuristic approaches and multi-attribute decision making techniques is a proven and efficient way for solving assembly lines problems.

3. SYSTEM AND CONFIGURATION

We though for long time to define the correct number of stations in the new systems.

For an assembly line the station number can range between one - a line degenerates in one only station, perfectly balanced - to **N,** where **N** is the total number of tasks the assembly process has been divided in. A low number don't offer a large opportunity to have a sufficient space for a resource balancing, instead, a too large number enforce the indetermination of the configuration to test.

Then, we considered to define a line with a cycle time quite equal to the half of the Ideal Tack Time in the six station configuration. The new system keeps the proportion with the previous. We just tried to "dilate" the previous system and achieve an "**homothetic**" increased system, with an opportunity chances to rebalance the line based on resources.

We have the following constrains:

- At 8th station we have a chamfer machine, the strongest constrain, and task 17 (chamfering).
- At 9th station there is a pneumatic screw driver, and task 19 (screwing).
- At 12th station we find the equipment to apply the air test, and task 33 (air testing machine).

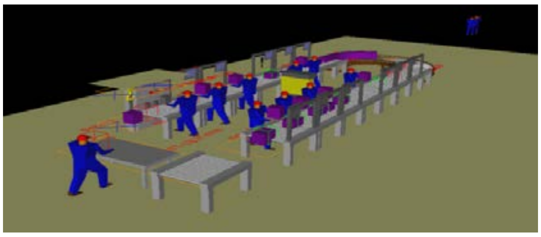


Fig. 2: screenshot of the doubled assembly line.

Execution times vary strongly, and can be described with lognormal o triangular distributions; in our case are described by triangular density functions with a large extension.

Model parameters (times in hundredth of minute):

Station Number $k \in [1, n]$ (1)

Task Number $n \in [1, 34]$ (2)

N° of tasks assigned to a single station

$i \in [1, h]$ (3)

Task Time *Top* (4)

Station Time $TStat = \sum_{i=1}^{h} T_{Op}$

$$T_{Stat} = SUC * RTT \tag{5}$$

Operation Unbalancing Coefficient

$$UC_{Op} = \frac{T_{Op}}{Tm} \tag{6}$$

Station Unbalancing Coefficient

$$SUC = \frac{T_{Stat}}{TT_{Line}} \%$$

$$SUC = \sum_i UC_{Op} = \frac{\sum_i T_{Op}}{TT_{Line}} \tag{7}$$

Line Lead Time $LLT = \sum_{i=1}^{k} T_{Stat}$ (8)

Line Tack Time or Cycle Time

$$LTT = Max(\sum_{1=i}^{h} T_{Op}) = Max(T_{Stat}) \tag{9}$$

Mean/Ideal Tack Time

$$ITT = \frac{\sum_{1=1}^{n} T_{Op}}{k} \tag{10}$$

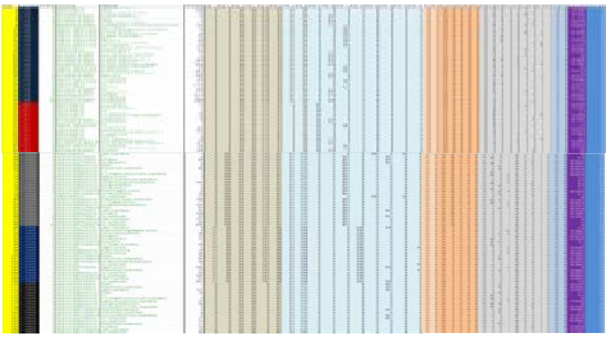


Figure 3: whole mix production plan with parameters values and task times.

3.1. MALB algorithm

Our assembly line is a multi-mixed model, then we face with a **MALBP** (Mixed-Model Assembly Line BP).

We define a cost function, Direct Assembly Cost, DAC, as the product of the manpower direct cost, multiplied by the station number (or operators when more than one is assigned to a station), multiplied by the volume for the OL, multiplied by RTT, to be representative of both the RTT of the line for each row, and for each model, both the number of resources used:

$$Direct\_Assembly\_Cost$$

$$DAC = (RTT \bullet ResNum \bullet Lot\ Quantity \bullet Man\ Work\ Cost) \tag{11}$$

Our heuristic algorithm is a mix of Work Content and Resource Balancing, that, with the objective function, takes their role and weight, very freely.

We will configure our situation as a **MALB-E** problem, given number of K stations, the aim is to maximize the efficiency E$_{\text{line}}$, i.e. minimizing the direct cost of assembling the lot.

First, we will allocate tasks to stations trying to fulfil the referable cycle time, moreover, respecting task constrains, to achieve cost minimization, and a better charge balancing.

First heuristic logic, called **config_1**, try to assign and redistribute tasks to stations in dynamic and balanced way, under the respect of all constrains, as the sequence diagram, with the aim of minimize a whole cost function, an objective function, and both to increase the Efficiency and the Balancing Level.

Line will result better balanced when maximized

$$MAX(E_{linea})\ where\ E_{linea} = \frac{\sum_i UC_{stat}}{\sum_i S} \tag{11}$$

Efficiency is calculated as sum of all SUC divided by current number of stations defined for each item, and, at the starting time, equal to resource number, then as we maximize efficiency, is maximized utilization rate and is minimized UC’s for each stations.

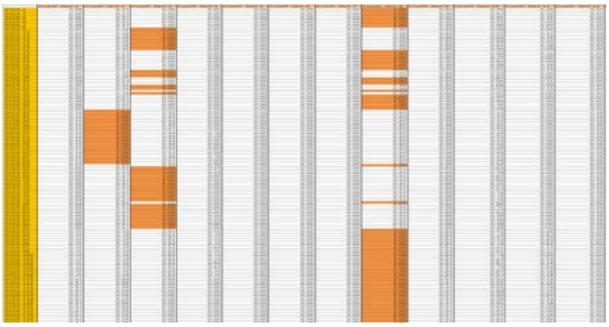


Figure 4: Unbalancing Coefficient for stations at the initial assignment configuration config_1.

After first dynamic task assignment phase, we can outline following considerations about task times.

One among stations 2, 3 or 8 is the one responsible for the line tack time, that is equal to RTT for each OL, because in this configuration, with 12 initial stations, there is more often a task time that overpass the ideal tack time for the line. Quite always, stations 6 and 7 results as empty, then needless. With a lower frequency, the same happens to stations 4 and 5, instead stations 8 and 9, that usually show high unbalancing coefficients, that is clear because those stations are constrained both.

We remember that, in the 6 stations configuration, usually, was the station number 5 to affect LTT.

Furthermore, station number 11 is charged but with a very lower SUC, ad in the following application of strategies, that station become discharged. Station number 12, also, has very low unbalancing coefficients, but unlike of the 11 ones, usually is charged with assigned tasks because there is the air test equipment.

Initially, each operator has been considered bounded to his station, and tasks allocation was made balancing on the content of work. The primary action the algorithm was designed for, is to allocate tasks dynamically to stations, with the aim of minimizing the DAC.

An important parameter has been RTT:

Reference Tack Time

$$RTT = MAX\left(T_{mean} = \frac{\sum_i T_{Op}}{n};\ MAX\left(T_{Op_n}\right)\right) \qquad (3)$$

Minimum time, in hundreds of minute, used as a limitation roof in the allocation of tasks, maximum threshold the Station Time cannot overcome. It is the higher value between the ideal tack time (**ITT**), and the maximum values of the specific various tasks (T_{OP}) for each line of the production plan. RTT represents 100% of work time that can be assigned to each station. In our case, a further control has been added to assure task allocation to the correct stations.

Other parameter, already declared, is SUC, that is the percentage value calculated as the sum of the Unbalancing Coefficient of any operation assigned to the station. It’s value is lower or equal of the RTT one, that is the 100%.

Another specific problem is the highly variable size of the tasks. Some of them, in fact, affect dramatically the Line Tack Time, and, when compared to the value of ITT, they are often even larger.

The increase of the original RTT, that is a read on a CSV file is of 5%. Any station value of the precedent step is put to zero, and again, is tried to reassign tasks to stations with the new RTT, until all tasks are assigned.

The simulation code will be used first to apply the heuristic rules and logic cascade, and later as a validation tool by testing any winning configuration with the emulation of the line. At any step all relevant parameters have been calculated and compared. Chutima P. et alter, (2004), Jolai F., et alter, (2008).

3.1.1. Model Description

AP and configuration data, are read from external files.

The code process part that takes care of reading data is called "P_read" program.

Transferring times are included in the average time of execution of the task, and result very lower when compared to operational times, with no statistical significance.

A first piece of code initializes the model and its parts, to load the variables with the values of the external file and configures the same in accordance with the structured algorithm for assigning tasks to stations.
In this phase, there will be defined the values of the Line Tack Times, of the SUCs and the parameters of cost and inefficiency. The logic routine dynamically assigns tasks to the stations respecting the allocation of joined tasks to the stations of belonging.

The file "***SpeadSheetWorkDataCSV_line_U***" contains information about a large number of parameters reported in array variables. In another reading file, "***ConfigurationCSV***", are the values of the variables to configure the system, such as the percentage increase value for RTT, the threshold value to divide tasks between two operators when more than one is assigned to a station, the percentage of tack time that defines when a station is under used, the limit value to accept an UC for the resources, and so on.

Moreover, other code portions manage the initialization of operators, of their disposal on the line, and to define the daily and weekly shifts.

As the reading process ends, the assignment process, "***P_allocation***", of tasks to stations starts. A control regards the constraints, in fact check if the task is not constrained so that could be freely assigned, or, on the other hand, if we are in the station it belongs to, so that it could be attributed.

Another control checks whether or not, the addition of the task you are trying to give, does not lead to a Station Tack Time exceeding the reference limit. In this case, the station is closed and the allocation try to assign the current task to the next available station. In case last station gets overcharged, over passing reference task time, before last task should be assigned, the code logic increase the reference by a defined percentage and set all row array values, containing stations tack time, and all others parameters to zero, and again, goes to try allocation again, till it reaches.

When an OL is already processed, because all tasks are finished, the following row is considered till the last one in the AP. When you have no more tasks to be allocated, assignment process for that line ends, and the next one in the production plan is considered, just after saving the data for each station of the completed order row within the appropriate variables: the RTT, as well as, the number of operations assigned to each station for each line, etc. are saved to array variables.

3.1.2. Reallocation of “spontaneous” idle operator and strategy of under-used stations emptying.

Now, we can observe yielded data and first conclusions and analysis: we can see many stations showing markedly under – charged station time, or even empty.

The first improving strategy provides that the operators that are already uncharged, “**spontaneously**”, are reassigned to the station with the highest tack time.

This last configuration (**config_2**) will be compared with the previous scenario (config_1), without considering resources associated with empty stations. In same case, is not possible allocate all uncharged operators because in some station there is just a long task: it doesn’t make sense schedule two operators for just one task.

The config_2 is calculated by two procedures that control the stations without assigned operations, storing them on array of pointers, to define the amount of the resulting uncharged operators to free. The second process chooses the most charged station to assign the operator and calculates the decremented tack time.

At this point the situation is photographed by saving the various parameters in appropriate variables.

As mentioned earlier, the reallocation of uncharged workers will follow.

In the second one, all stations suitable to have a assigned another operator are defined.

In "**P_undercharged_Op_forced_assignment**" procedure, all the stations for each line are checked, to trace the presence of under used stations. Once that has been identified a station of this type, it is emptied, and its operator, released by force: RTT is increased by a defined percentage, all assignment values are placed to zero, a new tasks reallocation starts till under charged station is empty.

The tack time reduction follows a procedure that care of dividing tasks operators in the best balanced manner. The procedure is repeated until there are more operators to be assigned. In previous versions we divided station tasks time by two, and half was charged to one operator and half to another. In the present model, 12 stations instead of 6, there is an increased difficulty to share tasks between two operators, because of the lower number of them that could affect the error approximation, much more than before. Four sharing strategies are been introduced, that tray to assign tasks to each of operators, looking for the best one.

Successively, through conditional cycles "if…then", all results are compared and just the best sharing is chosen for each case and the maximum time of the operators become the station time, i.e. the closest to the 50% of the station tack time.

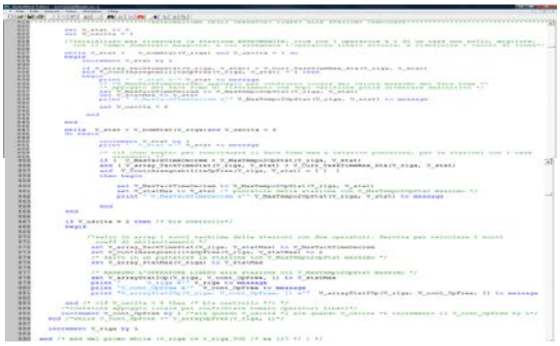


Fig. 5: logic to assign freed operators and to calculate new RTT snapshot.

A while loop choose the most decremented tack time among all stations that tried the assignment of another operator as the one we confirm, and the logic is applied till any freed operator be not assigned.

The assignment of operators to overcharged stations is done in three distinct steps: first one, called **config_3**, where in case of undercharged stations, RTT is increased recursively, and undercharged operators are just freed but not reassigned, **config_4** where just operators "spontaneously" uncharged are assigned, and then, in the **config_5**, also freed operators will be assigned again. In both strategies all logics to calculate new tack time, to define the station to help are similar, and, once again all the characteristic parameters of each situation will be saved for later comparison with those from previous situations in appropriate variables, with the same name distinct just for the suffix.

The highest tack time among all various stations, is saved for each OL as LTT

3.1.3. Resource balancing and the multi station assigning process

Just after first simulation runs, when the winning configurations for each OL, when no other opportunity to improve the balancing performance seemed possible, we tried to achieve an adding opportunity, the resource balancing.

On the same layout, for each OL we consider just the already final assignment configuration, and we tried to assign operators more than one station, in two distinct way.

In the first one, we apply a logic that calculate the RTT in the winning final configuration, **config_final**, and then, recursively, look for the station with the minimum value of the station time, not already considered. The algorithm try to match this station with the one that at this step shows the higher value for the station time with exception of the one that define the RTT. In case the station time sum doesn't overpass the RTT, the two stations are coupled and assigned to the first scheduled operator. If not, the station with the current minimum station time is tried to be coupled with the station with the second maximum station time, and so on. Then, next two station are evaluated.

In the first approach, this rule is applied without increase the RTT, **config_6**. In the second approach, **config_7**, instead, to favorite the coupling opportunity, a recursive increase of the RTT is allowed till the 200% of the initial value, or if RUCs, considering all tasks for any station assigned them, result higher than a defined percentage of RTT. At the end of all step, any performance parameter is recorded and compared.

3.2. ANALYSIS AND COMPARISON OF EXPERIMENTAL RESULTS

Single Line

First, in next tables it is possible to observe the low presence, in the best configuration case, of stations under-utilized. Second, we can see that many of the stations, under charged at the first instance, have been depleted and erased, as those already empty since first tasks allocation. This means that the final winning situations appear to be those which have a lower number of resources, except in those cases in which stations are all well filled.

Table shows as the config_final results in an overall improvement in the efficiency and cost, without worsening, in fact, but at least, values remain the same. Infact, it means that the best configuration is the first, the one obtained after the first dynamic allocation. Any way, our new dynamic assignment results a large improvement for cost and efficiency values, if we should compare these to those corresponding to the first situation, or that provided by the company.

Observing data on tables 3, it's possible note many differences.

First, in the final case there are few undercharged stations that means a better balancing level, so that config_final is with a general lower number of resources.

Figure 6: snapshot of UCS in config_1 vs config_final configuration. In blue, station with maximum UC, in green, undercharged stations, in red, empties stations.

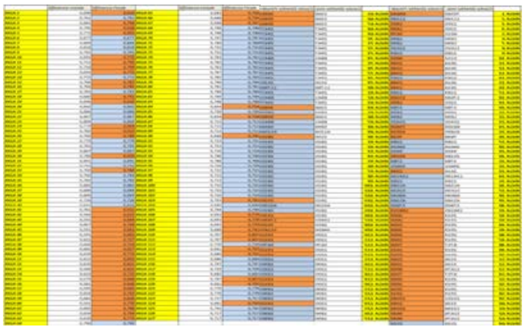


Figure 7: Efficiency and Direct Cost in the config_1 vs config_final. In orange, improved values, in blue, unchanged values, in red, worsened values.

Generally, stations that change from the config_1 to the final one, are those with one or more than undercharged stations in the initial situation.

In the table 4 UCs for the config_1 vs config_final show the gained improvement: at least, in the config_final, for a single OL, there is the same situation as in config_1, but never a worse one.

Table 1: number of winning configurations for all lines.

Config.1	Config.2	Config.3	Config.4	Config.5
69	0	58	0	0

From Table 5 it is evident that the winning configurations are config_1, corresponding to the initial dynamic distribution, and config_3, when RTT is amplified, before tasks of the under charged station are reassigned.

In the following figure 5 is showed the whole OLs Efficiency value for all considered configurations.

The whole mean efficiency value of the line after the first dynamic reassignment is of 72%.

The following phase, knew as Config_2, with the redistribution of the operators will cause a decrease, while to 52%, where, again, in the Config_3, the increase of the tack time makes possible the elimination of the undercharged stations which lead to a vertex efficiency of 78%, no more over passed also by Config_4, with redistribution of released workers, and config_5.

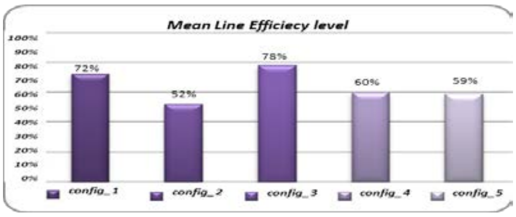


Fig. 8: Mean Line Efficiency level for all configuration.

The cost trend is still fairly close to that one of efficiency, this trend is shown in Figure 6. The Average Direct assembly Cost undergoes a substantial decrease, both in the case of dynamic allocation, equal to 16.7%, both in the case of depletion of the stations under - utilized, when compared to the initial situation.

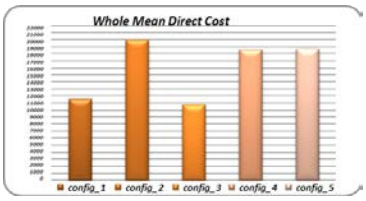


Fig. 9: Average Direct Assembly Cost for all single line configurations

Many further evidences, much more strong, arise from the observation of the UCs, when grouped just for OL with the same number of scheduled stations/resources. We will show just some situation.

We are in presence of situations with a variable number of /stations/operators, and therefore, it will be better outline results based on the number of remaining stations/operators.

Now it is clearer the balancing advantages that we can achieve with our proposed heuristic.

Moreover, we propose just means values, just aggregated in some way for shortness needs, but if we observe the single OL better results can be noted.

In fact, in the above figures never is reached the 100% value, since they are averaged.

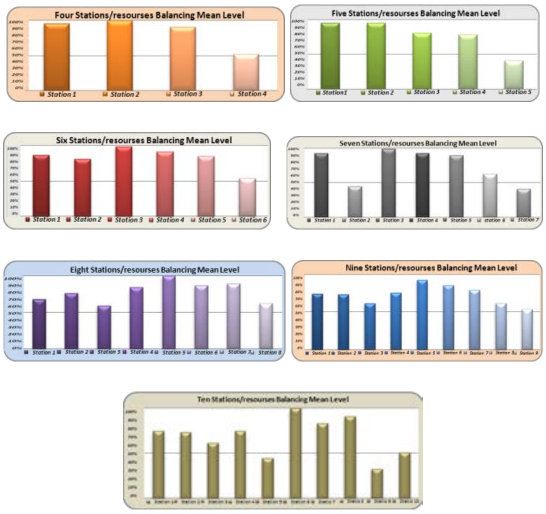


Fig. 10: Mean UCs just for OL with xx stations.

Balancing Strategy based on Resources

In the following figures, we outlines results for the resources balancing approaches.

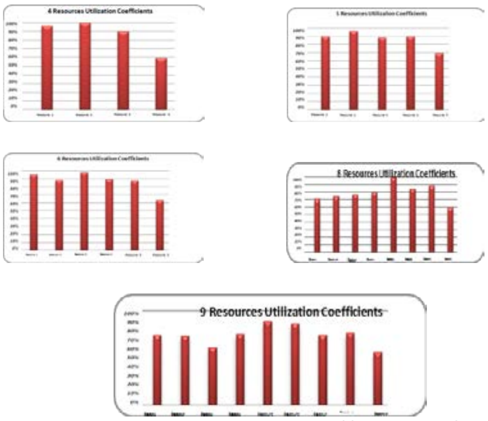


Fig. 11: Mean Utilization Coefficients for line configurations with xxx Operators and coupled stations in the first approach.

Observing UCs, a good improvement can be noted, both for the average efficiency, and both for the internal SUCs, that are much more homogeneous than before.

In the following lines we show a comparison of the average Line Efficiency among the pre coupling approach, in the config_final, in the first coupling approach, and, finally in the second coupling approach.

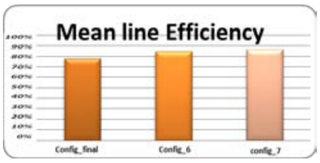


Fig. 12: Mean RUCs for line configurations.

Line efficiency presents a relevant increase of 7% since the first coupling approach, and one smoother passing to the second.

Similar consideration can raise observing the mean Direct Cost for the whole AP.

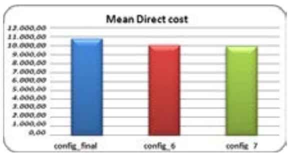


Fig. 13: Mean Direct Cost for line configurations.

In no one OL we can observe a worsening of the performances of efficiency and Direct Cost.

In fact, in config_6, without increase RTT, is just possible the RUCs improve, and in the second approach, config_7, just more opportunities to have a resource balance come, and the increase of RTT is compensated by the eventual resource reduction.

The resource balancing operates on 95 OL on 127.

4. CONCLUSIONS

In this paper a new step of thoroughly research was conducted regarding possible improvements of heuristics logics to be applied to the case of an manual assembly line.

The most critical issues were identified and then addressed the, through an multiphase algorithm definition and consequent simulation of the process.

We based this new step based on previous models and on related outputs. We oriented our attention to a more general solution, in terms of flexibility, variability of mix, number of resources and stations, number and position of some constrains.

Some deeper solutions have been evaluated to define time savings when more than one resource is assigned to the same station, and to decide the station where assign more than one operator to.

The ultimate strategy based on resource balancing seems be much more better to face a large typology of situations, with any values for tasks time and constrains positions.

The opportunity of improvement can be obtained with the layout shape and with a very low investment cost, and with a really general, versatile and flexible algorithm, under the dimension level, but also with easy configuration of data and parameters values.

The aim of this study was to define a global strategy to apply o a wide range of assembling systems, to optimize production.

All strategies have been defined respecting any of the main constraints and considering an appropriate production volume, which could give validity to the model.

Then, a cascade of ameliorative approaches were evaluated, structured as algorithms and heuristics so that they could then translate into a programming language for the implementation and verification of their actual goodness, to the computer.

Future subsequent optimization approaches could include a new data collection and the variation of data of the system randomly with logic, to have a greater validation of the algorithm.

It could be possible also refer to an advancement of multiple products simultaneously on the same line, similar to what we saw in the last part of this work, but without another line, but by simply extending the existing one, with U-Shaped layout, and evaluate, a scheduling strategy but on the double of the stations, with possible assignment of stations even at the same operators, in order to obtain a greater opportunity to balance based on the scheduling of resources, but also to be able to feed as a double line, alternately.

This opportunity is under evaluation.

Finally you could structure the analyzes concerning the study of the cost of any delays on deliveries or completion of the off-line products, evaluating solutions for the optimization of these parameters and the creation of configurations can prevent the emergence of such issues.

REFERENCES

Amen, M., 2000. Heuristic methods for cost-oriented assembly line balancing: *A survey. International Journal of Production Economics 68*, 1–14.

Amen, M., 2001. Heuristic methods for cost-oriented assembly line balancing: A comparison on solution quality and computing time. *International Journal of ProductionEconomics 69*, 255–264.

Bautista J., Pereira J., 2008. "*A Dynamic Programming Based Heuristic for the Assembly Line Balancing Problem*", International Journal of Production Economics.

Baybars, I., 1986. A survey of exact algorithms for the simple assembly line balancing problem. *Management Science 32*, 09–932.

Becker, C., Scholl, A., 2006. A survey on problems and methods in generalized assembly line balancing. *European Journal of Operational Research 168,* 694–715.

Chutima, P. Suphapruksapongse, H., 2004. Practical Assembly-Line Balancing in a Monitor Manufacturing Company, *Tharnmasat Int. J. Sc. Tech.*, Vol. 9, No. 2

Gallo, S. A., Davoli G., Govoni A., Melloni R., Pattarozzi G., Simulation models to support GALB heuristic algorithms and to evaluate multi objective performance index, The 24th European Modeling & Simulation Symposium, September, 19-21, 2012, Vienna, Austria.

Ghosh, S., Gagnon, R.J., 1989. A comprehensive literature review and analysis of the design, balancing and scheduling of assembly systems. *International Journal of Production Research 27*, 637–670.

Erel, E., Sarin, S.C., 1998. A survey of the assembly line balancing procedures. *Production Planning and Control 9*, 414–434.

Gökçen, H K. Ağpak, R. 2006. "Balancing of Parallel Assembly Lines", International Journal of Production Economics.

Kottas, J.F., Lau, H.S., 1973. A cost oriented approach to stochastic line balancing. *AIIE Transactions 5*, 164–171.

Kottas, J.F., Lau, H.S., 1981. A stochastic line balancing procedure. *International Journal of Production Research 19*, 177–193.

Jolai, F., Jahangoshai REZAEE M., Vazifeh, A. 2008. Multi-Criteria Decision Making for Assembly Line Balancing, *Springer Science Business Media*.

Moodie, C.L., Young, H.H., 1965. A heuristic method of assembly line balancing for assumptions of constant or variable work element times. *Journal of Industrial Engineering 16*, 23–29.

Pierreval, H., Caux, C., Paris, J.L., Viguier, F., 2003. Evolutionary approaches to the design and organization of manufacturing systems. *Computers & Industrial Engineering 44*, 339–364.

Raouf, A., Tsui, C.L., 1982. A new method for assembly line balancing having stochastic work elements. *Computers & Industrial Engineering 6*, 131–148.

Rekiek, B., 2000. Design of assembly lines. Memoire presente en vue de l'obtention du grade de docteur en sciences appliquees. *Universite libre de Bruxelles*, Brussels, Belgium.

Rekiek, B., De Lit, P., Delchambre, A., 2002. Hybrid assembly line design and user's preferences. *International Journal of Production Research 40,* 1095–1111.

Salveson, M. E., 1955. The assembly line balancing problem. *Journal of Industrial Engineering* 6, 18–25.

Sarin, S.C., Erel, E., Dar-El, E.M., 1999. A methodology for solving single-model, stochastic assembly line balancing problem. *OMEGA—The International Journal of Management Science 27,* 525–535.

Scholl, A., 1995. Balancing and Sequencing of Assembly Lines. *Physica-Verlag, Heildelberg*.

Scholl, A., Boysen, N., 2009. Designing Parallel Assembly Lines with Split Workplaces: Model and Optimization Procedure. *International Journal of Production Economics*.

Silverman, F.N., Carter, J.C., 1986. A cost-based methodology for stochastic line balancing with intermittent line stoppages. *Management Science 32,* 455–463.

Sculli, D., 1979. Dynamic aspects of line balancing. *OMEGA— The International Journal of Management Science 7*, 557–561.

Sculli, D., 1984. Short term adjustments to production lines. *Computers & Industrial Engineering 8*, 53–63.

Shtub, A., 1984. The effect of incompletion cost on the line balancing with multiple manning of work stations. *International Journal of Production Research 22*, 235–245.

Süer G.A., 1998. Designing Parallel Assembly Lines, *Industrial Engineering Department,* University of Puerto Rico-Mayagüez.

Suresh, G., Sahu, S., 1994. Stochastic assembly line balancing using simulated annealing. *International Journal of Production Research 32*, 1801–1810.

Suresh, G., Vinod, V.V., Sahu, S., 1996. A genetic algorithm for assembly line balancing. *Production Planning & Control 7*, 38–46.

Van Oyen, M.P., Gel, E.S., Hopp, W.J., 2001. Performance opportunity for workforce agility in collaborative and noncollaborative work systems. *IIE Transactions 33*, 761–777.

Vrat, P., Virani, A., 1976. A cost model for optimal mix of balanced stochastic assembly line and the modular assembly system for a customer oriented production system. *International Journal of Production Research 14*, 445–463.

CROWDSOURCING SUPPORTED MODELLING AND ANALYSIS INFRASTRUCTURE FOR INTELLIGENT MONITORING OF NATURAL-TECHNOLOGICAL OBJECTS

Andrejs Romanovs[a], Boris V. Sokolov[b], Arnis Lektauers[c], Julija Petuhova [d]

[a] [c] [d]Riga Technical University, Kalku Street 1, LV-1658 Riga, Latvia
[b]St. Petersburg Institute For Informatics and Automation of Russian Academy of Sciences,
14 Linia V.o., 39, SPIIRAS, St. Petersburg, 199178, Russia

[a]andrejs.romanovs@rtu.lv, [b]sokol@iias.spb.su, [c]arnis.lektauers@rtu.lv, [d]julija.petuhova@rtu.lv

ABSTRACT

The effective use of the results of ground-space monitoring and its integration with the processes of national economic management, within the intensification of human activities and recent years natural disasters and major accidents, becomes important strategic factor for accelerating socio-economic development of any region of the world. This article discusses the possibility of combining modern social technologies and the process of ground-space monitoring of natural and technological objects, as well as improving the efficiency and social importance of this process, by involving public representatives to the dissemination and use of the monitoring data.

Keywords: social technologies, crowdsourcing, ground-space monitoring, natural-technological objects, intelligent technological platform

1. INTRODUCTION

In recent years, mankind has increasingly faced with natural and technological disasters, which may be explained by the intensification of economic activities of people in the scientific and technological progress, is causing unwanted and dangerous natural and technological phenomena. At the same time as the development of science and technology society becomes more protected from natural and other disasters, the number of victims is reduced, but the total damage from disasters is rising. To reduce the risk of developing dangerous situations, a global monitoring of risk areas and facilities is needed, as well as the creation of a common information space to provide objective information to all interested parties.

Remote sensing from space provides a unique opportunity to obtain information about the objects and phenomena on a global scale with high space and time resolution. The criteria for the appropriateness of space systems in the solution of a problem are the relevance of its solutions, economic efficiency or the impossibility of solving by the traditional technologies. Most effective for the monitoring of the majority of natural and technological objects is the solution, integrating traditional and space monitoring tool.

The monitoring information regarding incidents and disasters is received typically from different data sources (e.g. biometric systems, aerospace systems, etc.), and, therefore, it is heterogeneous by nature (e.g. electrical signals, graphical, audio, video information, text, etc.). Thus, since modern natural-technological objects are very complex and multifunctional ones, their monitoring should be performed in conditions of large-scale heterogeneous data sets. Currently, the monitoring and control of natural and technological systems are still not fully automated.

Developed within the project INFROM "Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems" technology (Merkuryev, Sokolov and Merkuryeva 2012) involves the creation of an intellectual platform for the processing and use of the results of both ground-and space monitoring. Project provides the development of a common information space to monitor natural and technical objects-border states, providing for the government and the public topical environmental information for use in education, science, business, case management, and will also provide additional independent source of operational information on natural and technological hazards processes.

Another important result is to attract people to the development of innovative technologies and the active use of space activities. Developed intellectual platform will also help to reduce the risk and minimize the impact of natural or technological disasters by helping the timely notification of the population in the case of the disaster and its prognosis. To achieve this, it is proposed to use modern social technologies (crowdsourcing) that have been widely spread in many areas of the economy.

2. CROWDSOURCING AS A MODERN SOCIAL TECHNOLOGY

Recently conducted by a well-known Gartner Inc. company studies have shown the need for entrepreneurs, willing to win the competition in the market, choosing the new business models (instruments and processes), which primarily will be based on social networks and media (Bradley and McDonald 2011). One of the such tools is the crowdsourcing, which is a conceptual part of the Human Computing, which can take various forms (Participatory Sensing, Urban Sensing, Citizen Sensing), in accordance with the scale of involvement of the people, tasks, they are addressed to design, and incentives that are designed to facilitate their participation.

The term "Crowdsourcing" is derived from the words crowd and outsourcing; this is a process required people, who are not organized in any other system, to perform a specific job. The creator of the term Jeff Howe considered crowdsourcing as a new social phenomenon that is beginning to emerge in certain areas (Howe 2006), as a phenomenon of bringing people together for the solution of the problem without any reimbursement, and the consequences of such groups/associations for business, solving similar tasks professionally. The method consists in the fact that the task is offered to an unlimited number of people, regardless of their professional status or age. Participants of a crowdsourcing project form the society that chooses by discussing the most successful solution of the given problem. For businesses, this method is an inexhaustible resource for finding solutions to solve their own problems and issues, a powerful tool that allows to adjust the cost-effectively development, including the development of the most customer-oriented products.

Currently a number of social tools are ranked as a crowdsourcing; researchers from Crowdsourcingresults (Dawson 2010) proposed a comprehensive classification of modern methods of crowdsourcing (see Fig. 1), the most popular of which are:

1. Reference Content, when everyone who knows more, improves reference resource. The most popular such resource is Wikipedia;

2. Content Markets, when visitors locating and evaluating some content, and site owners are allowed to produce of its best examples;

3. Crowdfunding is a collaboration of people who pool their resources (money) to support projects initiated by other people or organizations;

4. Competition Platforms, when the customer announces, placing job online platform; actors offer their solutions and evaluate the proposals of colleagues, as the result the best work is chosen, which is usually rewarded; one of the most famous examples of such resource is the site Zooppa.com;

5. Micro-tasks, when the customer announces the use of human intelligence to perform small tasks that cannot be formalized and solve by computers (Human Intelligence Tasks). The most famous site is the platform Amazon Mechanical Turk;

6. Crowdsourcing Aggregators, when the performers take on a client project, divide it down into individual tasks that are offered in the form of micro-projects for crowdsourcing workforce, and then aggregated with ethyl results. This approach allows solving large-scale, automated by hard task. Indicative of this type of platform is the site of CrowdFlower;

7. Cycle Sharing, realizing of the idea of using computers for volunteers distributed computing.

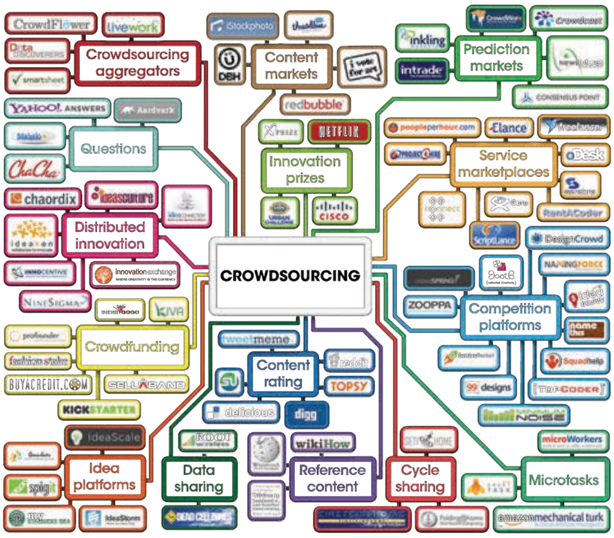


Figure 1: Methods of crowdsourcing (Dawson 2010)

3. CROWDSOURCING IN THE TASKS OF MONITORING

Remote monitoring has a long history of use for collection of environmental measurements. Many sensor networks have been deployed to monitor Earth’s environment, and more will follow in the future. Environmental sensors have improved continuously by becoming smaller, cheaper, and more intelligent. Due to the large number of sensor manufacturers and differing accompanying protocols, integrating diverse sensors into observation systems is not straightforward. A coherent and integrated infrastructure is needed to treat sensors as interoperable, platform-independent and uniform way. The concept of the Sensor Web reflects such a kind of infrastructure for sharing, finding, and accessing sensors and their data across different applications. It hides the heterogeneous sensor hardware and communication protocols from the applications built on top of it. The Sensor Web Enablement initiative started by Open Geospatial Consortium defines the term Sensor Web as “Web accessible sensor networks and archived sensor data that can be discovered and accessed using standard protocols and application programming interfaces”. Thus, the Sensor Web is to sensor resources what the WWW is to general information sources – an infrastructure allowing users to easily share their sensor sources in a well-defined way.

Environmental management and monitoring systems provide an important application for the crowdsourcing-based paradigm, particularly in the area of integrated planning and management.

More specifically, crowdsourcing can be integrated in environmental planning, management and monitoring at three different levels:

1. Setting up a social network, for a better comprehension of the underlying social system:
   - Network identification;
   - Interest characterization;
   - Stakeholder clustering and representative selection;
   - Social disambiguation of interests.
2. Putting humans in the loop, in order to exploit human potential as sensors, task solvers and decision makers:
   - Human sensing;
   - Human judgment for task solving;
   - Co-deciding.
3. Eliciting collective knowledge on the environmental systems by exploiting situated and distributed knowledge and expertise, i.e., the so-called social capital.

Thus, crowdsourcing can be applied not only to monitor the status of the selected object or area, but at the same time, increase the awareness of people about the behaviour of the monitoring object. The motivation for engaging the public in monitoring is two-fold (Stevens and D'Hondt 2010). On the one hand, the system of crowdsourcing can complement modern assessment methods to achieve a high degree of spatial-temporal granularity at lower costs. On the other hand, the active involvement of citizens in the processes of decision-making control and increases their self-awareness and sense of responsibility. Not surprising that numerous international reports (European Parliament and Council (EPC) 2002; United Nations Environment Programme (UNEP) 1992) show in the participation of all concerned citizens, at all levels for sustainable socio-economic development. For example, the introduction of smartphones as a personal instrumentation reduces barriers to achieve the democratization process monitoring.

We are starting to see the impact of emerging these technologies on information security - 14% of large organisations had a security breach relating to social networking sites and 9% had a breach relating to smartphones or tablets (PWC 2013), thus assuring security of industrial and private information assets is becoming extremely sensitive and topical issue. There is huge number of available free-ware and paid methods of information protection from unauthorized access by unwanted individuals (Dorogovs and Romanovs 2012).

4. INTEGRATION OF TRADITIONAL AND SOCIAL DATA

Mobile phones increasingly become multi-sensor devices, accumulating large volumes of data related to our daily lives. These trends obviously raise the potential of collaboratively analysing sensor and social data in mobile cloud computing (Yerva, Jeung and Aberer 2012).

In the same time, there exists a growing fleet of various robotic sensors (e.g., robotic fishes) coupled with the emergence of new and affordable monitoring technology that increases exponentially the amount of data collected from the world's geo-spheres. This puts decision-makers and researchers who work with these data in a completely fresh situation.

The two popular data types, social and sensor data, are in fact mutually compensatory in various data processing and analysis. Participatory / citizen sensing (Boulos, Resch, Crowley, Breslin, Sohn, Burtner, Pike, Jezierski and Chuang 2011; Fraternali, Castelletti, Soncini-Sessa, Ruiz and Rizzoli 2012), for instance, enables to collect people-sensed data via social network services (e.g., Twitter, Waze, Ushahidi) over the areas where physical sensors are unavailable. Simultaneously, sensor data (Figure 2) is capable of offering precise context information, leading to effective analysis of social data. Obviously, the potential of blending social and sensor data is high; nevertheless, they are typically processed separately, and the potential has not been investigated sufficiently. Therefore, there is an urgent need for fusing various types of data available from various data sources.

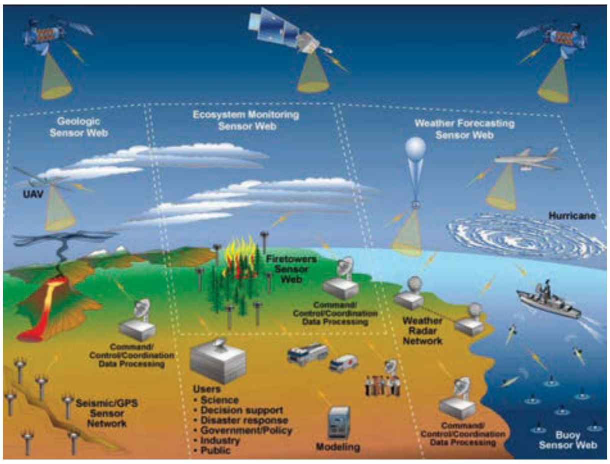

Geologic Sensor Web
Ecosystem Monitoring Sensor Web
Weather Forecasting Sensor Web
UAV
Hurricane
Firetowers Sensor Web
Command/ Control/Coordination Data Processing
Weather Radar Network
Command/ Control/Coordination Data Processing
Seismic/GPS Sensor Network
Users
• Science
• Decision support
• Disaster response
• Government/Policy
• Industry
• Public
Modeling
Command/ Control/Coordination Data Processing
Buoy Sensor Web

Figure 2: Various sensor data sources (NASA 2008) arranged in a Sensor Web (Fraternali, Castelletti, Soncini-Sessa, Ruiz and Rizzoli 2012)

Data fusion is the process of combing information from a number of different sources to provide a robust and complete description of an environment or process of interest (Durrant-Whyte and Henderson 2008). Automated data fusion processes allow essential measurements and information to be combined to provide knowledge of sufficient richness and integrity that decisions may be formulated and executed autonomously.

The existing projects and platforms for data collection and processing, e.g., GOOS (GOOS 2013), Marinexplore (Marinexplore 2012), Social.Water (Fienen and Lowry 2012), show that the bottleneck of the data market is not in collecting the data, but in the processing the data. Most available data is disconnected, often archived, and sometimes never used again (Marinexplore 2012).

5. CROWDSOURCING MODEL FOR THE INTEGRATED INTELLIGENT PLATFORM OF MONITORING THE CROSS-BORDER NATURAL-TECHNOLOGICAL OBJECTS

Existing space-ground monitoring information processing platform can be without significant cost supplemented with an application that processes the data of social sensors. At a minimum, this social application could consist of two components: a mobile application and server public knowledge base.

Mobile applications are for free downloaded and installed on smartphones, to turn them into mobile monitoring systems social sensors. Smartphones collects information from various sensors (microphone, GPS, descriptive or qualifying user-typed information), and in real time sends to the servers public knowledge base.

Public knowledge base (called also Web-based Community Memory) can be defined as a resource of information and communication technologies that enable the public to record and archive information relating to the management of common property (Steels and Tisseli 2008). Thus, it is part of the software that operates on a central Web server, collects and processes all data received from mobile social sensors, supports a website that allows users to search, analyse and visualize data.

5.1. System Architecture

The objective of the proposed crowdsourcing-supported software platform is to allow blending the heterogeneous social and sensor data for integrated analysis, extracting and modelling environment-dependent information from social and sensor data streams.

The general system architecture consists of four coupled layers (Figure 3):

1. *External data sources*. Environment monitoring is based on data gathered externally by sensors, from structured and unstructured data sources. Data and information providers include researchers, non-researchers, companies, universities, students, communities, primary / secondary schools' pupils.
2. *High-performance computing layer*. High-performance computing layer includes the grid computing cluster, GPU-based computing cluster, environmental modelling subsystem.
3. *Storage layer*. Storage layer is intended for storing and managing high volumes of raw and aggregated data.
4. *Presentation / Service layer*. The presentation / service layer of monitoring system architecture is designed as a set of extendable services. Services are flexible and configurable for various data sources (sensors, structured and unstructured data). Services can be multimodal having a capability to work in automatic, semiautomatic and manual modes.

5.2. Modelling Scenarios

The developed application has a wide range of use, mainly in the form of two scenarios. First scenario: Citizen-led initiatives. Because of the low barrier, in terms of both cost and complexity, concerned individuals can use the platform to study noise pollution in their neighbourhood. The participants can be self-organized citizens with varying levels of organizational involvement: ranging from total strangers that happen to live in the same area; over loosely organized groups of neighbours facing a shared problem; to well-organized previously existing activism groups. The motivation for such initiatives can be diverse: from curiosity about one’s daily environment to the gathering of evidence on concrete local issues. These can be long-term issues (such as the problems faced by people living close to airports, highways, factories or nightclubs); short-term ones (such as roadwork’s or nearby construction sites); or accidental annoyances (such as manifestations).

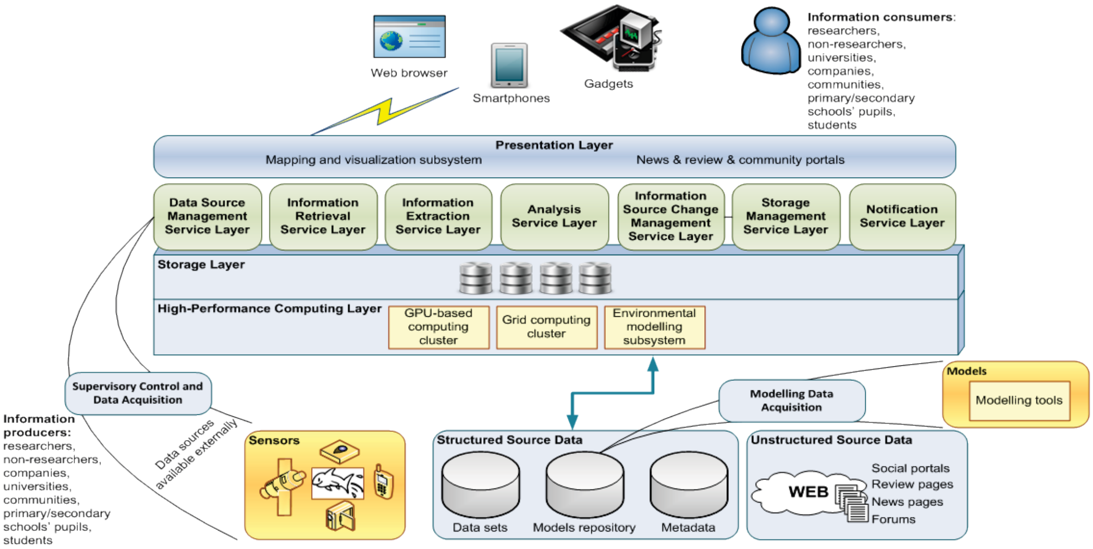


Figure 3: General crowdsourcing-supported system architecture and its components

Second scenario: Authority-led initiatives. Social application can be used by the authorities and public institutions - usually at the municipal or regional level - to collect data on the behaviour of natural and technological objects in their territory. These data can be used to support decision-making and policy-making in areas such as health and urban planning, environmental protection and mobility. When used alongside an existing monitoring system a participatory sensing platform could make up for missing data, help to estimate error margins of simulation models, add semantics (e.g. identification of pollution sources), etc.

In the most effective way social application can be used to control and rapid dissemination of information on natural disasters, major accidents, etc. The prototype of the social application, developed under the supervision of the RTU and SPIIRAS researchers J. Petuhova and S.A. Potryasaev, allow to effectively implementing both of the above described scenarios by the example of the Daugavpils City (Republic of Latvia) (Figure 4).

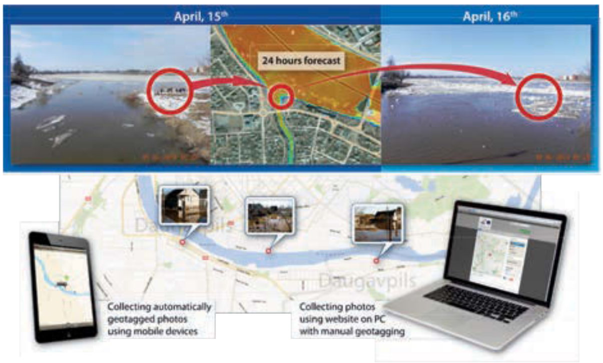

April, 15th
April, 16th
24 hours forecast
Daugavpils
Daugavpils
Collecting automatically geotagged photos using mobile devices
Collecting photos using website on PC with manual geotagging

Figure 4: Crowdsourcing for flood modelling: social application prototype

6. CONCLUSIONS

On the basis of research results presented in this paper it can be marked the importance and effectiveness of integration of remote sensing data, as well as data of other sources with social information in the context of the monitoring of natural-technological systems by focusing on the issues of changing ecosystems, geo systems, climate and providing services for sustainable economy, healthy environment and better human life.

The integration of social and information technologies allows to effectively solving the three biggest issues around managing environment-related data:

- How to access the vast amount of data that is available in different data formats, has different spatial and temporary resolution and quality, as well that reside in isolated silos, segregated and disconnected from each other;
- How to make the time-consuming handling and processing of all this data more efficient;
- How to make the available data, modelling and analysis results publicly available in an efficient and a user-friendly way to facilitate the social interest and responsibility in the environmental monitoring and research processes.

ACKNOWLEDGMENT

This research is supported by the project 2.1/ELRI-184/2011/14 «Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems» as a part of «Estonia-Latvia-Russia cross border cooperation Programme within European Neighborhood and Partnership instrument 2007-2013».

REFERENCES

Boulos, M.N.K, Resch, B., Crowley, D., Breslin, J., Sohn, G., Burtner, R., Pike, W., Jezierski, E., Chuang, K.-Y.S., 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, vol. 10, no. 1, p. 67.

Bradley, A.J., McDonald, M.P., 2011. The social organization: how to use social media to tap the collective genius of your customers and employees // GARTNER,INC. – Boston: Harvard business review press.

Dawson, R., 2010. *Crowdsourcing Landscape – Discussion. Crowdsourcingresults*. Available from: http://crowdsourcingresults.com/competition-platforms/crowdsourcing-landscape-discussion [Accessed 14 May 2013].

Dorogovs, P., Romanovs, A., 2012. Modelling and evaluation of IDS capabilities for prevention of possible information security breaches in a Web-based application. *Proceedings of the 14th International Conference on Harbor Maritime and Multimodal Logistics M&S, HMS 2012*, pp.165–170. September 19-21, 2012, Vienna, Austria.

Durrant-Whyte, H., Henderson. T.C., 2008. Multisensor Data Fusion. In: Sicilliano B., Oussama K., eds. *Springer Hand-book of Robotics*. Springer, 2008, p. 1611.

European Parliament and Council (EPC), 2002. Directive 2002/49/EC relating to the Assessment and Management of Environmental Noise. *Official Journal of the European Communities*, 18.7.2002, pp. 12–26.

Fienen, M.N., Lowry, C.S., 2012. Social.Water—A crowdsourcing tool for environmental data acquisition. *Computers & Geosciences*, vol. 49, pp. 164–169.

Fraternali, P., Castelletti, A., Soncini-Sessa, R., Ruiz, C.V., Rizzoli, A.E., 2012. Putting humans in the loop: Social computing for Water Resources

Management. *Environmental Modelling & Software*, vol. 37, pp. 68–77.

GOOS, 2013. *The Global Ocean Observing System*. Available from: http://www.ioc-goos.org [Accessed 5 July 2013].

Howe, J., 2006. The Rise of Crowdsourcing. *Wired*, Iss. 14.06, June 2006.

Marinexplore, 2012. Marinexplore: Cutting Ocean Data Processing Time Fivefold. *Marine Technology Reporter*, no. 11/12, pp. 30–35.

Merkuryev, Y., Sokolov, B., Merkuryeva, G., 2012. Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems. *Proceedings of the 14th International Conference on Harbor Maritime and Multimodal Logistics M&S, HMS 2012*. pp. 7–10. September 19-21, 2012, Vienna, Austria.

NASA, 2008 Report from the Earth Science Technology Office (ESTO) Advanced Information Systems Technology (AIST) Sensor Web Technology Meeting.

PWC, 2013 Information security breaches survey. *PWC, InfoSecurity Europe, UK Department for Business Innovation and Skills.*

Stevens, M., D'Hondt, E., 2010. Crowdsourcing of Pollution Data using Smartphones. *UbiComp '10*, 2010, Copenhagen, Denmark

Steels, L., Tisseli, E., 2008. Social Tagging in Community Memories. *Social Information Processing – Papers from the 2008 AAAI Spring Symposium* (March 26-28, 2008; Stan-ford University), pp. 98–103, Menlo Park, California, USA, March 2008. AAAI Press.

United Nations Environment Programme (UNEP), 1992. Rio Declaration on Environment and Development. Proclaimed at the United Nations Conference on Environment and Development, June 1992.

Yerva, S.R., Jeung H., Aberer, K., 2012. Cloud based Social and Sensor Data Fusion. *15th International Conference on Information Fusion*, 2012.

AUTHORS BIOGRAPHY

Andrejs Romanovs, Dr.sc.ing., assoc. professor and senior researcher at Information Technology Institute, Riga Technical University. He has 25 years professional experience teaching post-graduate courses at the RTU and developing as IT project manager and system analyst more than 50 industrial and management information systems in Latvia and abroad for state institutions and private business. His professional interests include modelling and design of management information systems, IT governance, IT security and risk management, information systems for health care, integrated information technologies in business of logistics and e-commerce; as well as education in these areas. A. Romanovs is senior member of the Institute of Electrical and Electronic Engineers (IEEE), Latvian Simulation Society (LSS), and member of the Council of RTU Information Technology Institute; author of 2 textbooks and more than 40 papers in scientific journals and conference proceedings in the field of Information Technology, participated in 25 international scientific conferences, as well as in 7 national and European-level scientific technical projects.

Boris V. Sokolov is a deputy director at the Russian Academy of Science, Saint Petersburg Institute of Informatics and Automation. Professor Sokolov is the author of a new scientific field: optimal control theory for structure dynamics of complex systems. Research interests: basic and applied research in mathematical modelling and mathematical methods in scientific research, optimal control theory, mathematical models and methods of support and decision making in complex organization-technical systems under uncertainties and multiple criteria. He is the author and co-author of five books on systems and control theory and of more than 270 scientific papers. Professor B.Sokolov supervised more than 50 research and engineering projects. Homepage: www.spiiras-grom.ru.

Arnis Lektauers, Dr.sc.ing., is assistant professor at the Department of Modelling and Simulation of Riga Technical University (RTU). His main professional interests include the development of interactive hybrid modelling and simulation algorithms with an application to complex systems analysis and the research of industrial, economic, ecological and sustainable development problems. A. Lektauers is the Secretary of Latvia section of the Institute of Electrical and Electronics Engineers (IEEE), a member of the Council of RTU Faculty of Computer Science and Information Technology, and a member of Latvian Simulation Society, System Dynamics Society and European Social Simulation Association (ESSA); author of 1 textbook and more than 30 papers in scientific journals and conference proceedings in the field of Information Technology.

Julija Petuhova, Dr.sc.ing., is lecturer at the Institute of Information Technology at Riga Technical University. Her research interests include simulation methodology of logistics systems, supply chain dynamics, practical applications of simulation modelling, training and education via simulation-based business games. She is a member of the Latvian Simulation Society and has a wide experience in performing research projects in the simulation area at national level. She is a member of Latvian Simulation Society (LSS) since 2002 and a member of Technical Program Committee (TPC) of the International Conference on „Computer as a tool (EUROCON)" (2007). She authors 20 scientific publications, including 1 book and 1 international journal.

MODELING A WRONG MAINTENANCE POLICY

Diego D'Urso

Università degli Studi di Catania - Dipartimento di Ingegneria industriale
ddurso@diim.unict.it

ABSTRACT

The behavior of a single unit system, which is maintained according to the preventive policy, despite it fails with a constant failure rate, is focused.

A discrete event simulation (DES) enables to compare the expected number of failures that belong to alternative maintenance scenarios: the corrective and the preventive one.

The results comparison shows, at a first glance, that the preventive maintenance has to be preferred.

A deeper analysis, based on the information content registered at each step of a simulation process, on the Skellam function properties and on the Small Number Law, helps to clarify this strange behavior.

The real aim of modeling such a system would be to refuse the constant failure rate as an operations statement but only as a missing-information maintenance state. The simulation model enables to find more comprehensive information content about the behavior of stochastic single unit maintenance.

Keywords: Single item maintenance, discrete event simulation, Skellam function, small number Law, learning by simulation.

1. INTRODUCTION

Maintenance has ever been a critical issue for management of industrial processes. Since early sixties, for military purposes, maintenance literature was already extensive and rapidly growing. Barlow and Hunter (1960) and Barlow and Proschan (1967) started to evaluate the state probabilities of a complex system. Few years later, Jorgenson et al. (1966) edited a comprehensive report about optimal maintenance policies for stochastically failing equipment: the corrective, preventive and preparedness maintenance policies were defined as well the uselessness of use the preventive maintenance to items which fail according to constant failure rate if compared with the application of the corrective one. They also unified the principle optimal preventive models both for preventive and preparedness maintenance.

Henceforth preventive maintenance policy gained growing attention: economic models of optimization were performed over a horizon of thirty years and as a consequence new comprehensive focusing review about specific maintenance techniques have tried to make a picture of the state of the art (Sheut and Krajewski, 1994, Dekker 1996).

Due to the constant improvement of technology, processes have become more complex while service levels and most of all higher reliability performances are required. As a consequence the cost of preventive maintenance became the most important for industrial companies (Jardine et al. 2006).

Therefore, Researchers focused analysis on more efficient maintenance approaches such as condition-based maintenance (CBM) which are being implemented to handle the situation.

The state of art confirms the early structure of maintenance policies and their suggested applications: preventive policy is to be dedicated to monotone failure rate items; corrective maintenance is the only answer to random failure occurrence.

Relatively little has been written about the limits of corrective maintenance.

Inspired by this literature background, I was trying to compare the results of maintenance policies applied to a system during a simulation exercise designed for students in logistics course of master degree.

Because of the educational nature of the experiment, the application of above mentioned maintenance policies was applied to a simple case study such as that is represented by a single item system; just to make clear the contribution of proper maintenance policy to economic outcomes, it was simulated the application of corrective and preventive maintenance although the single item fails according to a constant failure rate.

The experiment was designed, however, so as to change the failure mode of the component so as to determine which strategy is better suited to the conditions change scenario.

The study below reported describes how the simulation modeling of a system can led to gain a better understanding of the focused problem and to overcome critical issues and maths tricks.

The study subtends the importance of *learning by simulating.*

The paper is structured as follows: a case study is described in order to define the problem; it regards the modelling of a single unit system; the model belongs to the discrete event simulation type: it is at the same time the creator and the solver of an apparent paradox. A brief discussion of the features of the model enables to solve the counter current behaviour and shows how is deep the information content of a simulation model.

2. THE CASE STUDY

25.01.2013 – Department of Industrial Engineering, University of Catania – During a maintenance lecture –

The topic of items reliability has just been introduced as the probability that an item could work without failing until a fixed time of mission is reached and under working conditions which are similar to that are declared by the supplier.

The general reliability equation, dedicated to constant failure rate items, was presented:

$R(t)=e^{-\lambda Tm}$

where λ is the constant failure rate and *Tm* is the time of mission.

In order to discuss the principle maintenance policies, the memory-less property was defined as a special feature of constant failure rate items.

The elegant application of the Bayesian theorem was introduced to consider the constant failure rate as the joint probability that the single item fails during the next elementary time interval, *dt*, if a failure didn't yet occur.

Given a number of items, $N_{0,}$ which are subjected to a reliability test, at each control step the following equations can be written:

$N_0 = N_F(t)+N_S(t)$

$R(t) = N_S(t)/N_0$

$\lambda = dN_F(t)/dt\ l/N_S(t) = cost$

$\lambda\ dt = dN_F(t)/(dt\ N_0)\ N_0/N_S(t) = f(t)\ dt/R(t) = cost$

where $N_F(t)$ and $N_S(t)$ are respectively the number of failed and save units at each time *t*.

The memory-less property enables to understand how is fallacy to replace a constant failure rate item before it fails because, after the replacement, the failure rate doesn't change as the above mentioned joint probability suggests; so the corrective maintenance policy is the only one model which is to be taken in to account (Jorgenson et al. 1966).

A brief mathematics procedure was performed in order to compare the corrective and preventive maintenance results when they are applied to a constant failure rate single unit system.

So a single unit system was considered in order to define the conceptual maintenance model; it fails according to a constant failure rate, λ.

The corrective maintenance horizon can be evaluated by the following equation:

$H = NF_{CM}\ MTTF$
(1)

where $MTTF=\lambda^{-1}$ is the mean time to failure and NF_{CM} is the expected number of failures.

On the other hand, the preventive maintenance horizon leads to replace the single unit if it doesn't fail before the end of a fixed preventive maintenance period; the preventive maintenance horizon can be written as follows:

$H = NF_{PM}MTTF'+ N_{PM}T_{PM} = NI_{PM}\ MTTI$ (2)

where:

NF_{PM} is the number of failures which anyway are registered during the preventive scenario;

MTTF' is the mean time to failure of items which fail before the end of the preventive maintenance period T_{PM};

N_{PM} is the number of preventive maintenance replacements;

NI_{PM} is the expected number of interventions whether they belong to preventive or corrective type;

MTTI is the mean time to intervention along the preventive maintenance horizon.

The preventive maintenance features, as the mean time to failure, MTTF', the expected numbers of failures, NF_{PM}, and the expected number of preventive interventions, N_{PM}, can be evaluated by using the following further equations:

$MTTF'=\int_0^{T_{PM}} tf(t)\ dt$

$MTTF'= MTTF-T_{PM}R(T_{PM})/(1-R(T_{PM}))$ (3)

$NF_{PM} = NI_{PM}(1- R(T_{PM}))$ (4)

$N_{PM} = NI_{PM}\ R(T_{PM})$ (5)

The substitution of equations (3), (4) and (5) in the equation (2) allows finding the equivalence between the above mentioned NF_{CM} and NF_{PM} numbers of failures:

$H = NF_{CM}\ MTTF =$

$= NI_{PM}\ (1- R(T_{PM}))\ [MTTF-T_{PM}R(T_{PM})/(1-R(T_{PM}))]+$

$+NI_{PM}\ R(T_{PM})\ T_{PM}\ = NI_{PM}\ [(1- R(T_{PM}))\ MTTF+$

$-T_{PM}\ R(T_{PM})+ R(T_{PM})\ T_{PM}] = NI_{PM}\ (1- R(T_{PM}))\ MTTF =$

$= NF_{PM}\ MTTF.$

To carve the latter equivalence on the stone, a little exercise was designed and showed to the students.

Figure 1 shows the scheme of corrective maintenance model which was presented to the students and its space state representation; the single unit can assume only a working or a failure state; the item fails with a constant failure rate $\lambda=0{,}01\ h^{-1}$, the defined maintenance horizon is H= 400 h.

The time to replace the failed item is considered deterministic and null ($\mu=\infty$).

The single item system was then modeled by sampling a sequence of items which simulate the system according to the Monte Carlo method; let i denote the i-th working item and T_i the time to failure which is randomly sampled by using the inverse function of the cumulated exponential probability density:

$T_i = -\ln(1-R_i)/\lambda$

where R_i is the i-th random number belonging to]0..1[range.

The model was coded by using an Excel Microsoft® spreadsheet during the lesson; the well known WYSIWYG property enables to show each step of the coding process and to display it on the dashboard: every student can see what is modeled and how.

So the corrective maintenance policy is modeled, by iteration, evaluating the series of sampled times to failure, T_i, until $\Sigma^k_{i=1}T_i<H$; when $\Sigma^k_{i=1}T_i>H$ the corrective maintenance scenario is fully simulated and the number of failures can be counted as $NF_{CM} = k-1$.

Table 1 shows the spreadsheet model which was coded.

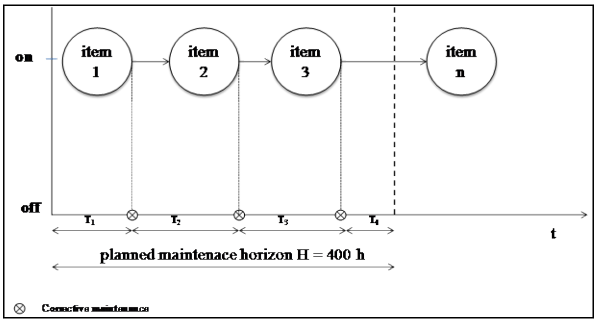


Figure 1: Single item corrective maintenance scheme

Table 1: Model of the corrective maintenance policy

Pos	Ri	T_i=-ln(1-R_i)/λ	$\Sigma^k_{i=1}T_i$	NF_{CM}
	[-]	[h]	[h]	[-]
1	0,93368139	271,33	271,33	1
2	0,49454915	68,23	339,56	2
3	0,54793473	79,39	418,95	3
4	0,46229076	62,04	418,95	3
5	0,00106236	0,11	418,95	3
6	0,14932840	16,17	418,95	3
7	0,23915160	27,33	418,95	3
8	0,95918196	319,86	418,95	3

The preventive maintenance scenario was simulated according to the age-dependent policy; the single unit is replaced at its age *t* or failure, whichever occurs first, where $t=T_{PM}$ is the preventive maintenance period.

Figure 2 shows the scheme of the preventive maintenance model which was presented to the classroom and its space state representation.

Each simulated unit operates as follows:

$$T_i' = T_i \text{ if } T_i < T_{PM}\text{, otherwise } T_i' = T_P \qquad (6)$$

T_{PM} was also set equal to the item mean time to failure, MTTF = 1/λ.

The time to the preventive replacement is yet consider deterministic and null.

The preventive maintenance policy is modeled, by iteration, evaluating a series of random sampled time to failure items, T_i', which respects the equations set (6), until $\Sigma^k_{i=1}T_i'<H$; when $\Sigma^k_{i=1}T_i'>H$ the preventive maintenance scenario is fully simulated and the number of failures, NF_{PM}, can be counted among items for which both the following equation are respected:

$T_j'<T_{PM}$ and $\Sigma^j_{i=1}T_i'<H$.

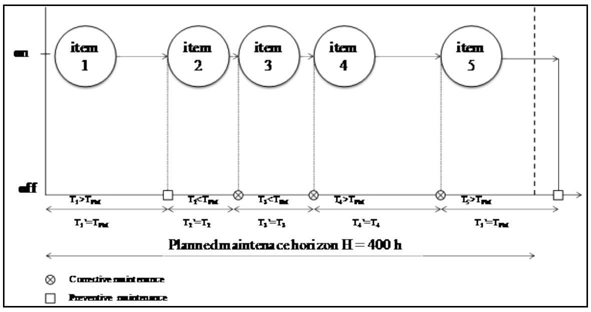


Figure 2: Single item preventive maintenance scheme

Table 2 shows the spreadsheet model which was coded.

Table 2: Preventive maintenance policy model

Pos	Ri	Ti	Ti'	$\Sigma^k_{i=1}T_i'$	NF_{PM}
		[h]	[h]	[h]	
1	0,93368139	271,33	100,00	100,00	0
2	0,49454915	68,23	68,23	168,23	1
3	0,54793473	79,39	79,39	247,62	2
4	0,46229076	62,04	62,04	309,67	3
5	0,00106236	0,11	0,11	309,77	4
6	0,14932840	16,17	16,17	325,95	5
7	0,23915160	27,33	27,33	353,28	6
8	0,95918196	319,86	100,00	453,28	6

To make more interesting the experiment it was evaluated the scenario in which the corrective cost was much more expensive than the preventive one; as a consequence, the maintenance cost function, both for corrective and preventive scenarios, depends only on the number of failures.

So the comparison of the two maintenance policies is based on the number of failures NF_{CM} and NF_{PM} in terms of average value and standard deviation.

The difference between NF_{CM} and NF_{PM} was also registered, for each simulation step, in order to calculate the relative distribution of frequency.

After few hundreds of replications the following results were discovered:

$f(NF_{PM} < NF_{CM}) \approx 40\%$;

$f(NF_{PM} = NF_{CM}) \approx 30\%$;

$f(NF_{PM} = NF_{CM}) \approx 30\%$.

These results didn't agree with the theoretical managerial implications.

It was late so students were asked to bright their doubts again the next time.

3. RESULTS

The model of the single unit was checked and a simulation process of 10^6 iterations was performed.

In order to verify the simulation process and the data input integrity, a comparison between theoretical and simulated results was calculated as regards to the following variables: mean value and standard deviation of random numbers which were generated in order to perform the Monte Carlo simulation process; mean value and standard deviation of times to failure (T_i, T_i').

A good agreement among theoretical and simulated results was found.

Figure 3, 4 and 5 depict further results which were obtained.

The simulation process enabled to verify the substantial identity between the distribution of number of failures f(NF_{PM}) and f(NF_{CM}*)* and how they fit very well the Poisson distribution (see figure 3); this behavior agrees to the Law of small number by which the number of rare event, along a fixed horizon, follows the Poisson distribution of the average expected number of events (in the focused case study the expected number of failures is H/MTTF=4,0) (Crathorne 1928).

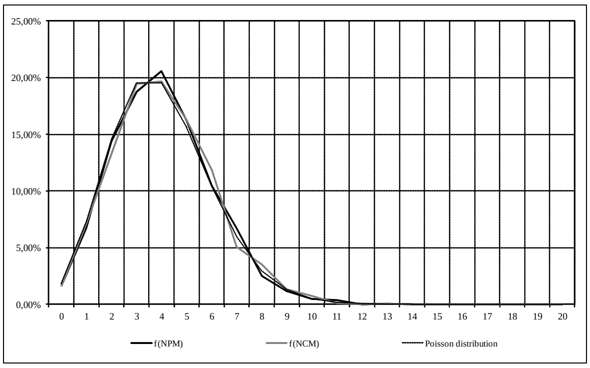


Figure 3: Failures distribution per maintenance policy

On the contrary, the difference between NF_{PM} and NF_{CM} follows an asymmetric distribution; furthermore the comparison between number of failures which are registered step by step of simulation process shows that the preventive maintenance scenarios has a more frequently number of failures which is lower than the one is registered by simulating the corrective one (see figures 4 and 5):

$F(NF_{PM}-NF_{CM} <=0) = 70\%$;

$F(NF_{PM}-NF_{CM} >0) = 30\%$.

Figure 4 reveals also a pseudo-Skellam behaviour; the Skellam distribution is the discrete probability distribution of the difference NF_{PM} - NF_{CM} of two statistically independent random variables NF_{PM} and NF_{CM} each having Poisson distributions with the same expected values (Skellam 1946).

The distribution is also applicable to a special case of the difference of dependent Poisson random variables, when the two variables have a common additive random contribution which is cancelled by the differencing (Karlis and Ntzoufras 2006).

The simulation process results are counter revolutionary: because they don't agree with the consolidate literature knowledge (the preventive maintenance appears to be preferred with respect to the corrective one) and because the difference between two Poisson distribution doesn't follow a Skellam function. This issue requires a deeper discussion in order to be solved. The simulation process can register data which can solve the rising issue.

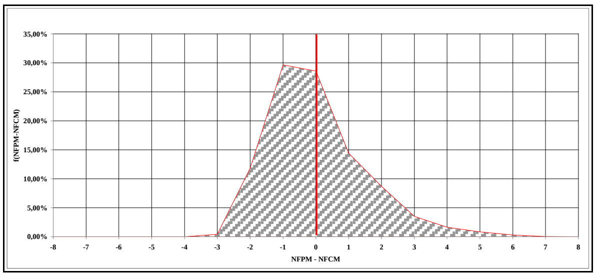


Figure 4: NF_{PM} - NF_{CM} simulated frequency distribution

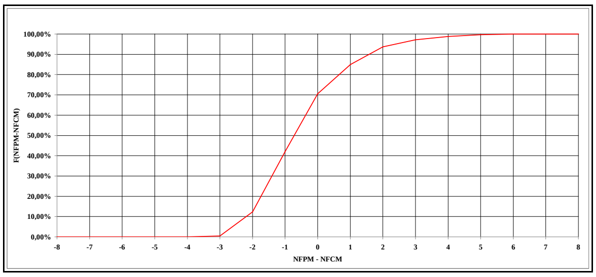


Figure 5: NF_{PM} - NF_{CM} cumulative frequency distribution

4. DISCUSSION

Figure 3 shows that the distribution frequency of failures, for both policies, seems to follow a Poisson distribution; the shape of this function is asymmetric: the probability of a number of failures which is higher than the expected one is F(NF>H/MTTF)=0,3712.

The Law of small number is respected and at the same time one can argue that the overall probability of T_i > MTTF is higher than the vice versa.

On the other hands, a high number of failures can happen with a lower overall probability, but they happen.

As regard to the difference between NF_{PM} and NF_{CM} it is to be noted that the time to failure which are random sampled for the above mentioned maintenance scenarios, Ti and Ti', are not independent:

$T_i' = T_i$ if $T_i < T_{PM}$, otherwise $T_i' = T_{PM}$

Tables 1 and 2 allow showing the dependency of the two set of variables.

A new simulation process was performed and the two maintenance models were provided with two different set of random numbers.

Figure 6 shows that when T_i and T_i' are independent variables, because they are sampled from different set of random numbers, the difference NF_{CM} - NF_{PM} follows a Skellam function.

This behaviour can be assumed as a validation of the simulation model which confirms the theoretical results when the theoretical condition of independency of input variables is established.

Although we found the reason of the pseudo-Skellam behaviour, the sequence by which items are procured and replaced in the single unit system is unique and the dependency between T_i and T_i' can not be overtaken; we can only register that a Skellam distribution doesn't occur.

In order to find the solution of the problem from a more holistic point of view was appointed.

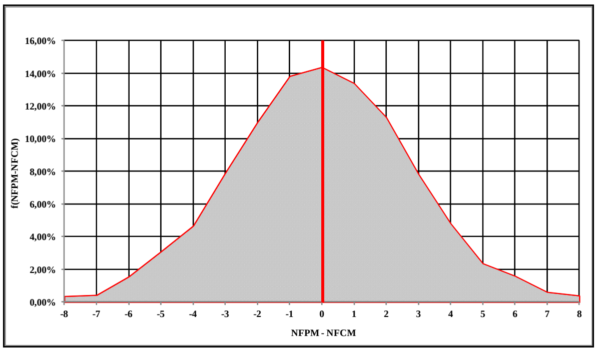


Figure 6: the Skellam NF_{CM}-NF_{PM} distribution.

Maintenance management should be defined as a risk management task: a maintenance policy is to be selected when it minimizes the risk of failures and not only the frequency of failures which belongs to a certain maintenance policy.

Let's define the risk of a maintenance policy, RoM, as:

$$RoM=f \cdot I$$

where f is the frequency according to which the scenario happens and I is the impact that can be calculated by counting the number of failures.

A final simulation process was performed; this time the model enables to register, at each step of Monte Carlo simulation, the number of failures for each kind of maintenance policy.

As regard to the preventive policy scenario, it's now possible to compute the average number of failure $NF_{PM}(NF_{PM}<NF_{CM})$ exclusively for those step of Monte Carlo simulation to which the number of failure NF_{PM} is lower than the number of failure NF_{CM}; the same calculation is performed for the opposite condition.

The comparison of the risks of maintenance policy was evaluated as follows:

$$RoM_{PM} = E(NF_{PM}(NF_{PM} < NF_{CM}))\ f(NF_{PM} < NF_{CM})$$
$$RoM_{PM}' = E(NF_{PM}(NF_{PM} > NF_{CM}))\ f(NF_{PM} > NF_{CM})$$

Figure 7 shows the evolution of the simulated risk of maintenance, RoM_{PM} and RoM_{PM}', and allows finding again that preventive and corrective maintenance policies have the same risk of maintenance when items fail according to a constant failure rate.

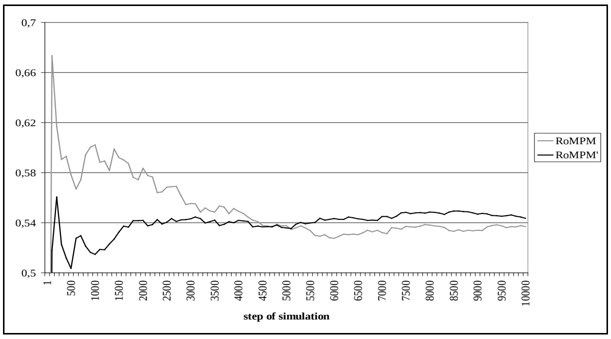


Figure7: progressive risk of maintenance RoM and RoM' (10^4 replications)

The simulation process allows discovering stochastic behaviours sometimes hidden in to the system's dynamics.

Further information could be pointed out from another high point of view: the problem was only failures dependent due the particularly relation between preventive replacement costs and failure costs; when preventive maintenance costs can be neglected if compare with the failures one, item redundancy must be considered.

When I came back to the student I was able with the same case study and the same model to make the previous doubts a new little knowledge.

CONCLUSIONS

The simulation model of maintenance policies, which was applied to a conceptual case study, allowed learning a comprehensive lesson about the behavior of the entire system.

The simulation model enabled to change point of view focusing before on the details of modeling and after on the general meaning of the process: the strident initial inconsistency of results, which appears considering from a too close point of view the problem, is overtaken trough a more general approach. This order of event seems to better lead students to learn the lesson; we would call it learning by simulating.

The nature of model and the software environment which enables to see what is get during each step of coding (Microsoft Excel®) help to capture attention from students and increases the learning empathy.

The original aim of the exercise was to point out the equivalence between the corrective and preventive maintenance of constant failure rate items.

The attempt to model and simulate the reliability system allowed pointing out some further information:

1. corrective and preventive maintenance follow equivalent risk of failure when those policies are applied to constant failure items;
2. borderline conditions, as in the case study presented which shows great failure costs, need a system assessment and not only a maintenance policy decision making;
3. the memory less property of items is rather a state of information missing than an antecedent of an elegant reliability calculation;
4. differencing two dependent Poisson distributed variables led to an asymmetric Skellam function with expected value equal to de difference of the expected values of the dependent functions.

A further analysis is requested in order to estimate which kind of failure distributions, for example the constant probability distribution, meet the small number law as the exponential one.

ACKNOWLEDGMENTS

A special thank to Professor Vittorio Romano and Professor Dimitris Karlis who set author on the right way.

REFERENCES

Barlow R.E. and Hunter, L.C.; 1960, Optimum preventive Maintenance Policies, Operation research 8, 90-100.

Barlow, R. E., and Proschan, F., *Mathematical Theory of Reliability*, Wiley, New York, New York, 1967.

Jorgenson, Dale Weldeau, John McCall and Roy Radner. Optimal Maintenance of Stochastically Failing Equipment. Santa Monica, CA: RAND Corporation, 1966.

R. Crathorne, (1928), The Law of Small Numbers, *The American Mathematical Monthly* Vol. 35, No. 4 (Apr., 1928), pp. 169-175.

Karlis D. and Ntzoufras I. (2006). Bayesian analysis of the differences of count data. Statistics in Medicine, 25, 1885–1905.

Skellam, J. G. (1946); The frequency distribution of the difference between two Poisson variates belonging to different populations. Journal of the Royal Statistical Society, Series A, 109 (3), 296.

Andrew K.S. Jardine, Daming Lin, Dragan Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing, Volume 20, Issue 7, October 2006, Pages 1483-1510

C. Sheut & L. J. Krajewski (1994): A decision model for corrective maintenance management, International Journal of Production Research, 32:6, 1365-1382

Dekker, R.; 1996; Applications of maintenance optimization models: a review and analysis; Reliability Engineering and System Safety 51 (1996) 229-240

AUTHORS BIOGRAPHY

Diego D’Urso is a mechanical engineer and PhD in mechanics of structures.

He is currently assistant professor at Department of Industrial Engineering of Catania University.

Principle topics of interest are: supply chain management and human behavior, industrial plant design, warehouse operations management, maintenance.

A STOCHASTIC APPROACH FOR SUPPLY SYSTEMS

Matteo Gaeta[(a)], Luigi Rarità[(b)]

Centre of Research for Pure and Applied Mathematics,
c/o Department of Information Engineering, Electric Engineering and Applied Mathematics,
University of Salerno, Via Giovanni Paolo II, 132, 84084, Fisciano (SA), Italy

[(a)]mgaeta@unisa.it, [(b)]lrarita@unisa.it

ABSTRACT

This paper focuses on a possible approach for supply systems modeled using queueing networks. According to Poisson processes, unfinished goods and control impulses arrive at the working stations, namely the nodes of the network. When the working process in a node ends, a good moves to another node with fixed probabilities either as a part to process or as a control impulse, or leaves the network. Each control impulse is activated during a random exponentially distributed time. According to some probabilities, activated impulses move an unfinished good from the node they arrive to another node, or destroy another unfinished part. For such a queueing network, a product form solution is found for the stationary state probabilities. The stability of the network, the stationary probabilities and the mean number of unfinished parts are studied via an algorithm. Such results are also useful to analyze a real system for assembling car parts.

Keywords: production systems, queueing networks, product form solution, simulation

1. INTRODUCTION

Scientific communities have always shown a great interest in modeling dynamics of industrial realities managed by supply networks and/or systems. This exigence has become deeper and deeper especially in last years, due to the growing necessity of having fast and safe processes, which could reduce, in some way, unwished phenomena, namely dead times, bottlenecks, and so on.

A great amount of mathematical approaches have been considered for this aim. Some models are continuous, mainly based on differential equations. Examples are Cutolo et al. 2011, Göttlich et al. 2006 and Pasquino et al. 2012 where, for a generic supply chain, parts dynamics is described by conservation laws, while queues, that are in front of each suppliers, are defined by ordinary differential equations. Beside continuous models, there are other ones, dealing with individual parts: some of them are based on exponential queueing networks. In this direction, a classical theoretical example is given by Jackson for waiting lines in Jackson 1957. Possible applications of queueing networks and systems are also in Yao et al. 1986, where stochastic equations are proposed for modeling supply systems, that are also analyzed in detail in Askin et al. 1993. In order to enrich the stochastic characterization of a great variety of systems, other possible variants of queueing networks have been studied. An example is given by the so called "G – networks" (see Gelenbe 1991, Gelenbe 1993), characterized by the simultaneous presence of positive customers, negative customers, signals and triggers. Positive customers are the usual ones, who join a queue in order to receive a service, and they can be destroyed by a negative customer arriving at the queue. The role of a trigger is to displace a positive customer from a queue to another one, while a signal can behave either as a negative customer or as a trigger. A vast review of G – networks is done in Artalejo 2000, Bocharov et al. 2004 and Bocharov et al. 2003, where exact solutions for queueing networks are found in "product form", which is very important as it permits the decomposition of the joint probabilities of the states of the model into products of marginal probabilities.

In this paper, considering some descriptions of G – networks in Bocharov 2002 and Gelenbe et al. 1999, we focus on a queueing network, that models a supply system, characterized either by parts dynamics or control impulses in the working stations. Unfinished parts and control signals, these last ones generated by a Central Elaboration Unit (CEU), arrive from outside the network at each node according to two independent Poisson processes. Goods are processed one by one (one server) at each node, and service times of the unfinished parts are exponentially distributed. After the working process, a good goes from a node to another one with fixed probabilities either as a part to process or a control impulse, or leaves the network. The activation time of a control impulse is exponentially distributed. Activated impulses with fixed probabilities either move a good from the node they are activated to another one or destroy an unfinished part.

For the just described queueing network, the stationary state distribution is computed in product form, and numerical results are then obtained. From simulations, we notice that the control impulses deeply influence the stationary probabilities and the mean number of parts in the network. In particular, the instability of the network easily occurs in case of low control impulses rates, although there is a high flow of unfinished goods arriving at each node.

The stability results for the queueing networks under consideration are also applied to describe a real system for assembling car parts. For such an industrial system, two separate material flows are considered: a primary flow, consisting of car skeletons, and a secondary flow for the little parts of cars, namely mirrors, glasses, wheel rims, and so on. The performances are studied via a cost functional J, that weights either the number of parts N_p, that are processed inside the system, or the amount of control signals inside nodes, N_s. A numerical analysis of J shows that it is possible to maximize N_p and to minimize N_s at the same time, with consequent advantages in terms of quality of industrial processes.

The outline of the paper is the following. Section 2 deals with the description of supply systems and its mathematical modeling. Section 3 presents the set of Chapman Kolmogorov equilibrium equations for the model. A product form solution is obtained for the steady state probabilities in Section 4. Section 5 reports some numerical results, concerning the stationary probabilities and the mean number of parts for a simple supply system and, finally, a numerical analysis of an industrial process for assembling car parts. The paper ends with Conclusions in Section 6.

2. A STOCHASTIC MODEL FOR SUPPLY SYSTEMS

We consider a supply system, which is modeled by a queueing network with the following characteristics:

- each node of the network is a working station, at which raw material flows arrive. Such flows can be either of external type, e.g. they come from outside the network, or of internal one, namely flows arrive from some inner nodes of the network;
- each node has its own working frequency, processes materials one by one, and has an infinite buffer for its own material queues;
- there exists a Central Elaboration Unit (CEU), whose aim is to give each node some electrical impulses, useful to guide dynamics in each working station;
- beside the electrical signals given by the CEU, each node has a set of non – active control impulses, that are activated if necessary. Such signals also have their own frequency action;
- if a node of the network is empty, namely there are not goods to process, the activation of a control impulse has no effect; the impulse is disabled and is not activated anymore;
- an unfinished part, once it has been processed in a given node i, either leaves the network or moves to another node j. Inside node j, the good can be further manufactured, or can behave like a control impulse. In this last case the unfinished part can destroy a good, which is inside node j, or move the good itself to another node k.

From a mathematical point of view, we deal with a queueing network with N nodes (working stations), having an infinite buffer. External arrival flows to the network are independent Poisson processes. We indicate, respectively, with a_{0i}^p and a_{0i}^c the arrival rates of external unfinished parts and electrical control signals, generated by the CEU, at node i, $i = 1,...,N$. Goods are processed one by one (one server) inside node i and the working process of a part is completed with probability $s_i^p \Delta + o(\Delta)$ in a time interval $]t, t+\Delta[$. An unfinished part, that leaves node i, moves to node j, $j = 1,...,N$, with: probability α_{ij}^p as a good that has to be processed at node j; probability α_{ij}^c as a control impulse for node j. Finally, the unfinished part leaves the network with probability $\alpha_{i0} = 1 - \sum_{j=1}^{N} \left(\alpha_{ij}^p + \alpha_{ij}^c \right)$. Indicate by $\mathbf{A^p}$ and $\mathbf{A^c}$, respectively, the matrices with elements α_{ij}^p and α_{ij}^c. The matrix $\mathbf{A} = \mathbf{A^p} + \mathbf{A^c}$, with elements $\alpha_{ij} = \alpha_{ij}^p + \alpha_{ij}^c$, is the transition matrix of a Markov chain for the dynamics of goods.

A control impulse is activated during a random time. An impulse, which is sent to node i, works in a time interval $]t, t+\Delta[$ with probability $s_i^c(c)\Delta + o(\Delta)$, provided that c non – activated control signals are present inside node i at the time instant t. When the activation period ends, a control impulse: with probability β_{ij}^p lets a good, that is inside node i, move to node j to continue the working process; with probability β_{ij}^c moves to node j an unfinished good, which belongs to node i, and the moved part behaves as a control impulse in node j. Moreover, we indicate by $\beta_{i0} = 1 - \sum_{j=1}^{N} \left(\beta_{ij}^p + \beta_{ij}^c \right)$ the probability that a control impulse destroys an unfinished good in node i. When this happens, the control impulse ends its own action and is not activated inside node i anymore. Define now the matrices $\mathbf{B^p} := \left(\beta_{ij}^p \right)$ and $\mathbf{B^c} := \left(\beta_{ij}^c \right)$. Then, the matrix $\mathbf{B} = \mathbf{B^p} + \mathbf{B^c}$, whose parameters are $\beta_{ij} = \beta_{ij}^p + \beta_{ij}^c$, is the transition matrix of a Markov chain, that describes all possible situations concerning control impulses.

The just described queueing network is identified by the couple $(\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ and $\mathcal{A}$ indicate, respectively, the set of nodes and arcs. We have that:

$\mathcal{N}=\{0,1,2,...,N\}$, where node 0 represents the external of the network, while node i, $i=1,...,N$, is a generic working station, which belongs to the queueing network; $\mathcal{A}=\bigcup_{i\in\mathcal{N},j\in\mathcal{N}}\{e_{ij}\}$, where e_{ij} is the arc that connects nodes i and j, from i to j. We further assume that arc e_{ij} exists if $\alpha_{ij}+\beta_{ij}>0$, namely if some dynamics of the network involves nodes i and j. A possible graph for the queueing network is represented in Figure 1.

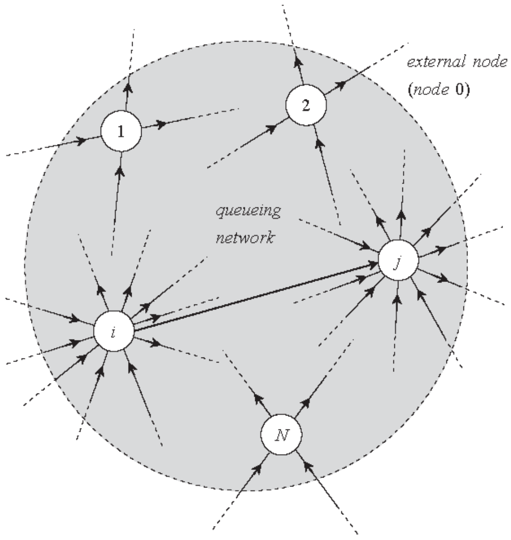


Figure 1: Possible topology for the considered queueing network

3. EQUILIBRIUM EQUATIONS

The queueing network, that models the supply system described in Section 2, is represented by a homogeneous Markov process $\{X(t),t\geq 0\}$, whose state space is $\chi=\{((p_1,c_1),\ (p_2,c_2),...,\ (p_N,c_N))\}$, with $p_i\geq 0,$ $c_i\geq 0,$ $i=1,...,N$. The state $((p_1,c_1),\ (p_2,c_2),...,\ (p_N,c_N))$ has the following interpretation: at a given instant of time, there are p_1 unfinished goods and c_1 non – active impulses inside node 1, p_2 unfinished goods and c_2 non – active impulses inside node 2, and so on. Define the following quantities:

$$\begin{aligned}&\mathbf{p}:=(p_1,p_2,...,p_N),\ \mathbf{c}:=(c_1,c_2,...,c_N),\\&(\mathbf{p},\mathbf{c}):=((p_1,c_1),\ (p_2,c_2),...,\ (p_N,c_N)),\end{aligned}\tag{1}$$

and let $\mathbf{e}_i$ be the vector, whose $i-$th component is equal to 1 while the other ones are zero. Moreover, set:

$$a_0^p:=\sum_{i=1}^N a_{0i}^p,\ a_0^c:=\sum_{i=1}^N a_{0i}^c.\tag{2}$$

Indicate by $\pi(\mathbf{p},\mathbf{c})$ the stationary probability of the state $(\mathbf{p},\mathbf{c})$, namely the probability that the queueing network has, for large times, p_i unfinished goods and c_i non – active impulses inside node i, $\forall$ $i=1,...,N$. If the steady state distribution $\{\pi(\mathbf{p},\mathbf{c}),\ \mathbf{p}\geq\mathbf{0},\ \mathbf{c}\geq\mathbf{0}\}$ of the process $\{X(t),t\geq 0\}$ exists, then the following Chapman Kolmogorov equations system holds:

$$\begin{aligned}&\pi(\mathbf{p},\mathbf{c})\left(a_0^p+a_0^c+\sum_{i=1}^N s_i^p\left(1-\alpha_{ii}^p\right)H(p_i)+\sum_{i=1}^N s_i^c(c_i)\right)=\\&=\sum_{i=1}^N\pi(\mathbf{p}-\mathbf{e}_i,\mathbf{c})a_{0i}^pH(p_i)+\sum_{i=1}^N\pi(\mathbf{p},\mathbf{c}-\mathbf{e}_i)a_{0i}^cH(c_i)+\\&+\sum_{i=1}^N\pi(\mathbf{p}+\mathbf{e}_i,\mathbf{c})s_i^p\alpha_{i0}H(p_i+1)+\\&+\sum_{i=1}^N\pi(\mathbf{p}+\mathbf{e}_i,\mathbf{c}+\mathbf{e}_i)s_i^c(c_i+1)\beta_{i0}+\\&+\sum_{i=1}^N\pi(\mathbf{p},\mathbf{c}+\mathbf{e}_i)s_i^c(c_i+1)\left(1-H(p_i)\right)+\\&+\sum_{i=1}^N\sum_{j=1,j\neq i}^N\pi\left(\mathbf{p}+\mathbf{e}_i-\mathbf{e}_j,\mathbf{c}\right)s_i^p\alpha_{ij}^pH(p_i+1)H\left(p_j\right)+\\&+\sum_{i=1}^N\sum_{j=1}^N\pi\left(\mathbf{p}+\mathbf{e}_i,\mathbf{c}-\mathbf{e}_j\right)s_i^p\alpha_{ij}^cH(p_i+1)H\left(c_j\right)+\\&+\sum_{i=1}^N\sum_{j=1}^N\pi\left(\mathbf{p}+\mathbf{e}_i-\mathbf{e}_j,\mathbf{c}+\mathbf{e}_i\right)s_i^c(c_i+1)\beta_{ij}^pH\left(p_j\right)+\\&+\sum_{i=1}^N\sum_{j=1,j\neq i}^N\pi\left(\mathbf{p}+\mathbf{e}_i,\mathbf{c}+\mathbf{e}_i-\mathbf{e}_j\right)s_i^c(c_i+1)\beta_{ij}^cH\left(c_j\right)+\\&+\sum_{i=1}^N\pi(\mathbf{p}+\mathbf{e}_i,\mathbf{c})s_i^c(c_i)\beta_{ii}^c,(\mathbf{p},\mathbf{c})\in\chi,\end{aligned}\tag{3}$$

where $s_i^c(0)=0$ and $H(x)$ is a unit Heavyside function. The system (3), useful to get a mathematical expression for the steady state probability $\pi(\mathbf{p},\mathbf{c})$, has been computed considering all transitions from and to the state $(\mathbf{p},\mathbf{c})$, and balancing incoming and outgoing flows for the state $(\mathbf{p},\mathbf{c})$ (various examples of such a procedure are in Gelenbe et al. 1999).

4. STATIONARY PROBABILITIES

We want to find a general product form solution of the equations system (3), which indicates the state transitions of the presented queueing network, whose nodes have one server. With this aim, define the following quantities: $\forall\ i=1,...,N,$ $x_i^c:=a_i^c+s_i^p,$

$$\rho_i:=\frac{a_i^p}{x_i^c};\qquad q_i^c(j):=\frac{a_i^c}{s_i^c(j)},\ \forall\ i=1,...,N,\ j=1,...,N.$$

Notice that ρ_i represents the stationary probability that the queue of the working station i is busy. Moreover, the following traffic equations hold (see Artalejo 2000, Askin et al. 1993, Bocharov 2000, Bocharov et al. 2003, Gelenbe 1991, Gelenbe 1993, Gelenbe et al. 1999, for more details):

$$\begin{aligned} a_i^p &= a_{0i}^p + \sum_{j=1}^{N} \rho_j \left(s_j^p \alpha_{ji}^p + a_j^c \beta_{ji}^p \right),\ i = 1,...,N, \\ a_i^c &= a_{0i}^c + \sum_{j=1}^{N} \rho_j \left(s_j^p \alpha_{ji}^c + a_j^c \beta_{ji}^c \right),\ i = 1,...,N. \end{aligned} \qquad (4)$$

Equations (4) are interpreted as follows: a_i^p and a_i^c are the total steady state rates of arrival of goods and control impulses, respectively, at node i. For traffic equations, we have the following:

Theorem 1 *(Solution of traffic equations). If matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are irriducible, there exists a unique solution* $\left\{ a_i^p, a_i^c \right\}$, $i = 1,...,N,$ *to equations (4).*

An exhaustive idea of the proof for Theorem 1 is in Gelenbe 1991 and Gelenbe 1993.

Theorem 2 *(Product form solution for stationary probabilities). If matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are irreducible and the following conditions hold:*

$$\rho_i < 1, \quad \delta_i = \sum_{c_i=0}^{+\infty} \prod_{j=1}^{c_i} q_i^c(j) < \infty, \quad i = 1,...,N, \qquad (5)$$

then the Markov process $\left\{ X(t), t \geq 0 \right\}$ *is ergodic and its stationary distribution is represented in product form as:*

$$\pi(\mathbf{p}, \mathbf{c}) = \prod_{i=1}^{N} \pi_i(p_i, c_i), \qquad (6)$$

where, $\forall \ \ i = 1,...,N,$

$$\pi_i(p_i, c_i) = (1 - \rho_i) \rho_i^{p_i} \delta_i^{-1} \prod_{j=1}^{c_i} q_i^c(j),\ p_i \geq 0,\ c_i \geq 0, \qquad (7)$$

and $\prod_{j=1}^{0} \equiv 1.$

Proof. The proof is based on verifying that (6) is a solution of (3). In particular, substituting the expressions of ρ_i and $q_i^c(j)$ and formulas (6) and (7) into the equilibrium equations system (3), we obtain:

$$\begin{aligned} & a_0^p + a_0^c + \sum_{i=1}^{N} s_i^p H(p_i) + \sum_{i=1}^{N} s_i^c(c_i) = \\ & = \sum_{i=1}^{N} \frac{a_{0i}^p}{\rho_i} H(p_i) + \sum_{i=1}^{N} \frac{s_i^c(c_i)}{a_i^c} a_{0i}^c + \sum_{i=1}^{N} \rho_i s_i^p \alpha_{i0} + \\ & + \sum_{i=1}^{N} \rho_i a_i^c \beta_{i0} + \sum_{i=1}^{N} a_i^c \left(1 - H(p_i)\right) + \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\rho_i}{\rho_j} s_i^p \alpha_{ij}^p H(p_j) + \\ & + \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_i \frac{s_j^c(c_j)}{a_j^c} s_i^p \alpha_{ij}^c + \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\rho_i}{\rho_j} a_i^c \beta_{ij}^p H(p_j) + \\ & + \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_i \frac{s_j^c(c_j)}{a_j^c} a_i^c \beta_{ij}^c . \end{aligned} \qquad (8)$$

Using some simplifications, we get that:

$$\begin{aligned} & \sum_{i=1}^{N} \frac{s_i^c(c_i)}{a_i^c} a_{0i}^c + \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_i \frac{s_j^c(c_j)}{a_j^c} s_i^p \alpha_{ij}^c + \\ & + \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_i \frac{s_j^c(c_j)}{a_j^c} a_i^c \beta_{ij}^c = \sum_{i=1}^{N} s_i^c(c_i), \end{aligned} \qquad (9)$$

and:

$$\begin{aligned} & \sum_{i=1}^{N} \frac{a_{0i}^p}{\rho_i} H(p_i) + \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\rho_i}{\rho_j} s_i^p \alpha_{ij}^p H(p_j) + \\ & + \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\rho_i}{\rho_j} a_i^c \beta_{ij}^p H(p_j) = \sum_{i=1}^{N} \left(a_i^c + s_i^p \right) H(p_i). \end{aligned} \qquad (10)$$

Then, from expressions (9) and (10), the equality (8) becomes:

$$\begin{aligned} & a_0^p + a_0^c + \sum_{i=1}^{N} s_i^p H(p_i) + \sum_{i=1}^{N} s_i^c(c_i) = \\ & = \sum_{i=1}^{N} s_i^c(c_i) + \sum_{i=1}^{N} \left(a_i^c + s_i^p \right) H(pi) + \sum_{i=1}^{N} \rho_i s_i^p \alpha_{i0} + \\ & + \sum_{i=1}^{N} \rho_i a_i^c \beta_{i0} + \sum_{i=1}^{N} a_i^c \left(1 - H(p_i) \right) = \\ & = \sum_{i=1}^{N} s_i^c(c_i) + \sum_{i=1}^{N} s_i^p H(pi) + \sum_{i=1}^{N} \rho_i s_i^p \alpha_{i0} + \sum_{i=1}^{N} \rho_i a_i^c \beta_{i0} + \\ & + \sum_{i=1}^{N} a_i^c = a_0^p + a_0^c + \sum_{i=1}^{N} s_i^p H(p_i) + \sum_{i=1}^{N} s_i^c(c_i), \end{aligned} \qquad (11)$$

hence we have just proved an identity. Under the theorem assumptions, the process $\left\{ X(t), t \geq 0 \right\}$ is irreducible. Therefore, according to Foster’s theorem (see Bocharov et al. 2004), the process is ergodic, and formulas (6) and (7) give its unique stationary distribution. This completes the proof.

5. SIMULATIONS

In this section, we examine two different simulation cases. In the first case, a general supply system is considered, for which the mean number of parts to process is computed and an analysis of stability conditions for nodes is made. In the second case, a real network for car parts is studied. Such last situation indicates that, although some instabilities can arise inside the nodes of the network, it is possible to optimize the performances of supply systems via a suitable cost functional.

5.1. A general supply system

We present some numerical results for a supply system, which is represented in Figure 2: there are five working stations (nodes). External flows of goods arrive at each node, while the CEU sends electrical impulses only at nodes 1 and 2. According to some fixed probabilities, unfinished parts can travel from node i to node $i+1$, $i=1,2,3,4$; from node 5, goods either leave the network or come back to node 1. For control impulses, the dynamics is the same of the unfinished parts.

For the just described supply system, we will consider some numerical results for the stationary probabilities and the mean number of parts in the network.

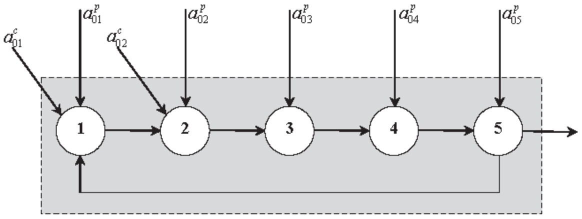


Figure 2: Scheme of the supply system

5.1.1. Numerical results

In what follows we consider some results for the queueing network of Figure 2. Assume that $a_{0i}^p = 10$ $\forall$ $i=1,...,5$, $s_1^p = 20$, $s_2^p = 40$, $s_3^p = s_4^p = 25$, $s_5^p = 30$, where all above quantities are intended as number of goods per minute; $a_{01}^c = a_{02}^c = 5$, $a_{03}^c = a_{04}^c = a_{05}^c = 0$, $s_1^c = s_2^c = s_3^c = s_4^c = 25$, $s_5^c = 30$, where a_{0i}^c and s_i^c, $i=1,...,5$, are measured as number of control impulses per minute;

$$\mathbf{A}^p = \mathbf{B}^p = \begin{pmatrix} 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \\ 0.2 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{12}$$

$$\mathbf{A}^c = \mathbf{B}^c = \begin{pmatrix} 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{13}$$

In Tables 1 and 2, we summarize some values of the stationary probability $\pi_i(p_i, c_i)$ for node i, $i=1,2$. We choose to analyze only the behaviour of nodes 1 and 2, as they are the only ones to be interested by external goods and control impulses rates. Notice that, although such rates are the same for both nodes, if the number of control signals increases, $\pi_i(p_i, c_i)$, $i=1,2$, decreases. This is not surprising, as controls in nodes provoke a variation of the ordinary goods dynamics, either in terms of movements to other nodes or destruction of parts.

Table 1: $\pi_1(p_1, c_1)$ for different values of p_1 (columns) and c_1 (rows)

$p_1 \setminus c_1$	1	2	3
1	0.0394809	0.00789617	0.00157923
2	0.0219893	0.00439787	0.00087957
3	0.0122472	0.00244944	0.00048989

Table 2: $\pi_2(p_2, c_2)$ for different values of p_2 (columns) and c_2 (rows)

$p_2 \setminus c_2$	1	2	3
1	0.0548667	0.0262527	0.0125614
2	0.0179102	0.0085697	0.0041004
3	0.0058464	0.0027974	0.0013385

In order to understand better how stationary probabilities depend on the number of goods, we define the probability $\widetilde{\pi_i}(p_i)$ that a certain node i, $i=1,...,5$, has p_i goods, namely:

$$\widetilde{\pi_i}(p_i) := \sum_{c_i=0}^{+\infty} \pi_i(p_i, c_i), \; i=1,...,5. \tag{14}$$

In Table 3, we collect some values of $\widetilde{\pi_i}$, $i=1,2$.

Table 3: $\widetilde{\pi_i}$ for node i (columns), $i=1,2$, assuming p_j unfinished goods (rows), $j=1,2,3$

$i \setminus p_j$	1	2	3
1	0.246755	0.137433	0.0765452
2	0.219874	0.0717737	0.0234292

Notice that $\widetilde{\pi_i}(p_i)$ increases when the number of goods decreases and, moreover, $\widetilde{\pi_1}(p_1) > \widetilde{\pi_2}(p_2)$, indicating that node 1 tends to have more parts than node 2. This is an evident influence of the possibility to reprocess some goods, coming from node 5, inside node 1.

Further studies can be done considering the mean number of parts in the network, namely:

$$N_p := \sum_{p_i=0}^{+\infty} p_i \left(\sum_{c_j=0}^{+\infty} c_j \pi_i(p_i, c_j) \right). \tag{15}$$

If we sketch N_p vs a_{01}^p (Figure 3, top) and vs a_{02}^p (Figure 3, bottom), we have a precise idea of the ergodicity condition of the network process. In particular, if the network is simulated with:

- a_{01}^{p} variable and other parameters equal to the ones used before, node 1 becomes instable when $a_{01}^{p} \simeq 24.88 = a_{01}^{p,*}$, leading to the instability of the overall network;
- a_{02}^{p} variable and other parameters equal to the ones used before, the network process is not ergodic anymore if $a_{02}^{p} \geq 48.89 = a_{02}^{p,*}$.

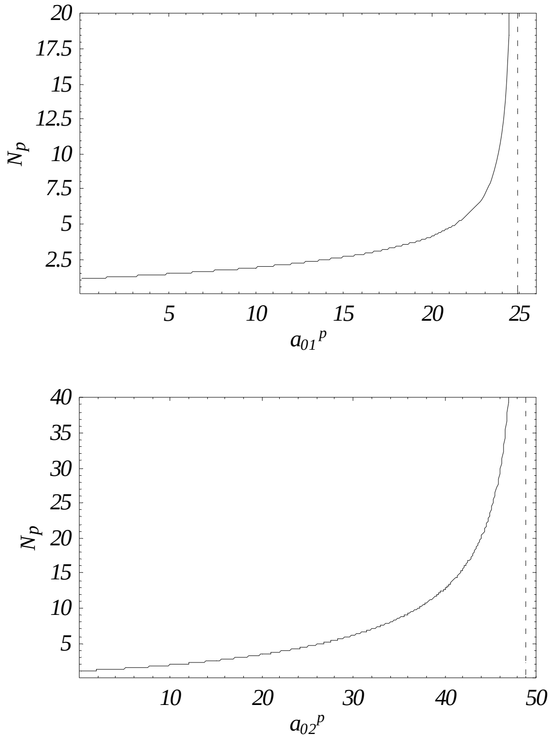


Figure 3: N_p vs a_{01}^{p} (top) and a_{02}^{p} (bottom)

A similar phenomenon happens considering the behaviour of N_p vs a_{01}^{c} (Figure 4, top) and vs a_{02}^{c} (Figure 4, bottom). We get that if:

- a_{01}^{c} is variable and the other parameters are equal to the ones used before, the condition of instability for node 1, and hence for the overall network, is achieved for $a_{01}^{c} \simeq 24.99 = a_{01}^{c,*}$;
- a_{01}^{c} varies while the other parameters are the same ones used before, node 2 is instable for $a_{02}^{c} \geq 18 = a_{02}^{c,*}$, and the network process is not ergodic anymore.

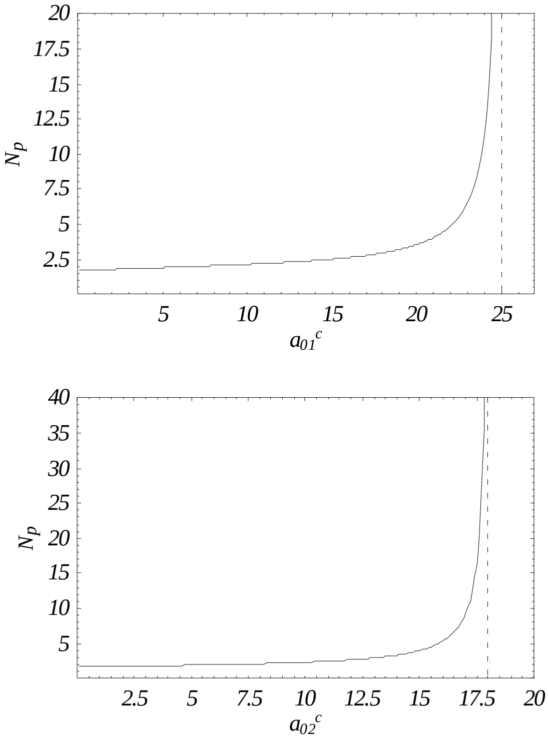


Figure 4: N_p vs a_{01}^{c} (top) and a_{02}^{c} (bottom)

Moreover, notice that, in analogy with the usual exponential queueing systems with one server, the shape of N_p in Figures 3 and 4 is the one of a hyperbolic function. In Figure 5, we represent: on the top, N_p as function of a_{01}^{p} and a_{02}^{p}; on the bottom, N_p vs a_{01}^{c} and a_{02}^{c}. In both cases, the other parameters, which are not assumed variable, are equal to the ones used for computing the stationary probabilities.

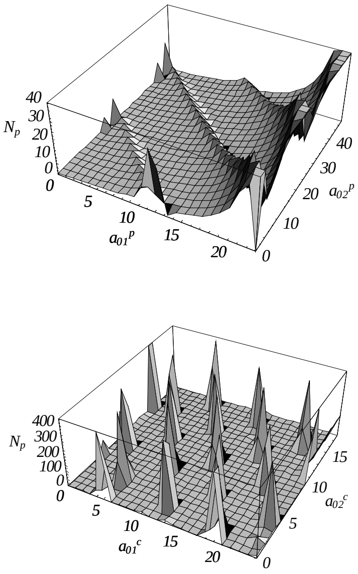


Figure 5: N_p vs a_{01}^{p} and a_{02}^{p} (top), and vs a_{01}^{c} and a_{02}^{c} (bottom)

Notice that: for Figure 5, top, N_p tends to infinity only if $a_{01}^{p} \simeq a_{01}^{p,*}$ and $a_{02}^{p} \simeq a_{02}^{p,*}$; for Figure 5, bottom, N_p approaches the infinity for various combinations of a_{01}^{c} and a_{02}^{c}, and not only for the critical values $a_{01}^{c,*}$ and $a_{02}^{c,*}$. Such effect indicates that the ergodicity of the network process is mainly influenced by control impulses rates, and this is not unusual, as controls always tend to create some natural discontinuities in the normal working processes of goods.

5.2. A real network for car parts

We describe some simulation results for the network in Figure 6, that represents a scheme of real industrial processes, that are commonly used for assembling car parts.

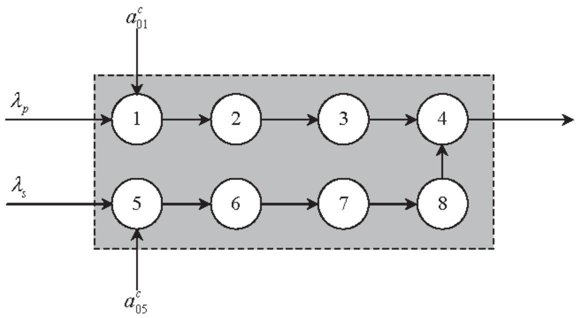


Figure 6: A network for assembling car parts

There are eight nodes and two external flows of goods. In particular, we distinguish: a primary flow, that has rate λ_p and consists of car skeletons; a secondary flow, with rate λ_s for little parts of cars, namely mirrors, glasses, wheel rims, and so on. At each node a precise activity is associated. As for car skeletons, in node 1 they are washed, dried in node 2 and then painted in node 3. For the secondary flow, instead, we have that little components of cars are washed, dried and painted in nodes 5, 6 and 7, respectively. Then, such parts are completely assembled in node 8, and a complete car is obtained in node 4. The just assembled cars go out of the network from node 4. Figure 7 sums up the complete assembling process. For such a system, we will consider some numerical results for a cost functional, that represents, using an opportune weight, the joint effect of the mean number of parts and controls inside the network.

5.2.1. Numerical results

Assume that primary and secondary flows have variable rates, respectively, $a_{01}^{p} = \lambda_p \in \left]0,30\right[$ and $a_{05}^{p} = \lambda_s \in \left]0,30\right[$. Moreover, we have: $a_{0i}^{p} = 0$, $i = 2,3,4,6,7,8$; $a_{01}^{c} = a_{05}^{c} = 2$, $a_{0i}^{c} = 0$, $i = 2,3,4$; $s_i^p = 10$ $\forall$ $i = 1,\dots,4$, $s_i^p = 20$ $\forall$ $i = 5,\dots,8$, $s_i^c = 1$ $\forall$ $i = 1,\dots,8$, where all above quantities are measured per minute; $\mathbf{A}^c = \mathbf{B}^p = \mathbf{B}^c = \mathbf{0}$, where $\mathbf{0}$ is the zero matrix of order 8; and $\mathbf{A}^p$ has elements:

$$\alpha_{ij}^{p} = \begin{cases} 1, & \text{if } j = i+1,\ i \in \{1,2,3,5,6,7\}, \\ & \text{or } i = 2j, \text{ with } j = 4, \\ 0, & \text{otherwise.} \end{cases} \qquad (16)$$

Notice that matrices indicate that all goods always travel as a “parts to process” from one station to the following one.

In order to describe the performances of the system, we define the following cost functional:

$$J\left(\lambda_p, \lambda_s\right) := wN_p - \left(1-w\right)N_s, \qquad (17)$$

where N_p is defined as in (15), N_s is the mean number of control signals inside the network, given by:

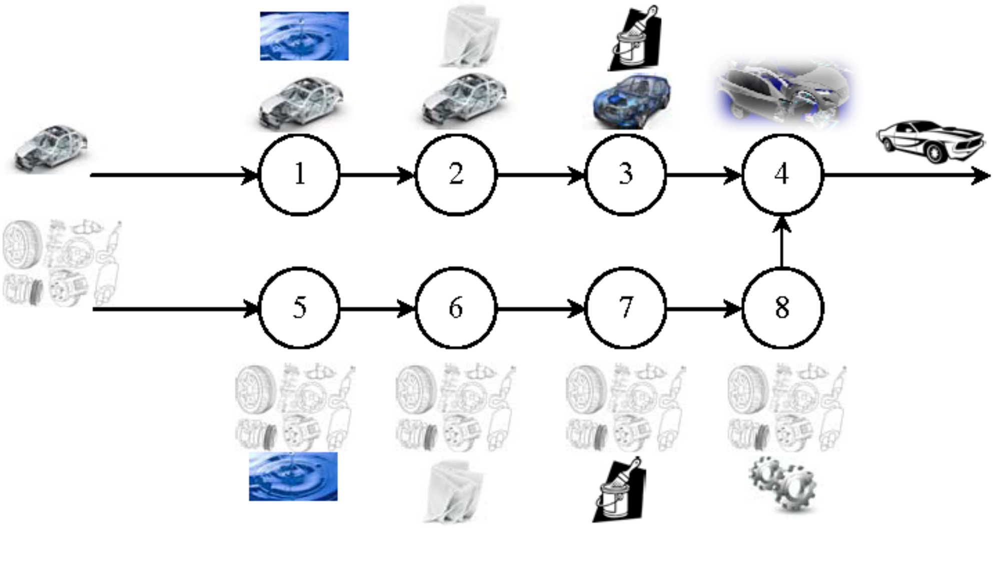


Figure 7: assembling process of car parts

$$N_s := \sum_{c_j=0}^{+\infty} c_j \left(\sum_{p_i=0}^{+\infty} p_i \pi_i \left(p_i, c_j \right) \right), \tag{18}$$

and $w \in \left]0,1\right[$ is a real number, that weights either the contribution of N_p or the one of N_s. The aim is to maximize J with respect to the couple $\left(\lambda_p, \lambda_s\right)$, namely we want to find the values of primary and secondary flows in order to: increase the mean number of parts inside the network, with consequent advantages in terms of the production itself; reduce the possibility of controlling nodes by signals. This aim is highly non-trivial as the ergodicity condition of the network process has also to be considered. Mathematically speaking, the problem is the following:

$$\begin{aligned} &\max_{\left(\lambda_p,\lambda_s\right)} J\left(\lambda_p, \lambda_s\right), \\ &\left(\lambda_p, \lambda_s\right) \in \left]0,30\right[\times \left]0,30\right[, \\ &\left(\lambda_p, \lambda_s\right) \text{ such that: } \rho_i < 1, \ \sum_{c_i=0}^{+\infty} \prod_{j=1}^{c_i} q_i^c\left(j\right) < \infty, \end{aligned} \tag{19}$$

where the last constraint of problem (19) indicates that $\left(\lambda_p, \lambda_s\right)$ has to be chosen in order to respect the stability condition for each node of the network.

As an analytical analysis of J is very complex, some numerical estimations have been made using the software Mathematica. For $w = \frac{1}{2}$, we have obtained that, if $\left(\lambda_p, \lambda_s\right) \in \left]\overline{\lambda}_p, 30\right[\times \left]\overline{\lambda}_s, 30\right[$, with $\overline{\lambda}_p = 22.3$ and $\overline{\lambda}_s = 24.7$, the network process is not ergodic as J tends to infinity. Indeed, for values of $\left(\lambda_p, \lambda_s\right)$ such that the network is stable, there exists a unique maximum point at $\left(\lambda_p^*, \lambda_s^*\right) \simeq \left(17.5, 16.5\right)$ for which $J\left(\lambda_p^*, \lambda_s^*\right) \simeq 6.4$, see Figure 8. Hence, the output for the car parts of the system is optimized for values of primary and secondary flows, that approach 30, the maximal possible rate.

Notice that, for other values of w, $\overline{\lambda}_p$ and $\overline{\lambda}_s$ are obviously different but $\max_{w \in]0,1[} \overline{\lambda}_p = 23.6$ and $\max_{w \in]0,1[} \overline{\lambda}_s = 25.8$, namely there is no meaningful difference with the case $w = \frac{1}{2}$. The same happens with the maximum point, for which minimal variations occur.

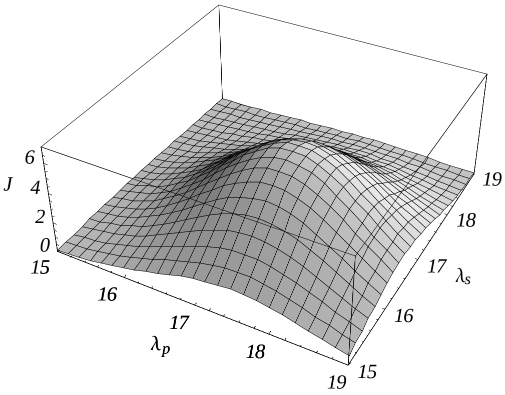


Figure 8: J vs λ_p and λ_s for $w = \frac{1}{2}$

6. CONCLUSIONS

In this paper, it has been described an exponential queueing network, which models a supply system, whose dynamics is determined either by unfinished parts or control impulses.

Steady state probabilities for such a queueing network have been found in product form.

A numerical analysis of the model has allowed to establish that the stationary probabilities are deeply influenced by control impulses, that also have a strong impact on the overall dynamics of the queueing network.

A real network for assembling car parts has been studied through a cost functional in order to maximize the mean number of parts inside the system with the minimal number of control signals.

REFERENCES

Artalejo, J. R., 2000. G - networks: a versatile approach for work removal in queueing networks. *European Journal of Operational Research*, 126, 233–249.

Askin, R. G., Standridge C. R., 1993. *Modeling and analysis of manufacturing systems*. Wiley and Sons, New York.

Bocharov, P. P., 2002. Queueing networks with signals and random signal activation. *Automation and Remote Control*, 63 (9), 1448–1450.

Bocharov, P. P., D'Apice, C., Pechinkin, A. V., Salerno, S., 2004. *Queueing Theory, Modern Probability and Statistics*. VSP, The Netherlands.

Bocharov, P. P., Vishnevskii, V. M., 2003. G – networks: Development of the Theory of Multiplicative Networks. *Automation and Remote Control*, 64 (5), 714–739.

Cutolo, A., Piccoli, B., Rarità, L., 2011. An Upwind – Euler scheme for an ODE – PDE model of supply chains. *SIAM Journal on Computing*, 33 (4), 1669–1688.

Gelenbe, E., 1991. Product Form Queueing Networks with positive and negative customers. *Journal of Applied Probability*, 28, 656–663.

Gelenbe, E., 1993. G – Networks with Triggered Customer Movement. *Journal of Applied Probability*, 30, 742–748.

Gelenbe, E., Pujolle, G., 1999. *Introduction to Queueing Networks, Second Edition*, Wiley and Sons, New York.

Göttlich, S., Herty, M., Klar, A., 2006. Modelling and optimization of supply chains on complex networks. *Communication on Mathematical Sciences*, 4, 315–330.

Jackson, J. R., 1957. Networks of waiting lines. *Operations Research*, 5, 518–521.

Pasquino, N., Rarità, L., 2012. Automotive processes simulated by an ODE - PDE model. *Proceedings of EMSS 2012 (24th European Modeling and Simulation Symposium, September 19th - 21th, 2012, Vienna, Austria)*, 352–361.

Yao, D. D., Buzacott, J. A., 1986. The exponentialization approach to flexible manufacturing systems models with general processing times, *European Journal of Operational Research*, 24, 410–416.

AUTHORS BIOGRAPHY

MATTEO GAETA was born in Salerno, Italy, in 1960. He graduated in Information Science in 1989. He is an Associate Professor of Information Processing Systems at the Engineering Faculty of the University of Salerno. His research interests include: Complex Information Systems Architecture, Software Engineering, Systems of Knowledge Representation, Semantic Web, Virtual Organization and Grid Computing. He has been an IEEE Member since 2005 and in 2008 also joined the Computational Intelligence Society. He is the Scientific Coordinator and Manager of several International Research Projects, Expert in Industrial Research and Innovation, Coordinator of the MIUR Working Group.
His e-mail address is mgaeta@unisa.it.

LUIGI RARITÀ was born in Salerno, Italy, in 1981. He graduated cum laude in Electronic Engineering in 2004, with a thesis on mathematical models for telecommunication networks, in particular tandem queueing networks with negative customers and blocking. He obtained PhD in Information Engineering in 2008 at the University of Salerno, discussing a thesis about control problems for flows on networks. He is actually a research assistant at the University of Salerno. His scientific interests are about numerical schemes and optimization techniques for fluid – dynamic models, queueing networks, and Knowledge models for the Cultural Heritage area.
His e-mail address is lrarita@unisa.it.

TOPIC EXTENSION USING THE NETWORK EXTRACTED FROM DBLP

Alisa Babskova[a], Pavla Draždilová[b], Jan Martinovič[c], Kateřina Slaninová[d]

[a,b,c]VŠB - Technical University of Ostrava,
Department of Computer Science,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[d]Silesian University in Opava,
SBA in Karvina, Department of Informatics,
Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic

[a]alisa.babskova@vsb.cz, [b]pavla.drazdilova@vsb.cz, [c]jan.martinovic@vsb.cz, [d]slaninova@opf.slu.cz

ABSTRACT

This article focuses on the topic extension in an area that is initially specified by the user through the topic's keywords. The extended area of interest defined by keywords is determined by a set of terms used by the community for which the selected keywords are significant. The extracted topic by selected communities can be used to update and broaden the area of interest. This new evaluation of edges depends on terms that appear in the titles of articles of two co-authors. The newly evaluated network more accurately describes the intensity of the relationships between co-authors. This network is suitable as an input to models, which are focused on prediction of future relationships and community structures in co-author networks. Moreover, the topic extension may be used in prediction models for the extraction of expected keywords which will be used in a given community.

Keywords: e-learning, topic extraction, DBLP, subnet-works, community detection, study materials

1. INTRODUCTION

In the field of e-learning, we come across a variety of electronic learning materials that allow students to familiarize themselves with the chosen topic. There are the syllabi, lecture notes, presentations and other educational texts available and they create a collection of documents.

These materials are mostly created by educators for the selected theme and provide a comprehensive overview of selected topics and areas. Our intention is not to provide comprehensive and pre-processed materials, but to offer students the opportunity to familiarize themselves with a self-selected and an interesting topic for them in the field of computer science. A student specifies the area in which she/he is interested in (topic) with one or more terms. Our approach finds the most important community described by the specified topic (for the selected keywords) in the co-authors' network DBLP (http://dblp.uni-trier.de/). These selected keywords are often used in headlines of articles, or are common to most members of the community, or the most commonly occur in posts on the blog. Authors who have written many articles that relate to the selected terms can be determined for this community. Thereafter, we are able to find other keywords from the document collection, which can not only describe the extension of the defined set of keywords for the given topic, but also can be significant for other topics, which are on the interest of the given community of co-authors. These new words may be useful for further selection of other keywords, and for further selection of other documents, which will lead to the extension of the initial topic. Some new keywords obtained by this way can be out of the initial topic, and can therefore extend the whole scope of users' interest (not only extend the narrow topic initialised at the beginning).

Another possibility is to choose articles from the collection of documents that were written in specified time period (e.g. the latest articles in the field). These articles can provide the newest information and additional terms, which may refine or expand students' topic and thus they can be independently and proactively involved in their education process.

This type of study materials searching is suitable for doctoral students or for undergraduate students in computer science, because our system works with large database of articles focused on this sphere - DBLP. Generally, this kind of study material searching can be used in any other area with documents that have titles or are briefly described (e.g. by abstracts). In the resulting network with newly evaluated edges, there are found communities of co-authors, who published together and who have strong relations to the selected terms. Additionally, we can choose the authors who have the most publications in the topic and whose work can be beneficial and inspiring for the students.

In the publication (Monachesi, Lemnitzer and Simov 2006) is presented The European project Language Technology for eLearning (LT4eL), which aim is to improve the efficiency and the availability of the static and dynamic content created for eLearning using Learning Management Systems (LMS). This problem was solved using language technologies based on the functionality and the integration of semantic knowledge, and might facilitate the management, distribution and retrieval of study materials.

Other possible approach to gaining the extended topics is usage of time information. The authors of publication (Chen, Luesukprasert and Chou 2007) deal with retrieval of actual topics from various collections of text documents published in a given time period. The presented method consists of two steps. The first step is focused on the extraction of the actual terms and on mapping of their distribution through the given time period. The second step consists of the identification of key sentences (based on the extracted actual terms), which are clustered. The clusters then represent the actual topics defined by multidimensional vector of sentences.

The article (Schirru, Baumann, Memmel, and Dengel 2010) is focused on the automatic identification of various topics, which are a scope of interest of source sharing platforms users.

The authors of article (Sun, Barber, Gupta, Aggarwal and Han 2011) focused on the prediction of the future relations between the co-authors in the heterogeneous bibliographic network (DBLP) using the heterogeneous topological characteristics. The community evolution is a scope of interest in (Brodka, Saganowski and Kazienko 2011), in which is presented a method Group Evolution Discovery (GED). The method uses not only the size and comparison of the group members, but also takes into consideration their significance and position inside the group, to find the progress of the group in the sequential time periods. In (Patil, Liu and Gao 2013) is presented the groups evolution and their stability. The analysis (for example of DBLP) showed that it is possible to predict the group stability with the high accuracy using various attributes which describe the group composition, the activities inside the groups and the group structural aspects.

Our developed network is suitable as the input for the models, which are focused on the prediction of the future relations and the community structures in co-author networks. Moreover, the topic extension can be usable for the extraction of the future keywords used within a given community.

The rest of this paper is structured as follows: Section 2 describes the related work in the social and co-author networks, and wikis and blogs as sources of documents. In Section 3, our proposed approach is presented. We depict our idea of relations between persons on the basis of term context and describe how the ContextScore as a new edge evaluation in the network can be obtained. Then, in Section 4, we present the comparison of co-authors' network with and without term context. Our method has been tested and a topic extension is presented in Section 5. In Section 6, we summarize our findings and present ideas for the future work.

2. DBLP AND OTHER SOURCES

DBLP (Digital Bibliography Library Project) is a computer science bibliography database hosted at University of Trier, in Germany. It was started at the end of 1993 and listed more than 2.1 million publications in January 2013. These articles were published in Journals such as VLDB, the IEEE and the ACM Transactions and Conference proceedings. DBLP has been a credible resource for finding publications, its dataset has been widely investigated in a number of studies related to data mining and social networks to solve different tasks such as recommender systems, experts finding, name ambiguity, etc. Even though, DBLP dataset provides abundant information about author relationships, conferences, and scientific communities. It has a major limitation that its records provide only the paper title without the abstract and index terms.

Wiki, blogs and different sources of information are useable to create a new edge evaluation of network, in which nodes are persons created page in wiki or blog and edges represent cooperation in project.

Many experts focused on the task of finding persons with the high level of experience in a specific topic. To achieve this objective, researchers approached this task mainly in three different ways. The first group applied information retrieval techniques to solve the mentioned problem (Deng, King, Lyu 2008).- The authors of this paper proposed a weighted language model, which introduced a document prior probability to measure the importance of the document written by an expert. The second group approached this task using social network analysis metrics (Zhang, Ackerman, Adamic 2007). In this study, the Java Forum, a large online help seeking community, was analysed using social network analysis methods and a set of network-based algorithms including PageRank and HITS. The third group used a hybrid approach of information retrieval and social network analysis for finding academic experts (Zhang, Tang, Li 2007). In (Zhang, Tang, Li 2007), the authors created a local information document for each person to measure his initial level of experience on a topic using information retrieval models. Then they applied propagation on the graph of experts to update his level of expertise according to his relations with the other nodes. In the article (Drazdilova, Martinovic, Slaninova 2013), the authors focused on the detection of communities using spectral clustering. This algorithm was used in the article (Minks, Martinovic, Drazdilova, Slaninova 2011) to find the communities in subnetworks that were defined by the selected terms (from the whole DBLP).

Wiki, blogs and different sources of information are useable to create a new edge evaluation of network, in which nodes are persons created page in wiki or blog and edges represent co-operation in project.

In (Yang and Ng 2008), the authors proposed an analytical system of web forum for the analysis of the content development and for the visualisation of the social relations in web forums. Our approach creates a new evaluation of network edges (relations) and head towards the usage of the documents from DBLP, wikis or blogs to extract and develop the initial topics.

The presented approach demonstrated for DBLP can be used also for other mentioned document resources, in which the relation between the persons (co-authors) is created on the basis of the common documents and its evaluation is dependent on the extracted terms from the document titles. Table 1 presents various document resources and other relevant information. Even from such different document resources, we are able to create the co-authors' network, who participated on the creation of common documents. Therefore, as well as in DBLP, the evaluation of the relations is dependent on the extracted terms from the documents.

Source	Documents with terms	Persons	Time
DBLP	Title of Paper	Authors	Year of Publishing
Wiki	Wiki page	Editors	Last Edited
Blog	Post	Bloggers	Last Sending
Codeplex	Project Description	Developers	Last Activity

Table 1: Mapping Different Information Sources to Person and Term Context

3. RELATIONS BETWEEN PERSONS ON THE BASIS OF TERM CONTEXT

In the paper, we propose a more precise evaluation of the intensity of person's relations to ascertain the context among persons (e.g. authors, editors, bloggers, developers) and the terminology they used in documents (for example terms in article titles in DBLP, terms in Wiki pages and blogs or terms in projects description).

Wang, Mccallum and Wei (2007) present topical n-grams, a topic model that discovers topics as well as topical phrases. Another area that utilizes text information is finding of expert in DBLP bibliography data (Deng, King and Lyu 2008), or the analysis of communities based on DBLP (Biryukov and Dong 2010).

In our approach, we use terms for the evaluation of the relation between persons. We extend a standard evaluation of the relation, which is based on the number of the common articles, by a factor that represents a context between persons and term selected from the term set.

Term set is understood as a collection of all keywords, which are extracted from the document. As the source of terms were used titles of articles from the DBLP dataset. A more detailed description of the term set was presented in article (Minks, Martinovic, Drazdilova, and Slaninova 2011).

3.1. Relations between Persons

Besides the computation of evaluated term set, we can compute *association strength* between the two persons. This method is not only interesting by itself, but it is also essential for extended evaluation of the term list by selected context.

Relevancy between persons is based on the participation on the same document. This relevancy is then approximated by Jaccard coefficient (Deza and Deza 2006).

Let A be a set of all persons in dataset. We define a single person A_i. For A_i, it is evaluated the strength of association with the other persons (co-participants).

The set of co-participants of person A_i is marked as C_{A_i}. Let set P be a set of all documents (papers) and P_{A_i}be a set of all documents of person A_i.

The *association strength* between the persons A_i and A_j can be defined with Jaccard coefficient that reflects mainly the proximity of both persons from number of their common documents:

$$Q(A_i, A_j) = \frac{|P_{A_i} \cap P_{A_j}|}{|P_{A_j}| + |P_{A_i}| - |P_{A_i} \cap P_{A_j}|}. \qquad (1)$$

If this method is applied to all the persons, we obtain weighted undirected graph that can be considered as a synthetic social network (with re-weighted edges between persons). This approach was inspired by (Ding 2011).

3.2. Persons and the Term Context

If we define a set T as the set of all terms in all documents and T_{A_i} as a set of all the terms that could be found in the documents of person A_i, then t_k is the term belonging to the person A_i ($t_k \in T_{A_i}$).

Thus, we define $(t_k\ in\ T_{A_i})$ as a number of occurrences of the term t_k in the documents of person A_i. Then, this number is divided by the number of occurrences of term $\mathrm{t_k}$ in the all project's description($\mathrm{t_k}$ in T). The higher value, the less relevant term $\mathrm{t_k}$ becomes. In addition, a number of terms of the author $\mathrm{A_i}$($|\mathrm{T_{A_i}}|$) is added to the number of occurrences of the term $\mathrm{t_k}$, because there is an assumption that $\mathrm{T_{A_i}}$, which has a high cardinality, lower the importance of the individual terms, while low cardinality indicates that the author has only one subject matter. Then, we can define the *relevance of author's terms* as:

$$R(T_{A_i}, t_k) = \frac{(t_k\ in\ T_{A_i})}{(t_k\ in\ T) + |T_{A_i}| - (t_k\ in\ T_{A_i})}, \qquad (2)$$

and in normalized form as:

$$R_{Norm}(T_{A_i}, t_k) = \frac{R(T_{A_i}, t_k)}{MAX(R(T_{A_i}, t_1), \ldots, R(T_{A_i}, t_{|T_{A_i}|}))}. \qquad (3)$$

Because we have defined the relation between the persons and we can express the relevance of the person's terms, we can assign the best suitable co-participant to a given term. We can demonstrate the usage of significance of each co-participant as well. Our reflections were inspired by associative memory, where one is able to better recall the event, which is associated with something significant (although it was already forgotten). For a given person, it is significant the term, which associates him the best co-participant in the selected topic.

The method extension, including the person's co-participant as a context, is constructed analogically. The context is calculated for a given person according to the Formula (1). Afterwards, the persons are selected from the evaluated list of co-participants.

The $ContextScore_k$ for selected term t_k is calculated by the equation:

$$ContextScore_k(A_i, A_j, t_k) = R_{Norm}(T_{A_i}, t_k)\ R_{Norm}(T_{A_j}, t_k)\ Q(P_{A_i}, P_{A_j}) \qquad (4)$$

The overall *ContextScore* between persons is given by the sum of particular $ContextScore_k$ related to the particular terms, which create the query.

$$ContextScore(A_i, A_j) = \sum_{t_k \in query} ContextScore_k(A_i, A_j, t_k) \qquad (5)$$

4. COMPARISON OF CO-AUTHORS NETWORK WITH AND WITHOUT TERM CONTEXT

The selected experiments focused on the comparison of the initial edge evaluation, which represents the amount of common publications of the two authors and the resulting node evaluation in 2010 within DBLP, and the new proposed one, which takes into consideration the selected terms. By this way, we obtain two different evaluations of the authors, which allow us to sort them and to select the most important authors, which are supposed to be described by other interesting terms, which extend a given topic.

In the experiments, the set of terms was selected, which represented a simplified topic. Therefore, these input terms defined a topic of our interest towards which we will evaluate the authors in DBLP. Using these terms we have selected from the complete graph of DBLP such subgraphs, in which the relations between the persons were based on term context. Of course, such subgraphs did not contain all the initial authors from the DBLP collection, but only these who had the required terms which defined the topic in the publication title (Babskova, Drazdilova, Martinovic, Svaton, and Snasel 2013).

During the experiment we start the creation of author subgraph set for the selected year 2010 based on input terms. We have chosen two types of queries for the experiment of which the second one specifies our area of interest. The queries are “social” and “social network”. Two different evaluations of the intensity of relationships between co-authors have been used to compare the results. The first subgraph features original evaluation of the edges representing the amount of joint publications of two authors in a given year and in a given area (labelled as ‘without term context’ in the text) and the second evaluation is ContextScore (see Formula 5) and it is labelled as “with term context” in the text.

As the next step in our approach, we have calculated the weighted degree (Newman 2004) for both methods of edge weight recalculation. Following that and based on these different evaluations, we have created author lists sorted according to relevant weighted degrees that demonstrate different significance of an author depending on the used edge evaluation. Information on the subgraphs retrieved for the entered terms of “social” and “social network” are in the Table 2.

Set of terms	Nodes	Edges	Components
Social	5599	8914	1537
Social network	2418	3715	669

Table 2: Subgraphs for Terms ‘social’ and ‘social network’

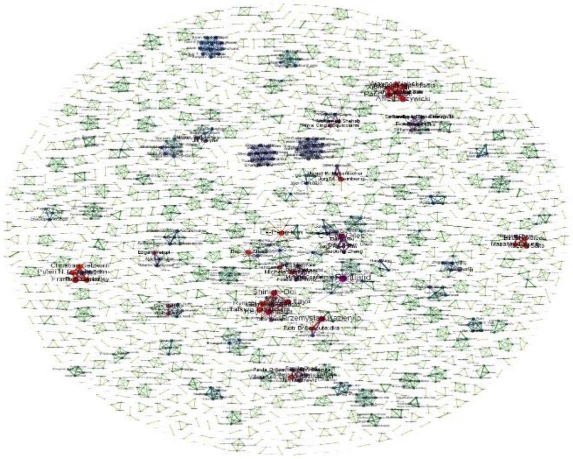


Figure 1: Subgraph (DBLP) for Term ‘social network’ – Weighted, without Term Context

Before starting to analyse the results, it is important to say that if we consider activity of an author in specific subgraph, then we always consider only the activity of an author in specific area which is specified by the input terms. The whole subgraph determined by the term “social” is shown in the Figure 1. Intensity of the relationships is determined by the original edge evaluation – i.e. by the number of joint publications in 2010. Re-evaluation of the edges by means of the term context resulted in different authors´ degrees in the subgraphs (see Figure 2).

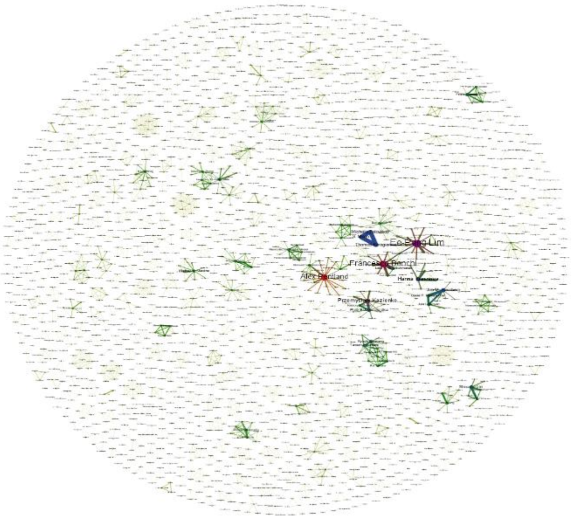


Figure 2: Subgraph (DBLP) for Term ‘social network’ – Weighted, with Term Context.

We focus on each subgraph separately in the next part of the experiment. We have calculated the weighted degrees for all the authors in the subgraph for the community with the edge evaluation both with and without term context. The examples of the authors´ weighted degree diagrams, always for both evaluation methods, are shown in Figure 3, Figure 4, Figure 5, and Figure 6.

The diagrams shown in Figure 3, Figure 4, Figure 5, and Figure 6 make it clear that the new edge evaluation has changed the node degree distribution which resulted in lower number of authors with the highest new weighted degree. That enables us to constrict the set of authors in a given area that are significant from our point of view.

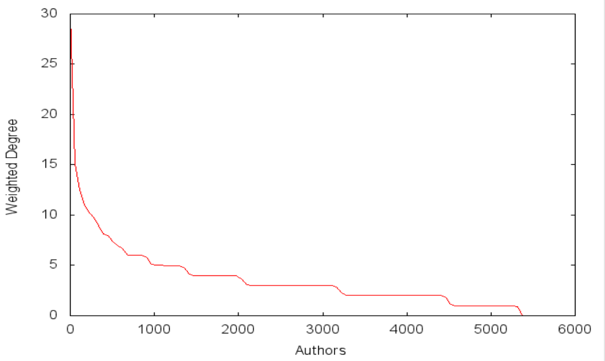


Figure 3: Histogram of Sorted Weighted Degree of Subgraph for Term ‘social’ - without Term Context

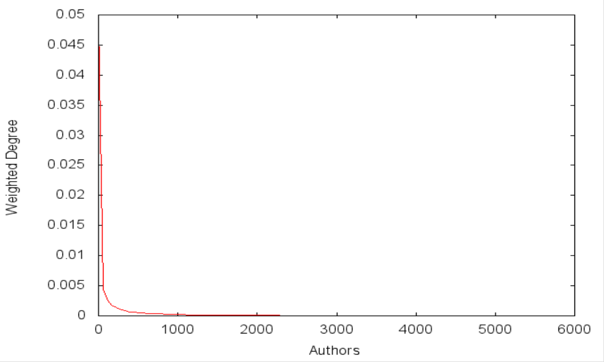


Figure 4: Histogram of Sorted Weighted Degree of Subgraph for Term ‘social’ - with Term Context

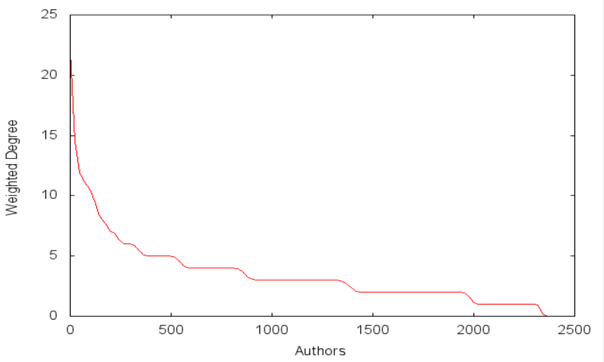


Figure 5: Histogram of Sorted Weighted Degree of Subgraph for Terms ‘social network’ - without Term Context

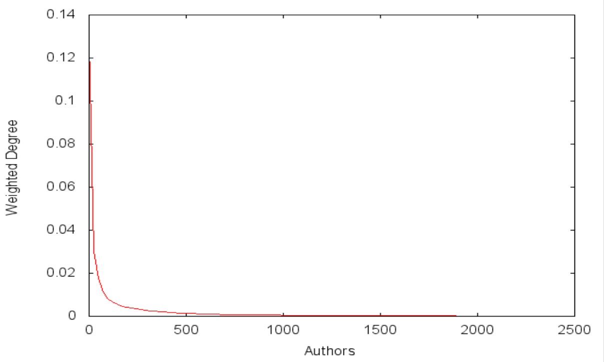


Figure 6: Histogram of Sorted Weighted Degree of Subgraph for Terms ‘social network’ - with Term Context

Further we spot the authors with relatively high degree when compared to the other authors in a given subgraph. We will find Top 10 authors with the highest degree value. Table 3 and Table 4 show Top 10 authors for the ‘social network’ and ‘social’ subgraph. These tables demonstrate that the Top 10 authors with the highest degrees, calculated on the basis of the original evaluation, mostly did not appear in Top 10 of authors with the degree calculated with term context. Only the very active or the significantly publishing authors occur in both subgraph without term context and subgraph with term context.

Subgraph, Weighted without Term Context		Subgraph, Weighted with Term Context	
ID of Author	Name of Author	ID of Author	Name of Author
49669	Alex Pentland	56679	Ee-Peng Lim
57581	C. Lee Giles	57835	Francesco Bonchi
28700	Przemyslaw Kazienko	49669	Alex Pentland
690549	Satoko Itaya	28700	Przemyslaw Kazienko
119660	Shinichi Doi	260831	Hanna Krasnova
252705	Keiji Yamada	225725	Thomas Karagiannis
735639	Xiongcai Cai	181414	S, Moon
698479	Alfred Krzywicki	96113	Michalis Faloutsos
357960	Wayne Wobcke	38231	Jon M, Kleinberg
333282	Yang Sok Kim	57593	Shou-De Lin

Table 3: Top 10 of Authors of Subgraph ‘social network’ without Term Context and with Term Context

Top 10 comprises the authors with the highest degree values in the subgraph. These authors have been detected as the most active in the area defined by the input terms. It is possible to retrieve other terms with frequent occurrence in the publications of each author. As we focus on the most active authors, there is high probability that the retrieved terms by this way describe the topic defined by the input terms set in a greater depth or even expand the pre-defined topic with other areas of interest being currently explored.

Figure 7 and Figure 8 are shown parts of subgraphs, which present significance of particular authors for communities of co-authors. These graphs provide information not only about the activity of the significant members, but about the whole community.

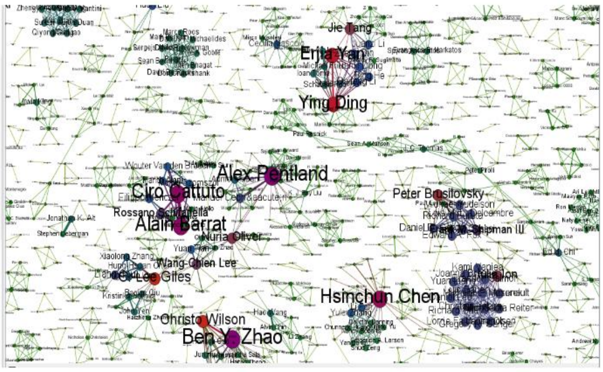


Figure 7: Part of Subgraph for Term ‘social’ – Weighted, without Term Context

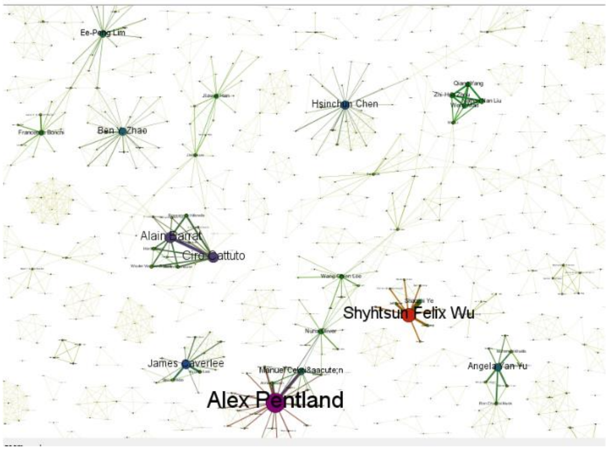


Figure 8: Part of Subgraph for Term ‘social’ – Weighted, with Term Context

Subgraph, Weighted without Term Context		Subgraph, Weighted with Term Context	
ID of Author	Name of Author	ID of Author	Name of Author
49669	Alex Pentland	49669	Alex Pentland
111861	Ben Y. Zhao	120967	Shyhtsun Felix Wu
182741	Alain Barrat	182741	Alain Barrat
360307	Ciro Cattuto	360307	Ciro Cattuto
46738	Hsinchun Chen	57006	James Caverlee
45474	Ying Ding	46738	Hsinchun Chen
45635	Erjia Yan	170122	Angela Yan Yu
57581	C. Lee Giles	111861	Ben Y. Zhao
543984	Christo Wilson	57593	Shou-De Lin
57006	James Caverlee	252892	Munmun De Choudhury

Table 4: Top 10 of Authors of Subgraph ‘social’ without Term Context and with Term Context

Table 55 and Table 66 for ‘social’ and ‘social network’ subgraphs show other frequently occurring terms for Top 10 users, retrieved by means of weighted degree without and with term context. These terms allowed us other possibilities, to which we can concern. For example, term ‘social’ has the most frequently occurred other term ‘network’. Other extended terms are ‘signal’, ‘online’, ‘movement’, ‘ontology’, ‘spam’, ‘socialtrust’, etc.

Subgraph, Weighted without Term Context		Subgraph, Weighted with Term Context	
ID of Author	Other Terms	ID of Author	Other Terms
49669	social;network; interaction;signal; processing;sensing	49669	social;networks; interation;signal; processing;sensing
111861	social;graph;online; networks;detecting; characterizing	120967	social;online; networks;systems; estimating
182741	social;link; live; semantics;creation; profile;network	182741	social;link; live; semantics;creation; profile;network
360307		360307	
46738	social;movement; network;cyber; research;web	57006	social;socialtrust; spammers;information;communities
45474	social;tagging;tag; ontology;integrating	46738	social;movement; network;cyber;

45635		170122	social;learning; online;networking;knowledge
57581	social;network; ranking;document	111861	social;graph;online; networks;detecting; characterizing
543984	social;detecting; spam;campaigns	57593	social;information; network;egocentric
57006	social;socialtrust; spammers;information;communities	252892	social;information; media; communication

Table 5: Other Terms for Top 10 of Authors of Subgraph ‘social’

Subgraph weighted without term context		Subgraph weighted with term context	
ID of author	**Other terms**	**ID of author**	**Other terms**
49669	sensor; time; face; composite; predicting; mobile; apps; installation; dynamics	56679	modeling; mining; dynamic; spatio; temporal; discovery; link; formation; visualizing; semantic
57581	based; analysis; ranking; temporal; scientific; document; sndocrank; video; snakdd; mining; report	57835	influence; mining; analysis; learning; propagation; data; perspective
28700	multi; layered; based; analysis; extraction; group; evolution; online; label; dependent; feature; node; method; complex; discovery; email	49669	sensor; time; face; composite; predicting; mobile; apps; installation
690549	incentive; rewarding; services; analysis; communication	28700	multi; layered; based; analysis; extraction; group; evolution; online; label; dependent; feature; node; method; complex; discovery; email
119660		260831	privacy; sites; online; calculus; germany; empirical; usa; trust; study
252705		96113	online
735639	people; recommendation; collaborative; filtering	225725	
698479		181414	
357960		38231	predicting; problem; directed; closure; process; hybrid; analysis
333282		57593	

Table 6: Other Terms for Top 10 of Authors of Subgraph ‘social network’

5. DISCUSSION AND CONCLUSION

Our experiments proved that there is benefit in creating subgraphs based on the context, specified by a topic (term set). It is also obvious that by using the context, we retrieve other terms to expand the topic which enables new ways of searching for other articles or capable authors.

In our future work, we want to implement the outlined approach into the recommended system described by the Figure 9 in order to make the processes as automated as possible. Let’s assume that the results presented in this article can be further implemented into the modelling instruments to provide them with the graph evaluated in a different way based on which they can make predictions.

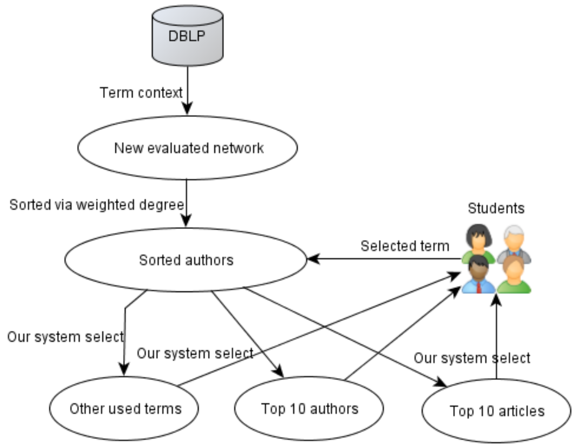


Figure 9: Description of Proposed Approach

ACKNOWLEDGMENTS

This work was supported by the SGS, VŠB–Technical University of Ostrava, Czech Republic, under the grant No. SP2013/167 ‘Analysis of Users' Behaviour in Complex Networks’ and by grant of Silesian University no. SGS/6/2013 ‘Advanced Modeling and Simulation of Economic Systems’.

REFERENCES

Babskova, A., Drazdilova, P., Martinovic, J., Svaton, V., Snasel, V., 2013. Evolution of co-authors communities formed by terms on DBLP. In *Proceedings of the Databases, Texts, Specifications and Objects 2013,* pp. 109-118.

Biryukov, M. and Dong, C., 2010. Analysis of computer science communities based on DBLP. *Research and Advanced Technology for Digital Libraries*, pp. 6273-6279.

Brodka, P., Saganowski, S., Kazienko, P., 2011. Group Evolution Discovery in Social Networks. *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp.247-253.

Chen, K.-Y., Luesukprasert, L., Chou, S.-C.T., 2007. Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling. *IEEE Transactions on Knowledge and Data Engineering*, 19(8), pp.1016-1025.

Deng, H., King, I., Lyu, M.R., 2008. Formal models for expert finding on DBLP bibliography data. *Eighth IEEE International Conference on Data Mining*, pp. 163–172.

Deza, M. M. and Deza, E., 2006. Dictionary of Distances. Elsevier Science, Amsterdam, The Netherlands.

Ding, Y., 2011. Scientific collaboration and endorsement: Network analysis of co-authorship and citation networks. *Journal of informetrics*, pp. 187-203.

Drazdilova, P., Martinovic, J., Slaninova, K., 2013. Spectral Clustering: Left-Right-Oscillate algorithm for detecting communities. In *Lecture Notes in Computer Science: Computer Information Systems and Industrial Management,* pp. 278-289.

Minks, S., Martinovic, J., Drazdilova, P., Slaninova, K., 2011. Author cooperation based on terms of article titles from DBLP. *Advances in Intelligent Systems and Computing*, pp. 281–290.

Monachesi, P., Lemnitzer, L., Simov, K., 2006. Language technology for eLearaing. In *Lecture Notes in Computer Science including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*. pp. 667-672.

Newman, M.E.J., 2004. Analysis of weighted networks. *Physical Review E*, 70(5), p.9.

Patil, A., Liu, J., Gao, J., 2013. Predicting group stability in online social networks. In *Proceedings of the 22nd international conference on World Wide Web* (WWW '13). International World Wide Web Conferences Steering Committee, pp. 1021-1030.

Schirru, R., Baumann, S., Memmel, M., Dengel, A., 2010. Extraction of Contextualized User Interest Profiles in Social Sharing Platforms. *Journal Of Universal Computer Science*, pp.2196-2213.

Sun, Y., Barber, R., Gupta, M., Aggarwal,Ch. C., Han, J., 2011. Co-author Relationship Prediction in Heterogeneous Bibliographic Networks. *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp.121-128.

Wang, X., Mccallum, A., Wei, X., 2007. Topical n-grams: Phrase and topic discovery, with an application to information retrieval. *Seventh IEEE International Conference on Data Mining ICDM 2007*.

Yang, C.C. and Ng, T.D., 2008. Analyzing content development and visualizing social interactions in web forum. In *IEEE International Conference on Intelligence and Security Informatics 2008 IEEE ISI 2008*. pp. 25-30.

Zhang, J., Ackerman, M.S., Adamic, L., 2007. Expertise Networks in Online Communities: Structure and Algorithms. W. W. W. C. Committee, ed. *Forum American Bar Association*, pp. 221-230.

Zhang, J., Tang, J., Li, J., 2007. Expert Finding in a Social Network. *Network*, LNCS 4443, pp.1066-1069.

OPTIMAL DESIGN, BASED ON SIMULATION, OF AN OLIVE OIL MILL

Juan-Ignacio Latorre-Biel[(a)], Emilio Jiménez-Macías[(b)], Julio Blanco-Fernández[(c)], Juan Carlos Sáenz-Díez[(d)]

[(a)] Public University of Navarre. Deptartment of Mechanical Engineering, Energetics and Materials.
Campus of Tudela, Spain
[(b and d)] University of La Rioja. Industrial Engineering Technical School.
Department of Electrical Engineering. Logroño, Spain
[(c)] University of La Rioja. Industrial Engineering Technical School.
Department of Mechanical Engineering. Logroño, Spain

[(a)]juanignacio.latorre@unavarra.es, [(b)]emilio.jimenez@unirioja.es,
[(c)]julio.blanco@unirioja.es, [(d)]juan-carlos.saenz-diez@unirioja.es

ABSTRACT

Global concurrence is a topic that affects many companies of most of the sectors of the economy. In particular, the improvement in the manufacturing, packing, storage, and transportation of food has allowed farming companies from all over the world to compete for customers of a global market. In order to achieve success in this complex environment it is convenient for the companies to be efficient even before the beginning of their business activity. This paper presents a decision support methodology for improving the design and management of an olive oil manufacturing facility based in the development of a Petri net model of the system, the simulation of its behavior under a selected set of alternative configurations and the choice of the most promising one by means of an optimization algorithm.

Keywords: Petri nets, farming company, alternatives aggregation Petri nets, simulation, olive oil.

1. INTRODUCTION

Global market provides farmers with excellent opportunities for increasing the potential customers of their products. However, the present customers of a given agricultural company can also decide to buy the farming products abroad.

For this reason, it is very convenient for farmers to improve their competitiveness by increasing the quality of their products, and by reducing the cost of their production, while improving their yield.

Key factors to achieve these objectives are the optimal design and management of a farm.

This issue has been researched by different authors, such as (Shikanai *et al.* 2008), where by simulating the operation of small sugar cane producers, an information system is developed to assist the management of agricultural companies.

The modelling and simulation of farming processes to produce atlantic salmon fish farm is discussed by (Melberg and Davidrajuh 2009), showing the suitability of the Petri nets formalism for this purpose.

The scheduling of farm work is analysed by (Guan *et al.* 2010) using a methodology based in the development of a model of the system and the application of an optimization process. Again, the formalism is chosen among the paradigm of the Petri nets. The optimization process is based in the use of the metaheuristic of the genetic algorithms to perform a search in the solution space.

(Cicirelli *et al.* 2010) explores the application of models based on the Petri nets paradigm to the stages of the production process of wine.

The specific issue of crop protection is discussed by (Léger *et al.* 2011) aiming the reduction in the use of chemical pesticides. A decision workflow tool for crop protection in the production of wheat is described.

The formalism of the Petri nets is applied by (Wang *et al.* 2011) for the management of the production of food processes, aimed at obtaining a certification of pollution-free agricultural products.

The development of a decision support system is described in (Latorre et al., 2012) and refined in (Latorre et al., 2013) with an application in the wine-making industry. A Petri net model of a winery in process of being designed is developed with the formalism of the alternatives aggregation Petri nets with the purpose of supporting the design of a winery, as well as its future management.

In particular, the production of olive oil has not received so much attention by the research community than other agricultural sectors. This fact, together with the need of developing tools for the efficient decision support of farmers and producers of olive oil has motivated the present research.

In this paper, a description of the development of a decision support system for the optimal design and management of an olive oil manufacturing company has been presented. This system is based in obtaining a model of the system by using the formalism of the alternatives agregation Petri nets (Latorre et al., 2011) and simulating it under different configurations of the system to select the configuration that bettr fits with the objective of the design and management of the company.

The rest of the paper is organized as follows. Section 2 describes the generalities on the manufacturing of olive oil. The formalism considered to develop the model of the system, the paradigm of the Petri nets, is introduced in section 3. In section 4, the model of the olive oil manufacturing facility in process of being designed is presented and explained. Section 5 deals with the decision-making methodology that can be constructed by the integration of the model of the system, a simulation tool and an optimization algorithm. The following section presents the conclusions and future research lines, while the last section lists the bibliographical references.

2. PRODUCTION OF OLIVE OIL

The olive oil is an important ingredient of the Mediterranean diet, being one of the healthiest vegetable oil from a nutritional point of view. It provides with vitamin A. D, and E, equilibrates the cholesterol, contributes to the development of the bones, as well as of the brain, and the nervous system, among other advantages.

The main producers of olive oil belong to the Mediterranean Basin, being Spain the largest world producer (43% of the world production) and Italy the second (14%) (IOC, 2011).

The production of olive oil began around 5000 years ago in Greece. Since then, a discontinuous method of extraction has been applied for obtaining a product of good quality but in small quantities.

Nowadays, the industrial manufacturing of olive oil is performed usually in a continuous process, which allows obtaining olive oil of different qualities, along with several other products able for feeding cattle, or as combustible in biomass furnaces.

The market of the products obtained from the olive is large and diverse as it is the variety of products themselves. The olive oil of the finest quality should present specific organoleptic characteristics and composition; hence, the production should be performed with a maximum care and precision. Moreover, different qualities of olive oil can be produced for different tastes and with diverse prices.

The three grades of olive oil are extra virgin, virgin, and olive oil, which can be divided into subcategories, such as "premium extra virgin" and "extra virgin" for the first grade, "fine virgin," "virgin," and "semifine virgin" for the second one, and "pure olive oil" and "refined oil" for the third one.

This paper deals mainly with the production of olive oil of the highest quality, for other inferior classes of olive oil, additional stages for processing the pomace and waste waters should be considered.

The process for obtaining olive oil consists basically in the separation of the oil from the rest of the fruit, and it is composed of several stages. There are variations in the manufacturing stages, according to the type of process chosen for producing the olive oil. Nevertheless, the main stages can be found in the following list:

a) Harvesting. This stage begins at the beginning of November and finishes at the end of February in the northern hemisphere. Depending on the month of harvesting the organoleptic properties of the oil may vary.

b) Leaf-removal. This is the first phase of cleaning up the olives.

c) Washing. This stage removes dust, soil, insecticide, and others by the use of a washing machine. The duration of this stage depends on the degree of dirtiness of the olives.

d) Weighing. A scale permits to control the weight of the olives before the production process begins.

e) Storage and measure out. The quantity of olives required by the following stage is provided, the rest of the clean olives are stored in hoppers.

f) Grinding. This stage produces a paste from the original olives.

g) Malaxation. The olive paste is stirred for mixing the small droplets of oil and producing the separation of the water phase and the oil phase.

h) Pumping. The product is pumped into the following machines.

i) Decantation. A screw conveyor introduces the product into the decanter, where by centrifugation three products are obtained: pomace (fragments of the pits of the olives), oil, and vegetation water (water and vegetal substances).

j) Filtering. The solid phase is separated from the liquid one. In the oil, the solid phase is the pulp, whereas in the vegetation water, the solid phase are small fragments of pits called pomace.

k) Centrifugation. A vertical centrifuge removes the impurities from the oil, as well as the last remnants of water.

l) Centrifugation of the vegetation water. The same process applied to the oil is performed on the vegetation water, in order to obtain a fat-free waste water, as well as dirty oil.

m) Storage. The obtained oil is stored until its expedition to the sellers.

n) Drying. The waste water and the pomace may be mixed and dryed for obtaining other useful products from the olives, such as pulp for fodder and pits for combustible.

o) Processing the oil. This processing permits obtaining oil of different qualities for diverse uses and customers and with a diversity of prices.

Several manufacturing processes can be chosen for producing virgin olive oil:

1) Traditional method. This methodology is appropriate for obtaining small quantities of high quality oil. It is a discontinuous process and used by a few small producers.

2) 3-phase decanter centrifugation. In this case, the first separation of the constituents of the olive paste (obtained after crushing and malaxing the raw olives) is performed in a horizontal decanter with three outputs for pomace, vegetation water, and oil.
3) 2-phase decanter centrifugation. The horizontal decanter has only two outputs, in this case for wet pomace and olive oil.
4) 2½-phase decanter centrifugation. The decanter requires a smaller amount of input water, hence the moisture content of the output pomace is not larger than 50%, and the amount of vegetation water is reduced.
5) Sinolea method. In this methodology, the first stage for the extraction of the olive oil from the olive paste is performed by the use of stainless steel discs, where the oil adheres. Scrapers can remove the oil from the discs. The rejected paste can be processed by horizontal decanter centrifugation for obtaining more oil.

The decision making on the design of an olive oil manufacturing facility implies the definition of the type of manufacturing process, as well as its capacity, including the number of machines to be installed in every stage of the process.

Furthermore, the decision making on the management of an olive oil mill, is related to deciding the time and amount of olives to be supplied to the mill, as well as the production rate, the temperature and amount of water to be added to the process, etc.

In the following sections, these decisions will be discussed in the frame of an automatic decision supply system.

3. THE PARADIGM OF THE PETRI NETS

The decision support system to be described in this paper is based in the use of a simulation model. This model has to be represented by an appropriate formalism.

The Petri net paradigm is suited as modeling formalism, as well as for the analysis, simulation, and optimization of discrete event systems.

The Petri nets can be applied by means of a double representation. Their graphical representation shows, in an intuitive and explicit way, complex behaviors, such as combinations of concurrence, synchronization, and competition for limited resources.

Furthermore, a second feasible representation of a Petri net model can be obtained by means of a matrix-based description, which can be obtained directly form the graphical representation and vice-versa.

The matrix-based representation allows the development of a structural analysis of the model, as well as to the performance evaluation required to solve the decision making processes that will be used in this paper to develop a decision support tool.

Definition 1. A Petri net system is a 5-tuple $R = (P, T, \text{pre}, \text{post}, \mathbf{m}_0)$ such that:

i) P is a non-empty set of places.
ii) T is a non-empty set of transitions and $P \cap T = \emptyset$.
iii) pre and post are functions that associate a weight to the directed arcs between the elements of the sets P and T, in the following way:
iv) pre: $P \times T \rightarrow \mathbb{N}^*$ and post: $T \times P \rightarrow \mathbb{N}^*$, where $\mathbb{N}^*$ is the set of natural numbers, excluding zero.
v) $\mathbf{m}_0$ is the initial marking, such that $\mathbf{m}_0$: $P \rightarrow \mathbb{N}^*$.

□

A discrete event system in process of being designed contains a set of freedom degrees. As the decisions are made, the freedom degrees in the system are converted into specific numerical values.

The freedom degrees of the system can be represented in the model of the system by means of controllable parameters. These controllable parameters can play different roles in the model of the system. For example they can be marking, structural, or transition-firing parameters.

A marking parameter can represent a freedom degree related to the number of resources, such as number of crushers, industrial decanters, or pumps. Moreover, a structural freedom degree may represent the feasible decision of choosing a methodology of olive oil manufacturing, such as traditional or 2-phase decanter centrifugation.

The definition 1 does not specify explicitly the mechanisms of the Petri net formalism to represent the controllable parameters, which have a significant influence in the behaviour and performance of the system, and the possibility of making a choice for them among their feasible set of combinations of values.

The design process of a system, modelled by the formalism of the Petri nets, usually requires the development of so many models as alternative structures can be selected for representing different solutions for the structure of the system. These models can be called alternative Petri nets (Latorre *et al.* 2011).

Nevertheless, this approach uses to be ineffective, since the whole set of alternative Petri nets usually shows large amounts of redundant information among the models (Recalde *et al.*, 2004*)*. This redundant information corresponds to shared subsystems.

This paper presents a methodology to develop a decision support system based on a formalism belonging to the paradigm of the Petri nets, called the alternatives aggregation Petri nets. This formalism has the ability of describing in a single model, a complete set of alternative Petri nets removing the redundant information, which correspond to shared subsystems.

The mechanism for the decision making associated to the choice of one of these alternative structural configurations is represented explicitly by means of the so called choice variables.

A definition of a set of choice variables for a certain Petri net model can be given once it is known the number of the alternative structural configurations for the system to be modelled.

Definition 2. Given a discrete event system with *n* alternative structural configurations, a set of choice variables can be defined as $S_A = \{ a_i$ Boolean $\mid \exists!\ k \in \mathbb{N}^*, k \leq n$, such that $a_k = 1 \wedge \forall j \in \mathbb{N}^*, j \leq n, j \neq k$ it is verified $a_j = 0 \}$, where $|S_A| = n$, and the assignment $a_k = 1$ is the result of a decision.

□

Based on the previous considerations, it is possible to state a definition for the formalism to be used in the decision support tool presented in this paper: the alternatives aggregation Petri nets.

Definition 3. An alternatives aggregation Petri net can be defined as a 10 tuple $R^A = (P, T$, pre, post, $\mathbf{m}_0$, S_α, $S_{valnstr\alpha}$, S_A , f_A, $R_{val\gamma}$), where

i) S_α is a set of undefined parameters.
ii) $S_{valnstr\alpha}$ is the set of feasible combination of values for the undefined parameters in S_α.
iii) S_A is a set of choice variables, $S_A \neq \emptyset$ and $|S_A| = n$.
iv) f_A: $T \rightarrow f(a_1, \ldots, a_n)$ is a function that assigns a function of the choice variables to each transition t such that type[$f_A(t)$] = boolean.
v) $R_{val\gamma}$ is a binary relation between $S_{val\gamma}$ and R_A.

On the other hand, f_A: $S_A \rightarrow T$, assigns a choice variable to a single or several transitions of the Petri net, and if $S_A' = \{ a_1 , a_2 , \ldots , a_k \}$ is the set of every choice variables associated to the transition t, then the guard function of the mentioned transition is $g_A(t) = a_1 + a_2 + \ldots + a_k$.

□

This formalism, can be used to develop the model of an olive oil mill in process of being designed, that is to say before a decision on the methodology for manufacturing the olive oil is made.

4. MODEL OF AN OLIVE OIL MILL

The process of constructing a Petri net model of a system, in this case an olive oil mill, can be tackled by means of different approaches. These approaches can be classified into two main methodologies, broadly used in the field of modeling discrete even systems by means of Petri nets (Silva, 1993).

The first one is called top-bottom modelling and consists of defining a simple and global model with a reduced level of detail. In a second stage, the subsystems present in the model are expanded to incorporate a larger amount of information.

The second methodology is the bottom-up modelling, which requires the development of independent models for the subsystems of the discrete event system and the subsequent link of the different submodels with the purpose of obtaining the final and complete model.

In the model presented in this paper, a bottom-up modelling has been considered, in several stages.

A first stage in the modelling has been carried out by constructing the models of the different methodologies for manufacturing olive oil. As it has been seen in section 2, a number of five different methodologies have been taken into account.

However, as a result of this phase of modelling, more than five Petri net models have been constructed, since there are different decisions, called structural decisions, which may lead to different configurations of the models that correspond to a given manufacturing technology.

Before entering into detail, it is convenient to take into account the structural decisions that have been considered in the model of the olive oil mill in process of being designed:

d_1. Traditional method.
d_2. Configuration alternative to the traditional method for obtaining high quality olive oil.
d_3. Decantation for separating the oil and the vegetation water.
d_4. Vertical centrifuge for separating the oil and the vegetation water.
d_5. Hammer crusher.
d_6. Disc crusher.
d_7. Depitting machine.
d_8. Knife crusher.
d_9. Three phase decanter centrifugator.
d_{10}. Two phase decanter centrifugator.
d_{11}. Sinolea method.
d_{12}. Two and a half phase decanter centrifugator.

As it has been seen, a sixth methodology for manufacturing olive oil can be considered. It is an alternative configuration to the traditional method. Where the last stages of the process are substituted by the use of modern separation technologies.

Moreover, some of the manufacturing methodologies can be implemented by using four different types of crushers and the sinolea method requires the addition of a horizontal decanter characteristic of others of the production methods.

As a result of these structural decisions, it is possible to define a set of choice variables, one for every alternative system that can be built up from the mentioned decisions.

The size of the set of choice variables S_A can be calculated in the following way:

$|S_A| = 1 + 6 + 24 = 31$.

In the previous calculation, the addend "1" comes from the traditional method, the value "6" of the six configurations for the alternative method, which can be complemented with other manufacturing technologies. Finally, the value "24" is related to the other 4 methodologies with decanter centrifugators, including the sinolea method, since every one of them can be implemented with different crushers.

In fact, easily, it is possible to develop more alternative configurations, which will not complicate largely the decision making process thanks to the use of the formalism of the alternatives aggregation Petri nets to construct the model of the system.

In addition of the previous considerations, it is possible to take into account other controllable parameters, such as the marking parameters that are listed below, in this same section.

The consideration of both the choice variables and the marking parameters increases largely the number of feasible solutions for the construction of an efficient olive oil mill based in the decision support tool.

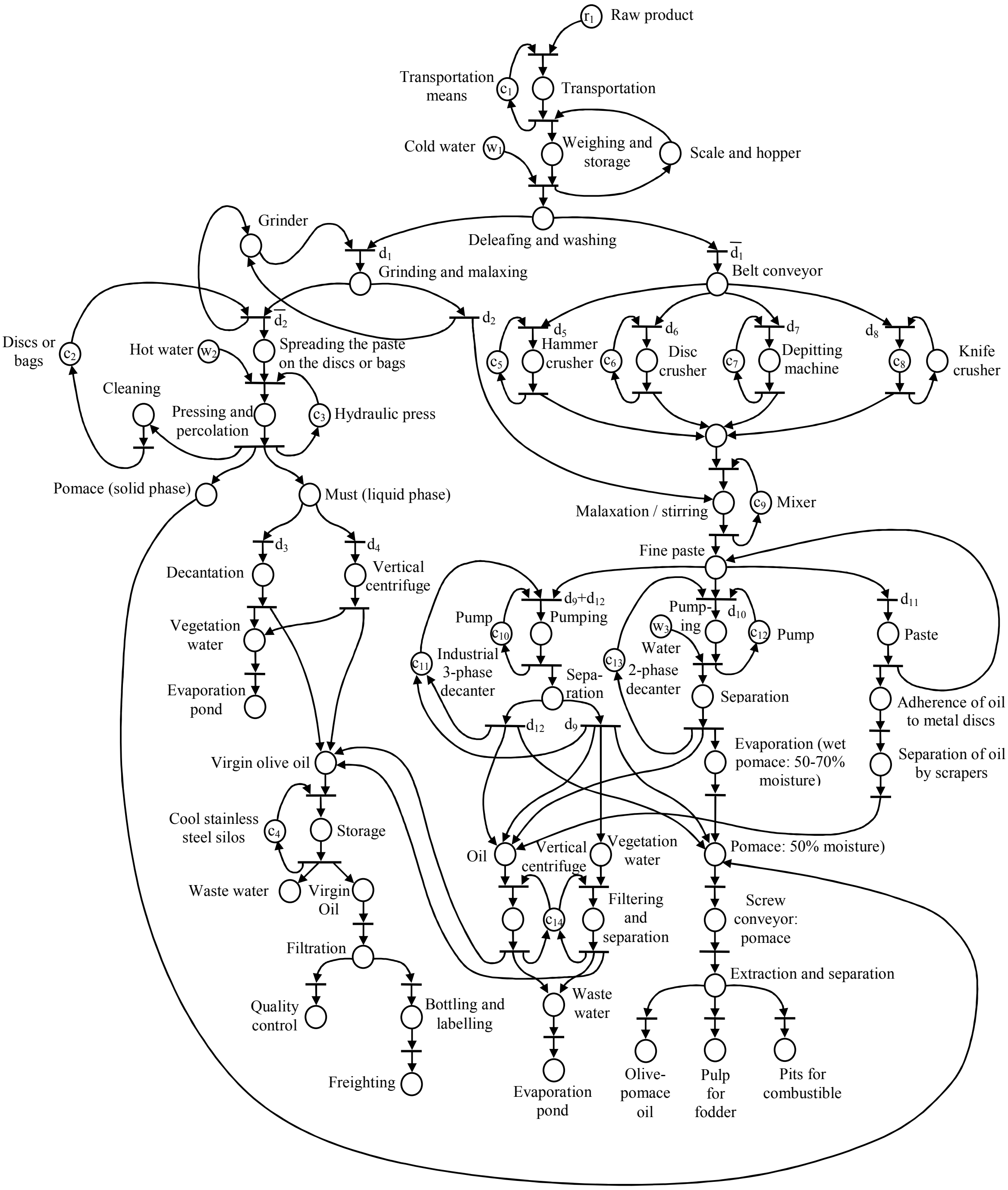


Figure 1. Model of an olive oil mill based on the formalism of the Petri nets

- **c_1**. Number of available lorries, or equivalent transportation means, for supplying the raw olives to the olive oil mill.
- **c_2**. Number of discs or bags for pressing the olive paste in the traditional method.
- **c_3**. Number of hydraulic presses available in the traditional method.
- **c_4**. Number of cool stainless steel silos for storing the resulting virgin olive oil.
- **c_5**. Number of hammer crushers.
- **c_6**. Number of disc crushers.
- **c_7**. Number of depitting machines.
- **c_8**. Number of knife crushers.
- **c_9**. Number of mixers for malaxation.
- **c_{10}**. Number of pumps for conveying the olive paste to the horizontal decanter (3-phase decanter centrifugator).
- **c_{11}**. Number of 3-phase decanter centrifugators.

c_{12}. Number of pumps for conveying the olive paste to the horizontal decanter (2-phase decanter centrifugator).

c_{13}. Number of 2-phase decanter centrifugators.

c_{14}. Number of vertical centrifuges.

Depending on the feasible value considered for the mentioned marking parameters, the size of the solution space may increase considerably, increasing subsequently the optimization time required to make a decision or, in the same way, reducing the quality of the best solution found in the search.

5. DECISION MAKING METHODOLOGY

A methodology for designing a facility for producing olive oil is presented in this section. This methodology requires the development of a Petri net model of the system to be designed, which has been depicted in figure 1.

The methodology is based in the statement of an optimization problem. The goals of the optimization are quantified in the so called objective function and the model is simulated under different promising scenarios with the purpose of choosing the best one.

This best solution will be provided to the designer in order to support the decisions to be made in the process with a forecast of the behavior of the system, thanks to the described optimization methodology (Latorre et al., 2010).

The choice of the best scenarios to be considered in the optimization process may be performed by means of a metaheuristic, such as the genetic algorithms, due to the large amount of feasible configurations that an industrial olive oil meal can have.

As a consequence of all these configurations, the exhaustive exploration of the complete set of feasible solutions is not practical; hence a non-deterministic methodology to guide the search is chosen.

6. CONCLUSIONS

As a conclusion, it can be stated that the application of a mathematical methodology for designing an olive oil mill, based on the simulation of a model of the system, can lead to a useful tool for the support of the decisions performed during the different stages of the design and the management of an olive oil mill.

The future research effort will be focused in the application of the presented methodology to a large pool of practical cases to improve the decision support tool.

REFERENCES

Cicirelli F.D., Furfaro A. and Nigro L., 2010. A Service-Based Architecture for Dynamically Reconfigurable Workflows. Journal of Systems and Software , Vol. 83, n. 7, pp. 1148-1164.

Guan, S., Nakamura, M., and Shikanai, T. 2010. Hybrid Petri Nets and Metaheuristic Approach to Farm Work Scheduling. In Aized, T. (Ed.) Advances in Petri Net Theory and Applications. Chapter 8. Sciyo.

IOC, 2011. *2011/2012 Forecast Reports*. The International Olive Council (IOC).

Latorre, J.I., Jiménez, E., Blanco, J., Sáenz, J.C. 2013. Integrated Methodology for Efficient Decision Support in the Rioja Wine Production Sector. International Journal of Food Engineering. (In press).

Latorre, J.I., Jiménez, E., Blanco, J., Sáenz, J.C. 2012. Decision making in the Rioja wine production sector. In Proceedings of the 24nd European Modelling and Simulation Symposium (EMSS 12). Vienna, Austria, pp. 452-457.

Latorre, J.I., Jiménez, E., Pérez, M., 2011. Petri nets with exclusive entities for decision making. International Journal of Simulation and Process Modeling, Special Issue on the I3M 2011 Multiconference.

Latorre, J.I., Jiménez, E., Pérez, M., 2010. On the Solution of Optimization Problems with Disjunctive Constraints Based on Petri Nets. In Proceedings of the 22nd European Modelling and Simulation Symposium (EMSS 10). Fez, Moroco, pp. 259-264.

Léger, B., Naud, O., Gouache, D., 2011. Specifying a strategy for deciding tactical adjustment of crop protection using CPN tools. Congress of the European Federation for Information Technology in Agriculture, Food and the Environment Efita 2011. Pages 1 – 6.

Melberg, R., Davidrajuh, R. 2009. Modeling Atlantic salmon fish farming industry. In Proceedings of the IEEE International Conference on Industrial Technology. ICIT 2009. Pages 1-6.

Recalde, L.; Silva, M.; Ezpeleta, J.; and Teruel, E. 2004. Petri Nets and Manufacturing Systems: An Examples-Driven Tour. In Desel, J.; Reisig, W.; and Rozenberg, G. (Eds.), Lectures on Concurrency and Petri Nets: Advances in Petri Nets, Lecture Notes in Computer Science / Springer-Verlag. Volume 3098, pp. 742-788.

Shikanai, T.; Nakamura, M.; Guan S.; Tamaki, M. 2008. Supporting system for management of agricultural corporation of sugarcane farming in Okinawa Islands. World conference on agricultural information and IT, IAALD AFITA WCCA 2008, Tokyo University of Agriculture, Tokyo, Japan, 24 - 27, pp. 1101-110.

Silva, M. 1993. Introducing Petri nets. In Practice of Petri Nets in Manufacturing, Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall.

Wang, F., Duan, Q., Zhang, L., and Li,G. 2011. Modeling and Analysis of Pollution-Free Agricultural Regulatory Based on Petri-Net. Computer and Computing Technologies in Agriculture IV. IFIP Advances in Information and Communication Technology, Volume 347/2011, 691-700.Agutter, A.J., 1995. *The linguistic significance of current British slang*. Thesis (PhD). Edinburgh University.

MANAGEMENT OF RESOURCES AND WASTES IN A NETWORK OF SCHOOLS MODELLED BY PETRI NETS

María Otero-Prego[(a)], Juan-Ignacio Latorre-Biel[(b)], Diego Azofra-Rojo[(c)], Emilio Jiménez-Macías[(d)]

[(a,b,c,d)] Industrial Engineering Technical School.
University of La Rioja. Logroño, Spain.
[(a,b,c)] Department of Mechanical Engineering.
[(d)] Department of Electrical Engineering.

[(a)] mariaoprego@googlemail.com, [(b)] juan-ignacio.latorre@unirioja.es,
[(c)] azofra_diego@hotmail.com, [(d)] emilio.jimenez@unirioja.es

ABSTRACT

The management of public resources, such as educative institutions is usually associated to making decisions affecting to a large number of schools, in changing environments, and with large amount of information. In order to ease the management of these complex systems, this paper describes a methodology for modeling a network of schools, simulating the model under a given configuration, and selecting the best decisions for assisting the educative authorities in their managing responsibilities.

Keywords: education, Petri nets, optimization, decision support system.

1. INTRODUCTION

The management of public resources is usually performed under the influence or approach of crisis. Environmental and economic crisis might require the educative authorities to make a special consideration for maximizing the social and educative benefits obtained from invested public resources, as well as minimizing the impact the educative activities might have in the environment.

Giving autonomy of decisions to the educative institutions or centralizing their management is an issue of debate and research, where there still remains a large number of open questions. Among the advantages of autonomous decision making can be found the swiftness in responding to non-expected situations, the possibility to negotiate directly to local suppliers, the detailed knowledge by the managing board of a school of their students, teachers, suppliers, and specificities of the school and its social and economic environment.

Nevertheless, a centralized management of educative centers in a situation of crisis seems to be especially suitable due to the convenience of sharing resources and services. Among the advantages of a centralized management it can be mentioned the economy of scale when negotiating large volumes of products or services or the better utilization rate of resources such as public transport, or communication networks.

A step further in the centralized management of schools can be the formation of clusters with schools of different educative levels, from basic school to secondary education and vocational training. In this case, more services and products can be shared, increasing the efficiency in their use, the reduction in their cost, the improvement in the quality and the satisfaction of members of the educative community with special needs, such as accessibility for handicapped people or special diets for allergic or intolerant students or members of the staff.

Clustering different educative levels into “campuses” may lead to a very efficient use of resources, as well as to better services to society, including not only education but also transportation, counseling, accessibility, catering, long-life learning, or leisure, just to give a few examples.

More examples, this time regarding resources that can be shared are the administrative staff and offices, central heating, catering service, waste management and recycling, communication networks and services, or even central heating.

However, the more centralized the management, the more complex the decision making becomes. This fact arises as a consequence of an increase on the information, as well as on the variability of the schools, which determine that a given decision may derive to different outcomes, depending on the educative level, the location of the school, the social background of the members of the educative community, etc.

As a consequence of the previous considerations, the centralization of the management of educative institutions implies the need of a decision support system for the assistance of the managers in the decision making.

In this paper, the development of a decision support system for the management of a network of schools is discussed.

The following section deals with the characteristics of such a tool, designed for helping in the process of decision making. Section 3 describes briefly the educative institutions, whose assisted management is the objective of the present research. The formalism chosen for developing the model of a network of schools, the Petri nets, is presented in section 4, while a model of a school is shown in the following section. In the same section 5, it is detailed the modeling of the decision making process in a given school, which has been decomposed in three levels: operational, tactical, and strategic one.

The modeling approach bottom-up is used in section 6 for constructing a Petri net model of a network of schools. The following section describes the development of a decision support system, based in the model shown in the previous section. A section of conclusions and another one dedicated to the bibliographical references relevant to the paper conclude the present paper.

2. DECISION SUPPORT SYSTEM

A decision support system assists a human decision maker in complex and difficult decisions. The application field of this kind of tools is very broad and grows every year, ranging from medical diagnosis to manufacturing management.

There are diverse approaches for producing a decision support system. One of them, the one followed in the research presented in this paper, is the one based in the construction of a model of the system of interest (Swanepoel, 2004).

In the field of education, the use of decision support systems has been broadly used for the development of timetables and, in general, for allocating resources, such as classrooms, to students and teachers. Nevertheless, its use in the management of educative institutions is much more limited (Otero et al, 2012b).

As it has been mentioned in the previous section, the more educative institutions in the scope of the management board, the more complicated is the resulting administration of the required resources, the provided services, and the generated wastes.

The management of a network of schools requires dealing with a large number of actors and variables, which are difficult to take into account in an appropriate manner by classic and manual methodologies.

Techniques applied commonly, such as the ones based on spread sheets or even simulation not based on models of the network but on information gathered at the beginning of the academic year or even in the precedent year, may lack of realism and prompt reaction to new variables or non-expected situations, which might arise at any moment of the academic year (Otero et al., 2012a).

In order to overtake the limitations of the mentioned classic approaches for managing a network of schools, while earning the benefit of a centralized administration of resources, services, and wastes, it is possible to consider a methodology of optimization based on simulation, where the simulation is performed by using a Petri net model of the network of educative institutions.

This approach is not new, but has been applied to diverse sectors with success. For example (Jiménez et al, 2006) discusses the application of modeling and simulation in the industry, while (Tuncel, 2007) presents a scheduling heuristic rule that aims to allow choosing the best operating policy and system configuration for a flexible manufacturing system. Also in the manufacturing field (Mušič, 2009) discusses the Petri net based job-shop scheduling by means of a combination of dispatching rules with a local search guided by a metaheuristic.

Furthermore, (Latorre et al., 2012) applies the methodology proposed in this paper to the Rioja wine production sector. On the other hand, (Latorre et al., 2013) presents a refinement of this methodology for the design of a discrete event system, instead of its management.

The mentioned approach, based on the simulation of a model of the system, will provide the human decision makers with predictions for the time to come, which can be adapted on-the-run by adding new variables or modifying the existing ones. This powerful tool may provide with confidence to the decision makers, who will be able to test the outcomes of different decisions.

By means of the information obtained from the analysis of the decisions by simulation, it will be possible that the decision makers choose the most promising option, improving in this way the management of the educative institutions.

3. EDUCATIVE INSTITUTIONS

In Spain, regional authorities have competences in education. Furthermore, there is a department in every regional government devoted to the management of schools, including primary education, secondary education, and vocational training.

There is little autonomy for the educative centers. This practice implies that the approach presented in this paper has a very appropriate application at this regional level of management.

In fact, as it will be described in section 5, operational decisions are made at the level of the educative center, whereas tactical decisions are made usually by regional authorities or by the managing board of the schools with the approval of the regional authorities. The strategic decisions on the management of the educative centers are made by the national or regional authorities.

4. FORMALISM OF THE PETRI NETS

The choice of the Petri nets as the modeling formalism considered in this research has been made because they are especially suited to model and analyze discrete event systems showing parallel evolutions and whose behavior are characterized by concurrency, synchronization and resource sharing.

The behavior of a network of schools falls under this category. In fact, the different schools present parallel evolutions, sometimes with a small contact with other schools. However, they compete for the limited public resources provided by the government, such as money, teachers, spaces for teaching, or courses for teachers.

On the other hand, the graphic nature of the net allows them to be self-documented specifications, which can make easy the communication among designers and users (Silva, 1993).

A definition of a Petri net is presented in the following:

Definition. A Petri net system is a 5-tuple $R = (P, T, \text{pre}, \text{post}, \mathbf{m}_0)$ such that:

i) P is a non-empty set of places.
ii) T is a non-empty set of transitions and $P \cap T = \emptyset$.
iii) pre and post are functions that associate a weight to the directed arcs between the elements of the sets P and T, in the following way:
iv) pre: $P \times T \to \mathbb{N}^*$ and post: $T \times P \to \mathbb{N}^*$, where $\mathbb{N}^*$ is the set of natural numbers, excluding zero.
v) $\mathbf{m}_0$ is the initial marking, such that $\mathbf{m}_0$: $P \to \mathbb{N}^*$.

□

The marking of the Petri net is an essential element of the model. In fact, the structure of a Petri net is something static, while the behavior of the system can be described in terms of system state and its changes, which is modeled by defining a marking and the marking evolution rule.

A token is represented as a black dot in a place, essentially to indicate the fact that the condition described by that place is satisfied.

In the model of the network of schools that is presented in this paper every actor in the educative community will be represented by an individual token. The mentioned actors include every student, teacher, as well as the members of the managing staff, the administrative staff, and the maintenance staff of every school.

In order to assign at every moment the appropriate activity to every actor, it is convenient that the model of every actor presents some identification to personalize it. In this way, a token representing a teacher will not perform the same activities than a student. Analogously, a student of primary education will not perform the same activities than a student of vocational training.

In order to personalize the tokens representing students, teachers, and other staff, it is possible to assign to them some attributes. There is a special kind of Petri net that is suited to include the mentioned attributes. It is the colored Petri nets, a very well-known formalism, broadly used, and provided with powerful analysis and simulation tools.

A formal definition of a coloured Petri net is given by (Jensen and Kristensen, 2009):

Definition. Coloured Petri net.
A non-hierarchical coloured Petri net is a nine-tuple

$CPN = \langle P, T, F, \Sigma, V, \text{c}, \text{g}, \text{e}, \text{i}\rangle$, where:

1. P is a finite set of places.
2. T is a finite set of transitions T such that $P \cap T = \emptyset$.
3. $F \subseteq P \times T \cup T \times P$ is a set of directed arcs.
4. Σ is a finite set of non-empty colour sets.
5. V is a finite set of typed variables such that type[v] $\in \Sigma$ for all variables $v \in V$.
6. c : $P \to \Sigma$ is a colour set function that assigns a colour set to each place.
7. g : $T \to EXPR_V$ is a guard function that assigns a guard to each transition t such that type[g(t)] = Boolean.
8. e : $F \to EXPR_V$ is an arc expression function that assigns an arc expression to each arc a such that type[e(a)] = c(p)$_{MS}$, where p is the place connected to the arc a.
9. i : $P \to EXPR_\emptyset$ is an initialisation function that assigns an initialisation expression to each place p such that type[i(p)] = c(p)$_{MS}$.

□

Notice that MS stand for multiset over S. A multiset is an ordered pair (S,f) where S is a set and f:$S \to$N is a function, called the frequency or weight function. (Joshi, 1989).

5. MODEL OF THE DECISION-MAKING IN A SCHOOL

The development of a Petri net model for a network of schools can be performed following a variety of methodologies. There are two main groups of methodologies, which are called bottom-up and top-down (Silva, 1993).

The bottom-up approach begins with a detailed model of every subsystem and is followed by a stage, where all the submodels are linked together to obtain a detailed model of the complete system.

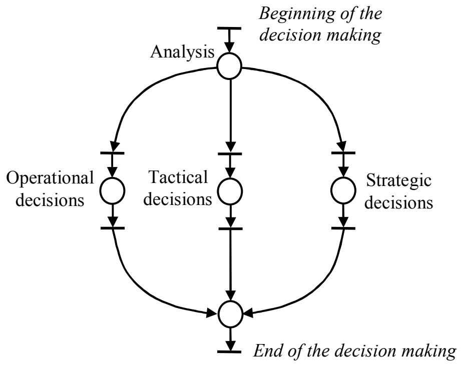


Figure 1: Decision-making by the managing board

On the other hand, the top-down methodology is implemented by constructing a low-detailed model of the complete system, which is refined and expanded by detailing different parts of it. This second approach is the one, which has been considered for the development of the present model. The low-detailed model of the network of schools is shown in the next section, as well as a general model for a school, which corresponds with an intermediate level of detail.

In this section, it has been represented detailed models of the decision making, which has been made correspond to the managing board of every school.

This approach does not prevent the simulation of centralized methodologies. On the contrary, it allows the simulation of both a centralized or a decentralized management for the educative centers.

In figure 1, it can be seen the model of the decision-making that correspond to the managing board. Considering that any set of decisions is made sequentially, a stage of analysis of the information required to make a decision is followed by either an operational decision, a tactical strategic one.

In figure 2, 3, and 4, the different levels of decisions have been modeled by means of Petri nets. In order to develop these models, it has been considered that operational decisions are those, whose influence is extended to a time window of weeks or few months. The tactical decisions are considered to range from few months to a year, while the strategic decisions range from a year to up to a decade.

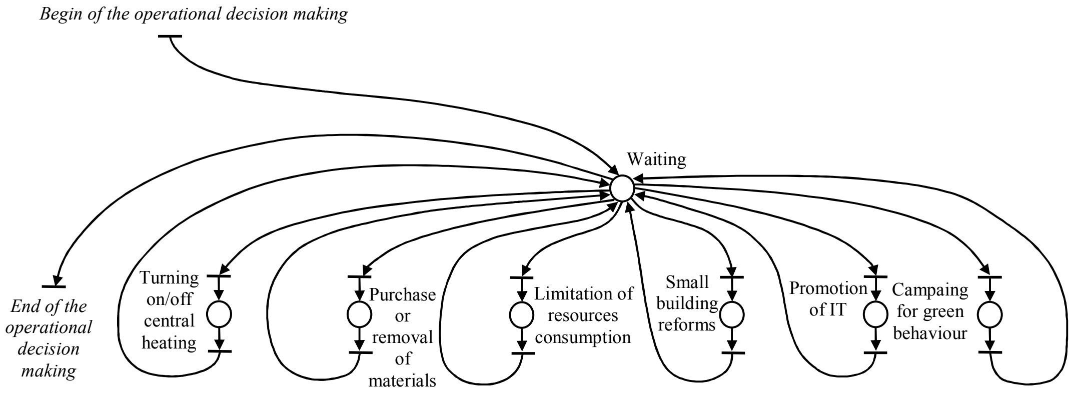

Begin of the operational decision making
Waiting
End of the operational decision making
Turning on/off central heating
Purchase or removal of materials
Limitation of resources consumption
Small building reforms
Promotion of IT
Campaing for green behaviour

Figure 2: Operational decision making by the managing board

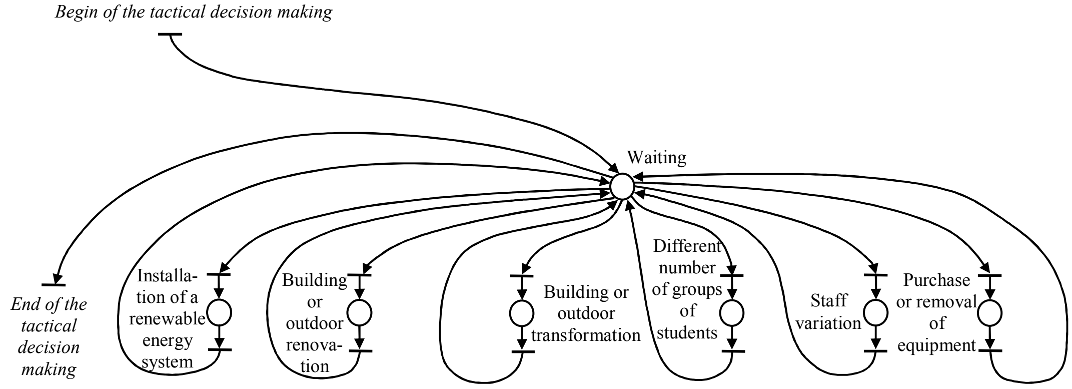

Begin of the tactical decision making
Waiting
End of the tactical decision making
Installa-tion of a renewable energy system
Building or outdoor renova-tion
Building or outdoor transformation
Different number of groups of students
Staff variation
Purchase or removal of equipment

Figure 3: Tactical decision making by the managing board

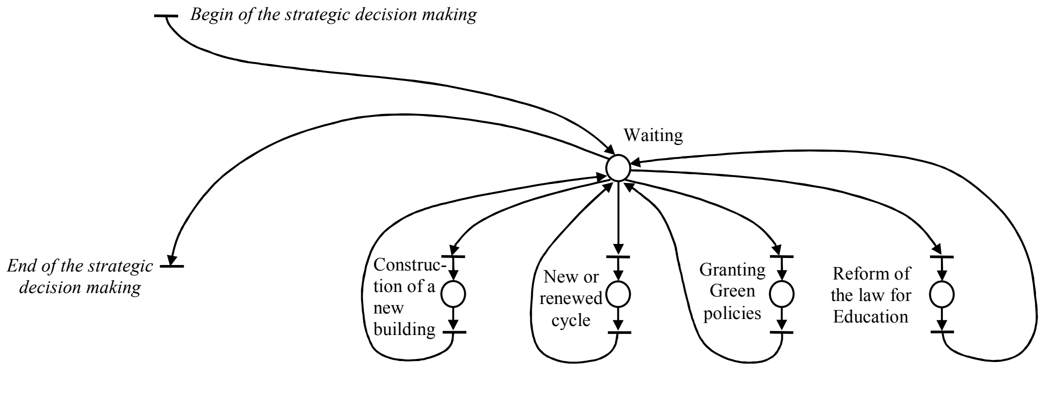

Begin of the strategic decision making
Waiting
End of the strategic decision making
Construc-tion of a new building
New or renewed cycle
Granting Green policies
Reform of the law for Education

Figure 4: Making strategic decisions by the educative authorities

6. MODEL OF A NETWORK OF SCHOOLS

This section discusses the model of a network of schools to be implemented in a simulation-based decision support system.

Figure 5 represents a model of an individual school, where the school day is divided into three groups of classes interrupted by two breaks.

During the class time, the different actors may perform diverse activities according to their role. This role is represented in the model by means of the attributes of the tokens of the marking. Every token represents an actor of the educative community (student, teacher, member of the managing staff, administration, or maintenance).

The model of the network of educative institutions should take into consideration the natural simultaneous evolution of the different schools, as well as the competition for limited resources or sharing common resources.

Moreover, in such a system, with discrete states distributed into the different educative centers, and with discrete number of actors, such as students, teachers, administrative staff, or external services, a very adequate formalism to develop an accurate model are the Petri nets, as it has been stated in section 4.

This formalism offers a very intuitive and easy to draw graphical representation, yet the underlying mathematical considerations permit a matrix-based representation, very appropriate to implement simulation and optimization algorithms to analysis the evolution of the model.

The development of a Petri net model of a complicated system, such as a network of educative institutions, has been undertaken by means of successive refinements from a low-detailed representation shown in figure 6. The model of the network of schools represents on the top, two places where the staff and students are hired and enrolled respectively, usually for a year. The rest of the model is composed by the schools belonging to the network.

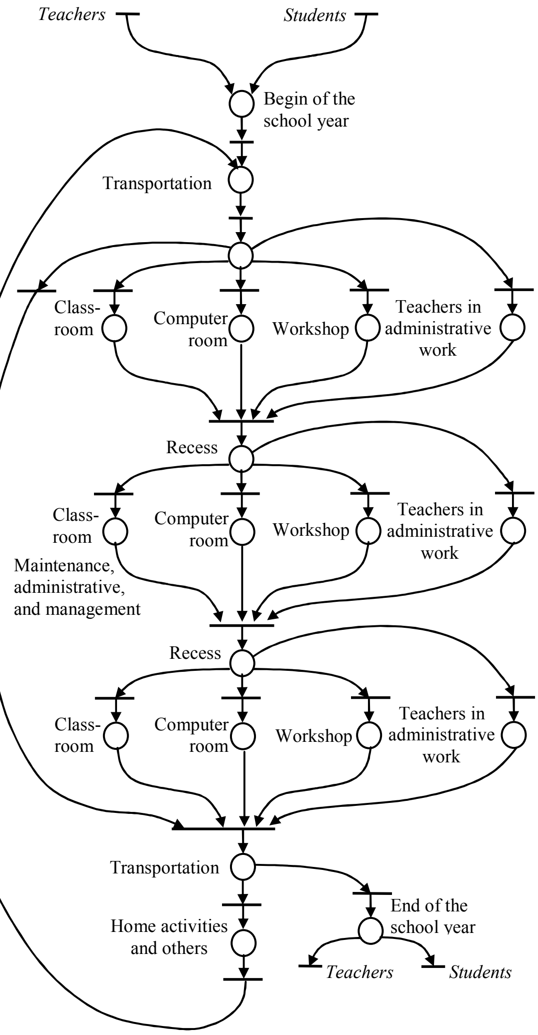


Figure 5: Petri net model of a school to be integrated in a network

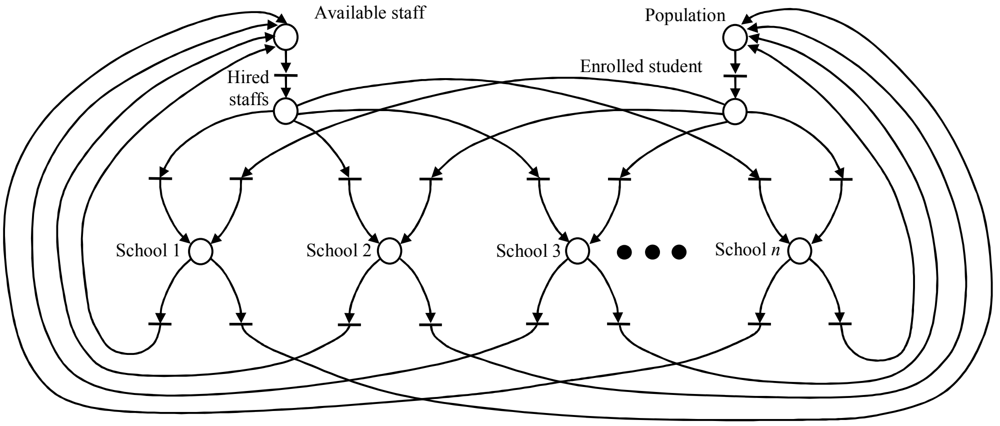


Figure 6: Network of *n* schools

7. OPTIMIZATION METHODOLOGY

Once the Petri net model of the network of educative institutions has been concluded, it is possible to use it for supporting the decisions made by the educative authorities. The methodology proposed to perform this activity is composed by the following stages:

a) Making a decision:

The model of the network of schools has freedom degrees, usually in the form of conflicts, which provide the system with flexibility and the administrative authorities with the possibility to make decisions. For this reason, it is necessary to choose a solution for every decision to be made in the period of the educative process to be analyzed.

The choice of the decisions to specify the freedom degrees of the model of the system can be performed by different ways, ranging from a random choice to the use of metaheuristics.

The selection of a method for choosing a solution for the problem of making decisions is of vital importance and will determine the efficiency of the decision support system, or even its effectiveness, since some problems do not allow an exhaustive search in the solution space of the problem in a reasonable amount of time.

b) Testing the decision:

The decisions made in the previous stage should be tested by simulation of the model to determine its suitability to solve the problem of management. As it can be seen, this approach is an application of the "what-if" analysis.

An important issue in this stage consists in the calculation of a parameter to represent with a numerical value the quality or suitability of the tested solution. In order to perform this task it is necessary to define the criteria that will determine the quality of a solution. In other words, the objectives of the educative authorities should be defined clearly and formally: social impact, invested resources, environmental impact, etc.

Once the objectives to be achieved by the management of the network of schools are clear it is necessary to quantify them, usually in the form of a multiobjective function. The calculation of this function should be done during the simulation. In this form, a tested solution can be "labeled" with a numerical value representing its quality. Steps (a) and (b) should be repeated iteratively.

c) Choosing the best decision:

From the pool of tested solutions, the one with the highest quality can be chosen as solution of the decision making of the educative authorities.

8. CONCLUSIONS

In this paper, a methodology of modeling and simulation of a network of schools has been discussed in order to construct a decision support system for assisting the educative authorities in their duties.

This methodology is based in the use of the paradigm of the Petri nets, which has been shown as a formalism very suitable to model educative institutions integrated in a network. The process of choosing the best decision is complemented by an iterative process of choosing feasible solutions and testing them to measure their quality. This methodology will provide with a quantitative tool for supporting the management of the educative institutions that may be very useful.

The next steps in the research will be to implement the tool and test it on a real network of schools.

REFERENCES

Jensen, K. and Kristensen, L.M. 2009. *Coloured Petri nets. Modelling and Validation of Concurrent Systems*. Springer.

Jiménez, E., Pérez, M., Latorre, I. 2006. Industrial applications of Petri nets: system modelling and simulation. *Proceedings of European Modelling Simulation Symposium*, pp. 159-164. Barcelona.

Joshi, K.D. 1989. *Fundations of discrete mathematics*. New Age International (P) Ltd, Publishers. New Delhi.

Latorre, J.I., Jiménez, E., Pérez, M. 2013. The optimization problem based on alternatives aggregation Petri nets as models for industrial discrete event systems. *Simulation: Transactions of the Society for Modeling and Simulation International. Special Issue M&S Optimization Applications in Industry and Engineering*. March 2013 vol. 89 no. 3, pp. 346-361.

Latorre, J.I., Jiménez, E., Blanco, J., and Sáenz, J. C. 2012. Decision making in the Rioja wine production sector. *Proceedings of the 24th European Modelling and Simulation Symposium.* pp. 452-457. Vienna, Sept. 2012.

Mušič, G. 2009. Petri net based scheduling approach combining dispatching rules and local search. *Proceedings of the 21st European Modelling and Simulation Symposium (EMSS 09)*. Puerto de la Cruz, Spain, vol. 2, pp. 27-32, September 2009.

Otero, M., Latorre, J. I., Jiménez, E. 2012. Renewable Energies in Vocational Training. *International Conference on Renewable Energies and Power Quality*. Santiago de Compostela, 2012.

Otero, M., Latorre, J. I., Jiménez, E., Pérez, M. 2012. Life cycle assessment applied to a vocational training centre. *Proceedings of the 4th International Conference on Engineering for Waste and Biomass Valorisation*, Porto (Portugal), September, 2012.

Silva, M. 1993. Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall.

Swanepoel, K. T. 2004. Decision support system: real-time control of manufacturing processes. *Journal of Manufacturing Technology Management*, Vol. 15, issue 1, pages 68 – 75.

Tuncel, G. 2007. A Heuristic Rule-Based Approach for Dynamic Scheduling of Flexible Manufacturing Systems. In Levner, E. (Ed.): *Multiprocessor Scheduling: Theory and Applications*, Itech Education and Publishing, Vienna, Austria.

SUSTAINABLE DESIGN FOR A CENTRE OF VOCATIONAL TRAINING. A PETRI NET APPROACH

María Otero-Prego[a], Juan-Ignacio Latorre-Biel[b], Eduardo Martínez-Cámara[c], Emilio Jiménez-Macías[d]

[a][d] University of La Rioja. Industrial Engineering Technical School.
Department of Electrical Engineering. Logroño, Spain
[b] Public University of Navarre. Department of Mechanical Engineering, Energetics and Materials.
Campus of Tudela, Spain
[c] University of La Rioja. Industrial Engineering Technical School.
Department of Mechanical Engineering. Logroño, Spain

[a] mariaoprego@googlemail.com, [b] juanignacio.latorre@unavarra.es,
[c] eduardo.martinezc@unirioja.es, [d] emilio.jimenez@unirioja.es

ABSTRACT

The management, design, and redesign of educative centers are performed usually by means of criteria that do not consider in detail the social, environmental, and financial impact of the daily activities in the institution. This fact may move away the design and management processes from daily reality, which can provide with a huge amount of data to be considered by human decision makers. The purpose of this paper is to describe a decision support system for the managers of a vocational training center that may help them make the best decisions for improving the effectiveness of the educative process. This tool arises as a combination of the Life Cycle Assessment methodology and the simulation of a Petri net model of the educative institution.

Keywords: vocational training, Petri nets, decision support system, life cycle assessment.

1. INTRODUCTION

Nowadays, the design and redesign of new educative centers require taking into consideration topics such as sustainability, efficient use of the different resources, for example energy, recycling, and waste management. This trend is based in a broad social concern on these aspects, in exigent national regulations and international agreements, as well in common sense, due to the high cost and limited availability of resources and restrictions in the elimination of wastes.

Different approaches are followed by designers, architects, and engineers, when designing educative centers. Among them, it is possible to let experts to advise them, such as commercial executives from specialized companies, to implement technologies that are familiar to the designers making them confident, to improve or modify in a more or less extent previous designs, etc.

Nevertheless, a sustainable design, where a large number of alternatives, technologies, and variables have to be considered, should lead to the best solution, achieved not only from strategic considerations but also from the day-to-day management, based on operational decisions. Only with such an approach, it is possible to include the level of detail in the future operation of decisions that a fine design should take into consideration for avoiding errors that will arise when the construction of the building has finished and the educative activities begin (Latorre and Jiménez, 2012).

In the next section, the life cycle assessment will be introduced as a methodology to achieve the sustainable design of an educative institution. Section 3 deals with the main characteristics of a vocational training center. The following section presents the formalism chosen for the representation of the model of the educative center: the Petri nets. Section 5 discusses the model developed by the paradigm of the Petri nets of the educative center. The next section describe the simulation-based methodology for the decision support system to the design of the educative center. The paper continues with a section describing the conclusions of the research performed so far. The last section lists the bibliography referenced in this paper.

2. LIFE CICLE ASSESSMENT

LCA or Life Cycle Assessment is a tool to evaluate how human activities impact on the environment. In consists on a methodology which comprises the complete life cycle of a product, process or service, what is called the “cradle to grave” approach (Curran, 1996). In fact, the application of the LCA requires a complete study from the raw materials requirement and the energy needs, to the return of the materials to the earth (EPA, 2006).

The LCA methodology has experienced a significant growth in the last decades. Its success has led to the standardization of LCA, such as ISO 14044, (Baumann and Tilman, 2004). Nowadays, LCA is considered as one of the most consolidated tools for the analysis of the environmental profile of products, processes, and services, as well as an interesting methodology for achieving sustainability goals in a certain institution (Hertwich, 2005).

The methodology of the LCA can be applied to the analysis of the activities of an institution, such as a vocational training center. In order to achieve this objective it focuses on the energetic efficiency and the evaluation of the environmental impact.

The choice of a vocational training center, among other institutions, is based in the fact that this kind of educative center is very influential on society, with regard to the development of environmentally-friendly activities, since they contribute notably to the education and the attitudes of new generations of citizens. In particular, present and future professionals and experts in many activities belonging to the industrial, agricultural, and service sectors are trained in this type of educative institution (Otero et al., 2012a).

The application of the LCA lists the inputs and outputs of resources and waste, providing with a precise vision of the advantages and drawbacks of the decision-making process in the development of products, processes, or services, from a point of view of environmental impact.

The LCA is a systematic methodology composed of four stages: (Jiménez et al., 2012)

a) Definition of the objectives and scope of the analysis.

b) Analysis of the inventory (ICV). Input resources and output waste are identified and quantified for the significant activities.

c) Impact assessment (EICV). The effects on the environment and society of the use of resources and generation of waste is evaluated.

d) Interpretation of the results and proposals for improvement.

3. VOCATIONAL TRAINING SCHOOLS

A vocational training school in Spain consists of an educative institution, where the students enroll non-compulsory education courses for becoming technicians in diverse sectors. These courses belong to the initial vocational training, while other courses are offered in the same institutions to unemployed people, to help them to find a job, as well as to workers, with the purpose of improving their knowledge, performance at their workplace, and their employability.

The initial courses of vocational training are classified into medium degrees and higher degrees, being the first ones a terminal way, which offer the graduates a qualification for their professional practice.

However, graduates of medium degrees of vocational training can enroll a course of higher degree of vocational training if they pass an examination that is organized once a year for this purpose.

Higher degrees provide the graduates a professional qualification as higher technicians. Moreover, they may continue their studies in the university.

The duration of the cycles of vocational training depends on the specialty and varies from one year in the educative center and three months of professional practice in a company to one year and seven months in the educative institution and three months of professional practices in a company.

The range of specialties of the courses is very wide and includes industrial subjects, such as mechanics, maintenance, electricity, and electronics, or other more related to the sector of services, such as configuration of networks of computers, hairdressing, commerce, marketing, assistant of nurse, sports, etc.

Sustainability in the design and management of a vocational training center has a special impact in society due to the fact that these institutions are the places where lots of citizens and future professionals are educated and trained for serving the society where they will perform their activities and which has invested in their education.

The design of a center, which is a model in the management of resources and wastes, will teach the students how to perform professional activities respecting the environment and the society, as well as to assess in their context the scarcity and cost of the available resources and the subsequent rational use of them (Otero et al., 2012b).

4. FORMALISM FOR THE MODEL

For this purpose, a very powerful approach consists of developing a model of the educative institution in process of being designed. The activity performed in a vocational training center corresponds to a large number of individuals, such as students, teachers, administrative staff, members of the management board, and external services. Some of these individuals act in parallel and converge with others to synchronize and compete for limited resources.

Regarding the previous characteristics that the model of the system should comply with, an appropriate formalism to be considered in this process is the paradigm of the Petri nets. This formalism allows representing the model of a discrete event system with a double representation: a graphical one and a matrix-based one. Moreover, this formalism presents a large body of knowledge and a number of available tools and techniques for validation, verification, structural and performance analysis, as well as for simulation and optimization (Silva, 1993).

Definition. A Petri net system is a 5-tuple $R = (P, T, \text{pre}, \text{post}, \mathbf{m}_0)$ such that:

i) P is a non-empty set of places.

ii) T is a non-empty set of transitions and $P \cap T = \emptyset$.

iii) pre and post are functions that associate a weight to the directed arcs between the elements of the sets P and T, in the following way:

iv) pre: $P \times T \rightarrow \mathbb{N}^*$ and post: $T \times P \rightarrow \mathbb{N}^*$, where $\mathbb{N}^*$ is the set of natural numbers, excluding zero.

v) $\mathbf{m}_0$ is the initial marking, such that $\mathbf{m}_0$: $P \rightarrow \mathbb{N}^*$.

□

5. MODEL OF THE VOCATIONAL TRAINING

The model of the educative center should be able to represent all the possible alternative configurations or scenarios for the vocational training center.

These scenarios can vary according to the decisions made by the different actors taking part in the educative process (students, teachers, members of the managing board, administrative staff, maintenance staff, families, etc).

Choosing the best scenario implies selecting the best sequence of decisions made by the actors; hence, the methodology developed to make the mentioned choice, described in this paper, can be used as a decision support system for the managing staff of the educative institution. This tool can be used to get advice in a complex environment for decision making as is an educative institution, where the number of actors can totalize more than a thousand people.

As it has been mentioned, the decision support system should choose a scenario for the educative center. For this reason it is a vital stage in the application of the methodology to define with precision the way to measure the "goodness" or quality of a given scenario.

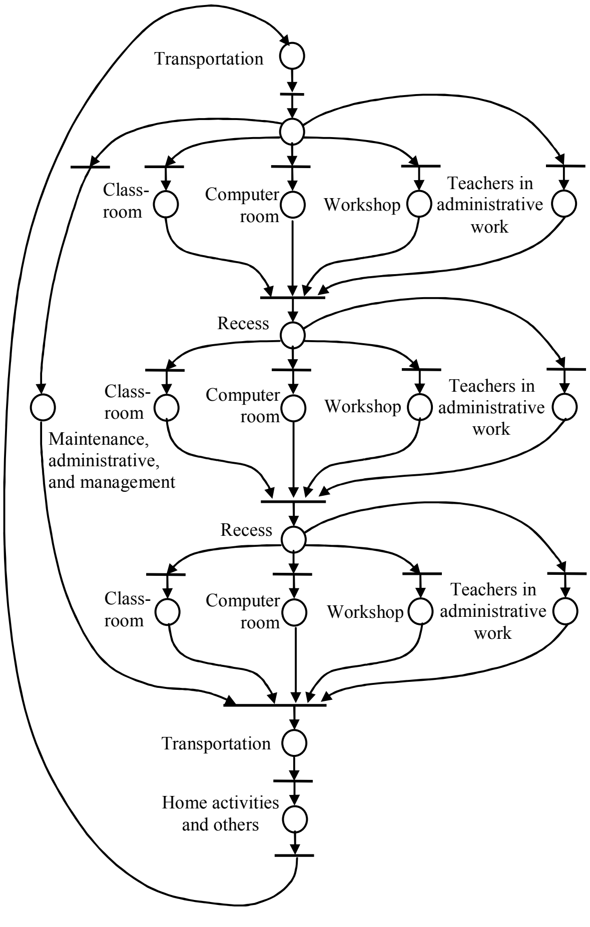

Transportation
Class-room
Computer room
Workshop
Teachers in administrative work
Recess
Class-room
Computer room
Workshop
Teachers in administrative work
Maintenance, administrative, and management
Recess
Class-room
Computer room
Workshop
Teachers in administrative work
Transportation
Home activities and others

Figure 1: Petri net model of a vocational training center

In order to achieve this objective, it is convenient to have in mind the main objectives of the educative institution or, at least, the objectives where the decision support system should focus. These objectives may be to lead the educative institution to its highest social impact, by using the smallest financial resources, and reducing at a minimum the environmental impact.

The present methodology quantifies the contribution of every individual achievement to the global objective and integrates these assessments in the model of the system.

Subsequently, the model is set up with a specific scenario which has been made correspond to a given sequence of decisions of the different actors belonging to the educative community. The choice of a given scenario can be performed randomly or by means of a technique to search in the solution space of the problem that is being solved.

The following step consists of performing a simulation of the Petri net model of the system by adjusting the desired duration of the educative activities to be simulated (from a day to a year could be common choices). During the simulation, a parameter that measures the quality of the considered scenario is constantly updated by the educative activities simulated and the previous quantification that measure the contribution of each activity to the global objective of the center.

Once a simulation has been completed, and its quality parameter calculated, another scenario should be simulated. When a number large enough of scenarios have been simulated, it is possible to choose the one associated to the highest quality parameter. This scenario would be the one proposed to the managing staff of the educative institution, together with the appropriate decisions related to it.

A model of the system, appropriated to the described methodology, is presented in the following paragraphs and figures. In this model, the actors are represented by means of individual tokens of the current marking of the system, while actions performed by the actors are associated to the places of the net. The firing of a transition leading to a place, which is associated to a given educative action, updates the quality parameter of the simulated scenario.

In figure 1, it has been represented a low-detailed model of an educative center, where the main daily activities of the educative process have been made explicit.

For the development of the model, some assumptions have been made. Among them, the educative activities are organized in three daily period of classes separated among them by means of two breaks. This schedule affects the activities of students and teachers. Nevertheless, the rest of the staff, that is to say the managing staff, administrative staff, and maintenance staff follow a different timetable.

According to the developed model, the activities of the actors belonging to the educative community are classified into three groups:

a) Transportation to and from the educative center.
b) Activities in the educative institution.
c) Activities out of the educative center.

In the first group, the transportation of the actors to the educative center at the beginning of the workday and from this institution at the end of the workday are considered. Statistical information particular to a given educative center should be included in the model to evaluate the impact of this activity, considering the percentage of actors that use public transport, shared cars, non-shared cars, motorcycles, bicycles, etc. It is also important to know the distance covered by the actors, as well as if the displacements are performed in urban areas with dense traffic, rural areas, etc.

The second group of activities, the ones developed in the educative institution, is presented in the model as the type of activities performed by students and teachers in different spaces, such as conventional classes, classes with Internet and a network of computers, and workshops, with professional equipment for practices of vocational training.

Moreover, teachers can have time, free of classes, to perform administrative work, such as preparing and correcting exams, preparing the classes, meeting with parents of students, meeting with other teachers, doing paperwork, etc.

The activities of the management, administrative and maintenance staff are represented together by the same single place in the Petri net model.

The third group of activities corresponds to the ones performed by the actors outside the educative center. They can be related or not with the educative process and are performed in the period of time from the end of the transportation of the actors from the center of vocational training to the beginning of the transportation to the educative institution at the beginning of the following day of work.

In order to simulate the educative process with a higher level of detail, for calculating the value of the quality parameter with a high accuracy, it is convenient to refine the model by expanding the different activities considered in the Petri net model of figure 1 into more specific tasks.

According to this idea, in figure 2, it has been represented in a bit more level of detail, the activities performed in a class developed in a conventional classroom, in a classroom with a network of computers, or in a workshop.

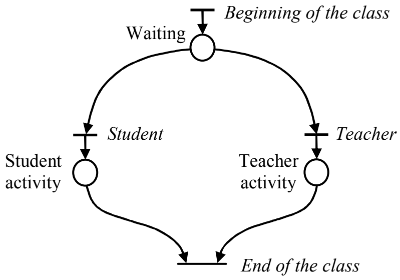


Figure 2: Tasks developed during a class.

In fact, the purpose of the Petri net presented in figure 2 is to split the tasks developed during a class into two groups: the ones performed by the teacher or teachers and the ones performed by the students. In fact, the mentioned Petri net corresponds to the places of the model depicted in figure 1 labeled “classroom”, “computer room” and “workshop”.

A further step in detailing the educative activities that correspond to a class will lead to two Petri net models. One of them corresponds to the tasks of the teacher and the other one the tasks performed by the students.

According to this idea, the Petri net model of the activities performed by a teacher in a class developed in a workshop has been represented in figure 3. The tasks developed by a teacher in a conventional classroom or in a computer room are similar and in general more restricted, due to a limitation in the available educative material and equipment, than the ones carried out in a workshop. For this reason the formers will not be detailed in this document.

Figure 3 shows a place representing the teacher waiting for the next task in the class to begin. From this place, the teacher can take the decision of beginning an explanation, helping the students individually or in small groups, while the other students perform a given educative task, correcting exercises or performing a brief task which can lead to significant influence in the social, financial, and environmental impact of the class: turning on/off the heating or the lighting.

Moreover, an explanation by the teacher can be developed using different technologies, implying diverse environmental impacts. The teacher can use a conventional blackboard, a projector and a computer presentation, or the explanation can be developed using professional equipment such as industrial machinery, chemical devices, or electrical systems.

It has to be said that more activities than the ones presented in figure 3 can be performed by a teacher during a class. However, the ones included in the model have been considered by the authors as the most representative ones, since they can be the most common ones and the environmental impacts can be significant.

Furthermore, the model of the educative institution developed for constructing the decision support system described in this paper should be detailed enough to allow the calculation of the quality parameter of a given configuration of the educative center with an appropriate accuracy but it should not be too large, since it might compromise the effectiveness of the methodology for requiring too much computer resources such as computing time to provide a suggestion of adequate decisions in a reasonable amount of time.

In the same way as it has been described for the activities performed by a teacher during a class, the place labeled “Student activity” in the Petri net model depicted in figure 2 can be expanded as it can be seen in figure 4. In this figure 4, there is a place with a conflict. In this place, labeled “Student waiting”, a student and the teacher, with the limitations imposed by the will and enthusiasm of the student, can choose the following task to be performed by the student.

Among the educative tasks that can be performed by the students, the model includes the use of ink and paper, writing notes or solving exercises, the use of a computer, the use of professional equipment or not using anything but their brains by paying attention to an explanation or by studying from a book.

Some of the potential tasks of the student can be against the rules of the class leading to a loss of time. Among them, only one, loosing time by using a smartphone, has been considered in the model, since in recent times it has become a very common, and difficult to control, activity of the students. Moreover, on a large scale, the social impact, and even the environmental and financial impact can be significant for the educative institution and the families of the students.

The rest of the places of the Petri net model depicted in figure 1, not expanded so far in this paper, can be detailed in the same way as the previously mentioned ones.

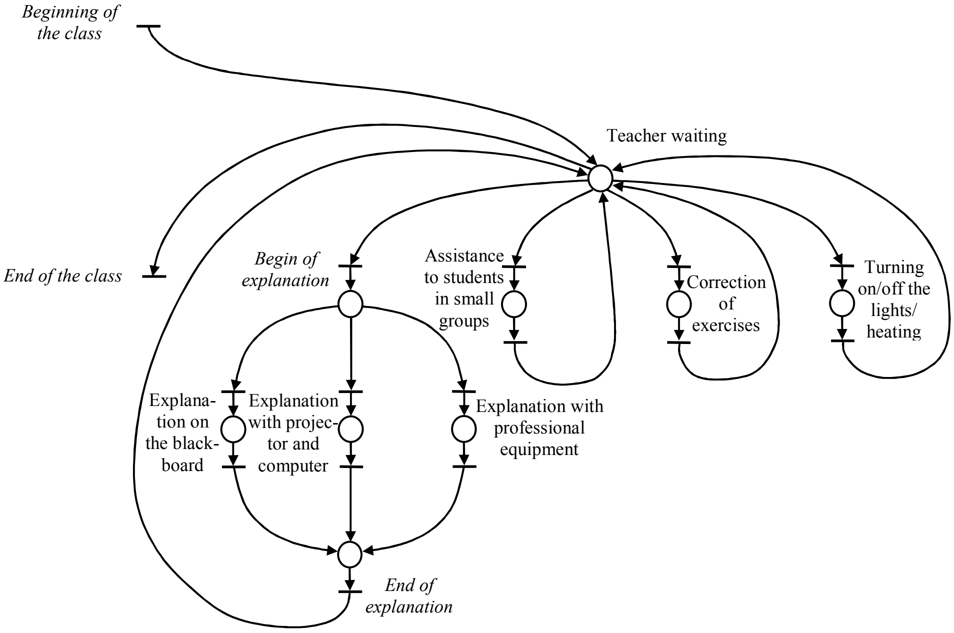

Beginning of the class
Teacher waiting
End of the class
Begin of explanation
Assistance to students in small groups
Correction of exercises
Turning on/off the lights/ heating
Explana-tion on the black-board
Explanation with projec-tor and computer
Explanation with professional equipment
End of explanation

Figure 3: tasks developed by a teacher during a class in a workshop.

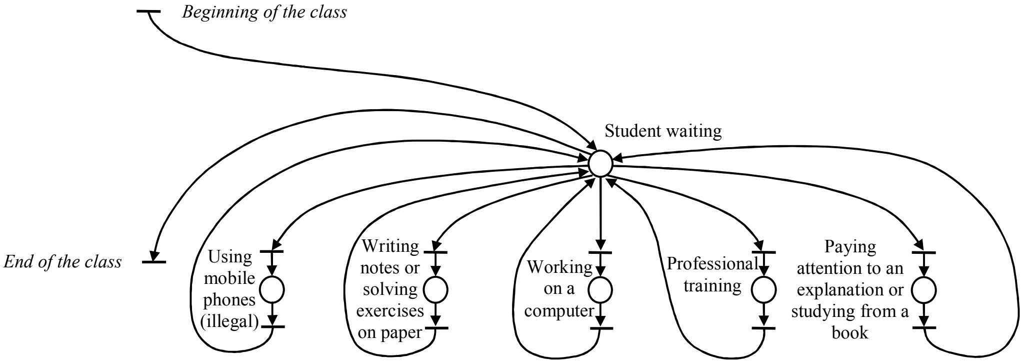

Beginning of the class
Student waiting
End of the class
Using mobile phones (illegal)
Writing notes or solving exercises on paper
Working on a computer
Professional training
Paying attention to an explanation or studying from a book

Figure 4: Activities of the students during a class in a workshop.

6. SIMULATION

The process of design of an educative institution based on a Petri net model can be supported very efficiently by means of simulation.

This methodology can be used in a variety of manners. One of the most common ways of using simulation is the technique called "what-if". According to it, it is possible to choose a diversity of scenarios for determining which one of them is the best option for solving a problem such as the design of a system. Simulation is then applied to mimic the evolution of the real system under every chosen scenario.

The comparison of the simulations is performed generally by means of quantitative assessments. In this methodology, the scenarios to be tested by means of simulation can be chosen manually, requiring highly trained and costly experts or automatically, using an algorithm to choose the most promising options from a pool of feasible solutions (Latorre et al., 2013).

In case that an automatic procedure is chosen for the selection of the scenarios to be tested by means of simulation and an objective function is defined to quantify the quality of every one of them, the methodology belongs to the category of optimization. In this methodology, the diverse choices performed in the process of design of a vocational training center represent different feasible solutions for the stated problem.

The quantification of the quality of a given scenario for the educative center is usually performed by means of an objective function. In case there are several competing objectives simultaneously, a multiobjective function can be constructed. The influence of the different objectives in the quality of the simulated solution can also be quantified in the multiobjective function by means of weighing coefficients.

The quantification of the environmental impact of a certain scenario is performed by combining the simulation of the evolution of the Petri net model with the methodology of the LCA.

In particular, the second stage in the application of the LCA, the analysis of the inventory, requires that every single activity considered in the model of the system is analyzed to quantify the consumed resources and generated waste. This information can be stored together with the Petri net model. The simulation of the model allow the addition of the contribution of every single activity to the total amount of resources and waste that correspond to the educative process.

Furthermore, the third stage in the application of the LCA consists of the impact assessment. The effects on the environment and society of the use of resources and generation of waste are evaluated for every single action represented in the Petri net model. The simulation process allows calculating the global effects of the educative process for any scenario involved in the simulation.

7. CONCLUSIONS

The efficient design of a vocational training center has been the subject of the research line presented in this paper. The main objective aimed with this paper has been to present a methodology to build up a decision support system, appropriate for the managing staff of the educative center.

It has been described the procedure to build up a detailed model of the system by using the paradigm of the Petri nets. This model allows a quantitative representation of the educative institution able to perform a numerical analysis of the evolution of the educative process.

Furthermore, it has been shown that the LCA, a reputed methodology for the analysis of the environmental impact of human activities, can be applied in conjunction with the Petri net model, in order to provide with a flexible tool for simulating the behavior of the educative institution on a daily basis for a variable amount of time.

The resulting decision support system may perform a simulation of a selection of a predetermined set of configurations of the educative institution and calculate the corresponding environmental impact of every one of them, compare the results and determine the best one of them. Moreover, the decision support system can provide with the sequence of decisions of the managing staff that lead to the most successful configuration.

In the following stages of the research, the methodology will be applied to a diversity of centers of vocational training in order to refine the model of the system, increase its versatility, and test it as decision support tool.

REFERENCES

Baumann, H. and Tilman, A. 2004. The Hitch Hicker's Guide to LCA: An Orientation in *Life Cycle Assessment Methology and Application*. Lund: Student Litterature,.

Curran, M. A. 1996. Environmental Life Cycle Assessment. McGraw-Hill.

Environmental Protection Agency. 2006. *Life Cycle Assessment: Principles and Practice*. National Risk and Research Laboratory. Cincinnati, Ohio, USA. 2006.

Hertwich, E. G. 2005. Life Cycle Approaches to Sustainable Consumption: a critical review. *Environmental Science and Technology*, 2005. Vol. 39, n# 13, pp. 4673-4684.

Jiménez, E., Martínez, E., Blanco, J., Pérez, M., Graciano, C. 2012. Methodological approach towards sustainability by integration of environmental impact in production system models through LCA. Application to the Rioja wine sector. Simulation: *Transactions of The Society for Modeling and Simulation International. Special Issue of Simulation: Modelling Sustainability for Third Millennium*.

Latorre, J.I., Jiménez, E., Pérez, M. 2013. Simulation-based Optimisation of Discrete Event Systems by Distributed Computation. *Simulation: Transactions of the Society for Modeling and Simulation International. Special Issue: Advancing Simulation Theory and Practice with Distributed Computing* (In press).

Latorre, J.I., and Jiménez, E. 2012. Automatic design based on the Petri nets paradigm. *Proceedings of the 24th European Modelling and Simulation Symposium.* pp. 446-451. Vienna, Sept. 2012.

Otero, M., Latorre, J. I., Jiménez, E. 2012. Renewable Energies in Vocational Training. *International Conference on Renewable Energies and Power Quality*. Santiago de Compostela, 2012.

Otero, M., Latorre, J. I., Jiménez, E., Pérez, M. 2012. Life cycle assessment applied to a vocational training centre. *Proceedings of the 4th International Conference on Engineering for Waste and Biomass Valorisation*, Porto (Portugal), September, 2012.

Silva, M. 1993. Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall.

ANALYSIS OF INFORMATION PARTIAL ENCRYPTION OPTIONS FOR EXCHANGING PETRI NETS SYSTEMS

Iñigo León Samaniego[a], Emilio Jiménez-Macías[b], Juan Ignacio Latorre-Biel[c]

[a]University of La Rioja. Computer Science Faculty. Logroño, Spain
[b]University of La Rioja. Industrial Engineering Technical School.
Department of Electrical Engineering. Logroño Spain
[c]Public University of Navarre. Deptartment of Mechanical Engineering,
Energetics and Materials. Campus of Tudela, Spain

[a]inigo.leon@gmail.com, [b]emilio.jimenez@unirioja.es, [c]juanignacio.latorre@unavarra.es

ABSTRACT

The aim of this work is to create a framework of definitions and notations to hide part of a Petri net, facing a possible delivery, maintaining the privacy of the critical, secret, or complex parts of the system.

From these definitions and notations we work with the incidence matrices and we analyze the implications of hiding. In this work only the structure of the network is processed. The study of markings and properties of networks with hidden pieces will we analyzed in further works.

1. INTRODUCTION

Petri nets are widespread for modeling many classes of systems, such as manufacturing, logistics, processes and services [3] [5], and in general discrete concurrent systems [4]. However, all these nets are described in a comprehensive manner and must have the information of the entire net to determine their evolution. It would be interesting to take a Petri net and hide a part of it. This can be useful, for example, when distributing a process with some secret [6], or simply to be a part of complex net that is not interested to be handle globally for any reason [5].

In advanced work, we studied the possibilities of Petri nets reduction [10], grouping in one place or transition a subnet, so that what happens on this subnet is encapsulated in a single point of execution. However, we want to go further by defining parts of the net that are hidden, not clustered, and even the implications within the network properties. The aim of this work is the creation of the theoretical basis for a further study of Petri nets in which certain parts are hidden.

So we setup a generic framework of definitions and notations that allows us to deepen in the study of the characteristics and properties of Petri nets. We will expand the vision of Petri nets, providing them with greater functionality in an interesting way for practical applications.

The first part of this paper studies the state the art in this field. We are going to deepen in the basic Petri nets definitions and properties [7] related with hidden information. All this will be necessary to create the framework that allows us to study occultation in PN.

For this paper we will always deal with ordinary networks and pure, unless otherwise expressly.

2. PETRI SUBNETS. DEFINITIONS AND PROPERTIES

Let be P and T the non-empty finite sets of places and transitions, respectively. Let $|P|$ = n (the number of places network) and $|T| = m$ (number of transitions).

Let be α and β pre and post incidence matrices respectively. Let $R = \langle P, T, \alpha, \beta\rangle$ be a Petri net and let C the incidence matrix of R

Definition 1 (Subnet [8]). A subnet of $R = \left\langle P, T, \alpha, \beta \right\rangle$ is a net $\overline{R} = \left\langle \overline{P}, \overline{T}, \overline{\alpha}, \overline{\beta} \right\rangle$ such that $\overline{P} \subseteq P$ and $\overline{T} \subseteq T$, $\overline{\alpha}$ and $\overline{\beta}$ are restrictions of α and β over $\overline{P} \times \overline{T}$.

In other words, a subnet is a subset of places and transitions with their arcs, joined together.

Let's look at the implications of the latter definition since it is one of the most important with regard to this work.

A subnet corresponds [6], from the matrix point of view, with the resulting submatrix obtained by keeping only the rows corresponding to transitions and places columns for the selected subnet.

Example 1. We take the Petri net which has the following incidence matrix:

$$C = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \end{array} \\ \left(\begin{array}{cccccc} -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array} \right) \end{array}$$

If we stay with places p_1, p_3, and p_5 P_4 and transitions t_1, t_2, and t_3 we have the subnet defined by this incidence matrix.

$$C' = \begin{array}{c} \\ p_1 \\ p_3 \\ p_4 \\ p_5 \end{array} \begin{array}{c} \begin{array}{ccc} t_1 & t_2 & t_3 \end{array} \\ \begin{pmatrix} -1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

In [6] is shown that the set of all possible permutations of rows and/or columns of a matrix of incidence corresponding to a network, either the previous or subsequent actual Incidence call, make an equivalence relation. In other words, given an incidence matrix can be rearranged both rows and columns and this rearrangement end is perfectly describing the original network.

In this way, we can study the incidence matrices reordering rows and columns as preferred one at any time, without loss of generality.

From all these definitions and proofs we can draw several trivial conclusions:

1. A subnet, like generic network does not have to be square.
2. If a row or column of the incidence matrix is all zeros, no mean that that place or that transition is isolated. this only occur with pure networks.
3. It does not matter the number of places and / or transitions are chosen for the subnet, if they are not empty sets.

3. SPLITTING A NETWORK INTO SUBNETS

Let $R = \langle P, T, \alpha, \beta\rangle$ a Petri net where $|P| = n$ and $|T| = m$. So $P = \{p_1, p_2 \ldots p_n\}$ and $T = \{t_1, t_2 \ldots t_m\}$. Select two subsets $P' \subseteq P$ and $T' \subseteq T$ so that $|P'| = r \leq n$ and $|T'| = s \leq m$. With these premises divide into two subnets the original one.

We have seen that we can identify a subnetwork simply removing rows and columns (places/transitions) of an incidence matrix. Taking advantage of the equivalence relation defined in [6], we reorder the incidence matrix so that they are in the top places and transitions of the subnet defined. Rename also the places and transitions (without loss of generality, and for convenience) so that the incidence matrix is as follows:

$$C = \begin{array}{c} \\ p_1 \\ \vdots \\ p_r \\ p_{r+1} \\ \vdots \\ p_n \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & \cdots & t_s & t_{s+1} & \cdots & t_m \end{array} \\ \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1s} & a_{1(s+1)} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rs} & a_{r(s+1)} & \cdots & a_{rm} \\ \hline a_{(r+1)1} & \cdots & a_{(r+1)s} & a_{(r+1)(s+1)} & \cdots & a_{(r+1)m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ns} & a_{n(s+1)} & \cdots & a_{nr} \end{array}\right) \end{array}$$

We now have the network divided into two disjoint and complementary subnets. They are disjoint because there is no place and no common transition, and complementary because the union of the two we gives the complete network. At this point note that the incidence matrix is divided into four blocks $C = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. The interpretation is as follows:

- A subnet made up of places $p_1 \ldots p_r$ and transitions $t_1 \ldots t_s$
- D subnet is complementary to A, made up of the places $p_{r+1} \ldots p_n$ and transitions $t_{s+1} \ldots t_m$.
- B is the matrix that defines the interaction of the places of A with D transitions
- C is the matrix that defines the interaction of D places with A transitions

This can be generalized to multiple disjoint and complementary subnets without further to re-apply the same process to any of the subnets already defined. Thus, generically we can divide a network into *i* subnetworks, so we'll have a matrix of this style:

$$\left(\begin{array}{ccc|ccc|c|ccc} a_{11} & \cdots & a_{1s} & a_{1(s+1)} & \cdots & a_{1t} & & a_{1u} & \cdots & a_{1m} \\ \vdots & SN_1 & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{ps} & a_{p(s+1)} & \cdots & a_{pt} & & a_{pu} & \cdots & a_{pm} \\ \hline a_{(p+1)1} & \cdots & a_{(p+1)s} & a_{(p+1)(s+1)} & \cdots & a_{(p+1)t} & & a_{(p+1)u} & \cdots & a_{(p+1)m} \\ \vdots & \ddots & \vdots & \vdots & SN_2 & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qs} & a_{q(s+1)} & \cdots & a_{qt} & & a_{qu} & \cdots & a_{qm} \\ \hline & \vdots & & & \vdots & & \ddots & & \vdots & \\ \hline a_{r1} & \cdots & a_{rs} & a_{r(s+1)} & \cdots & a_{rt} & & a_{ru} & \cdots & a_{rm} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & SN_i & \vdots \\ a_{n1} & \cdots & a_{ns} & a_{n(s+1)} & \cdots & a_{nt} & & a_{nu} & \cdots & a_{nm} \end{array}\right)$$

In this situation, if we select two subnets SN_j and SN_k , we locate the zones of influence of each with respect to the other:

$$\left(\begin{array}{c|c|c|c|c} \ddots & \cdots & \cdots & \cdots & \cdots \\ \hline \vdots & SN_j & \cdots & IM_1 & \cdots \\ \hline \vdots & \cdots & \ddots & \cdots & \cdots \\ \hline \vdots & IM_2 & \cdots & SN_k & \cdots \\ \hline \vdots & \vdots & \cdots & \vdots & \ddots \end{array}\right)$$

Thus, the submatrix IM_1 represents the arcs that connect places of the submatrix SN_j with SN_t*ransitionsk* and the matrix IM_2 represents the arcs that connect places of SN_k to Sn_t*ransitionsj*.
Arcs that are in one way or another indicates the sign of the corresponding element of A or B.

Definition 2 (Partition of a network into subnets). We say that a set $P = \{R_1 R_2 \ldots R_k\}$ is a partition into subnets of R if the following holds:

- $R_1 \cup R_2 \cup \ldots \cup R_k = R$
- $\forall i, j | 1 \leq i, j \leq k \Rightarrow R_i \cap R_j = \emptyset$

ie, the binding of the total network subnets and subnets are pairwise disjoint.

$$\begin{array}{c|cccccc} & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\ \hline p_1 & -1 & 0 & 1 & 0 & 0 & 0 \\ p_2 & -1 & 0 & 0 & 1 & 0 & 0 \\ p_3 & 1 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 1 & -1 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 1 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & -1 & 0 & 0 & 0 \\ p_7 & 0 & 0 & 0 & -1 & -1 & 1 \\ p_8 & 0 & 0 & 0 & 0 & 1 & -1 \end{array} \cong \begin{array}{c|cccccc} & t_1 & t_6 & t_3 & t_5 & t_4 & t_2 \\ \hline p_8 & 0 & -1 & 0 & 1 & 0 & 0 \\ p_1 & -1 & 0 & 1 & 0 & 0 & 0 \\ p_3 & 1 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & -1 & 0 & 0 & 0 \\ p_4 & 1 & 0 & 0 & 0 & 0 & -1 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 1 \\ p_2 & -1 & 0 & 0 & 0 & 1 & 0 \\ p_7 & 0 & 1 & 0 & -1 & -1 & 0 \end{array}$$

Figure 1 – Two equivalent incidence matrices to describe the same a Petri net.

4. DESCRIPTION OF THE PARTS OF A MATRIX ONCE DEFINED THE SUBNETS

As can be reordered places and transitions smoothly, we study a network N divided into 2 subnets, for simplicity and without loss of generality.

For consistency with [6] we will follow this notation:

$$\left(\begin{array}{c|c} H & HP \\ \hline HT & V \end{array}\right)$$

where

- H (Hidden Subnet) is the subnet you want to hide.
- V (Visible Subnet) is the subnet that is visible.
- HT (Hidden Transitions Submatrix) are the relationships between places of V and H transitions
- HP (Hidden Places Submatrix) are the relations between transitions of V and H sites

Note. Following this notation can be convenient because it is clear what is each of the submatrices. However, elsewhere in the document be referenced as *R*1 and *R*2 for be more clarifying or being something generic and independent networks concealment. However, using *R*1 and *R*2 the notation of subnets of influence with respect to the other is more diffuse.

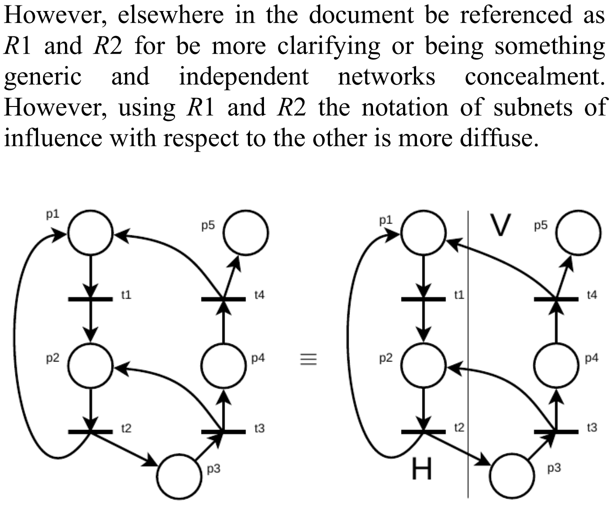


Figure 2 – Selecting subnet to hide

Example 2. Consider the Petri net of the figure 2 with the next incidence matrix:

$$\begin{array}{c|cccc} & t_1 & t_2 & t_3 & t_4 \\ \hline p_1 & -1 & 1 & 0 & 1 \\ p_2 & 1 & -1 & 1 & 0 \\ p_3 & 0 & 1 & -1 & 0 \\ p_4 & 0 & 0 & 1 & -1 \\ p_5 & 0 & 0 & 0 & 1 \end{array}$$

The subnet we want to hide is formed by sites 1and 2 and 1 and 2 transitions. Graphically, separate places and transitions to hide (H) from the rest of the network (V)

The incidence matrix is already sorted by the places and transitions to the top of it. Here’s the four parts described above.

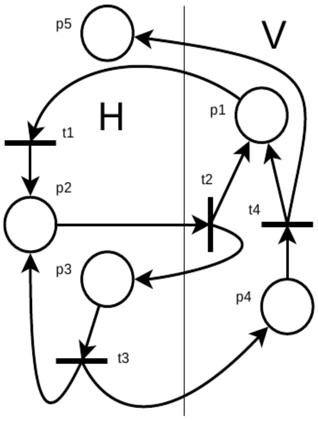
$$\begin{array}{c|cc|cc} & t_1 & t_2 & t_3 & t_4 \\ p_1 & -1 & 1 & 0 & 1 \\ p_2 & 1 & -1 & 1 & 0 \\ \hline p_3 & 0 & 1 & -1 & 0 \\ p_4 & 0 & 0 & 1 & -1 \\ p_5 & 0 & 0 & 0 & 1 \end{array}$$

In this matrix we can see the four described parts:

- $H = \begin{array}{c|cc} & t_1 & t_2 \\ p_1 & -1 & 1 \\ p_2 & 1 & -1 \end{array}$ is the subnet we want to hide.
- $V = \begin{array}{c|cc} & t_3 & t_4 \\ p_3 & -1 & 0 \\ p_4 & 1 & -1 \\ p_5 & 0 & 1 \end{array}$ is the subnet that is visible.
- $HP = \begin{array}{c|cc} & t_3 & t_4 \\ p_1 & 0 & 1 \\ p_2 & 1 & 0 \end{array}$ are the relationships between transitions of V and H places.

- $HT = \begin{array}{c} \\ p_3 \\ p_4 \\ p_5 \end{array} \begin{array}{cc} t_1 & t_2 \\ \hline \end{array}$ $\begin{array}{c|cc} & t_1 & t_2 \\ p_3 & 0 & 1 \\ p_4 & 0 & 0 \\ p_5 & 0 & 0 \end{array}$ are the relationships between places of V and H transitions.

Example 3. In the previous example we have seen a fairly simple option selection subnet and we have chosen the locations 1 and 2 and the transitions 1 and 2. However, we can choose any other subset of places and transitions. In this example we will select locations 2, 3 and 5 and the transitions 1 and 3. Thus, in the graph of the previous example move the locations and transitions to hide on one side and the rest on the other.



Although more confusing, can be seen that the graph is the same as the incidence matrix is the same (not just part of the equivalence class, it is exactly the same).
Now, in this matrix move places 2, 3 and 5, and 1 and 3 transitions at the beginning of the matrix:

$$\begin{array}{c|cc|cc} & t_1 & t_3 & t_2 & t_4 \\ p_2 & 1 & 1 & -1 & 0 \\ p_3 & 0 & -1 & 1 & 0 \\ p_5 & 0 & 0 & 0 & 1 \\ \hline p_1 & -1 & 0 & 1 & 1 \\ p_4 & 0 & 1 & 0 & -1 \end{array}$$

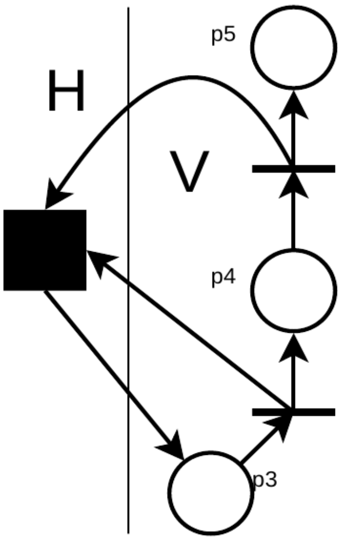
Interpreting each of the chunks of the matrix is similar to the previous example.

5. HIDING THE SUBNET

Once you select the subnet to hide we proceed to the occultation as such [6]. Graphically, it seems simple. Just replace the subnet to hide by a black box and modify some arcs according to the following rules:

1. The arcs originating in a place or transition within the black box, and target a place or transition out of it will have the black box as the source.
2. The arcs originating in a place or transition out of the black box, and target a place or transition within it, are replaced by the black box as a destination.

Example 4. We consider the Petri net of the Figure 2. The result of hiding the part of the graph H is the following:



In the associated incidence matrix also replace the subnetwork H by a black box:

$$\begin{array}{c|cc|cc} & t_1 & t_2 & t_3 & t_4 \\ p_1 & \multicolumn{2}{c|}{\blacksquare} & 0 & 1 \\ p_2 & & & 1 & 0 \\ \hline p_3 & 0 & 1 & -1 & 0 \\ p_4 & 0 & 0 & 1 & -1 \\ p_5 & 0 & 0 & 0 & 1 \end{array}$$

However, in this matrix notation is given information should also be hidden: it gives us information about the number of places and transitions of the hidden subnet, besides indicating hidden places and transitions with which it interacts. To solve this problem we proceed as follows. We can group all rows for the screened subnet into one. In each row position examine all elements of the original rows corresponding to that position, and will put:

- If all these elements are zero, in the grouped row will be a zero.
- If one and only one of those elements is nonzero, will put that item.
- If there are several non-zero elements, we will post a list of these items separated by commas, creating a d-dimensional element (in *d* dimensions).

In the same way we have done with the rows, proceed with columns. Thus, if the hidden subnet has *i* columns and *j* rows, we will get a matrix like this:

$$\left(\begin{array}{c|ccc} \blacksquare & a_{1(i+1)} & \cdots & a_{1m} \\ \hline a_{(j+1)1} & & & \\ \vdots & & V & \\ a_{n1} & & & \end{array}\right)$$

Where ∀p, ∀q|$i + 1 \leq p \leq m \wedge j + 1 \leq q \leq n$

$$a_{1p} = \begin{cases} 0 & \text{if } \forall r|1 \leq r \leq j, c_{rp} = 0 \\ c_{rp} & \text{if } \exists! r, 1 \leq r \leq j | c_{rp} \neq 0 \\ (c_{r1p}, c_{r2p}, \ldots) & \text{if } \exists! r_1 \neq r_2 \neq \ldots, 1 \leq r_1, r_2, \ldots \leq j \\ & |c_{r1p}, c_{r2p}, \ldots \neq 0 \end{cases}$$

$$a_{q1} = \begin{cases} 0 & \text{if } \forall s | 1 \leq \ s \ \leq \ i, c_{qs} = 0 \\ c_{qs} & \text{if } \exists! s, 1 \leq \ s \ \leq \ i | c_{qs} \neq \ 0 \\ (c_{qs1},\ c_{qs2},\ \ldots) & \text{if } \exists! s_1 \neq s_2 \neq \ldots, 1 \leq s_1, s_2, \ldots \leq \ s \\ & | c_{qs1},\ c_{qs2},\ \ldots \neq \ 0 \end{cases}$$

So we hide the number of places and transitions of the hidden subnet and their relationships. Yes, some information is given about the hidden network. Really if this resulting matrix some node that is *d*-dimensional, at least in the hidden network must exist *d* nodes of this type.

Example 5. We consider the Petri net defined by the following incidence matrix, separated into H,V ,HT and HP .

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

After applying the above steps for the group, we would have the following:

$$\left(\begin{array}{c|ccc} \blacksquare & (1,-1,1) & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

Here we see that the information about the number of hidden places and transitions is minimized. So we know that at least there is a hidden transition and at least three hidden places (there is a transition of dimension 3). However, we do not know the exact number of either.

6. HIDING VS. REDUCTION

Both Silva works [8] [9] as in the article by Xia [10] discusses possible Petri nets reductions for grouping and simplifying, under certain circumstances, places and / or transitions. These reductions can be structural (only dependent on the structure and initial marking of the net) or depending on the interpretation of the Petri net.

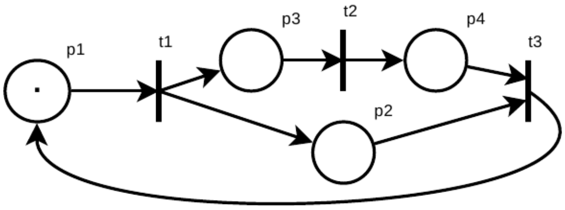
Should be clear that these reductions are not the same thing we are describing. We do not try to simplify the network together elements to have more or fewer places or transitions or to make it easier. What we want is to hide part of the network, regardless of how simple or complicated it is.

Here we have an example of what a reduction is.

Example 6 (Reduction of an implicit place [8]). In a marked Petri net, an implicit place is one that meets the following:

1. its marking can be calculated from other points marking
2. never is the only place that prevents the enabling of its output transitions

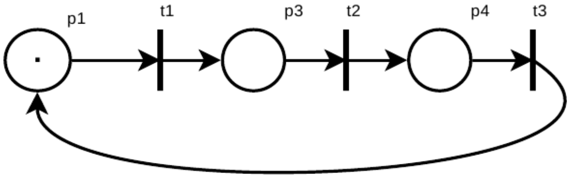
If we consider the following Petri net



we can notice that p_2 is an implicit place because its marking can be calculated as a function of p_3 y p_4:

$$M\,(p_2\,) = M\,(p_3) + M\,(p_4)$$

Moreover, by this same formula, it is clear that $M\,(p_2) \geq M\,(P_4)$ (marking cannot be negative) so the only place that can prevent enabling of T_3 is P_4 . Thus eliminating p_2 does not alter the behavior of the network, which would be as follows:



In this network elements have been removed, no hidden. This example helps us to see the difference between hiding and a reduction.

7. CLASSIFICATION BY TYPE OF SUBNET HIDING

We have seen how to hide part of a network. We have also studied how to make relations between the visible and hidden parts of the network providing minimal information about the network structure.

Then we see occultation special cases with special features. Suppose we take a pure network and want to hide part of it. Depending on how they are each of the four pieces of matrix (H, V, HP and HT) we can see some special cases.

7.1. Disjointed subnets

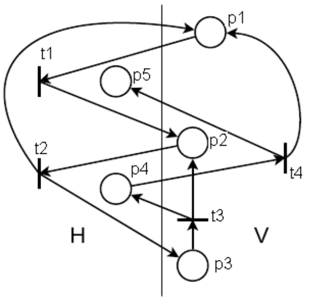
Suppose that in the incidence matrix divided into the four pieces explained, are H or V be a null matrix. In this case the interpretation is that there arcs between places and transitions of the subnet, which would simply places and / or no transitions related to each other but with the additional subnetwork. Subnet talk then disjointed.

Definition 3 (Disjointed subnet). Pure subnet said disjointed if there is no arc between places and transitions of that subnet, ie if its incidence matrix is zero.

Example 7. Consider the Petri net of figure 2. The incidence matrix is:

$$\begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{array} \begin{array}{c} \begin{array}{cccc} t_1 & t_2 & t_3 & t_4 \end{array} \\ \left(\begin{array}{cccc} -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right) \end{array}$$

We assume that we select as subnet formed by 4th and 5th places and transitions 1 and 2. Then the graph and the incidence matrix are thus:



$$C = \begin{array}{c} \\ p_4 \\ p_5 \\ p_1 \\ p_2 \\ p_3 \end{array} \begin{array}{c} \begin{array}{cc|cc} t_1 & t_2 & t_3 & t_4 \end{array} \\ \left(\begin{array}{cc|cc} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \\ \hline -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right) \end{array}$$

Here we can see that although really p_4, p_5, t_1 and t_2 are not isolated, there is no arc that connects them together. In the incidence matrix, the corresponding submatrix is the zero matrix. Therefore, whether or not there are elements isolated in the net, total subnet formed by p_4, p_5 , t_1 and t_2 is a disjointed net.

7.2. Macroplace

Suppose now that the incidence matrix divided into the four pieces explained, HT appears to be the zero matrix. Then we conclude that the subnet H is only related by arcs with places of subnet V. All arcs entering H come from transitions of V and all arcs coming out from H go to transitions of V. Stated another way, the subnet V behaves like a spot, but may contain places and transitions.

Definition 4 (Macroplace). A macroplace is a subnet H or V that meets the following:

1. arcs entering any node of the subnet from an external node come from a transition.
2. arcs leaving any node on the subnet to an external node go to a transition.

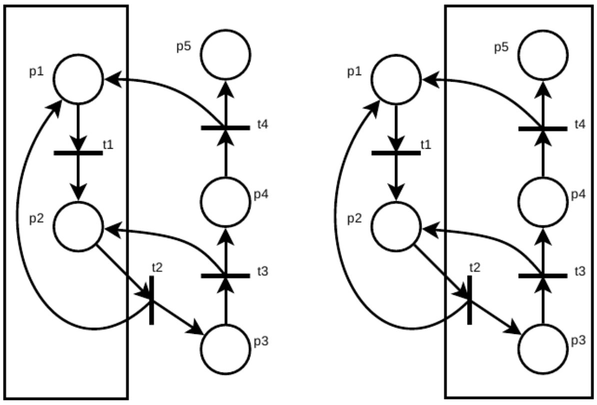
Note that this is not really a place, and that the subnet has not marked as such. The marking is on the places within the subnet and depend on the arches of arrival.

7.3. Macrotransition

Another option that can happen is that in the incidence matrix, HP appears to be the zero matrix. Then we conclude that the subnet H is only related by arcs with transitions of subnet V. All arcs entering H come from places of V and all arcs coming out from H go to places of V. Stated another way, the subnet V behaves like a transition, but may contain places andtransitions.

Definition 5 (Macrotransition). A macrotransition is a subnet H or V that meets with the following:

1. arcs entering any node of the subnet from an external node come from a place.
2. arcs leaving any node on the subnet to an external node go to a place.



a) Macroplace b) Macrotransition

Figure 3 – Macroplace y macrotransition

Like macroplaces, macrotransitions are not transitions as such, it is not necessary that all entries are marked to fire the macrotransition, and not all output places are marked after entering it. Everything depends on the inner workings of the macrotransition.

7.4. Sinkhole subnet and Source subnet

Another thing that can happen is that the hidden subnet reach only arcs. We then find that you can not leave the subnet. We speak then of a sinkhole subnet.

Definition 6 (Sinkhole subnet). It is said that a subnet is a sinkhole subnet if no arc has its origin in an internal node (place or transition) of the subnet.

It is easy to see that a subnet is sinkhole if and only if all elements of HP are greater or equal to zero and all elements of HT are less than or equal to zero.

H is sinkhole $\Leftrightarrow \forall a_{ij} \in HP, a_{ij} \geq 0 \wedge \forall a_{pq} \in HT, a_{pq} \leq 0$

If instead of this what happens is no arc gets into the subnet, we have a source subnet. In a source subnet we can not enter.

Definition 7 (Source subnet). It is said that a subnet is a source subnet if no arc has its destination in an internal node (place or transition) of the subnet.

It is easy to see that a subnet is a source if and only if all elements of HT are greater than or equal to zero and all elements of HP are less than or equal to zero.

H is source $\Leftrightarrow \forall a_{ij} \in HT, a_{ij} \geq 0 \wedge \forall a_{pq} \in HP, a_{pq} \leq 0$

8. FRONT-END INTERACTION WITH THE SUBNET. INPUT AND OUTPUT FUNCTIONS

Once you have defined all this environment, we will try to go a little further. Let’s assume that we want to export a subnet we have hidden in another network, like a black box. Our intention is to connect this hidden network to another network, and can thus be reused subnets. For example, let’s assume that we have a process modeling with Petri net modeling and in this there is a subnet we want to hide, but, at the same time, we want to reuse it in other Petri nets.

In this case we have a problem, and once hidden net work disappears half the information input or output arcs of the same. In particular, we do not know the source nodes and arcs that leave the target nodes of the arcs that enter the network until no visible again. But if we want to reuse it on other networks, can not wait to make it visible. Should remain hidden, but should be able to connect to other networks.

We will try to solve this problem. This way we can reuse hidden networks like plug-in modules on other networks. However, we will not need the actual implementation of the source or destination nodes of the arcs that leave or enter the network, respectively. The solution is to define a facade or front-end input and output of the network. This front-end will contain the information needed to interact with the network hidden, but hide the specifics of implementation. To define this behavior going from some assumptions.

8.1. Previous definitions

Let $R = \langle P, T, \alpha, \beta\rangle$ be a Petri net and let $P = \{R_1, R_2\}$ be a partition of R.

Definition 8 (Input place). Let p_i a place of R_1. p_i is an input place of R_1 if it is the destination of an arc coming from a R_2 transition, ie,

p_i is an input place of R_1 if $\exists t_j \in R_2 | c_{ij} > 0$

Definition 9 (Input transition). Let t_i a transition of R_1. t_i is an input transition of R_1 if it is the destination of an arc coming from a R_2 place, ie,

t_i is an input place of R_1 if $\exists p_j \in R_2 | c_{ji} < 0$

Definition 10 (Input node). An input node of R_1 is an input place or transition of R_1.

Definition 11 (Output place). Let p_i be a place of R_1. p_i is an output place of R_1 if an arc leaves it towards a transition of R_2, ie,

p_i is an output place of R_1 if $\exists t_j \in R_2 | c_{ij} < 0$

Definition 12 (Output transition). let t_i be a transition of R_1. t_i is an output transition of R_1 if an arc leaves it towards a place of R_2, ie,

t_i is an output place of R_1 if $\exists p_j \in R_2 | c_{ji} > 0$

Definition 13 (Output node). An output node of R_1 is an output place or transition of R_1.

After defining these concepts, we can define the sets thereof.

Notation. We denote the sets of the elements defined above:

- Let *IP* (*R*) $\subseteq \overline{P}$ (Input Places) be the set of input places of a subnet.
- Let *IT* (*R*) $\subseteq \overline{T}$ (Input Transitions) be the set of input transitions of a subnet.
- Let *IN* (*R*) $\subseteq \overline{P} \cup \overline{T}$ (Input Nodes) be the set of input nodes of a subnet.
- Let *OP* (*R*) $\subseteq \overline{P}$ (Output Places) the set of output places of a subnet.
- Let *OT* (*R*) $\subseteq \overline{T}$ (Output Transitions) be the set of output transitions of a subnet.
- Let *ON* (*R*) $\subseteq \overline{P} \cup \overline{\mathrm{T}}$ (Output Nodes) be the set of output nodes of a subnet.

Note. Recall that a node in a Petri net can be both a place and a transition, depending on the context.

Notation. Denote as n_i to a node of a Petri net.

As we have generic definitions, no problem in applying to a network divided into *H*, *V*, *HN* and *HT*, as the set $\{ H, V \}$ is a partition of R.

8.2. Subnet Front-end

Once all these concepts, we create the front-end input/output of a Petri subnet. A front-end of the Petri net will be a intermediate facade that allows us to physically divide that subnet from the rest of the net. Thus, in order to enter or leave the subnet, you need to make it through this front-end.

Let *IA* (input arcs) the set of arcs that enter the subnet R_1 and let *OA* (output arcs) the set of arcs leaving R_1.

Definition 14 (Input gate of a net). Let $a_i \in IA$ an arc of entrance to R_1. We define an input gate to R_1, and denote by ig_i, as a new logical node that is identified with an arc of entrance to the net. For each input arc, defines an input gate, regardless of the origin and destination of the arc. If the source is a transition, we denote igt_i and if a place, igp_i.

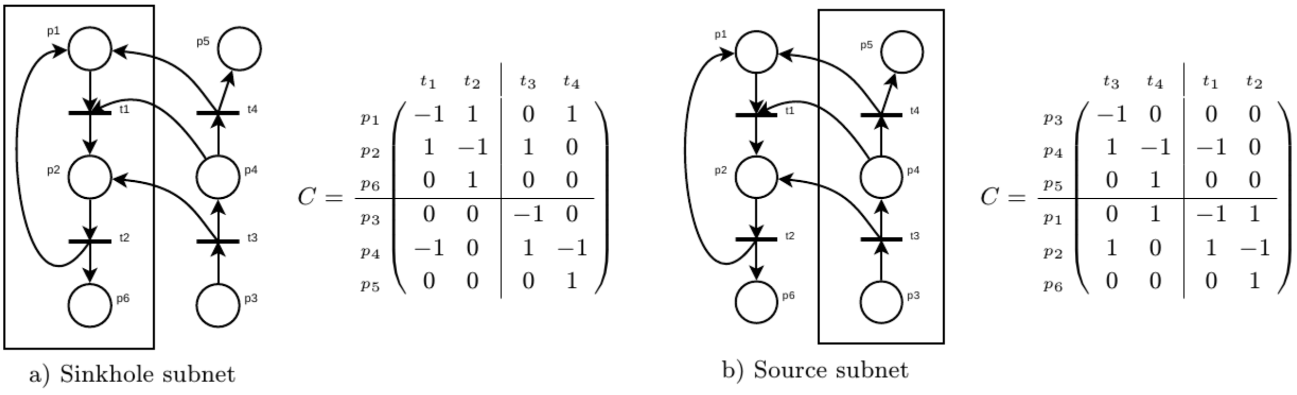
Definition 15 (Output gate of a net). Let a_i *in*OA output arc R_1. We define an output gate of R_1, and denote by og_i, as a new logical node that is identified with an exit arc of the net. For each exit arc is defined an output gate, regardless of the origin and destination of the arc. If the source is a transition, we denote ogt_i and if it is a place, ogp_g.

In this way we can divide the input arcs and output into two parts: a R_1 internal and external to R_1. If we take an arc of entrance ai that has an origin in n_j and destination in n_k, we define an input gate through a point of entry so that the original arc ai is divided into two parts.

- a_{i1} (external to R_1) with origin in n_j and destination in igt_i or igp_i depending on if n_j is a transition or a place.
- a_{i2} (internal to R_1) with destination in igt_i or igp_i depending on if n_j is a transition or a place respectively.

Similarly, if we take a exit arc a_i that has an origin in n_k and destination in n_j, we define an output gate og_i so that the original arc a_i is divided into two parts:

- a_{i1} (internal to R_1) with origin in n_k and destination in igt_i or igp_i depending on if n_j is a transition or a place.



$$C = \begin{array}{c} \\ p_1 \\ p_2 \\ p_6 \\ p_3 \\ p_4 \\ p_5 \end{array} \begin{array}{cc|cc} t_1 & t_2 & t_3 & t_4 \\ -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & -1 & 0 \\ -1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{array}$$

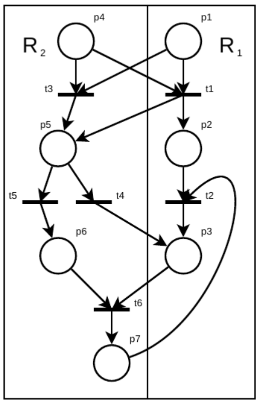
a) Sinkhole subnet

$$C = \begin{array}{c} \\ p_3 \\ p_4 \\ p_5 \\ p_1 \\ p_2 \\ p_6 \end{array} \begin{array}{cc|cc} t_3 & t_4 & t_1 & t_2 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{array}$$

b) Source subnet

Figure 4 – Sinkhole and Source subnets

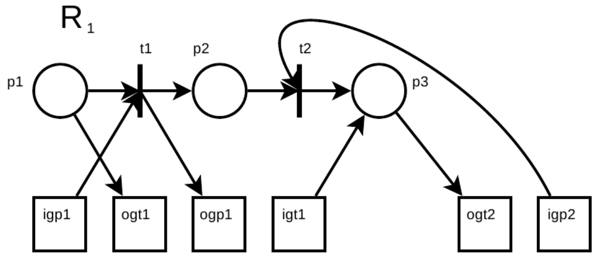
- a_{i1} (internal to R_1) with origin in n_k and destination in igt_i or igp_i depending on if n_j is a transition or a place.
- a_{i2} (external to R_1)with destination in n_j and origin in igt_i or igp_i depending on if n_j is a transition or a place respectively.



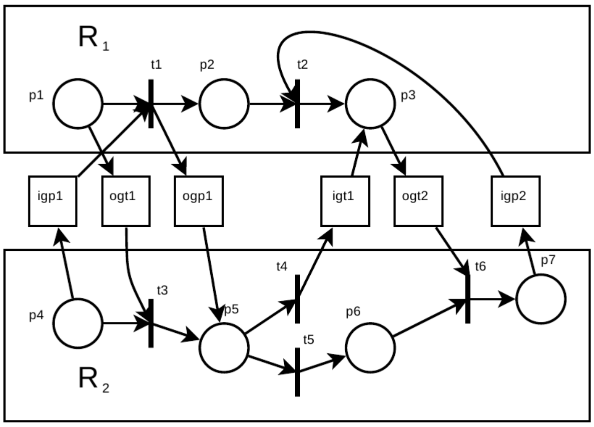
$$C = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \end{array} \begin{array}{cc|cccc} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\ -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ \hline -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 0 & 0 & 1 \end{array}$$

Figure 5 – Subnets with input and output nodes

Example 8. Consider the net in figure 5. In this network we have three arcs entering and leaving three arcs. For each of those emerging define output gates and each coming, we define input gates. The subnet R_1 becomes:



and in the complete net, arcs entering and leaving are divided into two pieces:



Definition 16 (Input Front-end of a net). The input front-end (or input interface) of a subnet R_1 is the set of all input gates of R_1. We denote by *IF* of R_1.

Definition 17 (Output Front-end of a net). The output front-end (or output interface)of a subnet R_1 is the set of all output gates of R_1. We denote by *OF* of R_1.

Definition 18 (Front-end of a net). The front-end (or interface) of a net R_1 is the pair of *IF* and *OF* of R_1. We denote by *F* of R_1.

$$F = \langle IF, OF \rangle$$

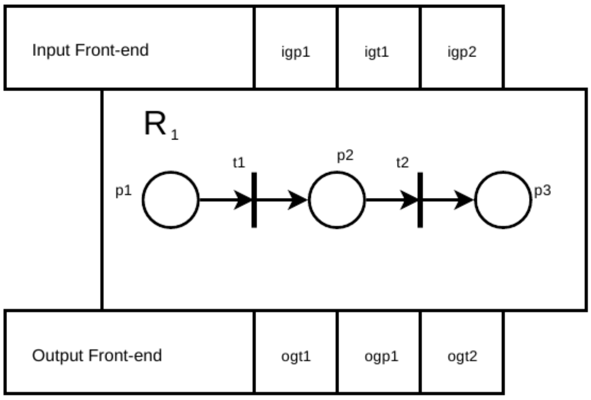


Figure 6 – Front-end of a net

Example 9. Taking the net of the example 8 and applying these new definitions, we would have R_1 net along with its front end as shown in Figure 6.

8.3 Input/output functions

Once all these input and output concepts defined, we will introduce a few key concepts for our purpose.

Let R a Petri net and let $\{R_1, R_2\}$ a partition of R. Let $F = \langle IF, OF\rangle$ the front-end of R_1.

Definition 19 (Petri net Input function). We define the input function f_i of R_1 as:

$$f_i : F \longrightarrow IN$$

such that for each input gate igt_i you mapped one or no input place R_1 and each input gate igp_j you mapped one or no input transition R_1.

$$f_o : ON \longrightarrow F$$

such that each output place R_1 you mapped one or no output gate ogt_i of R_1 and each output transition R_1 you mapped one or no output gate ogp_j to R_1

The input function can be defined for all the input gates and the output function should be surjective because if not, some door would not be connected. Anyway that is not essential. If a front-end door is not connected with any element of your network, simply by solving the final network, the arcs connected to that door disappear. Note also that the input function is not necessarily injective: Multiple input gates can be associated to the same node of R_1.

Example 10. Consider the net R_1 in figure 5 with its front-end in figure 6. The input and output functions are:

- Input function:

F	igp_1	igt_1	igp_2
IN	t_1	p_3	t_2

- Output function:

ON	p_1	t_1	p_3
F	ogt_1	ogp_1	ogt_2

8.4. Attachable net

By joining the subnet R_1 along with its front-end and its input and output functions f_i and f_o we grouped both the internal network with external communication. This way we can " extract" a subnet and "implant" it in another net. You only need this destination network is to communicate with the front-end. So naturally appears the following definition.

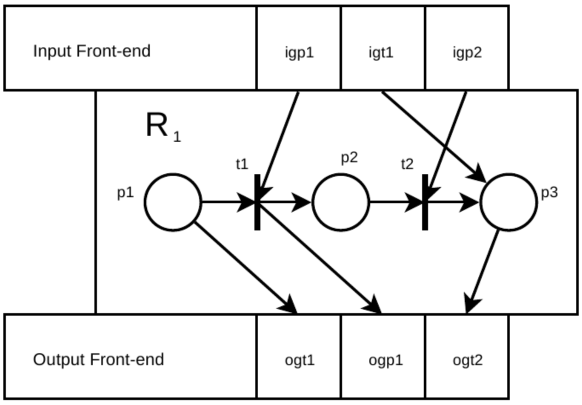
Definition 21 (Attachable Petri net). An [Attachable Petri net is a quadruple $R_a = \langle R, F, f_i, f_o\rangle$

From these definitions, it is clear that you can create attachable subnets taking a subnet of another given and applying the whole process we have defined. But it is also possible to create from scratch, starting from a network, defining a front end for that network and declaring the input and output functions. So you can create Petri nets modules providing functionality and out through a front-end without requiring the actual implementation.

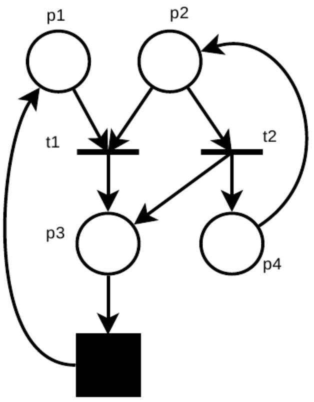
Example 11. The attachable net in figure 5 would be the next:

It can be seen as a private black box with visible input and output connectors that are "plugged" to other networks. In a attachable net, the private part would be R, f_i and f_o. The public part of the front-end would be F. All a net need to know is the input/output front-end.

A utility of these nets is that its definition is simple, since only the front-end is needed to define its operation. This makes possible to create nets using attachable nets in certain areas where they do not know their actual implementation, but its behavior. Additionally, it is possible to use different implementations of "network providers" of the same attachable nets, using at each moment the most appropriate one.



Example 12. Consider now the following Petri net



to which we want to connect an attachable net in the black box. Let's assume we have two equivalent alter natives described in Example 6:

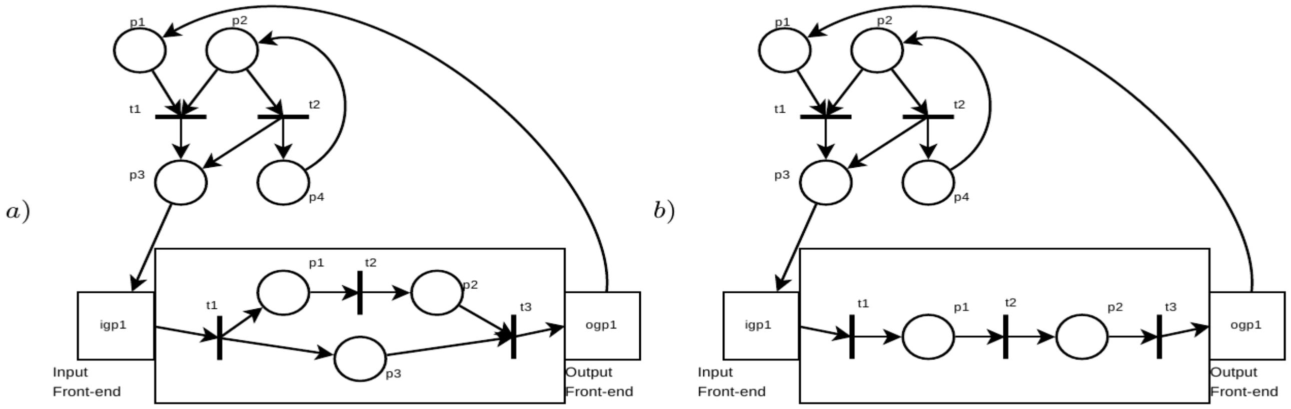
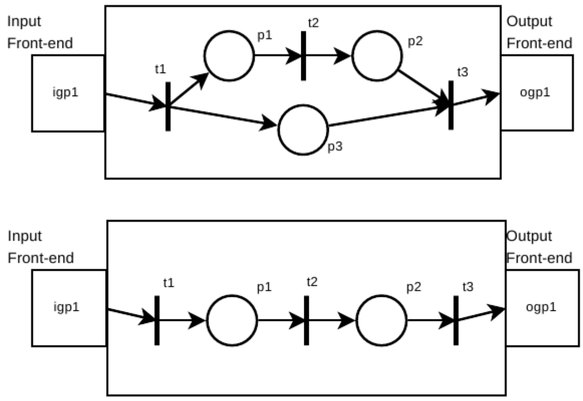


Figure 7 – Two different implementations of attachable nets



We Can ”plug” either because their front ends are equivalent and remains in figure 7.

In this case, the behavior of the net will be the same, but does not have to be. That will decide who connects nets. For example, you could create a ” silly” net that does nothing at first and replace it later by the real one.

9. CONCLUSIONS

Throughout this paper we have presented Petri nets with definitions and basic properties. From this initial presentation, a series of elements have been building as a basis for further investigation. In particular, a type of subtents has been defined, the subnets classifications have been studied, and the front-ends (interfaces) for those subnets have defined.

From this point a further study of these subnets (their properties, utilities, ...) is possible, and constitutes the line of continuity of this piece of research.

Therefore, the main contribution of this work has been to establish the basis for the methodological study of hiding parts of Petri nets.

REFERENCES

David, R. and Alla, H. (2010) Discrete, Continuous and Hybrid Petri Nets. Springer, Berlin. (1st ed., 2004)

Desrochers, A. and Al-Jaar, R.Y. (1995) Applications of Petri Nets in Manufacturing Systems. Modeling, Control ad Performance Analysis. *IEEE Press*, New York.

Guasch, T., Piera, M.A., Casanovas, J. and Figueras, J. (2002) Modelado y Simulación. Aplicación a procesos logísticos de fabricación y servicios. *Edicions UPC*, Barcelona.

Jensen, K. and Kristensen, L.M. (2009) Coloured Petri Nets. Modelling and Validation of Concurrent Systems. *Springer*, Berlin.

Jiménez Macías, E. and Pérez de la Parte, M. (2004) Simulation and optimization of logistic and production systems using discrete and continuous Petri nets. *Simulation*, 80(3), 143-152.

León, I., 2011. Seguridad y protección en envío y almacenamiento de datos. Firmado y cifrado. Aplicación a Redes de Petri y Gestión de Residuos con E3L. *Universidad de La Rioja Eds.*, Logroño, Spain.

Murata, T. (1989) Petri Nets: Properties, Analysis and Applications. *Procs. of the IEEE*, 77(4), 541-580.

Silva, M. (1985)Las Redes de Petri: en la Automática y en la Informática. *Ed. AC*, Madrid.

Silva, M. (1993) Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, 1-62. Chapman and Hall.

Agutter, A.J., 1995. The linguistic significance of current British slang. Thesis (PhD). Edinburgh University.

Xia, C. (2011) Analysis and Application of Petri Subnet Reduction. *Journal of computers*, 6(8), 1662-1669.

A BAYESIAN NETWORK ANALYSIS FOR SAFETY MANAGEMENT

Ciro D'Elia[(a)], Fabio De Felice[(b)], Paola Mariano[(c)], Antonella Petrillo[(d)], Simona Ruscino[(e)]

[(a) (c) (e)]Department of Electrical and Information Engineering - University of Cassino and Southern Lazio, Italy
[(b) (d)] Department of Civil and Mechanical Engineering - University of Cassino and Southern Lazio, Italy

[(a)]delia@unicas.it, [(b)]defelice@unicas.it, [(c)]p.mariano@unicas.it, [(d)]a.petrillo@unicas.it, [(e)]s.ruscino@unicas.it

ABSTRACT

This paper presents a methodology for safety analysis at workplace. The methodology incorporates Bayesian approach to assess the safety associated with safety requirements specifications. In this paper we propose a particular bayesian network, called Knowledge Driven Bayesian Network (KDBN), able to solve the problem of data availability thanks to the particular structure of the network itself. A case study based on marble industry is used to demonstrate the methodology.

Keywords: bayesian network, safety, prevention, decision support analysis

1. INTRODUCTION

The 19th century and first half of the 20th century is one of those periods in history of rapid economical, technical and social changes (Swuste et al., 2010). In this period occupational safety is developing into a professional field. Since then, the concept of safety culture has attracted a great deal of research attention from a range of academic disciplines (Parker et al., 2006; Falcone *et al.*, 2007 a).

Empirical research on safety climate and safety culture has developed considerably but, unfortunately, theory has not been through a similar progression (Guldenmund, 2000; Falcone *et al. 2007* b). Techniques used to manage accident prevention in companies include accident analyses, accident investigations, safety inspections and incident recall, etc. (Martín et al., 2009; Silvestri *et al*., 2012; De Felice and Petrillo, 2012).

Effective approaches to defining the interplay between variables have been developed by authors, for example, using structural equation models (Paul and Maiti, 2007). In the present work we use an approach based on Bayesian Networks (BNs) to describe the circumstances (and relationship between circumstances) associated with tasks performed.

In particular our aim is to provide a "system" for the automatic control of the safety of workers, that, from one side, may lower the cost of development of the security project, supporting the operator, and on the other side may ensure a higher quality of the final result.

Assuming that the effectiveness of a security project strongly depends on the *know-how* of the company that produces it, the existence of mechanisms for the exchange of know-how is of great importance for a company: the stratification of know-how can occur in various ways, such as through the experience of the staff or through the mere storage of past projects.

The proposed system regulates the stratification and the exchange of know-how; it is based on a database, called the knowledge-base, which contains aspects of know-how related to the activities, subject to the risk of dangerous events, which can generate different types of damage; in more detail the database contains the know-how organized as a catalog of predictors of risk associated with work activities.

The application part of the system, based on the written information in the database, automatically calculates the risk to which a worker is subjected when performing a certain activity.

The objective is therefore to allow, on one side, to stratify the experience of the operators, and on the other side to make "repeatable" and "less subjective" the risk assessment of an activity; moreover the use of computational methods for the risk assessment makes the effectiveness of the security project more measurable, both a priori, with the predictors, and a posteriori, with a matching between the statistical data and prediction models.

The paper is structured in section 2 in which literature review is presented; section 3 in which problem statement is analyzed; section 4 in which methodological approach is defined and finally conclusions and results are presented.

2 LITERATURE REVIEW

There has been a steady growth of interest in the application of Bayesian Network (BN) to risk analysis due to its capability to model complex system (Lu *et al*., 2011).

The BN is "a theory of reasoning from uncertain evidence to uncertain conclusions" because it can conduct the factorization of the joint distribution of variables according to the conditional dependencies (Dempster, 1990). BNs have been applied in several knowledge areas.

In Table 1 is shown a brief report on some papers present in literature.

Table 1: Major works on BNs

Authors	Year	Topic
Zhu and Deshmukh	2003	Business risk and product life-cycle analysis
Matías et al.	2008	construction and mining accidents
Adriaenssens et al.	2004	Ecology
Marcot et al. Baran and Jantunen Matías et al.	2001 2004 2006	Environmental assessment impact
Flage et al.	2012	Maintenance optimization model
Antal et al	2007	Medicine
Martín et al. Papazoglou et al.	2009 2006	Risk of falls
Huang and Abdel-At Miranda-Moreno et al.	2010 2013	Traffic and Road safety analysis
Galán et al. Zhang et al.	2007 2013	Workplace risk area

3. PROBLEM STATEMENT

According to the guidelines of the Legislative Decree 81/08, the risk to the health and safety of workers in the performance of their duties can be evaluated through the Equation (1):

$$R = PxD \qquad (1)$$

where D is the magnitude of damage value and P is the probability of occurrence of a dangerous event.

The probability P depends on many factors: it depends on the activity the worker is doing, so it may depend from the equipment used, from the working environment, etc. Be x a generic activity that the worker is doing, the Equation 1 can be rewritten as follow:

$$R_o \ = P(D|x \in X) * Damage\ Value \qquad (2)$$

where X is the set of all the activities that a worker, depending on the role and the working field, is called upon to perform; R_o is the original risk, i.e. the risk to which a worker is subjected if requirements for safety will not be respected.

Furthermore, we can say that, fixed a certain working field and a certain role, a worker is potentially exposed to a set of dangerous events (a vector $\underline{\varepsilon}$), which can produce a damage to the worker.

When some requirements for safety are respected, the risk decreases, hence we introduce $R_t(x)$ as the risk at time t, which is a function of requirements for safety; but the risk also depends on other factors, such as wrong behavior or improper training of workers, or wrong organization of the working area.

We can consider these last three factors as some of causes that can potentially result in a dangerous condition to the worker. Consequently the probability of damage depends on the same factors, hence the whole problem is well described by the joint probability, see Equation 3:

$$R_t(x) = P\left(D|\underline{\varepsilon}, \underline{C}, \underline{DB}, x \in X\right) * Damage\ Value \ (3)$$

where $\underline{C}$ is a the vectors of causes and $\underline{DB}$ is the vectors of duties and bans, i.e. requirements for safety.

In this context, the Bayesian networks are a useful tool for calculating the joint probability. However, the Bayes networks require a huge amount of statistical data to be reliable, which could be not available on the first use of the system.

The statistical data are usually accumulated in long time intervals, more over they are difficult to find in the literature; our proposal is a particular Bayes network, called Knowledge Driven Bayesian Network (KDBN), that solves the problem of availability of data, since, as said before, it allows operators experts in security to transfer part of their know-how in the model of risk assessment, due to the particular structure of the network itself.

The network KDBN exploits the a-priori knowledge of the experts on security and requires a much smaller amount of data to be operative. The result of our study is a model useful to identify the circumstances that have the greatest bearing on workplace accidents during working activities. A real case study will be analyzed.

4. METHODOLOGICAL APPROACH

This paper aims to address two related issues when applying hierarchical Bayesian models for marble industry. A simulation framework was developed to evaluate the performance of alternatives.

4.1 Mathematical model

Face a real problem, with a growing number of variables in relationships between them, requires tools that allow us to manage and assess uncertainty. A quantitative approach that allows the integration of uncertainty in the reasoning, comes from the Bayesian networks: powerful mathematical and conceptual tools that allow applications to manage complex problems with a large number of variables, bound together by probabilistic and deterministic relationships.

Bayesian networks (BNs), also known as belief networks (or Bayes nets for short), belong to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables.

A Bayesian network specifies a joint distribution, which describes the problem, in a structured form, represented dependence/independence via a directed graph; in general, given the bayesian network, the full joint probability is defined as follows, see Equation 4:

$$p\ (X_1, X_2, \dots X_N) = \prod_i p\ \ (X_i|parents(X_i\ )) \qquad (4)$$

where parents (X_i) are all parents nodes influencing the child node X_i.

In this section we describe the proposed model by applying it to a problem of safety in the marble industry. Table 2 shows some statistical data. The proposed method adopts a Bayesian probabilistic model to efficiently manage various uncertainties that can occur in the sector of stone extraction and processing, subjected to dangerous events related to the use of explosives or to cutting tools etc.

As said before, BNs allow the understanding of a complex problem, thanks to the definition of the links between the involved random variables. BNs require two components:

1. The graph structure (conditional independence assumptions).
2. The numerical probabilities (for each variable given its parents).

Thanks to the a-priori knowledge of the problem we can define the structure of the network, depicted in Figure 3.

Set a particular kind of dangerous event (ε = Rising, Slip, Tumble, Crash), the diagram of Figure 3 correlates all the variables involved. The variables in green circles are predisposing the dangerous event, then they represent the causes of the dangerous event. When a dangerous event occurs, it can result in a damage, that can be very serious, serious, slight or very slight; in the white circles the cases that the occurrence of a dangerous event does not involve in damage (Near messes and No Damage) are represented, that are fortunately much more recurring. The variables in gray circles represent the requirements of safety and, if applied, decrease the probability of the variables (in green circles) predisposing the danger. In the proposed model safety requirements are called duties and bans (DB). In the gray circles there are also Personal Protective Equipments (PPE), like gloves, safety shoes and safety glasses, which decrease the probability of the cause, but they also decrease the entity of damage (this type of relationship is represented by arrows from DBs, in gray circles, to damages, in red circles).

The model just described is contextualized in a particular case. Our proposal is a generic model, shown in the Figure 4, which is valid for all cases concerning safety at work. We assume relationship between events and damages is always the same for different events and conditional probabilities can change in relation to the events. In the green circles the generic causes predisposing a dangerous event: for example "poor illumination" and "structural deficiencies" of Figure 4 are represented, in the generic model, with "working area not-adequate" and "organization".

Table 2: Marble industry – Italian statistical data for Accident (2011) - Number of cases 33,178

Events	Damages	Causes	Probability of Damage	Severity of the damage	Personal protective equipment (PPE)
Pick up Slip Falling Bump	Dislocation, distortion, distraction	Structural deficiencies Poor Illumination	Very serious damage	Serious	Safety Shoes Education / Information Safety Signs Ergonomics
Pick up Slip Falling Bump	Contusion	Structural deficiencies Poor Illumination	Very serious damage	Medium	Education / Information Ergonomics
Pick up Slip Falling Slice	Injuries	Equipment non complying	Serious damage	Serious	Education / Information Ergonomics
Pick up Slip Falling Bump	Fracture	Structural deficiencies Poor Illumination	Serious damage	Medium	Education / Information Ergonomics
Skin contact Slice	Foreign objects	Equipment non complying Presence of irritant or flammable substances	Slight damage	Slight	Education / Information Safety goggles Protective gloves
Pick up	Strain injury	Structural deficiencies	Slight damage	Medium	Education / Information Ergonomics
Skin contact Inhale	Injuries caused by other agents	Presence of irritant or flammable substances	Slight damage	Medium	Education / Information Safety goggles Protective gloves
Crush Inhale Ingest	Anatomical loss	Equipment non complying Presence of irritant or flammable substances	Very Slight damage	Medium	Education / Information Ergonomics

In the lower part of the figure, the DBs (i.e. the safety requirements) are represented in the gray circles, which, if respected, lower the probability of the causes and therefore the likelihood of dangerous event.

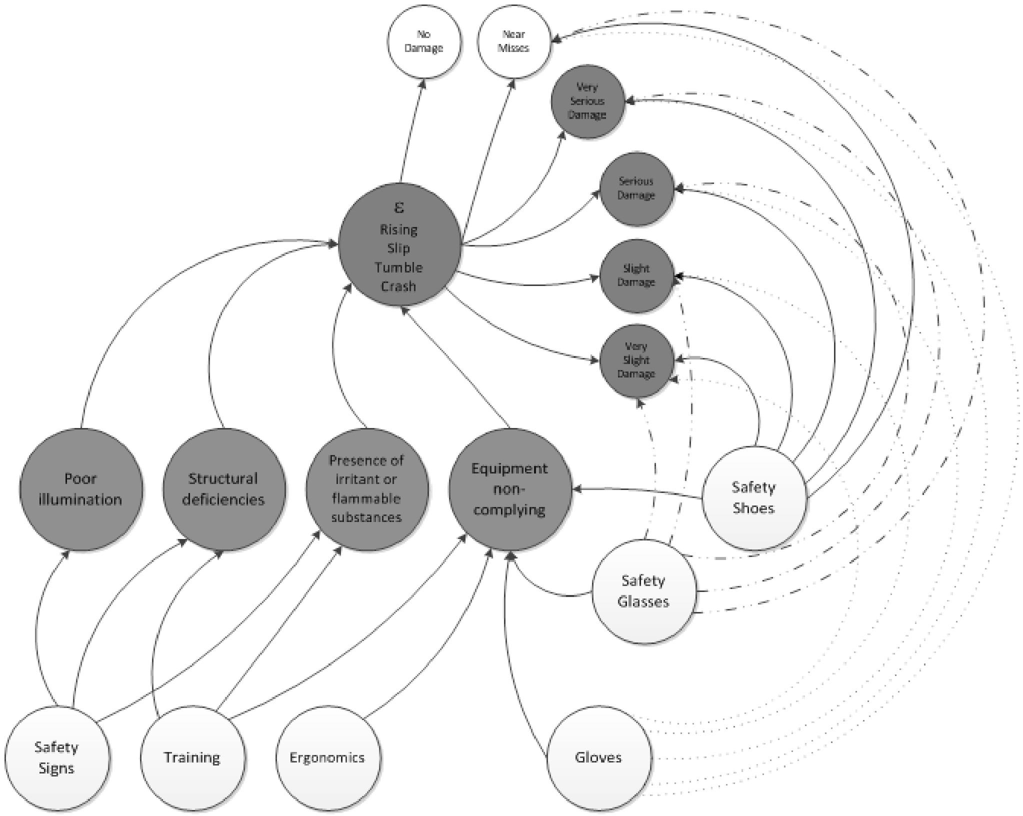


Figure 3: Specific Model

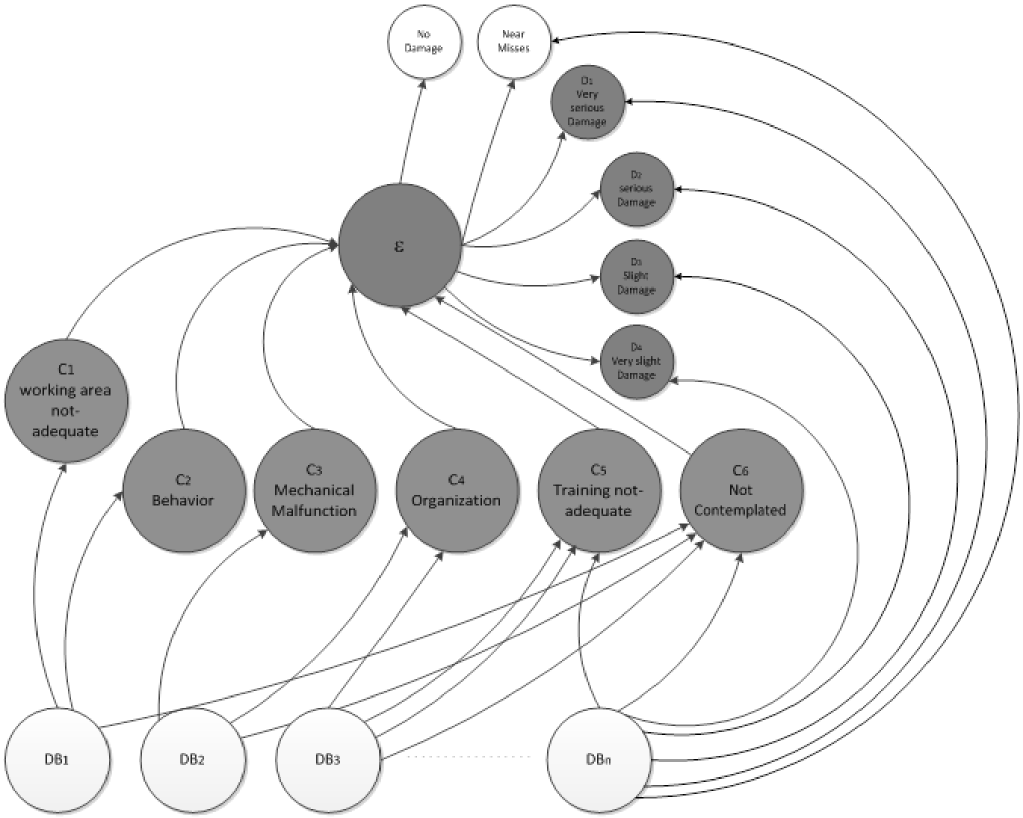


Figure 4: Generic Model

In this paper a risk assessment algorithm, which implements the Bayesian Network already illustrated, are proposed.

The network's learning is achieved by using a set of parameters conveniently chosen, in order to take into account some special or not contemplated cases: for example, no dangerous events can happen, or there is no cause although dangerous events and damages are present, or, even more, no damage occurs even if a dangerous event takes place.

Table 3 illustrates the model's parameters, their physical meaning and the way in which they have been defined and evaluated.

Table 3: Parameters used in the Dynamic risk assessment algorithm

PARAMETERS	PHYSICAL MEANING
$\alpha = P(NotC = 1)$	It represents the probability of the *not contemplated cause*, which corresponds to any cause not comprised into the set C_1, ..., C_6 previously defined. In other words, α indicates how the model is adherent to the reality: as lower the value of α is, as more corresponding to the real situation the model is.
$\beta = p$ where $p = P(C_i = 1) = \frac{\#causes}{\#observations}$	It represents the probability that a given causes, among the set C_1, ..., C_6, occurs. A possible evaluation of this parameter comes from the experience at working sites: it is defined as the ratio between the times in which the given cause C_i occurs and the number of total observations.
$\gamma = 1 - P(\varepsilon_R)$ where $\varepsilon_R = \bigcup_i \varepsilon_i$ represents the combination of all the dangerous events.	This parameter represents the probability that no dangerous events occur.
$\delta = 1 - \frac{\#not\ compliance}{\#observations}$	This parameter represents the company's reliability.
$\phi = \frac{\#days\ \ no\ damages\ occurred}{\#working\ days}$	The meaning of this parameter is expressed in terms of 1-ф, which is the accidents' rate, i.e. the frequency of accidents' occurrence, measured in days and number of injured workers.

4.2 Algorithm implementation

The aim of the risk assessment algorithm is to provide the probability of occurrence of one of the risk reported in Table 4:

Table 4: Types of risk

Ro	Original Risk	$R_o(x) = P(D \mid x \in X) * DamageValue$ It is defined by considering no DBs applied.
R$_t$	Risk at time t	$R_t(x) = P(D_i \vert x \in X, \underline{DB}) * DamageValue$ This kind of risk is computed at time t, by considering only the DBs satisfied (possibly not the DBs requested by the project safety plan.
R$_P$	Project risk	$R_P = P(D_i \vert x \in X, \underline{DB}) * DamageValue$ This kind of risk is evaluated by considering all DBs requested by the project safety plan.
R$_{RES}$	Residual Risk	This kind of risk represents the inferior limit of the risk's value, below which, even all the DBs are satisfied, the risk cannot assume any value.

x is the activity that workers are performing and that typically exposes them to a certain risk. The algorithm is conveniently adjusted on the model's parameters previously described in Table 3. The output of the algorithm is expected to be a dynamic risk R_t that, given the maximum permissible risk R_{MAX}, satisfying the following condition, see Equation 5:

$$R_{RES} < R_P < R_{MAX} < R_o \qquad (5)$$

where the worst (most dangerous) situation occurs when: $R_{MAX} < R_t < R_o$.

In order to calculate the risk at the time t, $\mathbf{R_t}$, given a certain activity x, we are interested in the calculation of the following probability (Equation 6):

$$\boldsymbol{P(D_i \vert \underline{DB})} = \sum\nolimits_{\mathbf{m}} \sum\nolimits_{\mathbf{n}} \mathbf{P(D_i, \boldsymbol{\varepsilon}_m, \underline{C_n} \,\vert \underline{DB})} \qquad (6)$$

which represents the probability of the damage D_i, when a vector of duty and buns are respected. The summation for m takes into consideration all possible dangerous events (ε_m), while the summation for n takes into consideration all possible combination of the causes predisposing the dangerous event ε_m. The probability in the second member of Equation 6, thanks to the relations set in the Bayesian network of Figure 4, can be written as follows (Equation 7):

$$P\left(D_i, \varepsilon_m, \underline{C_n} \middle\vert \underline{DB}\right) = P(D_i \vert \varepsilon_m, \underline{DB}) P\left(\varepsilon_m \middle\vert \underline{C_n}\right) P\left(\underline{C_n} \middle\vert \underline{DB}\right) \, (7)$$

Defined the structure of the network, we need the numerical probabilities, i.e. a sufficient number of statistics that allow us to define the conditional probabilities of the formula. The available data on the sector of marble industry, are unfortunately insufficient, frequently not reliable, incomplete and lacking in degree of detail. The selection and retrieval of data, both historical statistics or deducted through the control and measurements made by experts, is a very important aspect that influence the reliability of the system. A Bayesian network is able to accumulate, in a rather long time, the statistical data defining the problem, that are difficult to find in the literature. However, the system, in order to be reliable at the first start, needs of a great deal of data; in this paper we propose a particular bayesian network, called Knowledge Driven Bayesian Network (KDBN), able to solve the problem of data availability thanks to the particular structure of the network itself, which allows experienced operators (in this case in road safety) to transfer part of their know-how in the system.

Nowadays the term “know-how” means all technical, industrial and commercial knowledge, often secret, of a company, it is a competitive asset of extraordinary importance for any business. Because of its importance, its management must be careful; indeed the know-how is the most fragile asset, its value can be subjected to leaks of information, perhaps caused by disloyal employees; but its fragility is also linked to the difficulty encountered in the transfer of Know-How from experienced to less experienced employees inside the same company.

The KDBN exploits the a-priori knowledge and requires a much smaller amount of data to be operational; it is based on a database of specific knowledge (that could be empty at the first), and allows the expert operator to insert probability data based on its experience and personal evaluations (hence to insert part of its know-how) so as to allow the system to stratify and to share with all in the company.

We discuss below the calculation of each factor in the right side of Equation 7:

- $\boldsymbol{P\left(D_i \middle| \varepsilon_m, \underline{DB}\right)}$

The computation of this probability is related to the parameter Φ, which represents, as described in Table 3, the probability of “No Damage” (see Figure , i.e. $\Phi = \mathrm{P(NotD)}$. The term $\mathrm{P}\left(\mathrm{D_i} \middle| \varepsilon_\mathrm{m}, \underline{\mathrm{DB}}\right)$ must be weighted with (1-Φ) as follows (Equation 8):

$$P\left(D_i \middle| \varepsilon_m, \underline{DB}\right) = \begin{cases} (1-\Phi)\mathrm{P}' & \text{if } D_\mathrm{i} \neq \mathrm{NotDamage} \\ \Phi + (1-\Phi)\mathrm{P}' & \text{if } D_\mathrm{i} = \mathrm{NotDamage} \end{cases} \quad (8)$$

where P’ is given as follows (Equation 9):

$$P' = P(D_i|\varepsilon_m)P_R\left(D_i\middle|\underline{DB}\right) + P(D_{i+1}|\varepsilon_m)\left[1 - P_R\left(D_{i+1}\middle|\underline{DB}\right)\right] \quad (9)$$

The duty and bans (DB), which act directly on the damage (PPE), have the effect of reducing the extent of damage; in other words we can say that the PPEs reduce the probability of the damage D_{i+1}, of greater extent, and increase the damage D_i, of less extent. The equation 9 expresses this concept, where the term $P_R\left(D_i\middle|\underline{DB}\right)$ is the residual probability of the damage D_i, with the application of $\underline{DB}$ and having transfer his discount to the damage D_{i-1}.

- $\boldsymbol{P\left(\varepsilon_m \middle| \underline{C_n}\right)}$

In this case we have to consider the parameter γ, which is the probability of the Not-Dangerous event, in fact (Equation 10):

$$\gamma = P(\varepsilon_{NotD}) = 1 - P(\varepsilon_R) \quad (10)$$

A particular combination of the causes $\underline{\mathrm{C_n}}$ predisposes to a dangerous event ε_m, and, at the same time, reduces the $\mathrm{P}(\varepsilon_\mathrm{NotD})$, this implies that the probability $\mathrm{P}(\varepsilon_\mathrm{NotD})$ is reduced of a portion equal to the sum of all portions subtracted to it and transferred to the dangerous events.
Therefore, set a dangerous event $\varepsilon_\mathrm{m} \neq \varepsilon_\mathrm{NotD}$, the probability is as follows (Equation 11):

$$P\left(\varepsilon_m \middle| \underline{C_n}\right) = P_o(\varepsilon_m) + \sum_j P_o(\varepsilon_{NotD})P_o(\varepsilon_m)P_L(\varepsilon_m|C_j) \quad (11)$$

where $\mathrm{P_o}(\varepsilon_\mathrm{m})$ is the original probability of the event (ε_m) and we assume it is given by the formula below (Equation 12):

$$P_o(\varepsilon_m) = \begin{cases} \gamma & \text{if } \varepsilon_m = \varepsilon_{NotD} \\ \frac{1}{K-1}(1-\gamma) & if\ \varepsilon_m \neq \varepsilon_{NotD} \end{cases} \quad (12)$$

where K is the number of dangerous events.

- $\boldsymbol{P(\underline{C_n}|\underline{DB})}$

It represents the probability of the causes combination $\underline{\mathrm{C_n}}$ to occur given the application of a combination of duty and bans $\underline{\mathrm{DB}}$. The D&Bs, if applied, reduce the possibility that a certain union of causes generates a dangerous event. Therefore, also in this case the discount mechanism can be used: the reduction introduced by the application of the DBs does affect directly the so called ‘Not-Contemplated Cause’, that is any cause which does not belong to the set *C₁,…., Cₖ* (where K is the number of causes define for the network model). We may indicate with $P_R\left(C_i\middle|DB_j\right)$ the residual probability of the cause C_i given that the duty and ban DB_j – hence the discount - has been applied. If all the $\boldsymbol{\underline{DB}}$ are considered, the probability $P(C_i|\underline{DB})$ can be expressed as follows (Equation 13):

$$P\left(C_i\middle|\underline{DB}\right) = p \cdot \prod_{j=1}^{L} P_R(C_i|DB_j) \quad (13)$$

where L is the number of duty and bans applied. p, that represents the marginal probability of the occurrence of the cause C_i, is defined in relation to the α and β parameters previously described:

$$p = \begin{cases} \alpha & if\ C_i\ is\ the\ Not\ contemplated\ Cause \\ & \beta\ \ otherwise \end{cases}$$

In order to define the overall probability $P(\underline{C_n}|\underline{DB})$ a further parameter should be introduced. The expression of this probability is (Equation 14):

$$P\left(\underline{C_n}\middle|\underline{DB}\right) = (1-\delta') \cdot P'' \quad (14)$$

Let's define at first the parameter δ'. Supposing to model the situation in which some of the causes are *on* – i.e. some causes occurred – as a Bernoulli random process, we may define δ' as the probability of having an entire combination of null causes when the bernoullian process is off. The situation of having all causes equal to zero may happen either if the modeling process is off and if the modeling process is on but no cause occurs. This situation can be expressed with the following formulation (Equation 15):

$$\delta = \delta' + (1-\delta')\cdot\delta'' \quad \rightarrow \quad \delta' = \frac{\delta-\delta''}{1-\delta''} \quad (15)$$

where δ is the probability of having all null causes, while $(1-\delta')\cdot\delta''$ is the probability of having all causes equal to zero since the process is on but no causes occurs. δ'' is not a real parameter for the network model, since it depends by the probability $P(C_i|\underline{DB})$ as follows:

$$\delta'' = \prod_{i=1}^{K}(1 - P(C_i|\underline{DB}) \qquad (16)$$

where K is the number of causes of the network model. Turning back to the equation defining $P\left(\underline{C_n}\middle|\underline{DB}\right)$, we have to define what P'' represents. It corresponds to the product of the probabilities $P(C_i|\underline{DB})$ considered as follows:

$$P_c = \begin{cases} P(C_i|\underline{DB}) & if\ C_i = 1 \\ 1 - P(C_i|\underline{DB}) & if\ C_i = 0 \end{cases}$$

$$P'' = \prod_{i=1}^{K} P_c$$

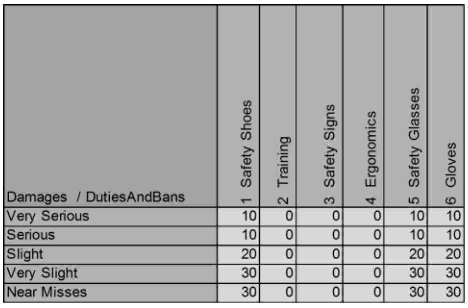
4.3 Data Collection

The risk assessment algorithm implements the computation of the risk's probability regarding a given network starting by the knowledge of the following conditional probabilities: 1) P(ε|C$_i$); 2) P(D$_i$|ε); 3) P(D$_i$|DB); 4) P(C$_i$|DB). These probabilities, whose values are archived in database, defined the KDBN – *Knowledge Driven Bayesian Network*.

They are provided as inputs to the KDBN network as a-priori knowledge coming from the expertise acquired by security-experienced operators.

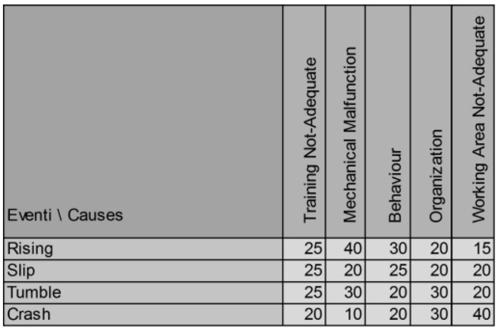
Thanks to this a-priori knowledge, the KDBN network needs a reduced amount of data in order to be operative. In fact it takes as inputs the conditional probabilities indicated in the points from 1 to 4, transferred by experts in the network. In the tables (Tables 5, 6, 7 and 8) below are reported the conditional probabilities used for testing the KDBN on the specific problem presented in this paper.

Table 5: Damages/Duties and Bans. (In the table the $P_R(D_i|\underline{DB})$ inserted in the KDBN by the expert)



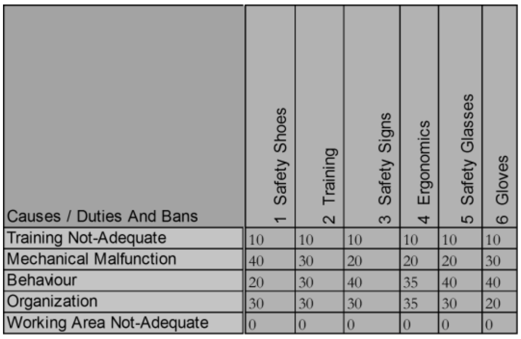
Damages / DutiesAndBans	1 Safety Shoes	2 Training	3 Safety Signs	4 Ergonomics	5 Safety Glasses	6 Gloves
Very Serious	10	0	0	0	10	10
Serious	10	0	0	0	10	10
Slight	20	0	0	0	20	20
Very Slight	30	0	0	0	30	30
Near Misses	30	0	0	0	30	30

Table 6: Events/Causes (In the table the $\boldsymbol{P_L\left(\varepsilon_m\middle|C_j\right)}$ inserted in the KDBN by the expert)



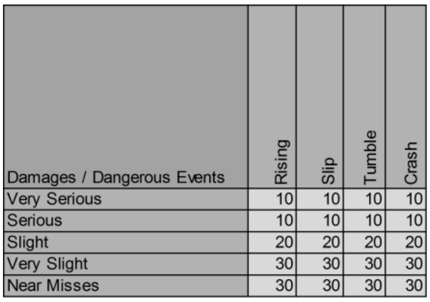
Eventi \ Causes	Training Not-Adequate	Mechanical Malfunction	Behaviour	Organization	Working Area Not-Adequate
Rising	25	40	30	20	15
Slip	25	20	25	20	20
Tumble	25	30	20	30	20
Crash	20	10	20	30	40

Table 7: Causes/Duties and Bans (In the table the $\boldsymbol{P_R(C_i|DB_j)}$ inserted in the KDBN by the expert)



Causes / Duties And Bans	1 Safety Shoes	2 Training	3 Safety Signs	4 Ergonomics	5 Safety Glasses	6 Gloves
Training Not-Adequate	10	10	10	10	10	10
Mechanical Malfunction	40	30	20	20	20	30
Behaviour	20	30	40	35	40	40
Organization	30	30	30	35	30	20
Working Area Not-Adequate	0	0	0	0	0	0

Table 8: Causes/Duties and Bans (In the table the $\boldsymbol{P(D_i|\varepsilon_m)}$ inserted in the KDBN by the expert.)



Damages / Dangerous Events	Rising	Slip	Tumble	Crash
Very Serious	10	10	10	10
Serious	10	10	10	10
Slight	20	20	20	20
Very Slight	30	30	30	30
Near Misses	30	30	30	30

4.4 Results

The model's parameters are set as in Table 7; remember that α indicates how the model is adherent to the reality, as lower the value of α is, as more corresponding to the real situation the model is, so we have supposed a good adherence.

Since statistics show that the stone industry is characterized by a very high accident rate compared to other sectors, we have also supposed a significant presence of the causes predisposing the dangerous event (parameter β), in consequence we have reduced the probability of no-dangerous event (parameter γ and Φ); furthermore, since we supposed safety in the stone industry unreliable, we set parameter δ' near zero.

Table 9: Values of the models parameter use for testing

Parameters	Value
α	0,1
β	0,3
γ	0,8
δ'	0
ϕ	0,8

The following (see Figure 1 and Figure 2) are two types of tests: the first having assumed all DBs respected, in the second we simulated a bad behavior by workers, assuming the breach of some PPE.

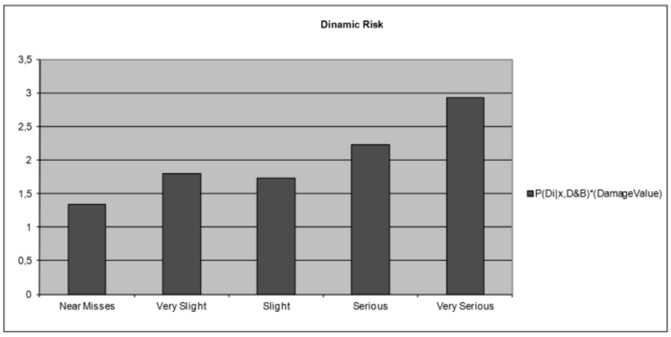


Figure 1: Risk obtained for each type of damage

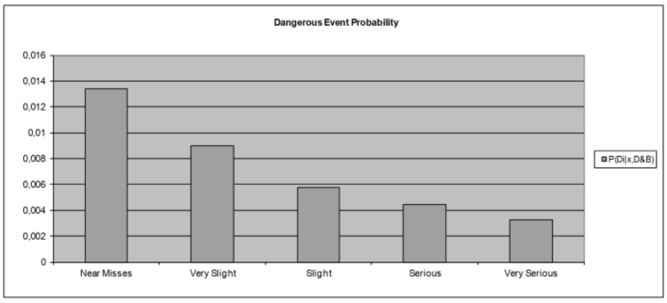


Figure 2: Probability of Damage having assumed all DB respected

In Table 10 is shown numerical results.

Table 10: Numerical results of the graphs in Figure 1-2

Damages	P(Di|x,DB)	Risk
No Damage	0,964120474	0
Near Misses	0,013406225	1,340622473
Very Slight	0,008959827	1,791965372
Slight	0,00578702	1,736106102
Serious	0,004468742	2,234370788
Very Serious	0,003257713	2,931941348

Having respected all the DB, in the first case we get a probability of damage, expressed in days, equal to the probability of having one very serious damage every year, and nearly two slight damage each year. Obviously, the situation gets worse if some DB are not respected, particularly if they are PPE. In the second test we have supposed that one PPE (safety shoes) is not respected, results are shown in Table 11.

Table 11: Numerical results of the second test, obtained supposing not respected the PPE "Safety shoes"

Damages	P(Di|x,DB)	Risk
No Damage	0,961685	0
Near Misses	0,013570838	1,357083815
Very Slight	0,009906712	1,981342369
Slight	0,006649711	1,994913208
Serious	0,004523613	2,261806358
Very Serious	0,003664126	3,29771367

In Table 12 the two cases (test 1 and test 2) are compared in terms of number of damage per year.

Table 12: Comparing results of Test1(all DB respected) and Test2(only one PPE not respected)

Damages	Test1: #damages per year	Test2: #damages per year
Very Slight	3,27	3,61
Slight	2,11	2,43
Serious	1,63	1,65
Very Serious	1,18	1,34

5. CONCLUSION

The model presented in this paper introduces a novel approach to assess safety at workplace. The BN model not only can perform risk assessment but also can help to simulate "critical" situation assessment during the work.

The model has taken into consideration a particular case study concerning safety in marble industry. The assessment results obtained by the BN model offered many useful suggestions to security work for the particular sector, and played a decisive role in reducing risk. In future work risk assessment algorithm could be improved in order to manipulate, with new parameters, the dependence with the structure of the network and the used statistics.

ACKNOWLEDGMENTS

The first application of this research to real context of safety on construction sites has been possible thanks to FILAS financing (http://www.filas.it/) to ISA (Ingegneria Sicurezza Ambiente s.r.l.) for a project called DCSS (Distributed Control Safety System, http://www.lit.ing.unicas.it/?q=ProjectDCSS).

REFERENCES

Adriaenssens, V., Goethals, P.L.M., de Pauw, C.N., 2004. Application of Bayesian belief networks for the prediction of macroinvertebrate taxa in rivers. Annales de Limnologie. *International Journal of Limnology*, 40 (3), 181–191.

Antal, P., Fannes, G., Timmerman, D., Moreau, Y., De Moor, B., 2007. Bayesian applications of belief networks and multilayer perceptions for ovarian tumour classification with rejection. *Artificial Intelligence in Medicine* 29, 39–60.

Baran, E., Jantunen, T., 2004. Stakeholder consultation for Bayesian decision support systems in environmental management. In: *Proceedings of the Regional Conference ECOMOD 2004*, 15– 16 September 2004. Universiti Sains Malaysia, Penang, Malaysia.

De Felice, F., Petrillo, A., 2012. Methodological approach to reduce train accidents through a probabilistic assessment. *International Journal of Engineering and Technology*. Vol. 4 No 6 Dec 2012-Jan 2013. pp. 500-509.

Dempster, A.P., 1990. Construction and local computation aspects of network belief functions[C]//R.M. Oliver, J.Q. Smith, Ed. Influence Diagrams, Belief Nets and decision Analysis. New York: Wiley, 1990: 121-141.

Falcone, D., Di Bona, G., De Felice, F., Silvestri, S., Duraccio, V., 2007 (a). Risk assessment in a co-generation system: Validation of a new safety allocation technique. In *Proceedings of the 16th IASTED International Conference ASM 2007*. Palma de Mallorca, Spain, August 20 – 22, 2007.

Falcone, D., Di Bona, G., Duraccio, V., Silvestri, A., 2007 (b). Integrated hazards method (IHM): A new safety allocation technique. *In Proceedings of the IASTED International Conference on Modelling and Simulation*, Montreal, Quebec, Canada, May 30 – June 1, 2007.

Flage, R., Coit, D.W., Luxhøj, J.T., Aven, T., 2012. Safety constraints applied to an adaptive Bayesian condition-based maintenance optimization model. *Reliability Engineering & System Safety*, Volume 102, June 2012, Pages 16–26.

Galán, S.F., Mosleh, A., Izquierdo, J.M., 2007. Incorporating organizational factors into probabilistic safety assessment of nuclear power plants through canonical probabilistic models. *Reliability Engineering and System Safety*, 92, 1131–1138.

Guldenmund, F.W., 2000. The nature of safety culture: a review of theory and research. *Safety Science*, 34 (2000) 215-257.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention*, Volume 42, Issue 6, November 2010, Pages 1556–1565.

Lu, S., Wu, D., Lu S.C., Zhang H.P., 2011. A Bayesian Network Model for the Asian Games Fire Risk Assessment. *2011 China located International Conference on Information Systems for Crisis Response and Management.*

Marcot, B.C., Holthausen, R.S., Raphael, M.G., Rowland, M.M., Wisdom, M.J., 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under mand management alternatives from an environmental impact statement. *Forest Ecology and Management*, 153, 29–42.

Martín, J.E., Rivas, T., Matías, J.M., Taboada, J., Argüelles, A., 2009. A Bayesian network analysis of workplace accidents caused by falls from a height. *Safety Science*, 47 (2009) 206–214.

Matías, J.M., Rivas, T., Martín, J.E., Taboada, J., 2008. A machine learning methodology for the analysis of workplace accidents. *International Journal of Computer Mathematics*, 85 (3), 559–578.

Matías, J.M., Rivas, T., Ordóñez, C., Taboada, J., 2006. Assessing the environmental impact of slate quarrying using bayesian networks and GIS. In: *Proceedings of the Fifth International Conference on Engineering Computational Technology*. Las Palmas de Gran Canaria, pp. 345–346.

Miranda-Moreno, L.F., Heydaria, S., Lord, D., Fu, L., 2013. Bayesian road safety analysis: Incorporation of past evidence and effect of hyper-prior choice. J*ournal of Safety Research,* Volume 46, September 2013, Pages 31–40

Papazoglou, I.A., Aneziris, O., Post, J., Baksteen, H., Ale, B.J.M., Oh, J.I.H., Bellamy, L.J., Mud, M.L., Hale, A., Goossens, L., Bloemhoff, A., 2006. Logical models for quantification of occupational risk: falling from mobile ladders. In: I*nternational Conference on Probabilistic Safety Assessment and Management*, New Orleáns, May 13–19, 2006.

Parker, D., Lawrie, M., Hudson, P., 2006. A framework for understanding the development of organisational safety culture. *Safety Science*, 44 (2006) 551–562

Paul, P.S., Maiti, J., 2007. The role of behaviour factors on safety management in underground mines. *Safety Science*, 45, 449–471.

Silvestri, A., De Felice, F., Petrillo, A., 2012. Multi-criteria risk analysis to improve safety in manufacturing systems. *International Journal of Production Research*. Vol. 50, No. 17, pp. 4806-4822, 1 September 2012, 4735–4737.

Swuste, P., van Gulijk, C., Zwaard, W., 2010. Safety metaphors and theories, a review of the occupational safety literature of the US, UK and The Netherlands, till the first part of the 20th century. *Safety Science*, 48 (2010) 1000–1018.

Zhang, L., Wu, X., Ding, L., Skibniewski, M.J., Yan, Y., 2013. Decision support analysis for safety control in complex project environments based on Bayesian Networks. *Expert Systems with Applications*, Volume 40, Issue 11, 1 September 2013, Pages 4273–4282

Zhu, J.Y., Deshmukh, A., 2003. Application of Bayesian decision networks to life cycle engineering in Green design and manufacturing. *Engineering Applications of Artificial Intelligence*, 16, 91–103.

DEVELOPMENT AND SIMULATION OF A NEW SCHEME FOR THE AIRCRAFT CLEANING SERVICE

Miguel Mújica [a], Mireia Soler [b], Idalia Flores [c]

[a] Amsterdam University of Applied Sciences
[b] Universitat Autonoma de Barcelona
[c]Universidad Nacional Autónoma de México

miguelantonio.mujica@uab.es, mireia.soler@campus.uab.es, idalia@unam.mx

ABSTRACT

During the last decade with the increase of competition, airlines have set up schemes to lower costs. Their present profit margin has narrowed to the point of not being able to compete with companies whose business model is similar to the low-cost ones forcing them to explore novel ways of managing the available resources in order to keep competitive.
One of the costs is the cleaning service generated by contracting this service and the delays that this operation can cause. The aim of this paper is to propose a new management system for scheduling the on board cleaning service, that lowers current costs, using tools such as modelling with coloured petri nets and simulation.

Keywords: Simulation, coloured Petri nets, cleaning services, aeronautics

1. INTRODUCTION

Years ago, airlines had enough capital to be able to have their planes cleaned on each leg of a journey. Moreover, plane tickets were much more expensive in those days, with flying being luxury and longer stopover times.

During the last decade, with the appearance on the scene of low-cost airlines, airlines have set up schemes to lower costs, as their present profit margin has narrowed to the point of not being able to compete with such low prices as these airlines offer for short and medium-haul flights.

One of the costs is the cleaning service and everything involved with cleaning a plane, such as the cost of hiring this service and the delays that this can cause.
The proposed system is based on modelling stopover times, by simulating an airline's flight schedule during a working day.

2. ECONOMIC STRUCTURE OF AIRLINES.

An airline is an organization or company, devoted to the transport of passengers, freight, mail and, in some cases, life animals, using airplanes for a profit.

The economic structure of the airlines in existence at the present time can be segmented as follows:
- Flag-carrying airlines: these are government-operated airlines. They have a wide variety of planes for short- and medium-and long-haul flights and tend to have a monopoly on domestic flights.
- Traditional airlines: these are private companies for passenger, freight or mail transport. They have a varied fleet of planes and their routes can be short- and medium-and long-haul. These are like the flag-carrying airlines but with the difference that, in this case, governments are not involved.
- Charter airlines: these are companies that transport passengers but on an occasion basis, their method of operation is to study the travel needs of a specific sector of customers. They organize a group of passengers and fit them up with a vacation package with the flight, hotel and excursions included. They usually have a small fleet of planes with capacity for approximately 180 passengers, per plane.
- Low-cost airlines: they supply the low-budget market in exchange for eliminating passenger services. Their strategy is to reduce operational and wage costs in order to be able to give their customers very low and affordable prices per route, thus achieving a broad customer base that goes from people with high net worth to people with a low level of purchasing power who would never been in a position to buy a plane ticket.

3. STOP OVER TIMES AND MAINTENANCE.

The stopover of a plane is the temporary space between consecutive flights when the plane is in the airport. Depending on the type of company, the time and space available, the plane's stopover will be more or less long (Basargan, 2004).
It is worth mentioning that every stopover takes a different length of time, as all the flight schedules are different. Moreover, it is impossible to homogenize the times of all the ground handling processes when the plane has already arrived at the airport.
The steps followed by a typical stopover of an aircraft are:

- Prior preparation for boarding: the lines of passengers are organized then all their hand luggage and documentation is checked.

- The plane arrives at the parking stand.

- Block-In is performed.

- The passengers and bags disembark.

- The plane is fuelled.

-Whether there is a scheduled cleaning, the cleaning team will proceed to clean the plane.

- When the last passenger leaves and the cleaning services have finished, the passengers for the next flight shall be boarded. Simultaneously the bags shall start to be loaded on the new flight.

- During the boarding of passengers, the coordinator shall deliver the necessary documentation to the captain.

- As soon as the plane is loaded with fuel, bags and passengers, the doors are closed.

- The chocks are removed.

- The plane performs the taxiing towards the corresponding runway for the take-off.

3.1. Maintenance of the Airplanes

There are three types of maintenance:

a) Daily check.

Inspect for obvious damage and check the general conditions and security.

b) Minor maintenance.

A-check: performed every 500-800 flight hours, consists in a general inspection of the systems, components and structure of the aircraft and it can take 20-100 man-hours.

B-check: is done every 4-6 months, this is a slightly more detailed check of components and systems and it can take 1-3 days.

C-check: is carried out every 15-21 months or after specific flight-hours determined by the manufacturer, this is a thorough inspection of the structures, the systems and the inside and outside areas of the plane and it can take 1-2 weeks.

c) Major maintenance.

Also called the “Heavy Maintenance Visit”. It covers the full structural inspection program for the airplane. This usually takes about two months and it should be done every 5 years or 30,000 flying hours.

On the other hand, it sometimes happens that a plane goes into AOG (Aircraft On Ground) which means that the plane has a problem that is sufficiently serious to stop it from making the next flight. In this case, the maintenance team needs to go to the plane to solve the fault.

4. OPERATION OF THE CLEANING SERVICE IN AIRLINES

There are two ways of delivering the service:

a) Subcontracting a company. Every week, they receive the stopovers schedule of each airplane and the pair of origin and destination of the flights. With this information, the cleaning service is scheduled, without any modification throughout the week.

b) Performed by flight attendants. They are in charge of cleaning the planes. The flight attendants have signed an agreement in which they agree to do these types of procedures and accept the conditions imposed by the airline. The aim of this method is to reduce the stopovers between one flight and another. This way the plane spends more time in the air during the day.

5. CASE STUDY OF A SPANISH AIRLINE

A new scheme for the cleaning operations during stopovers has been developed. The proposed scheme uses information that has been provided by a Spanish airline through a confidentiality agreement. We shall refer to this airline, when applicable, as “the airline”. The information of the schedule of one day has been used for the model. The proposed schema is a particular one for the case of the airline, but it can be extrapolated for the case of other airlines in a very straightforward way.

5.1. Current cleaning activities

The following are the cleaning operations currently under use by the airline.

Stopover cleaning.

This is the quickest way of cleaning and applies to stopovers that last for more than 40 minutes, as well as being the most common because, as the name says, it is done during stopovers and it takes 8-14 minutes.

Extra cleaning.

This type of cleaning is unscheduled. The crew or maintenance asks for some of the stopover cleaning jobs to be done. There can be an unexpected use of the temporary space of the stopover time. This type of cleaning does not share all the characteristics of the stopover cleaning. It only makes a required part of it. However the service is charged as a stopover cleaning.

There were 137 extra cleanings during the month of study.

Overnight cleaning.

This type of cleaning is done 4 or 5 times a week, when the plane spends the night in an airport. As this cleaning takes a long amount of time, it is done at night. It aims to improving the plane’s level of disinfection and cleanses places that cannot be reached during the stopover due to the lack of time.

Deep cleaning.

This is a type of cleaning designed to totally disinfect and clean the interior of the airplane. For this purpose, all the seats and luggage compartments are dismantled. As it takes too long, it is performed at night and once in a month.

5.2. Impact of Cleaning Operations

The delays in the aviation industry are one of the most important problems that the sector faces nowadays. Due to the complexity and precedence relationships of the aviation network one delay or primary delay caused in an airport will propagate easily to the rest of the network. Furthermore if more primary delays occur during the day, at the end of the day the accumulated delay would be sometimes huge (Jetzky 2009, Guest 2007). Every minute's delay in the departure of a flight signifies an increase in the different rates that the airport imposes on the airline.

The delays that directly affect the airline and a flight are mainly because of:

- Handling
- Airport authority
- Auxiliary Services
- Safety
- Meteorology

Cleaning service is a portion of the auxiliary services, in which it generates 65% of the delays in scheduled flight times for the airline.

The main characteristics of the current operation can be defined as:

- It is an inflexible system that does not adapt to the stopover times that airlines need under a fierce competitive market.

- The number of cleanings can and must be reduced.

- It does not make much sense to charge for an extra cleaning as if it were a stopover cleaning since the cleaning performed is more superficial.

- The delays caused by the cleaning operation can and must be reduced.

- More variables should be taken into account when a cleaning is assigned, such as, the number of passengers transported, number of previous cleaning among others.

- The current cleaning schedule is fixed and does not admit the variability produced by a plane breaking down or a request for an extra cleaning.

5.3. A novel operative schema for managing the stopover times

There is a very high cost in having the plane standing due mainly to the high tariffs demanded by the airports. Moreover, if the stopover times during the day are shortened, a plane can fly more hours, in other words the useful life of the plane would be maximized. The more hours a plane fly, the more flights it can do, the more passengers it can transport and the less expenditure on airport tariffs is incurred.

For these reasons, airlines seek to reduce the time their planes spend in airports and to increase the number of flights per plane.

However, shorter stopover times make the ground handling of the plane all the harder.

To better manage the stopover time, a cleaning system that fits with current needs must be designed.

New stopover times have been proposed and they are organized into 4 groups that are presented in Table1

Table1: Length of Stopover Time

Groups	Length of Stopover Time	Description
Group 1	Less than 41 minutes	A very short stopover is contemplated
Group 2	Between 41 and 50 minutes.	A short stopover is contemplated
Group 3	Between 51 and 60 minutes.	A medium/long stopover is contemplated
Group 4	Over 60 minutes.	A long stopover is contemplated.

Using the proposed segmentation, a cleaning management system has being designed for these new stop over times.

The proposed model has 5 cleaning types:

- Cleaning 1. This type of cleaning has been designed to give a basic and fast service, it takes 5-8 minutes. It shall be assigned in a very short stopover or when the last cleaning is type 4 or 5.

- Cleaning 2. This type of cleaning gives the same service as the stopover cleaning in the actual model. It shall be assigned in a short and medium stopover or when the last cleaning is type 4 or 5.

- Cleaning 3. This type has been designed to give a good level of cleaning in medium and long stopovers, and also to set back the cleaning number 4 and 5.

- Cleaning 4. This type of cleaning is just done once a week during long stopovers, to give a better level of disinfection and also set back the cleaning number 5.

- Cleaning 5. This type of cleaning is the same as the deep cleaning in the actual model, yet it can be done every month and a half.

6. DESCRIPTION OF THE CAUSAL MODEL

A causal model is proposed for evaluating the cleaning operations, in which stopover times are grouped according to the above division. The objective of the causal model is to assess the validity of the proposed schema while at the same time evaluate the magnitude of savings that can be achieved.

The model was developed in the coloured petri net formalism and tested using the CPNTools program.

6.1. Coloured Petri Nets

Coloured Petri Nets (CPN) is a simple yet powerful modelling formalism which allows to properly modelling discrete-event dynamic systems which present a concurrent, asynchronous and parallel behaviour (Moore et al. 1996, Jensen 1997, Christensen et al. 2001). CPN can be graphically represented as a bipartite graph which is composed of two types of nodes: the place nodes and the transition nodes. The entities that flow in the model are known as tokens and they have attributes known as colours.

The formal definition is as follows (Jensen1997):

$$CPN = (\sum, P, T, A, N, C, G, E, I)$$

Where

- $\sum = \{$ C1, C2, … , Cnc$\}$ represent the finite and not-empty set of colours. They allow the attribute specification of each modelled entity.

- P = { P1, P2, … , Pnp} represent the finite set of place nodes.
- T = { T1, T2, … , Tnt} represent the set of transition nodes such that P $\cap$ T = $\varnothing$ which normally are associated to activities in the real system.
- A = { A1, A2, … , Ana} represent the directed arc set, which relate transition and place nodes such as A $\subseteq$ P $\times$ T $\cup$ T $\times$ P
- N = It is the node function N(Ai), which is associated to the input and output arcs. If one is a place node then the other must be a transition node and vice versa.
- C = is the colour set functions, C(Pi), which specify for the combination of colours for each place node such as C: P $\rightarrow\sum$.

  $$C(P_i) = C_j \qquad P_i \in P, C_j \in \sum$$
- G = Guard function, it is associated to transition nodes, G(Ti), G: T $\rightarrow$ EXPR. It is normally used to inhibit the event associated with the transition upon the attribute values of the processed entities.
- E = these are the arc expressions E(Ai) such as E: A $\rightarrow$ EXPR. For the input arcs they specify the quantity and type of entities that can be selected among the ones present in the place node in order to enable the transition. When it is dealing with an output place, they specify the values of the output tokens for the state generated when transition fires.
- I = Initialization function I(Pi), it allows the value specification for the initial entities in the place nodes at the beginning of the simulation. It is the initial state of a particular scenario.
- EXPR denotes logic expressions provided by any inscription language (logic, functional, etc.)
- The state of every CPN model is also called the marking which is composed by the expressions associated to each place p and they must be closed expressions i.e. they cannot have any free variables.

6.2. Model Definition

The model is divided into two main modules:

1) Decision-making: the necessary information is collected to decide what the model is going to do. The results of this decision are:

The plane does not have to be cleaned.
The plane has to be cleaned.
The plane has suffered a problem.
An extra cleaning has been requested.

2) As soon as the decision has been taken, the plane shall be sent to the corresponding section of the model to execute the next task. The variability is integrated in the model through the use of two variables that simulate the situations that the plane undergoes a breakdown or an extra cleaning is requested.

The developed model in CPN is composed by 11 place nodes and 5 transition nodes. Table 1 describes the place nodes of the model.

Table 1: Place Nodes

Place	Colour	Description
Airplanes	airplane=product(ac*sa*p*h *te1*te2*te3*te4*nt*a*q*n*s)	The initial state of this place has 27 tokens with the information of the first flight of each airplane. This place will keep track of the status of the airplanes.
Next stopover	new=product(ac*sa*sa2*p1* h1*ne1*ne2*ne3)	This place has the flight schedule information for each airplane, except the first flight.
AOG	aog	This place has 170 tokens to generate the airplane-break-down probability
Extra Cleaning	le	This place has 252 tokens to generate the request –extra-cleaning probability.
Control	y	This place controls the activation of transition 1 or 2.
Decision	airplanes1=product(p*h*te1 *te2*te3*te4*nt*s*a*q*n*b*x* y*ne1*ne2*ne3*ne4*up)	This place receives and sends the information of the next step of the airplane process.
New stopover without cleaning	airplane=product(ac*sa*p*h *te1*te2*te3*te4*nt*a*q*n*s)	This place receives a token whether the airplane does not have to be cleaned, which means the airplane will do the next flight without the need of cleaning.
Aircraft in AOG	airplane=product(ac*sa*p*h *te1*te2*te3*te4*nt*a*q*n*s)	This place receives the token whether the airplane breaks down and needs major reparation.
Counter	ne=product (u*d*tr*cu*ci*ex)	This place keeps track of the number of times the aircraft has been cleaned.
Solution	airplanes1=product(p*h*te1 *te2*te3*te4*nt*s*a*q*n*b*x* y*ne1*ne2*ne3*ne4*up)	This place records the final state of the aircraft.
Cleaning	airplane=product(ac*sa*p*h *te1*te2*te3*te4*nt*a*q*n*s)	This place records the information necessary to decide if airplane has to be cleaned.

Table 2 presents the colour definition used in the CPN model of the new cleaning system.

Table 2: Colours and Definitions

Colour	Definition
Ac	Aircraft identification.
Sa	Flight identification.
H	Amount of minutes that the aircraft has flown since the last cleaning service.
P	The total of passengers that has been transported since the last cleaning service.
te1	Whether the stopover is in the first group of the table 1.
te2	Whether the stopover is in the second group of the table 1.
te3	Whether the stopover is in the third group of the table 1.
te4	Whether the stopover is in the fourth group of the table 1.
Nt	The type of cleaning that was done last time.
A	The amount of minutes that the aircraft has flown since the last cleaning number 5.
Q	Whether the plane can carry out all types of cleaning.

Table 2 (cont.)

Colour	Definition
N	The amount of minutes that the airplane has flown since the last cleaning number 4.
S	The number of flights has flown the aircraft, since the last cleaning service.
Up	The operational status of the aircraft.
E	Whether an extra cleaning has been requested.
sa2	Next flight identification.
h1	The duration of the next flight.
p1	The number of passengers will be transported on the next flight.
ne1	Whether the next stopover is in the first group of table 1.
ne2	Whether the stopover is in the second group of table 1.
ne3	Whether the stopover is in the third group of table 1.
ne4	Whether the stopover is in the fourth group of table 1.
U	Cleaning counter of type 1
D	Cleaning counter of type 2
Tr	Cleaning counter of type 3
Cu	Cleaning counter of type 4
Ci	Cleaning counter of type 5

The model has been run using the information of a particular day in which the airline had 27 operative aircrafts.

Figure 1 presents transition T1, which would receive the information related to the actual and future flights, the operational status of the incoming aircraft and whether an extra cleaning has been requested.

The outcome information of the transition will be used to decide the next step of the airplane.

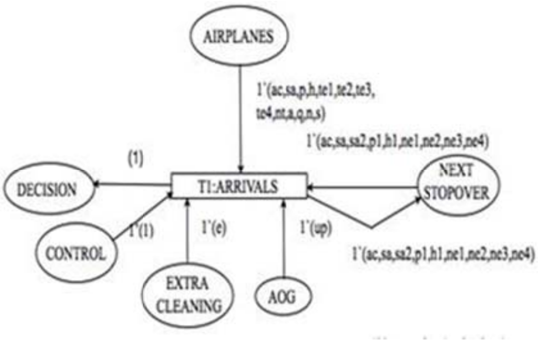


Figure1: Transition T1

Arc (1): This arc has the restrictions to decide the next step of the airplane. It evaluates the operational status of the aircraft, whether is necessary a cleaning service or it has being requested an extra cleaning. The outcome information will assign what the next step of the aircraft is. This information is evaluated by the second transition.

Figure 2 illustrates transition T2; it receives the information about what the next step of the aircraft will be and based on that information it will send the aircraft to the corresponding place.

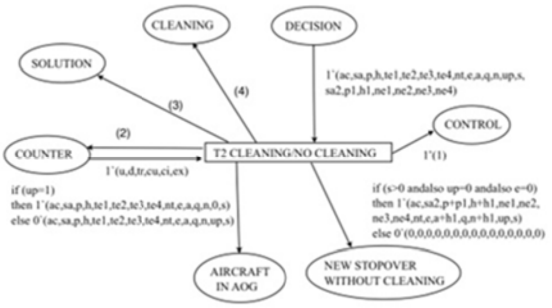

CLEANING
DECISION
SOLUTION
1`(ac,sa,p,h,te1,te2,te3,te4,nt,e,a,q,n,up,s,
sa2,p1,h1,ne1,ne2,ne3,ne4)
(4)
(3)
CONTROL
(2)
COUNTER
T2 CLEANING/NO CLEANING
1`(1)
1`(u,d,tr,cu,ci,ex)
if (up=1)
then 1`(ac,sa,p,h,te1,te2,te3,te4,nt,e,a,q,n,0,s)
else 0`(ac,sa,p,h,te1,te2,te3,te4,nt,e,a,q,n,up,s)
if (s>0 andalso up=0 andalso e=0)
then 1`(ac,sa2,p+p1,h+h1,ne1,ne2,
ne3,ne4,nt,e,a+h1,q,n+h1,up,s)
else 0`(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
AIRCRAFT
IN AOG
NEW STOPOVER
WITHOUT CLEANING

Figure 2: Transition T2

Arc (2): this arc evaluates the restrictions related to what type of cleaning will be performed in the airplane and it will increase the value of the correspondent cleaning counter. The place node contains the information about which flight must be cleaned.

Arc (3): this arc send the current status of the data information to the correspondent place node (SOLUTION). The income data contains the information of the flight that must be cleaned and the next flight. The SOLUTION place node keeps track of the current status of the system.

Arc (4): this arc evaluates the information of the tokens concerning what type of cleaning operation shall be performed. The decision takes into account the stopover time, the information of the airplane and the information of the flights. The outcome information will assign the type of cleaning to be performed. The information is used by the fourth transition.

The Figure 3 presents transition T3. This transition represents the outcome when the airplane does not need a cleaning service. The data will be updated with the information of the next flight and passed through with the token colours to the AIRPLANES place node.

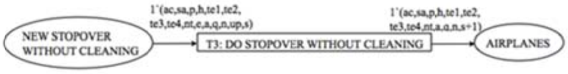


Figure3: Stopover without cleaning

Figure 4 presents transition T4, it evaluates the variables to assign a cleaning in the next stopover. The data will be updated using the token created in the AIRPLANES place node.

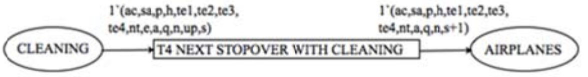

1`(ac,sa,p,h,te1,te2,te3,
te4,nt,e,a,q,n,up,s)
1`(ac,sa,p,h,te1,te2,te3,
te4,nt,a,q,n,s+1)
CLEANING
T4 NEXT STOPOVER WITH CLEANING
AIRPLANES

Figure 4: Stopover with cleaning

Finally Figure 5 presents transition T5. This transition evaluates the correspondent variables and simulates an AOG to the correspondent Aircraft. Once the AOG has been performed, the variables' data is updated through the correspondent token created in the AIRPLANES place node.

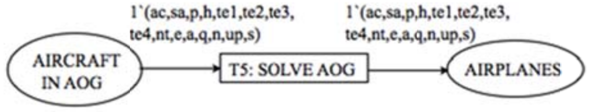

1`(ac,sa,p,h,te1,te2,te3, te4,nt,e,a,q,n,up,s)
1`(ac,sa,p,h,te1,te2,te3, te4,nt,e,a,q,n,up,s)
AIRCRAFT IN AOG
T5: SOLVE AOG
AIRPLANES

Figure 5: Solve AOG

6.3. Analysis of the causal model.

To evaluate the proposed system, the model was simulated 15 times. Table 3 presents the results obtained with the simulation.

Table 3: Simulated Results from the causal model

Results	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Average value
Cleaning 1	17	18	18	18	18	18	18	18	18	18	19	18	17	17	18	17,93
Cleaning 2	15	14	14	14	15	14	15	15	15	15	16	15	15	15	15	14,80
Cleaning 3	11	11	10	11	11	12	11	9	11	11	11	11	11	11	11	10,87
Cleaning 4	3	3	4	3	2	3	3	3	3	3	3	3	3	3	3	3
Cleaning 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Extra Cleaning	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

On the other hand, it is possible to evaluate the cost impact of implementing the new schema. The cost analysis can be appreciated in Table 4.

Table 4: Economic analysis

Economic Variables	Actual System	Proposed System	Difference between Systems
Number of cleaning flights	1978	1310	668
Number os extra cleanings	133	18	115
Percentage of cleaned flights	43,84%	29,03%	14,81%
Cost of the cleaning service *	€ 49.421	€ 24.675,8	€ 24.745,20
Airport Rates	€ 37.895,99	€ 34.102,69	€ 3.793,30
Percentage of delayed flights	19,32%	9,64%	9,68%
Number of delay flights	493	50	443
The cost of delay flights	€ 4.317,66	€ 340,87	€ 3.976,79
RESULTS	€ 91.634,65	€ 59.119,36	€ 32.515,29

Through the results, it can be concluded that the proposed model is less expensive than the actual model due to:

- Creating more types of cleaning with different durations makes the cleaning service more flexible which means the cleaning service has been adapted to the stopovers time. The number of cleaning operations has been reduced due to the flexibility achieved. With the proposed model only the 29,03% of flights were cleaned rather than 43,84% of the current schema.
- The amount of delay flights has been reduced. With the proposed model only the 9,64% of the flights were delayed by the cleaning service rather than the 19,32% of the current schema.
- The proposed schema decreases the number of extra cleanings. With the proposed model it is needed 18 extra cleaning rather than 133 extra cleaning in the current system.
- The total amount of cost in the proposed model for November's month is € 59.119,36 rather the € 91.634,65 of the current schema.

7. VALIDATION OF THE MODEL

The previous results have been validated using a discrete-event-oriented simulation program (SIMIO) in which the complete elements of the Turnaround of a A320 aircraft have been taken into account. The purpose of the simulation model is twofold, on the one hand to evaluate the results of the CPN causal model and on the other to include all the elements of an actual turnaround that could not be included in the causal model. The final goal is to obtain a better management for the turnaround process that allows mitigating the delays caused by the current management schema.
Figure 6 shows a snap shot of the graphical aspect of the model for the turnaround of the Aircrafts of the company (Airbus-A320).

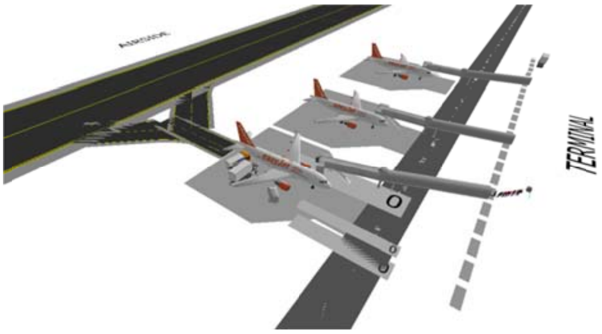


Figure 6: Virtual environment

Several operations occur during the stop over: Catering, Fuelling, Disembarking-Boarding of Passengers, Cleaning. In Figure 6 three trucks can be appreciated, 1 big truck performs the catering operation, the one under the wing is fuelling the aircraft and the one in the rear position of the aircraft is cleaning the system from organic disposals. It can also be appreciated that the passengers are deboarding the plane through the fingers. Is important to note that in the particular case of the fueling operation it does not start until all the passengers have left the aircraft; this is due to security reasons.
In the turnaround process some activities has been identified as being the critical path of the turnaround time. Figure 7 illustrates the total operations that can be performed in such an aircraft and the ones that are part of the critical path of the process (AIRBUS 2012).

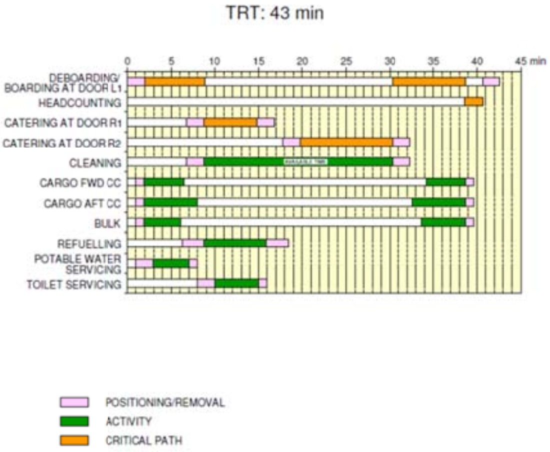


Figure 7: Operations of the turnaround for a A320

The airline of the study does not perform all the operations; in order to reduce the turnaround time the company perform only a few operations, namely boarding, deboarding, catering on door R2 only, cleaning, refuelling, cargo operations and toilet servicing. Under this operative schema the cleaning operation becomes part of the critical path that determines the turnaround time of the aircrafts.

Parameters of the Simulation Model

In order to assess the importance of the cleaning operations, the current process was simulated using information provided by the airline. Table 5 presents the values used for the model.

Table 5: Simulation parameters

Operation	Time
Opening/closing doors	2 min
Deboarding Rate	22 pax/min
Deboarding Rate/pax	Triangular (2.5,2.7,3) sec
Boarding Rate	18 pax/min
Boarding Rate/pax	Triangular (3,3.3,5) sec
Fueling Time	Triangular (7,8,9) mins
Cleaning Operation	Triangular (8,13,16) mins
Full size trolley equivalent (FSTE) to unload/load	7 for R2
Load Time of each Trolley	1.5 min/FSTE
Catering Equipment Position/Removal	2 min
Probability of Cleaning	0.4348
Probability of Extra Cleaning/P.of Cleaning	0.0664

The previous data was used for developing the turnaround model for the current and the proposed schema. The last two rows were obtained from the information provided by the causal model. The first value (P. of Cleaning) is the probability that the aircraft performs a cleaning operation; and the second value corresponds to the conditional probability of an extra cleaning once the cleaning has been performed. The rest of the values will be the same for the current scenario and the proposed one.

7.1. Evaluation of the Proposed Schema

The simulation model was used for analysing the current operations and at the same time obtaining different values that provide insight about the inefficiencies present in the current performance. The second scenario will be implemented assuming new values for the cleaning operations (based on the results provided by the causal model).

Current Operations

The simulation model was run with the aforementioned values and the turnaround times, number of extracleanings and delays were analyzed. Table 6 presents the results obtained with the current operations.

Table 6: Information from the current process

	Cleaning Operation			
	AVG	Min.	Max.	STD. Dev.
Max. No. of Extra Cleanings	6.7	3	12	2.306
Max.No. of Total Delays	37.43	13	81	15.904
Turnaround Times	38.59	37.17	40.9	0.8262
Max. Turnaround Times	54.38	49.14	59.31	1.8539

The previous values were obtained of a total of 240 flights and the simulator was run for 30 replications. As it can be appreciated the first row gives information about the maximal number of extra cleanings, the second row about the maximum number of total delays and the last two rows the average and maximal turnaround times for this scenario.
In the case of the delays it should be pointed out that the upper bound of delays is 81 out of 240 flights which correspond approximately to 33% of the scheduled flights incurred in a delay. On the other hand the maximal turnaround times which are the upper bound for the model mean that some aircrafts could have a turnaround time of 59 minutes which would be translated into a big cost penalty for the airline.

Proposed Schema

The new schema was tested using the same values of the standard operations but in this case the probability of the cleaning operation and the conditional probability of an extra-cleaning once the cleaning operation has been performed are 0.2903 and 0.0137 respectively.
The cleaning times in this new schema also change to a Triangular(5,7,8) since it is assumed that the aircraft in the simulation model are only of group type I. Table 7 presents the results obtained with the proposed schema.

Table 7: Proposed Scenario

	Cleaning Operation			
	AVG	Min.	Max.	STD. Dev.
Max. No. of Extra Cleanings	1.65	1	4	0.8846
Max.No. of Total Delays	12.42	1	56	18.69
Turnaround Times	37.57	36.03	39.47	0.9127
Max. Turnaround Times	40.4	38.57	43.47	1.235

From the previous table it can be appreciated that the mean average turnaround time has been reduced about a minute. As it will be clear with the next figure, the most important achievement is that the dispersion or variability is drastically reduced. As a consequence the probability of delays has been reduced as it can be appreciated in Figure 8.

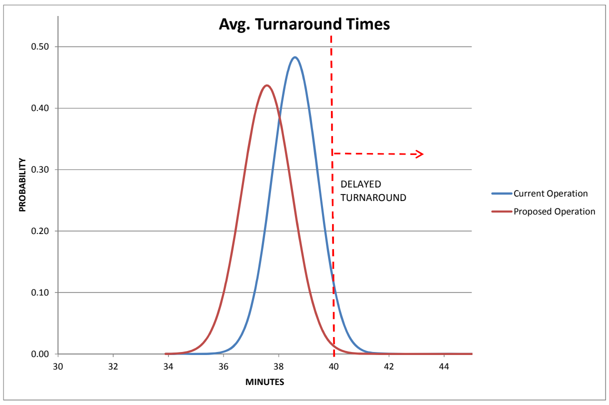


Figure 8: The reduction in the avg. turnaround times

With the new implementations and with the dispersion obtained from the simulation model, it can be appreciated that the curve of the new schema falls within the acceptable region while with the current operations approximately 33% of the flights incur in delays.

On the other hand, if the worst-case scenarios are analysed (i.e. the max. turnaround times) the improvements are more evident. As it can be appreciated from Figure 9, the worst-case values from the current operations fall out of the accepted region while with the new schema only approximately the 50% of the worst-case turnaround times would incur in a delay.

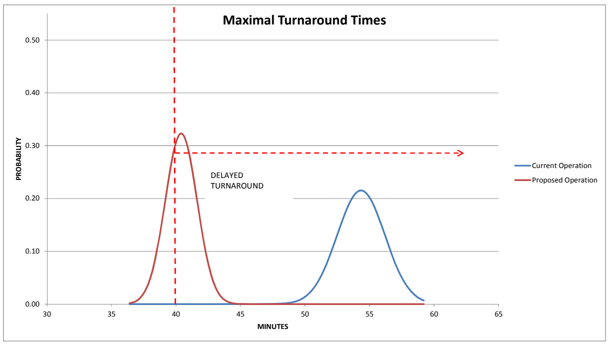


Figure 9: The worst-case scenarios

8. CONCLUSIONS

In this article a new cleaning schema for an airline was devised with the objective of reducing the costs of extra-cleanings and to avoid as much as possible the probability of delays in the turnaround time. The proposed schema has been analysed using a causal model developed using the coloured Petri net formalism and it has been validated with a more detailed simulation model that takes into account all the different operations that are critical for the turnaround time. The results clearly indicate that it is possible not only to reduce the extra-cleanings which is a common practice for a commercial airline but also reducing the possibility of incurring in delays due to the cleaning operations.

ACKNOWLEDGMENTS

The authors would like to thank the Mexican Council of Science and Technology (CONACYT) for supporting the research, and UNAM-DGAPA project PAPIIT IN116012-3.

APPENDIX A

Definitions

Chocks: A block or wedge placed under the aircraft wheels, to keep it from moving

Medium haul flights: is a flight between 3 and 6 hours in length.

On board cleaning service: is the main job in cabin service. They include task such as cleaning the passenger cabin, replenishment of on-board consumables or washable items such as soap, pillows, tissues and blankets, and do the sanitation service.

Short haul flights: is flight: is a flight under 3 hours in length.

APPENDIX B

Operational Costs

Airport Rates	Every 15 minutes	Every 30 mins	Per Flight	Monthly
Airport Use			€ 9,88	
Vehicle Parking		€ 0,02		
Workers				€ 33,90
Energy system 400HZ	€ 6,79			
Fingers use	€ 27,18			
rate for cleaning 1, 2 and stopover cleaning	€ 9,90			
rate for cleaning 3 and Overnight cleaning	€ 43,87			
rate for cleaning 4	€ 77,86			
rate for cleaning 5 and 3 and deep cleaning	€ 111,82			
Delays	€ 2,27			

REFERENCES

AENA air tariffs. Available from http://www.aena.es/csee/ccurl/124/479/guiaTarifasNA_2013-EN.pdf. [accesed May 20, 2013]

Airbus, "A320, 2012. Aircraft Characteristics Airport and Maintenance Planning",*Technical Report,* Airbus.

Bazargan, M., 2004. Airline operations and Scheduling. Burlington, USA, *Ashgate publishing company*.

Christensen, S., Jensen, K., Mailund, T., Kristensen, L.M., 2001. State Space Methods for Timed Coloured Petri Nets. *Proc. of 2nd International Colloquium on Petri Net Technologies for Modelling Communication Based Systems*, 33-42, Berlin.

Civil Aviation Department Hong Kong, China, 2012. *CAD 452 Aircraft Maintenance Schedules and Programmes, Information and Guidance*. Available from http://www.cad.gov.hk/english/pdf/CAD452.pdf uk [accessed June 15, 2013].

CPNTools Available from http://www.cpntools.org.

Guest, T. 2007. Air traffic delay in Europe. *Trends in Air Traffic Vol. 2, Brussels*-Belgium, EUROCONTROL.

Jensen, K., 1997. *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use*. Springer-Verlag, Berlin.

Jetsky, M, 2009. *The propagation of air transport delays in Europe*, Thesis, RWTH Aachen University.

Moore, K.E., Gupta, S.M., 1996. Petri Net Models of Flexible and Automated Manufacturing Systems: A Survey. *International Journal of Production Research*, 34(11), 3001-3035.

Rhodes, W., Lounsbury, R., Steele, K., Ladha, N., 2003. *Fatigue Risk Assessment of Aircraft Maintenance Tasks*. Transportation Development Centre Safety and Security Transport Canada.

Yeung, S.S.M, Yu, I. T. S., Hui, K.Y.L., 2005. World at Work:Aircraft cabin cleaning. *Occupational and Environmental Medicine*, 62:58-60.

AUTHORS BIOGRAPHY

Miguel Mujica Mota was born in Mexico City. He Studied Chemical Engineering in the Metropolitan Autonomous University of Mexico. He also studied a MSc. in Operations Research at the National Autonomous University of Mexico. After spending some years in industry he continued his studies and he obtained the PhD in Industrial Informatics in 2011 with the highest honors from the Autonomous University of Barcelona and the PhD in Operations Research at the National Autonomous University of Mexico.

Dr. Mujica has been awarded with the Candidate to Level I of the Mexican Council of Science and Technology where he also participates as a scientific evaluator. He is currently the sub director of the Aeronautical Studies at the Autonomous University of Barcelona and his research interests lie in the use of simulation, modeling formalisms and heuristics for the analysis of performance and optimization in manufacture, logistics and aeronautical operations.

Mireia Soler Grané was born in Barcelona, Spain. She studied Aeronautical Management in Air Transport Logistics. While she was studying, she was working in Barcelona's airport as Handling Agent, after two years she was the Assistant of Station Manager of Spanair S.A.

Idalia Flores received a Master with honors, being awarded the Gabino Barreda Medal for the best average of her generation, in the Faculty of Engineering of the UNAM, where she also obtained her Ph.D. in Operations Research. Dr. Flores is a referee and a member of various Academic Committees at CONACYT as well as being a referee for journals such as Journal of Applied Research and Technology, the Center of Applied Sciences and Technological Development, UNAM and the Transactions of the Society for Modeling and Simulation International. She is a full time professor at the Posgraduate Program at UNAM and her research interests lie in simulation and optimization of production and service systems.

COMPARISON OF INCIDENCE MATRICES TO DETECT COMMON PATTERNS IN PETRI NETS

Juan-Ignacio Latorre-Biel[(a)], **Emilio Jiménez-Macías**[(b)]

[(a)] Public University of Navarre. Department of Mechanical Engineering, Energetics and Materials.
Campus of Tudela, Spain
[(b)] University of La Rioja. Industrial Engineering Technical School.
Department of Electrical Engineering. Logroño, Spain

[(a)] juanignacio.latorre@unavarra.es, [(b)] emilio.jimenez@unirioja.es

ABSTRACT

Given a discrete event system modeled by an alternatives Petri net system, the identification of common patterns is required in the incidence matrices in order to transform the model into another minimum one necessary to develop a more efficient optimization. Transformations of set of the alternatives Petri nets to be considered are two: aggregation and fusion. Aggregation is used to obtain alternatives Petri nets, and is performed by means of the following operations: identifying of shared subnets on the alternative nets, identification of binding transitions and unshared blocks, and aggregation of the incidence matrices. Fusion is used for obtaining a composed Petri net, and is made by means of the following operations: application of swaps to rows or columns to achieve an optimal configuration, and overlay of matrices. Those types of transformations on Petri nets are based on the equivalence class of the incidence matrices that can be formed by permuting or swapping the rows and the columns. This paper constitutes such a basis, by means of the analysis of this equivalence class.

Keywords: Petri nets, Incidence matrix, equivalence classes

1. INTRODUCTION

The decisions on discrete event systems under design and alternative structural configurations can be addressed by applying a family of formalisms based on Petri nets (Silva, 1993; Alla and David, 2005; Jensen and Kristensen, 2009) that include a set of mutually exclusive entities. This decision-making can be addressed through an optimization process based on simulation of the system model under different valid configurations. The optimization process efficiency depends on the speed with which the simulation is performed, that in the case of a model expressed by the formalism of Petri nets requires the solution of the state equation. Simulation therefore be the more efficient the smaller the system model, and in particular the size of the incidence matrices (Zimmermann et al., 2001; Tsinarakis et al. 2005;Jimenez et al., 2006, 2009; Latorre et al., 2013a).

Given a discrete event system modeled by an alternatives Petri net system, the identification of common patterns is required in the incidence matrices in order to transform this model into another minimum one necessary to develop a more efficient optimization (Berthelot, 1987; Haddad and Pradat-Peyre, 2006). Transformations of set of the alternatives Petri nets to be considered are two: aggregation and fusion (Latorre et al., 2009, 2011a).

a) Aggregation:

The aggregation of alternative Petri nets is used to obtain an alternatives Petri net, and aggregation is performed by means of the following operations: a.1) Identifying shared subnets on alternative Petri Nets (matching columns in various incidence matrices). At this stage it is possible to exchange between two rows and between two columns for optimal arrangement of the elements of each incidence matrix.

a.2) Identification of the binding transitions (columns that do not match with other incidence matrices having some nonzero element in the same row in which a shared subnet of the same incidence matrix has nonzero elements).

a.3) unshared blocks (other columns, ie columns mismatched with other incidence matrices having non-zero elements only, in which the matching columns in various matrices have nulls).

a.4) Aggregation: Building an aggregate incidence matrix as follows:

* The first block in the aggregate matrix is the first incidence matrix assembly.

* For each new incidence matrix of the set, they are added to the aggregate matrix binding their transitions, placing the non-zero elements in the matrix rows correspond to aggregate shared subnets.

* Also unshared blocks are added so that the non-zero elements in rows match corresponding to the nonzero elements and link transitions or new rows inserted in the same way as in the original array.

* The aggregate matrix voids are filled with zeros.

b) Fusion.

Merging alternative Petri nets used for obtaining a composed Petri net is made by means of the following operations.

b.1) Application of swaps to two rows or two columns to achieve an optimal configuration of the elements of each of the incidence matrices.

b.2) Overlay of matrices to obtain the merged matrix, element by element. Each element of the resulting array will be associated with a single value if overlapping elements coincide (this element is called defined parameter) or a set of possible values that will have many elements with different values coming from the original elements (this element is called undefined parameter).

To manage efficiently those types of transformations on Petri nets, some intermediate goals are required, in concrete:

Algorithm 1 (optimization): search for common patterns in a set of matrices for minimizing the size of the aggregate matrix.

Algorithm 2 (multiobjective optimization): minimization of aggregate array size and the number of transitions link (decision variables).

Algorithm 3 (optimization) search for common patterns in a set of matrices to minimize the number of undefined parameters of the resulting array.

Algorithm 4 (multiobjective optimization): minimize the number of undefined parameters of the resulting matrix and the size of the containing sets of possible values for each parameter of the resulting matrix.

In addition, to developing the algorithms mentioned and provide evidence of correct operation (eg statistics) is included within the objectives to determine the computational complexity of the system.

Regarding distributed optimization, given a set of m matrices and a set of processors p, with p<m, the development of the following algorithms is tacked:

Algorithm 5 (optimization): determination of exchange operations pairs of rows and columns needed and the optimal partition of all the m matrices in p sets, to minimize the average size of the aggregate incidence matrices in each class partition and its variance.

Algorithm 6 (optimization): determination of exchange operations pairs of rows and columns needed and the optimal partition of all the m matrices in p sets, to minimize the average size of the incidence matrices resulting from the incidence matrices in each class of the partition and its variance.

Any of those algorithms constitute a goal and an advance in the state of the arte, with immediate applications and being the basis of other interesting issues. And all of them are based on the equivalence class of the equivalent matrices, that is, the equivalence class of the incidence matrices that can be formed by permuting the rows and the columns. This paper constitutes such a basis, by means of the analysis of the equivalence class of the incidence matrices ().

2. EQUIVALENCE CLASSES

2.1. Operations of incidence matrices

A decision problem based on an undefined Discrete Event System (DES) can be stated as an optimization problem based on an undefined Petri net. The performance of the optimization process can be influenced in a dramatic way by the representation considered for the undefined Petri net. Some operations allow transforming an alternative Petri net into another one that has an equivalent state space to the original PN and might be more appropriate for representing an undefined Petri net in an efficient optimization process. This relation of equivalence will guarantee that the equivalent Petri nets have isomorphous reachability graphs and the same set of reachable significant markings (Latorre et al., 2011b, 2013b, 22013c).

The transformation of one alternative Petri net into an equivalent one will be performed by means of the application of certain matrix-based operations to the incidence matrix. In fact, by the application of these operations to the alternative Petri nets it is possible to obtain adequate incidence matrices for their merging into a more compact compound Petri net. This compound Petri net will be equivalent representations of the same undefined PN in decision problems.

As a consequence, the equivalent Petri net of a certain alternative PN verifies that its incidence matrix can be obtained from the transformation of any other from the same set by means of the application of certain matrix-based operations. Any alternative Petri net of a well-constructed set define an equivalence class. This equivalence class is created by the application of the different feasible sequences of matrix-based operations to the incidence matrix of the alternative Petri net that creates it. From this alternative Petri net the equivalent Petri nets that can substitute a given alternative Petri net can be taken.

Definition 1. Operation of swapping two rows of a matrix.

The operation of swapping two rows of a matrix is defined as the following function:

swapr: $\mathbf{M}_{m\times n} \times \{1, 2, \ldots, m\} \times \{1, 2, \ldots, m\} \rightarrow \mathbf{M}_{m\times n}$

$(\mathbf{A}, i, j) \# \mathbf{B} \in \mathbf{M}_{m\times n}$

$$\text{where,}\quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathbf{M}_{m\times n},$$

$$\mathbf{B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathbf{M}_{m\times n}.$$

□

In other words, **definition 1**, describes the swapping of the *i*-th and *j*-th rows in a matrix **A**. This operation is denoted by swapr(**A**, *i*, *j*).

Definition 2 . Operation of swapping two columns of a matrix.

The operation of swapping two columns of a matrix is defined as the following function:

swapc: $\mathbf{M}_{m\times n} \times \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \rightarrow \mathbf{M}_{m\times n}$

$(\mathbf{A}, i, j)$ # $\mathbf{B} \in \mathbf{M}_{m\times n}$

where,

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mi} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \in \mathbf{M}_{m\times n},$$

and

$$\mathbf{B} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2i} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mi} & \cdots & a_{mn} \end{pmatrix} \in \mathbf{M}_{m\times n}$$

□

In other words, **definition 2**, describes the swapping of the columns i and j in matrix **A**, which is denoted by swapc(**A**, i, j)

Remark 1. The state equation of a Petri net requires representing the characteristic vector that summarizes the information contained in the sequence of transitions fired. The characteristic vector (also called firing count vector) contains elements that are different to zero in the positions that correspond to the transitions fired. If an operation swapc is applied to an incidence matrix and the state equation is represented, the characteristic vector should be modified according to this same swapc operation.

Definition 3. Operation of adding a row of zeros to a matrix.

The operation of adding a row of zeros to a matrix is defined as the following function:

addr: $\mathbf{M}_{m\times n} \rightarrow \mathbf{M}_{(m+1)\times n}$

A # **B**, such that

Given $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathbf{M}_{m\times n}$

$\Rightarrow$ addr(**A**) = $\mathbf{B} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \\ 0 & \ldots & 0 \end{pmatrix} \in \mathbf{M}_{(m+1)\times n}$

□

The operation described in **definition 3** is denoted by addr(**A**) and adds a row of zeros to the matrix **A**.

Remark 2. The operation addr applied to the incidence matrix of a Petri net implies the addition of a new place with a particular property: every input and output arc has weight zero. In other words, this new place is an isolated node of the Petri net. The marking of the Petri net that results from the application of this operation should include the marking of the new place, which will occupy the last position of the vector. However, being isolated, the place cannot experience any variation of its initial marking in the evolution of the Petri net. Furthermore, the marking of other places will not be influenced by the added place, hence the marking of the new Petri net, excluding the added place, will be the same to the original one. If the new place is considered in this comparison it is possible to say that the significant marking is the same in both Petri nets; hence the graphs of reachable markings are isomorphous.

Definition 4. Operation of removing a row of zeros of a matrix.

The operation of removing a row of zeros of a matrix is defined as the following function:

removr: $S \rightarrow \mathbf{M}_{(m-1)\times n}$

A # **B**, such that

$S = \{\mathbf{A} \in \mathbf{M}_{m\times n} \mid a_{m*} = (0\ 0\ \ldots\ 0)\ \}$, in other words, S is the set of matrices whose m-th (last) row is a row of zeros.

Given $\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \cdots & \cdots & \cdots \\ a_{m-1,1} & \cdots & a_{m-1,n} \\ 0 & \ldots & 0 \end{pmatrix} \in \mathbf{M}_{m\times n} \Rightarrow$

removr(**A**) = $\mathbf{B} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \cdots & \cdots & \cdots \\ a_{m-1,1} & \cdots & a_{m-1,n} \end{pmatrix} \in \mathbf{M}_{(m\text{-}1)\times n}$

□

The operation described in **definition 4** is denoted by removr(**A**) and removes the last row of a matrix **A**, which should contain only zeros.

Definition 5. Operation of adding a column of zeros to a matrix.

The operation of adding a column of zeros to a matrix is defined as the following function:

addc: $\mathbf{M}_{m\times n} \rightarrow \mathbf{M}_{m\times(n+1)}$

A # **B**, such that

Given $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathbf{M}_{m\times n} \Rightarrow$

addc(**A**) = $\mathbf{B} = \begin{pmatrix} a_{11} & \ldots & a_{1n} & 0 \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & \ldots & a_{mn} & 0 \end{pmatrix} \in \mathbf{M}_{m\times(n+1)}$

□

The operation described in **5** is denoted by addc(**A**) and adds a column of zeros to a matrix **A**.

Remark 3. The operation addc applied to the incidence matrix of a Petri net implies the addition of a new transition with a particular property: every one of its input and output arcs has weight zero. In other words, this new transition is an isolated transition of the Petri net. Moreover, the added transition will be associated to the last column of the incidence matrix.

Any characteristic vector associated to R should be modified before being associated as well to the Petri net R' that results from the application of the operation addc to its incidence matrix. This modification of the characteristic vector consists of the addition of a zero as the new last element. Thanks to this modification the size of the vector will fit with the one of the incidence matrix **B** in the state equation.

Definition 6. Operation of removing a column of zeros of a matrix.
The operation of removing a column of zeros of a matrix is defined as the following function:

removc: $S \rightarrow \mathbf{M}_{(m\text{-}1)\times n}$
  A # **B**, such that

$S = \{\mathbf{A} \in \mathbf{M}_{m\times n} \mid a_{*n} = [0\ 0\ \ldots\ 0]^{\mathrm{T}}\ \}$, in other words, S is the set of matrices whose nth (last) column is a column of zeros.

Given $\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,(n-1)} & 0 \\ \cdots & \cdots & \cdots & \cdots \\ a_{m,n} & \cdots & a_{m,(n-1)} & 0 \end{pmatrix} \in \mathbf{M}_{m\times n} \Rightarrow$

removc(**A**) = $\mathbf{B} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,(n-1)} \\ \cdots & \cdots & \cdots \\ a_{m,1} & \cdots & a_{m,(n-1)} \end{pmatrix} \in \mathbf{M}_{m\ \times(n\text{-}1)}$

□

The operation described in **definition 6** is denoted by removc(**A**) and removes the last columns of a matrix **A**, which should contain only zeros.

It can be proven that none of the matrix-based operations defined in this section modify the form of the reachability graph of the Petri net, when thay are applied to its incidence matrix. Furthermore, the significant markings are the same.

Proposition 1.
Let R be a Petri net with an incidence matrix given by **A** $\in \mathbf{M}_{m\times n}$ and let swapr(**A**, i, j) = **B** $\in \mathbf{M}_{m\times n}$. The Petri net associated to **B** is R'. The initial markings of R and R' are respectively $\mathbf{m}_0$ and $\mathbf{m}_0$'.

rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$')

□

Proposition 2.
Let R be a Petri net with an incidence matrix given by **A** $\in \mathbf{M}_{m\times n}$. and let swapc(**A**, i, j) = **B** $\in \mathbf{M}_{m\times n}$. The Petri net associated to **B** is R'. The initial markings of R and R' are respectively $\mathbf{m}_0$ and $\mathbf{m}_0$'.

rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$')

□

Proposition 3.
Let R be a Petri net with an incidence matrix given by **A** $\in \mathbf{M}_{m\times n}$ and let addr(**A**) = **B** $\in \mathbf{M}_{(m+1)\times n}$. The Petri net associated to **B** is R'. The initial markings of R and R' are respectively $\mathbf{m}_0$ and $\mathbf{m}_0$'.

rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$') for the marking of the connected places.

□

Proposition 4.
Let R be a Petri net with an incidence matrix given by **A** $\in \mathbf{M}_{m\times n}$ and let removr(**A**) = **B** $\in \mathbf{M}_{(m\text{-}1)\times n}$. The Petri net associated to **B** is R'. The initial markings of R and R' are respectively $\mathbf{m}_0$ and $\mathbf{m}_0$'.

rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$') for the marking of the connected places.

□

Remark 4. What **proposition 3** and **proposition 4** mean in fact is that the graphs of reachable markings are isomorphous. They are only the same for the significant marking or more specifically for the marking of the connected places. Furthermore, the isolated places are associated to constant markings and they do not influence the evolution of the Petri net (the valid sequences of transition firing and the markings of the connected places). For this reason this relation between the graphs of reachable markings of the original Petri nets and the ones resulting from the application of the operations addr and removr is approximated with a high degree of reliability and usefulness in the applications for decision problems as being the same: rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$').

Proposition 5.
Let R be a Petri net with an incidence matrix given by **A** $\in \mathbf{M}_{m\times n}$ and let addc(**A**) = **B** $\in \mathbf{M}_{m\times(n+1)}$. The Petri net associated to **B**, obtained from R is R'. The initial markings of R and R' are the same, in other words $\mathbf{m}_0$ = $\mathbf{m}_0$'.

rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$')

□

Proposition 6.
Let R be a Petri net with an incidence matrix given by **A** $\in \mathbf{M}_{m\times n}$ such that $a_{*j} = 0$ (the elements of the last column are zeros) and let removc(**A**) = **B** $\in \mathbf{M}_{m\times(n\text{-}1)}$. The Petri net associated to **B** is R'. The initial markings of R and R' are the same ($\mathbf{m}_0$ = $\mathbf{m}_0$').

rg(R, $\mathbf{m}_0$) = rg(R', $\mathbf{m}_0$')

□

3. WELL-CONSTRUCTED SETS OF ALTERNATIVE PN AND EQUIVALENCE CLASSES

The matrix-based operations described in the previous section may be used to substitute one or several alternative Petri nets to find more convenient representations for solving the optimization problem in a more efficient way.

In order to proceed as described in the previous paragraph it is interesting to classify the sets of alternative Petri nets into two categories, which will be called the well-constructed sets and the redundant sets. It is possible to describe a well-constructed set as the one containing alternative Petri nets able to define equivalence classes from the equivalence relation given by the application of the matrix-based operations. In this case, the alternative Petri nets of the original set cannot be transformed into another Petri net of the same set, since they belong to different equivalence classes, which verify the property of being disjoint sets. This property of exclusiveness is associated to the idea of set of exclusive entities, which is the signature of the decision problems where there are structural alternatives as feasible solutions.

The name "redundant" arises from the fact that at least one alternative Petri net can be transformed into another one from the same. subsequently both of them belong to the same equivalence class. As a consequence, the couple of PN is related to the same solution of the decision problem.

3.1. Well-constructed sets of alternative PN and equivalence classes: definitions

Definition 7. Set of equivalence operations.
The set of equivalence operations is S_{OP} = { swapr, swapc, addr, removr, addc, removc } set of all matrix-based operations defined previously in this chapter.

□

Definition 8. Feasible sequence of operations.
A feasible sequence of operations is a finite set of the form

$$S_{secop} = \{ \text{op}_1, \text{op}_2, \ldots, \text{op}_{nop} \mid \text{op}_i \in S_{OP}, 1 \leq i \leq n_{op} \}$$

□

Definition 9. Set of feasible sequence of operations.
The set of feasible sequence of operations is
$S_{SOP} = \{ S_{secop} \}$

□

Definition 10. Application of a sequence of operations to an incidence matrix.
Let $\mathbf{W}_a$ be the incidence matrix of a Petri net.
Let S_{secop} be a feasible sequence of operations such that $S_{secop} \in S_{SOP}$.
The application of S_{secop} to $\mathbf{W}_a$ is called $S_{secop}(\mathbf{W}_a)$ and is performed in the following way

$$S_{secop}(\mathbf{W}_a) = \text{op}_{nop}(\ldots\text{op}_2(\text{op}_1(\mathbf{W}_a)))$$

The way of applying this sequence of operations is the following: first of all it is calculated the operation $\text{op}_1(\mathbf{W}_a) = \mathbf{W}_a$', in a second step it is calculated the operation $\text{op}_2(\mathbf{W}_a') = \mathbf{W}_a'' = \text{op}_2(\text{op}_1(\mathbf{W}_a))$ and so on until the last operation op_{nop} is applied.

□

Definition 11. Equivalence relation between two Petri nets.
Let us consider the Petri nets R_a and R_b , whose incidence matrices are $\mathbf{W}_a$ and $\mathbf{W}_b$ respectively.
The equivalence relation ~ is defined in the following way:

$$R_a \sim R_b \text{ iif } \exists\, S_{secop} \in S_{SOP} \text{ such that } S_{secop}(\mathbf{W}_a) = \mathbf{W}_\text{b}.$$

□

Definition 12. Equivalence class defined by an alternative Petri net.
Given a set of alternative Petri nets $S_R = \{ R_1, R_{2, \ldots,} R_{nr}\}$, the binary equivalent relation ~ , defined in **definition 6.21**, allows to create an equivalence class for every alternative Petri net, such that

Let $R_i \in S_R$, the equivalence class created by R_i is $[R_i] = \{ R \text{ PN} \mid R_i \sim R \}$.

□

Definition 13. Well-constructed set of alternative Petri nets.
Given a set of alternative Petri nets $S_R = \{ R_1, R_{2, \ldots,} R_{nr}\}$. S_R is said to be well constructed iif $\forall\, R_i , R_j \in S_R , i \neq j$,it does not exist any sequence of operations $S_{secop} \in S_{SOP}$ such that $S_{secop}(\mathbf{W}_i) = \mathbf{W}_j$.

□

In other words, for a set of alternative Petri nets to be well constructed set it is a necessary and sufficient condition that none of the Petri nets of the set has an incidence matrix that can be transformed, by means of equivalence operations, into the incidence matrix of another Petri net of the same set.

On the other hand, it has been mentioned before as well that the definition of this category of well-defined sets of alternative Petri nets do not compromise the applicability of the methodologies developed in this thesis to real cases. On the contrary, the correct models developed for undefined DES will not contain isolated places or transitions (which do not contribute to the evolution of the system) or different order of the rows or columns between them (due to the assignment of different names for the same real items modelled by nodes in the PN). Subsequently, this condition do not prevent the representation of most of the different possible real or academic cases that can arise in a decision problem, but it is a guarantee of the correct development of a model of an undefined DES for a decision problem. Furthermore, this small restriction will have important implications in the improvement of the efficiency of the algorithms to solve decision problems in the scope of this thesis.

3.2. Sufficient conditions for a set of alternative Petri nets to be well constructed.

The next topic to be considered is how to check that a set of alternative Petri nets is a well-constructed one and, hence, able to create so many different equivalence classes as the cardinality of the set.

Proposition 7. Sufficient condition to identify a well-constructed set of alternative Petri nets.
Let D be a DES.
Let R^U be an undefined Petri net developed as model for D.

Let $S_R = \{ R_1, R_{2, \ldots,} R_{nr}\}$ be a set of alternative Petri nets developed as representation of R^U.

If the following conditions are verified

a) $\forall\ R_i \in S_R \wedge \forall\ p_j \in P_i$ it is verified that p_j is a connected place.
b) $\forall\ R_i \in S_R \wedge \forall\ t_j \in T_i$ it is verified that t_j is a connected place.
c) $\forall\ R_i, R_j \in S_R \wedge \forall\ p_{k1} \in P_i$, $p_{k2} \in P_j$ such that $X(p_k)$ is the item in D modelled by p_{k1} and p_{k2} then $k_1 = k_2$.
d) $\forall\ R_i, R_j \in S_R$ it is verified that $\mathbf{W}(R_i) \neq \mathbf{W}(R_j)$.

Then S_R is a well-constructed set of alternative Petri nets.

□

A sufficient condition for a set of alternative Petri net to have incidence matrices that cannot be transformed one into another by means of "add" or "remov" operations is that there is not any isolated place or transition in any of the alternative Petri nets. As it has been mentioned before, this is a usual case in the application of PN found in the literature so far, since in the definition of PN it is usually imposed the condition of connectivity in every node. A way to detect this situation is to search for rows or columns of zeros.

A sufficient condition for a set of alternative Petri net to have incidence matrices that cannot be transformed one into another by means of "swap" operations is to give in the model the same reference name to the same physical item that is modelled as a place or transition, to associate an alias to every reference name with the same subindex and to compare the incidence matrices element by element. If there is a pair (or more) elements which are different in the incidence matrices of different Petri nets, the set of alternative Petri nets is well constructed.

Another sufficient condition for a set of alternative Petri net to be well constructed is given below.

Proposition 8. Sufficient condition to identify a well-constructed set of alternative Petri nets.
Let $S_R = \{ R_1, R_{2, \ldots,} R_{nr}\}$ be a set of alternative Petri nets, where $\mathbf{W}_k$ is the incidence matrix of $R_k \in S_R$, $\mathbf{W}_k \in \mathbf{M}_{mk\times nk}$ and $a^k_{i,j} = \mathbf{W}_k[i,j]$

$$\text{Let } sumt_k = \sum_{i,j=1}^{\substack{i=m_k\\ j=n_k}} a^k_{i,j}$$

If $\forall\ R_{k1}, R_{k2} \in S_R$, $sumt_{k1} \neq sumt_{k2} \Rightarrow S_R$ is well constructed

□

As it has been proven, a sufficient condition for a set of alternative Petri nets to be well constructed is that the sums of all the elements of the incidence matrix are different for the different alternative Petri net.

3.3. Redundant set of alternative Petri nets.

It has been explained previously that a set of alternative Petri nets that is not well constructed may lead to the creation of the same equivalence class by several different alternative PN of the set.

The detection of this fact by means of the sufficient conditions described earlier can allow the removal of the redundant alternative Petri nets and hence the simplification of the statement of the decision problem based on a reduced set of alternative PN. A simplification in the representation of an undefined Petri net may lead to a more efficient optimization process based on a model of reduced size.

In case that a redundant set of alternative Petri net is not detected and used to state an optimization problem, the different alternative Petri nets that create the same equivalence class will be treated as if their respective equivalence classes were different. The behaviour and the structure of the Petri nets will not be different from one of these alternative PN and the others that create the same equivalence class. As a consequence, the optimization algorithm might duplicate the computational effort by considering twice the same alternative Petri net.

4. CONCLUSIONS

Given a discrete event system modeled by an alternatives Petri net system, the identification of common patterns is required in the incidence matrices in order to transform the model into another minimum one necessary to develop a more efficient optimization. Transformations of set of the alternatives Petri nets to be considered are two: aggregation and fusion.

This paper has analyzed the equivalence class of the incidence matrices, that is, of matrices that can be obtained by swapping rows and columns. This knowledge, and the basis of the definitions and properties, constitute the starting point to analyze the optimization of several incidence matrices that are wanted to be merged, as it happens with alternatives Petri nets.

ACKNOWLEDGMENTS

This paper has been partially supported by a grant of the University of La Rioja and Banco Santander.

REFERENCES

Berthelot, G., 1987. Transformations and decompositions of nets. In: Brauer, W., Reisig,

David, R., Alla. H., 2005. *Discrete, Continuous and Hybrid Petri nets*, Springer.

Haddad, S. and Pradat-Peyre, J.F., 2006. New Efficient Petri Nets Reductions for Parallel Programs Verification. *Parallel Processing Letters*, pages 101-116, World Scientific Publishing Company.

Jensen, K., Kristensen, L.M., 2009. *Coloured Petri nets. Modelling and Validation of Concurrent Systems*. Springer.

Jimenez, E., Perez, M., Latorre, J.I., 2006. Industrial applications of Petri nets: system modelling and simulation. *Proceedings of International Mediterranean Modelling Multiconference* 2006, pp. 159-164

Jimenez, E., Perez, M., Latorre, J.I., 2009. Modelling and simulation with discrete and continuous PN: semantics and delays. *Proceedings of 21st European Modeling and Simulation Symposium, Vol II*, pp. 14-19

Latorre, J.I., Jimenez, E., Blanco, J., Sáenz-Díez, J.C., 2013a. Integrated methodology for efficient decision support in the Rioja wine production sector. *International Journal of Food Engineering*, (In press).

Latorre, J.I., Jiménez, E., Pérez, M., 2011a. Matrix-based operations and equivalence classes in alternative Petri nets. *Proceedings of the 23rd European Modelling and Simulation Symposium* (EMSS 11). Rome, Italy, pp. 587-592.

Latorre, J.I., Jiménez, E., Pérez, M., 2011b. Petri net transformation for decision making: compound Petri nets to alternatives aggregation Petri nets. *Proceedings of the 23rd European Modelling and Simulation Symposium* (EMSS 11). Rome, Italy, pp. 613-618.

Latorre, J.I., Jimenez, E., Perez, M., 2013b. Simulation-based Optimisation of Discrete Event Systems by Distributed Computation. *Simulation-Transactions of the Society for Modeling and Simulation International*, (In press).

Latorre, J.I., Jimenez, E., Perez, M., 2013c. The optimization problem based on alternatives aggregation Petri nets as models for industrial discrete event systems. *Simulation-Transactions of the Society for Modeling and Simulation International*, 89 (3), 346–361.

Latorre, J.I., Jimenez, E., Perez, M., Blanco, J., 2009. The problem of designing discrete events systems. A new methodological approach. *Proceedings of 21st European Modeling and Simulation Symposium, Vol II*, pp. 40-46

Silva, M., 1993. Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall.

Tsinarakis, G. J., Tsourveloudis, N. C., and Valavanis, K. P., 2005. Petri Net Modeling of Routing and Operation Flexibility in Production Systems. *Proceedings of the 13th Mediterranean Conference on Control and Automation*, pages 352-357.

Zimmermann, A.; Freiheit, J.; Huck, A. 2001. A Petri net based design engine for manufacturing systems. *International Journal of Production Research*, Vol. 39, No. 2, pages 225-253.

SIMPHONY: AT THE PINNACLE OF NEXT GENERATION SIMULATION MODELING ENVIRONMENTS FOR THE CONSTRUCTION DOMAIN

Ronald Ekyalimpa[(a)], **Stephen Hague**[(b)], **Simaan AbouRizk**[(c)]

[(a)]University of Alberta, 5-047 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA
[(b)]University of Alberta, 5-048 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA
[(c)] University of Alberta, 3-015 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

[(a)] rekyalimpa@ualberta.ca, [(b)] steve.hague@ualberta.ca, [(c)] abourizk@ualberta.ca

ABSTRACT

This paper presents Simphony simulation system as a tool that is leading the way in the evolution of simulation systems within the construction domain. This discussion is introduced by presenting an overview of simulation, the different simulation methods and the tools that support this method. Simphony is then introduced as an environment that supports discrete event and continuous simulation. Other features such as its extensible API, calendar, data connectivity, special purpose development abilities etc., are also highlighted to show why Simphony is a powerful simulation system. Two practical problems (earth-moving and traffic light) that are solved using Simphony are presented to demonstrate the use of some of these features.

Keywords: Simphony, extensible API, calendars, simulation methods

1. INTRODUCTION

Simulation is a numeric method that has been in use for several years and has been applied in the analysis of complex dynamic systems. The simulation community has three well established methods to apply a simulation-based approach in solving their problems: System Dynamics (SD), Agent-Based Modeling (ABM), and Discrete Event Simulation (DES).

The use of each of these methods depends on the complexity of the system being analyzed and the level to which the modeller would like to abstract the system. System dynamics is famous for its precision in modeling systems that have numerous components that are dynamic, inter-related and interact with a feed-feedback behavior. This method supports a top-bottom approach to that analysis of systems e.g. the evaluation of the impact of different policies or strategies on the behavior of a system. Simulation systems build to support this method implement numeric integration algorithms (e.g. the family of Runga-Kutta equations) in a continuous fashion. It does not involve the flow of tokens but rather tracks rates of change in specified quantities with time using integration. Examples of such systems include AnyLogic, Vensim, PowerSim, STELLA (iThink), Simulink, DYNAMO etc.

Agent-based modeling on the other hand supports a bottom-up approach to the analysis of systems. This approach models a unit or a component within a system as an agent that has intelligent behaviors that are influenced by its peers (other agents) and the environment in which it operates. An agent exhibits different behavior by transitioning through different states. This behavior is controlled by an algorithm embedded within the agents. This algorithm is defined using concepts of state diagrams. Communication between agents and the environment is triggered by the events (in the computing science sense). Examples of simulation systems that support this modeling approach include Repast-Simphony, AnyLogic, A3/AAA, ABLE, Agent Builder, MASON, NetLogo, SimAgent etc.

Discrete event simulation is an approach in which a system is described using entities, resources, activities and other modeling constructs. These constructs interact with each other to define the state of the system and are responsible for its evolution at discrete points in time. In typical DES systems, this change of state is triggered by the flow of entities. Changes in the state of a system typically occur when resources are captured or released, activities are started or finished. DES is best suited for analyzing systems at an operations level. This explains why it has been extensively used in analyzing production systems, supply chain, medical facility operations and construction operations. Various general purpose discrete event simulation software systems have been developed for a wide range of industries: AweSim (Pritsker, 1997) and GPSS/H (Crain, 1997); for construction: Micro-CYCLONE (Halpin, 1973), STROBOSCOPE (Martinez, 1996), and Simphony (Hajjar and AbouRizk, 1999).

In construction, DES has been widely used to model and improve processes such as tunnel construction, earth-moving, fabrication shops, bridge construction, and scheduling.

In 1999, AbouRizk et al. developed a special purpose template in Simphony for analyzing tunnel construction using TBMs. The template was used to evaluate the effect of different site setup configurations at the working shaft through predictions of tunnel advance rates along the tunnel length. Work on this template is on-going and has resulted in it evolving into a more sophisticated yet easy to use simulation tool for analyzing tunnel construction processes. Examples of additions to the template include: (1) ability to model shifts through the use of calendars, (2) features for generating cost estimate reports for the simulated tunnel and (3) the ability of the template to participate in a larger distributed simulation system that is based on the HLA standards, so that it can support other components such as visualization of the tunnel construction. Later on, Zhou et al. (2008) used Simphony to develop a special purpose template for modeling tunnel shaft construction. They refined the way tunnel shafts are simulated so that most site processes and constraints are well represented. They validated their template by implementing a case study (NEST NL1-NL2 tunnel in Edmonton). Modeling constructs developed in this work are used in the current version of the template. Touran and Asai (1987) used CYCLONE simulation system to the advance rate of a TBM during the construction of a long, small-diameter tunnel. Ioannou and Martinez (1996) used STROBOSCOPE simulation system to compare two alternative construction methods for rock tunneling; a conventional verses the New Austrian Tunneling Method (NATM). They demonstrated effective ways of using simulation for comparing alternatives. Al-Bataineh et al. (2013) recently wrote a paper in which they used simulation to project planning and control in tunnel construction.

The earth-moving operation is one that had been extensively analyzed using simulation because of its repetitive nature and simplicity. A small portion of the work done in simulating earth-moving operations is discussed here. In 2002, Marzouk and Moselhi combined simulation and optimization (genetic algorithms) to get optimal cost and durations associated with earth-moving operations. Fu (2012) presented a paper in which he used Global Simulation Platform (GSP), a simulation system developed by Volvo CE, and CYCLONE to simulate and compare three loading scenarios for an earth-moving operation. He compared the options based on fuel cost per unit production. However, logic flaws can be identified in some of the CYCLONE model layouts presented by Fu because they don’t explicitly represent the loading of haulers with multiple buckets. In 2011, Cheng et al. proposed a simulation model for virtual simulation of earthmoving operations using petri nets. In 2009, Ahn et al. published a paper in which they presented a simulation-based sustainability analysis of earth-moving operations with respect to emissions. STROBOSCOPE and VITASCOPE were both used as simulation and visualization platforms for estimating omissions and visualizing simulated objects respectively. Rekapalli and Martinez (2011) also presented a recent paper on earth-moving operations.

Simulation has been extensively applied for modeling processes at the different stages of the delivery of industrial projects. They include: structural steel fabrication, pipe spool fabrication, module assembly. In 2008, Liu and Mohamed used an agent-based modeling approach to simulate the dynamics of resource allocation within a module assembly yard for a construction company in Edmonton, Canada. They used Repast-Simphony for their work. Song and AbouRizk (2003) developed Simphony general purpose template models (for a structural steel fabrication shop) which they integrated with CAD drawings to obtain attributes of steel pieces whose fabrication process was to be simulated. Their model used attributes of steel pieces obtained from CAD drawings and embedded artificial neural networks to predict durations for the different fabrication processes (cutting, fitting, welding and painting) (Song and AbouRizk, 2006). Alvanchi et al. (2012) developed a special purpose simulation template in Simphony for modeling the fabrication of structural steel within a shop. The template reads its input of steel pieces to be fabricated from an information management system (database) and simulates the fabrication process for different shop layouts and processing equipment so that the operational efficiency of the shop can be assessed. Sadeghi and Fayek (2008) developed a Visual Basic application that utilizes the Simphony discrete event engine behind the scenes to model operations in a pipe spool fabrication shop. Mohsen et al. (2008) used Simphony general purpose template to simulate the erection of a building that was constructed using a modularized approach. They used their model to determine the utilization of the different resources (crane, rigging crew, welding crew and delivery space) involved in the operation and compared these with those recorded on site. Wang et al. (2009) developed models that simulate the operations in a typical pipe spool fabrication shop. They compared the “traditional batch-and-queue fabrication system” to “the new cell-based work flow fabrication systems” by constructing a simulation model for each system. Cycle time for the fabrication of pipe spools was used as a statistic for comparing the two methods and the new method was found to be more efficient.

In bridge construction, a number of modeling studies have been done. Dulcy and Halpin (1998) pointed out that cable stay bridge construction provides enormous opportunity for the use of computer simulation in the analysis and design of the operations involved. They attributed this to the fact that this type of bridge involves many repetitive cycles of placing concrete segments and supporting cables. They constructed a CYCLONE model for the construction of a cable-stay bridge (Dame Point Bridge in Jacksonville, Florida) and used it to investigate different resource combinations that would result in higher utilizations and shorter construction durations. They came up with an optimum mix of resources for this problem. In 2007, Marzouk et al. also used simulation to model the construction of a bridge in Cairo, Egypt (“The 15th May Bridge”) that was constructed using the incremental launching construction technique. Marzouk et al. (2007) stated that concrete bridges can be placed into 6 categories based on the construction method. These include: 1) cast-in-situ on false-work, 2) cantilever carriage, 3) flying shuttering, 4) launching girder, 5) pre-cast balanced cantilever, and 6) incremental launching. They further stated that the incremental launching construction method is the preferred option when the spans being constructed are larger than usual. Marzouk and his colleagues used STROBOSCOPE to develop a special purpose template of the bridge construction process and resource constraints. They ran their model for one scenario similar to that used on site and obtained production results that were very close to actual values on site. They then used this model to experiment with different resource mixes to shorten the project duration. A number of authors have used simulation-based methods to visualize/animate the construction of bridges. Recent examples include: Dori and Borrmann (2011), and Chui-Te et al. (2011). Visualization helps with the verification of constructed simulation models and assists is displaying the evolution of the simulation to those not knowledgeable in simulation in an effective manner. Other studies that have involved the use of simulation for modeling bridge construction processes include work done by Ailland et al. (2010), Liu (2012), and Chuen-Tsai et-al (2013).

Simulation has been used to improve scheduling practices within the construction domain. For example, a state-based simulation approach was used by Hu and Mohammed (2010) to facilitate updates of schedules developed in Microsoft project. They used the Simphony simulation engine. Other studies done on simulation-based scheduling include: Chehayeb and AbouRizk (1998), Zhang et al. (2002) and Lu (2003).

The aforementioned studies demonstrate the vast simulation opportunity that construction processes offer to the simulation community. Simphony has been used significantly in solving construction problems using a simulation-based approach, especially the more complex ones.

2. SIMPHONY SIMULATION SYSTEM

Simphony is a discrete event simulation system that was originally developed by Hajjar and AbouRizk (1999) and is currently being extended and maintained by the 2nd and 3rd authors. Simphony provides an-easy-to-use User Interface (UI), core services (a simulation engine, resources, files, calendars etc.), modeling services and simulation templates.

Simphony is built using the Microsoft .NET framework in a fashion that makes it extensible. Its Application Programming Interface (API) can therefore be utilized within the Simphony UI or any other applications that is compatible with .NET APIs. This is what makes Simphony exceptionally powerful.

Furthermore, Simphony supports the development of custom special purpose templates. These templates provide for an efficient way to abstract complex processes in a manner that makes it easy for domain experts to make use of simulation without having in-depth knowledge of the science behind the method. Examples of special purpose templates previously developed and currently supported in Simphony include the tunneling template, aggregate crushing template, de-watering template, PERT template, earth-moving template, structural steel fabrication template, and range estimating template.

Simphony supports a general purpose template that has elegant graphical modeling elements, and directional arrows which are an essential feature of discrete event simulation modeling languages/environments. Constructing GPT models requires elements to be dragged and dropped onto the modeling surface and connected with directional arrows in a convenient way. Simphony GPT also provides advanced features such as attributes for entities and scenarios, and formula editors into which user written code can be embedded within the models to facilitate solving more sophisticated problems.

Simphony is built to support Monte-Carlo simulation experiments. Figure 1.0 summarizes the manner in which Simphony processes simulation models.

Other features or services that exist within Simphony giving it an edge over other systems include:

- It has discrete event simulation capabilities and supports combined simulation as well (discrete event-continuous simulation).
- It supports calendars.
- It provides for data visualization.
- It supports connectivity to data storage applications e.g. databases.
- It has a neat user interface that provides for model debugging features (a trace window).
- Templates developed in Simphony can easily be integrated into larger distributed simulation systems (developed in line with the HLA).

The rest of the paper is dedicated to demonstrating the use of Simphony for modeling typical simulation problems.

3. MODELING WORK SHIFT DYNAMICS USING CALENDARS IN SIMPHONY.NET

3.1. An Earth-Moving Operation

In order to demonstrate the use of calendars within Simphony's general purpose template, a simple earth-moving operation is described, modeled and experimented with. We shall investigate the effect of using different calendars on the total number of calendar days it will take 5 dump trucks (@ has a capacity of 20cy) to move 10,000 cubic yards of dirt from the source to a placement area. There is one loader at the source responsible for loading dirt onto trucks. The details of the load, haul, dump and return activities are summarized in Table 1.

A simple operation (earth-moving operation) is chosen to demonstrate the concepts of utilizing a calendar within a simulation model. A brief section detailing calendar features within Simphony is introduced and then their applications in modeling the problem at hand are presented.

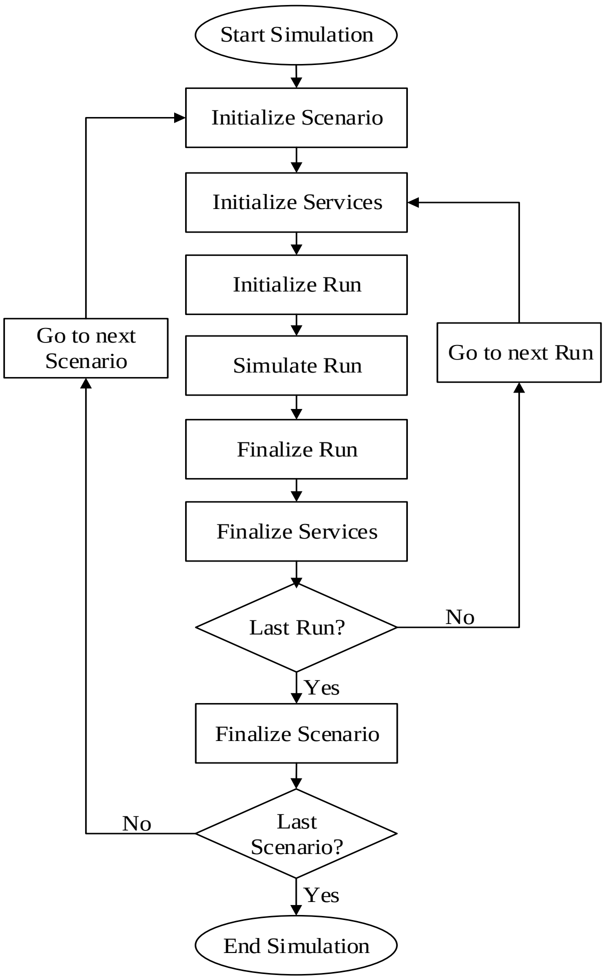


Figure 1: Schematic Layout of Simphony's Simulation Process

Table 1: Earth-Moving Activity Durations

Activity	Duration (Minutes)
Load @ Truck	Constant(8.0)
Haul to dumpsite	Constant(40.0)
Dump Truck load	Constant(5.0)
Return to source	Constant(55.0)

3.2. Calendars within Simphony.NET 4.0

Embedding and using calendars to constrain the execution of simulation models requires a clear understanding of the behavior of calendars from a simulation perspective. When activated, the calendar in Simphony continuously transitions through two states: (1) a working state and (2) a non-working state. The transition between any two calendar states (that are the same or different) gives rise to a calendar event. It is during the processing of a calendar event that a work shift (or the processing or the simulation model) gets turned ON or OFF through the *"SuspendEvent(entity)"* and *"ResumeEvent(entity)"* methods, respectively.

Simphony provides two constructs that facilitate the modeler to model work shifts: (1) a calendar, and (2) a calendar entity. The calendar keeps track of the working time periods, non-working time periods and their lengths. It is responsible for triggering calendar events whenever there is a transition in its working state. The calendar also provides methods that facilitate the modeler to get the total working or non-working times for a particular shift and pay type between specified dates. This is usually useful when the modeler is tracking costs for simulated operations.

The Calendar in Simphony is activated to start raising events as soon as the *"engine.SubscribeCalendar(...)"* method is invoked. The calendar then keeps continuously raising calendar events and will not stop until the simulation is halted by either (1) maximum criteria achieved, (2) maximum count achieved or (3) through an explicit halt invoked within an *"execute element."* The criteria for terminating the simulation through the engine running out of simulation events can never be achieved when the calendar is activated because the calendar keeps looping and raising calendar events infinitely. The calendar can be de-activated by invoking the *"engine.UnsubscribeCalendar(...)"* method. The modeller can obtain a calendar from the calendar list (defined in Simphony core services or in the *"calendar property"* of the scenario).

The calendar entity on the other hand carries with it information about the calendar event that has been triggered (summarized in Table 2). This entity is passed on to the calendar event handler so that this information can be used for implementing computations at the time that the calendar event is being processed. These properties are of a calendar entity are summarized in the schematic layout presented in Figure 2.0.

Table 2: Properties of a Calendar Entity

Calendar Entity Property	Purpose of the Property
Calendar	Gets or sets the calendar with which the entity will be associated
Entities	Avails the entities controlled by the calendar
IsWorking	Determines whether the calendar is currently in a working (true) or non-working (false) state
Time Remaining	Avails the time to the next calendar event (time span)

It is important to carefully track the work state of the calendar because it affects the action taken on the entities controlled by the calendar (suspends their processing, resumes their processing or does nothing). The modeller would like to act on the entities when a calendar event is associated with a change in the work state of the calendar (Figure 3.0) and not do anything when a calendar event is not associated with a change in the work state of the calendar (Figure 3b). An example of a situation where an action is warranted is when there is a transition between a working period and a work break and a work break and work period. This is illustrated in Figure 3a. At calendar event 1, all entities are suspended and at calendar event 2, processing of all entities is resumed. An example where there is no need for action on entities arises when there are two consecutive work periods, especially when there is a shift change without a break – i.e. from shift 1 to shift 2 (Figure 3b).

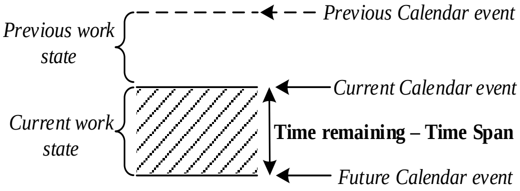


Figure 2: Schematic Layout for Occurrence of Simphony Calendar Events

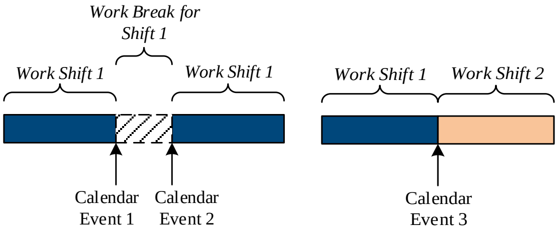


Figure 3: Different Transitions between Work Periods (3a: Work-to-Non Work-to-Work; 3b: Work-to-Work

Details of the working states associated with each calendar event are summarized in Table 3.0. Based on the explanations provided, it becomes necessary to keep track of the current work state associated with each calendar event (provided by the calendar entity) and the previous working state (tracked by the modeller).

Table 3: Shift Details Associated with Figure 3.0

Calendar Event	Current State	Previous State	Shift
1	Working	Non-working	1
2	Non-working	Working	2
3	Working	Working	1→2

At present, the intrinsic statistics reported in Simphony (such as resource utilizations and file length) are not reliable when calendars are enabled because they don't distinguish between working time and non-working time in the course of the simulation.

3.3. Simphony Model Layout, Discussion and Results

At the start of simulation, the model subscribes to a calendar when the initialize run method is invoked on the *"Subscribe to a Calendar"* execute element. Within this element, C# code snippet is written to achieve the subscription to the calendar. A method associated with the calendar event handle is written within the partial formulas class for this execute element. Figure 4.0.

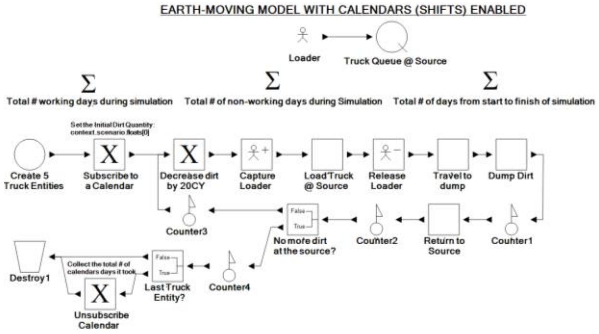

EARTH-MOVING MODEL WITH CALENDARS (SHIFTS) ENABLED
Loader
Truck Queue @ Source
Total # working days during simulation
Total # of non-working days during Simulation
Total # of days from start to finish of simulation
Set the Initial Dirt Quantity: context.scenario.floats[0]
Create 5 Truck Entities
Subscribe to a Calendar
Decrease dirt by 20CY
Capture Loader
Load Truck @ Source
Release Loader
Travel to dump
Dump Dirt
Counter3
No more dirt at the source?
Counter2
Return to Source
Counter1
False
True
Destroy1
Collect the total # of calendars days it took
Last Truck Entity?
Counter4
Unsubscribe Calendar

Figure 4: Earth-moving Model Layout in Simphony

Five truck entities are created at the start of the simulation which represent dump trucks. These entities flow through the model (Figure 4.0) emulating the movement of dirt from source to placement. When there is no more dirt to move, the simulation is terminated.

At the end of the simulation, the last truck entity flows through the *"Unsubscribe Calendar"* execute element where the the subscription to the calendar is undone and working days and non-working days retrieved from the calendar and saved in the respective statistics nodes. The C# code written within the execute element formula to achieve this is presented in Figure 5.0.

Three different calendars were experimented with in this model; a standard calendar, 24 Hour calendar, Night Shift and a calendar that was created with custom settings. The custom calendar used in this experiment was defined using the Simphony calendar editor (accessed through the *"Calendars"* property of the scenario – see Figure 7.0). The calendar was setup such that the work (or non-work) periods presented in Table 4.0 are utilized. For simplicity, all work periods were considered as regular time since the simulation did not model dynamics of work performance changes with work time or costs incurred due to different pay types.

This custom calendar considers Sunday as a non-working day, Saturday as a working day with one-eight-hour work shift and all other week days as working days with two-eight-hour work shifts.

Simulation results (see Table 5.0) indicate that the work scope can be completed earliest with the *"24 Hour"* calendar, followed by the *"custom"* calendar defined. Given that the *"24 Hour"* option is not a calendar per se, the*"custom"* calendar would be the most efficient option to complete the work in the shortest time. The modeller can experiment with different resource and shift configurations to obtain results that can be compared to pick a work plan that best suits their needs.

This section demonstrates that the Simphony simulation system fully supports the integration of calendars in simulation models.

```
public static partial class Formulas
{
  //Private fields for the methods to use in their
  calculations
  private static System.Boolean WasWorking = false;
  public static Simphony.Simulation.DiscreteEventEngine
  SimulationEngine;
  public static System.Boolean
  Formula(Simphony.General.Execute context)
   {
    //Set the Initial quantity of dirt to 10,000 CY
    context.Scenario.Floats[0] = 10000.0;
    //Get a Calendar to subscribe to from the scenario
    Simphony.Simulation.Calendar MyCalendar =
    context.Scenario.Calendars["My Custom Calendar"];
    //Create an instance of a calendar entity
    CalendarEntity MyCalendarEntity = new
    CalendarEntity();
//Make a global attribute point to this calendar
entity
    context.Scenario.Objects[0] = MyCalendarEntity;
    context.Scenario.Objects[1] = MyCalendar;
    //Get access to a calendar
context.Engine.SubscribeCalendar(MyCalendarEntity,MyCalend
ar,CalendarEvent);
    //Make reference to the simulation engine
    SimulationEngine = context.Engine;
    return true;
    }
  //Call the method that is related to the calendar events
  public                     static                     void
CalendarEvent(Simphony.Simulation.CalendarEntity e)
   {
    //Check if there is a change in the work state of the
calendar ---> If there is a change,Then do something
    if(e.IsWorking != WasWorking)
    {
      //It is a change from a non-work period to a work
period ---> Resume processing of events
      if(e.IsWorking)
      {
        System.Diagnostics.Trace.WriteLine("A  work  period
is      begining      now      at      time:      "      +
SimulationEngine.DateNow.DayOfWeek           +",          "+
SimulationEngine.DateNow + " and will end in "
        + e.TimeRemaining.TotalHours + " hours");
        foreach(var entity in e.Entities)
        {
          //Resume only suspended entities
          if(entity.IsSuspended)
          {
            SimulationEngine.ResumeEvent(entity);
            System.Diagnostics.Trace.WriteLine("Work    on
truck #" + entity.Id +" has just been resumed on " +
SimulationEngine.DateNow.DayOfWeek+","+SimulationEngine.Da
teNow);
           }
         }
       }
       else
       {  System.Diagnostics.Trace.WriteLine("A work break
is begining now at time: " +
         SimulationEngine.DateNow.DayOfWeek     +",     "+
SimulationEngine.DateNow + " and will end in "
         + e.TimeRemaining.TotalHours + " hours");
         foreach(var entity in e.Entities)
         {
      //Suspend all entities controlled by this calendar
          SimulationEngine.SuspendEvent(entity);
          System.Diagnostics.Trace.WriteLine("Work      on
truck #" + entity.Id.ToString() +" has
           just     been     suspended     on     "     +
SimulationEngine.DateNow.DayOfWeek + ", " +
           SimulationEngine.DateNow);
          }
         }
       }
       WasWorking = e.IsWorking;}
```

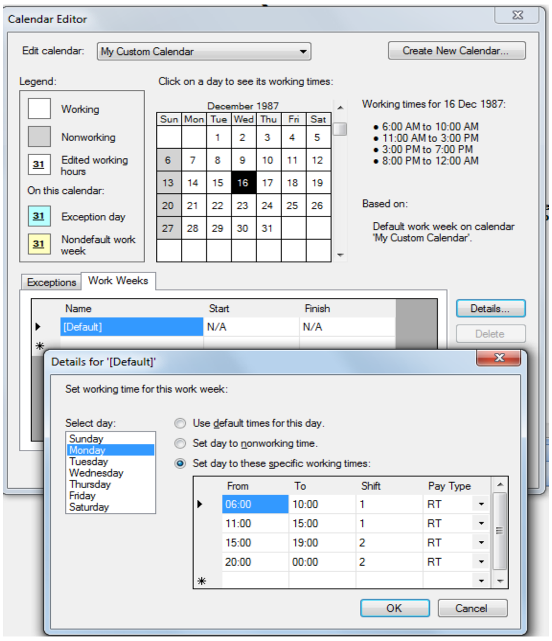
Figure 5: C# Code Snippet for Subscribing to the Calendar in Execute Shown in Figure 4.0

Table 4: Work Periods Used in the Custom Calendar

Day of Week	Work Times
Sunday	-
Monday - Friday	6:00 – 10:00
	11:00 – 15:00
	15:00 – 19:00
	20:00 – 00:00
Saturday	6:00 – 10:00
	11:00 – 15:00

```
public static partial class Formulas
{
  public              static              System.Boolean
Formula(Simphony.General.Execute context)
    {
     //Get the statistics node from the scenario for the
total work days
     Simphony.General.Statistic          S1            =
context.Scenario.GetElement<Simphony.General.Statistic>(
"Total # working days during
     simulation");
     //Get the statistics node from the scenario for the
total non-work days
     Simphony.General.Statistic S2 =
context.Scenario.GetElement<Simphony.General.Statistic>(
"Total # of non-working days
     during Simulation");
     //Get the statistics node from the scenario for the
total # of days from start to
     end of simulation
     Simphony.General.Statistic S3 =
context.Scenario.GetElement<Simphony.General.Statistic>(
"Total # of days from start
     to finish of simulation");
     //Get the calendar from the global attribute --->
So that we can get the working and
     non-working time
     Simphony.Simulation.Calendar  MyCalendar  =      \
(Simphony.Simulation.Calendar)(context.Scenario.Objects[
1]);
      //Collect   the   total   #   of   work   days
S1.Collect(MyCalendar.GetWorkingTime(context.Scenario.St
artDate,
      context.Engine.DateNow).Days);
      //Collect   the   total   #   of   non-work   days
S2.Collect(MyCalendar.GetNonWorkingTime(context.Scenario
.StartDate,
      context.Engine.DateNow).Days);
      //Collect the total # days from start to finish of
simulation
      System.TimeSpan  TS  =  context.Engine.DateNow  -
context.Scenario.StartDate;
      S3.Collect(TS.Days);
      //Trace the finish date of Simulation
      System.Diagnostics.Trace.WriteLine("The     earth-
moving operations has been completed
      on date:" + context.Engine.DateNow);
      //Get  the  Calendar  entity  -->  So  that  we  can
unsubscribe calendar
      Simphony.Simulation.CalendarEntity
MyCalendarEntity =
(Simphony.Simulation.CalendarEntity)(context.Scenario.Ob
jects[0]);
      //Unsubscribe              the              calendar
context.Engine.UnsubscribeCalendar(MyCalendarEntity);
      return true;
     }
 }
```

Figure 6: C# Code Snippet for Unsubscribing to the Calendar and Computing Work and Non-working Days



Edit calendar: My Custom Calendar

Click on a day to see its working times:

December 1987

Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Working times for 16 Dec 1987:

- 6:00 AM to 10:00 AM
- 11:00 AM to 3:00 PM
- 3:00 PM to 7:00 PM
- 8:00 PM to 12:00 AM

Based on:

Default work week on calendar 'My Custom Calendar'.

Name	Start	Finish
[Default]	N/A	N/A

Details for '[Default]'

Set working time for this work week:

Select day: Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday (Monday selected)

○ Use default times for this day.

○ Set day to nonworking time.

● Set day to these specific working times:

From	To	Shift	Pay Type
06:00	10:00	1	RT
11:00	15:00	1	RT
15:00	19:00	2	RT
20:00	00:00	2	RT

Figure 7: Dialogue for Creating or Editing Calendars in Simphony

Table 5: Experimental Results from Simulation Using Different Calendars

Calendar	Start Date	Finish Date	Total work days	Total non-work days
Standard	July 15, 2013	Aug. 13, 2013	22	7
24 Hours	**July 15, 2013**	**July 22, 2013**	**7**	**0**
Night	July 15, 2013	Aug.10, 2013	6	19
Custom	**July 15, 2013**	**July 27, 2013**	**7**	**5**

4. MODELING TRAFFIC LIGHT CONTROLS IN SIMPHONY

4.1. Problem Description

Simphony is an easy-to-use simulation system for modeling typical discrete event simulation problems regardless of whether they are within the construction domain or not. A sample problem identified from two popular simulation text books by Halpin (1992) and Pritsker (1997) is described here and used for purposes of demonstrating the modeling abilities of the Simphony environment.

One lane of a 500 m section of road is closed off for major repair work. The road comprises two lanes with traffic flowing in opposite directions (east bound traffic and west bound traffic). For this section of road, lights allow traffic to flow for a specified time interval from only one direction. This arrangement is depicted in Figure 8.0.

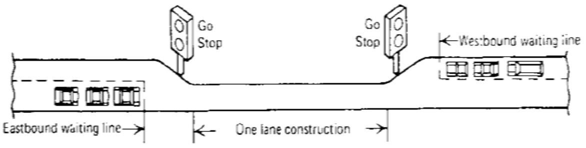


Figure 8: Schematic Layout of Site in the Traffic Light Problem (Halpin and Riggs, 1992)

When a light turns green, the waiting cars start and pass the light every 3 seconds. If a car arrives at the green light when there are no waiting cars, it passes through the light without delay. The car arrival pattern is such that there is an average of 10 seconds between cars from the east direction and 9 seconds between cars from the west direction. A light cycle consists of green for east bound traffic, both red, green for west bound traffic, both red, and then the cycle is repeated. Both lights remain red for 50 seconds to allow cars in transit to leave the repair section before traffic from the other direction can be initiated. The objective is to obtain green times for traffic lights that minimize waiting times for east and west-bound traffic.

4.2. Simphony Models, Discussion and Results

A traffic light cycle is perceived as involving a sequential process in which lights transition through different states (signals) represented by different light colors (*Green* → *All-Red* → *Red* → *Green*). Each traffic light is modelled as a resource so as to provide a convenient link between the traffic light cycle and the traffic flow (permit flows at right time, halt flows and track waiting statistics). An entity is used to loop through the cycle triggering the start and finish of each state. State change is triggered by capture or release of a traffic light resource. Higher priorities are given to the entity flowing within the *"traffic light control loop"* (for the capture of traffic light resources) compared to vehicle entities flowing in the *"traffic flow"* sub-models. The time that the system stays within a given traffic light state is modeled by task elements. *"All-Red"* time is set to 50 seconds and we are to experiment to determine an optimal value of the *"Green"* light times that minimize the waiting time of traffic flowing in both directions.

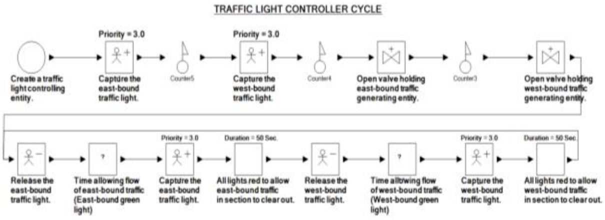

TRAFFIC LIGHT CONTROLLER CYCLE
Priority = 3.0
Priority = 3.0
Create a traffic light controlling entity.
Capture the east-bound traffic light.
Counter5
Capture the west-bound traffic light.
Counter4
Open valve holding east-bound traffic generating entity.
Counter3
Open valve holding west-bound traffic generating entity.
Priority = 3.0
Duration = 50 Sec.
Priority = 3.0
Duration = 50 Sec.
Release the east-bound traffic light.
Time allowing flow of east-bound traffic (East-bound green light)
Capture the east-bound traffic light.
All lights red to allow east-bound traffic in section to clear out.
Release the west-bound traffic light.
Time allowing flow of west-bound traffic (West-bound green light)
Capture the west-bound traffic light.
All lights red to allow west-bound traffic in section to clear out.

Figure 9: Model Layout of Traffic Light Controller Cycle

A discrete event model was developed in Simphony.NET 4.0 for the traffic system. The constructed model is comprised of 3 sub-models: (1) a traffic light control cycle, (2) an east-bound traffic flow model and (3) a west-bound traffic flow model. Each of these sub-models is discussed in detail. In these sub-models, the traffic lights for the *"east-bound"* and *"west-bound"* traffic lights are modelled explicitly as resources within the Simphony general purpose template. The entities in this model include: the traffic flowing in the east direction, west direction and a flow unit that triggers the traffic light signals (ON/OFF).

Figure 9.0 shows a layout of the sub-model that emulates a typical traffic light cycle. One entity *("traffic light controller entity")* is created at the start of simulation which captures the east and west bound traffic lights. Thereafter, the entity triggers opening of valves that were retaining entities created to generate east and west bound traffic, respectively. The traffic controller entity then releases the east-bound traffic resource so that east-bound entities arriving capture this resource and flow through the section. The resource is freed for a specific duration that emulates the time that the traffic light is green after which the traffic light is captured once again by the traffic controller entity (representing the east-bound traffic light turning red). At this point, the traffic controller entity has both traffic light resources in its possession (signaling "all-red" on traffic lights) and is transferred into a task element that holds it for 50 seconds. This 50 second delay mimics the time required by east-bound traffic (caught in the construction zone when the east-bound traffic light turns red) to clear out of this section. The traffic controller entity then flows into an element that releases the west-bound resource and then subsequently into a task element that delays it for a duration equivalent to that for which the west-bound traffic light is green. The west-bound traffic light resource is made available to west-bound traffic entities that were queued or are just arriving, hence, allowing them to flow through the construction zero. Thereafter, the "traffic controller entity" requests for the "west-bound traffic light resource" with a high priority (3.0). It is granted this resource after the current "west-bound traffic flow entity" utilizing it releases it. The "traffic controller entity" will once again have both traffic light resources in its possession and is transferred into an *"all-red"* task element for 50 seconds during which west-bound traffic currently flowing in the construction zero section is expected to clear out. The *"traffic light controller entity"* is then looped back to the start of the traffic light cycle where it resumes with the release of the *"east-bound traffic light resource."*

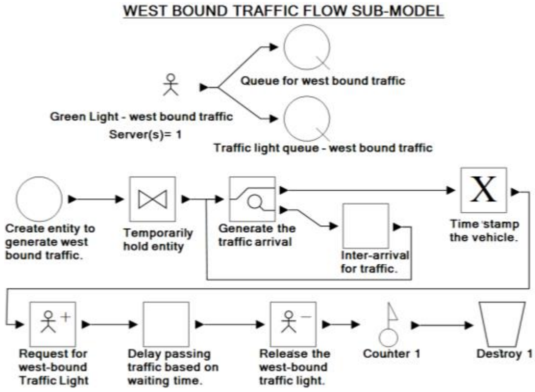

WEST BOUND TRAFFIC FLOW SUB-MODEL
Queue for west bound traffic
Green Light - west bound traffic
Server(s)= 1
Traffic light queue - west bound traffic
Create entity to generate west bound traffic.
Temporarily hold entity
Generate the traffic arrival
Inter-arrival for traffic.
Time stamp the vehicle.
Request for west-bound Traffic Light
Delay passing traffic based on waiting time.
Release the west-bound traffic light.
Counter 1
Destroy 1

Figure 10: Model Layout for West-Bound Traffic Flow

The east-bound (EB) and west-bound (WB) traffic flow sub-models represent the arrival, queuing and flow of traffic in the east and west directions, respectively. The model layouts (Figures 10 and 11) are identical but involve different resources *("East-Bound Traffic Resource"* and *"West-Bound Traffic Resource"*), waiting files (*"Queue for East-Bound Traffic", "Queue for West-Bound Traffic", "Traffic Light Queue-East Bound Traffic" and "Traffic Light Queue-West Bound Traffic")*, valves, tasks, capture and release elements. In these sub-models, the waiting files for traffic entities are separated from those of the *"traffic light controller entity"* so that the statistics on queued traffic are not distorted.

One entity is created in each sub-model at time zero and held behind a valve control until the "*traffic light controller entity*" has captured the *"East-Bound Traffic Resource"* and *"West-Bound Traffic Resource"* and triggered the valves to open. This entity in each sub-model serves as a *"traffic generating entity."* It is transferred into a *"generate element"* which clones it.

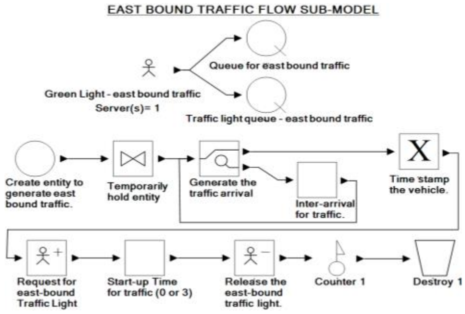
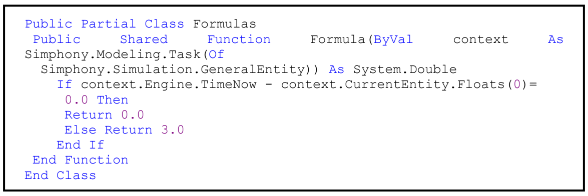

EAST BOUND TRAFFIC FLOW SUB-MODEL
Queue for east bound traffic
Green Light - east bound traffic
Server(s)= 1
Traffic light queue - east bound traffic
Create entity to generate east bound traffic.
Temporarily hold entity
Generate the traffic arrival
Inter-arrival for traffic.
Time stamp the vehicle.
Request for east-bound Traffic Light
Start-up Time for traffic (0 or 3)
Release the east-bound traffic light.
Counter 1
Destroy 1

Figure 11: Model Layout for East-Bound Traffic Flow

The entity flowing out of the top point of the *"generate element"* represents an arrival of a vehicle and is routed into an *"execute element."* The cloned entity is transferred out of the bottom output point of the *"generate element"* into a *"task element"* where it is delayed for the inter-arrival duration before being re-routed into the *"generate element"* to release another entity that represents another vehicle arrival. This cyclic process keeps going until the simulation is terminated.

Arriving traffic entities flow through the *"execute element"* where they are time-stamped with the time at which they arrive at the construction zone. Arriving traffic entities then proceed to a capture element where they request their respective traffic light resource. If the traffic light resource is available, the traffic entity proceeds on its journey without delay; otherwise, it is queued until the traffic light resource becomes available. Traffic entities that were queued and are allowed to travel through the construction zone when the light turns green are delayed by 3 seconds as they pass-by the traffic light. These 3 seconds represents start-up time for vehicles moving from a complete stop. This logic is modelled with the *"task element"* using the VB code snippet shown in Figure 12. This was inserted into the formula editor of the duration property for the *"task element."*

After the traffic entity passes by the traffic light, it releases the traffic light resource to the next entity. It then flows through the counter element where the traffic count is registered and then into a "destroy element" where it is removed from the simulation. The flow of traffic entities is halted when the green time is used up (and the *"traffic light controller entity"* captures the traffic light resource).



```
Public Partial Class Formulas
 Public     Shared     Function     Formula(ByVal     context     As
Simphony.Modeling.Task(Of
  Simphony.Simulation.GeneralEntity)) As System.Double
    If context.Engine.TimeNow - context.CurrentEntity.Floats(0)=
     0.0 Then
     Return 0.0
     Else Return 3.0
    End If
 End Function
End Class
```

Figure 13: VB Code for Generating a Delay for a Vehicle Passing a Traffic Light

Simulation settings used are summarized in Table 6.0. These were used because of the stochastic inputs used e.g. the inter-arrivals of traffic. Also, a seed was fixed to ensure that the same sequence of random deviates is used for scenarios that are compared.

To determine the green times to allocate to the east-bound and west-bound traffic, equal arbitrary values were used. This phase of the experiment was used to get the local minimum (waiting time for traffic flowing in each direction). These were different because traffic inter-arrivals in each direction were different. Results from this phase are summarized in Table 7.0 and are plotted in Figure 14.0. The acronyms EB-GT, WB-GT, EB-WT and WB-WT represent east-bound green time, west-bound green time, east-bound waiting time and west-bound waiting time in seconds respectively.

Table 6: Simulation Setting Used for the Traffic Problem

Simulation Setting	Parameter Value
Seed	5,000
Run Count	100
Time Unit	Seconds
Maximum Time	86,400 Seconds (1 day)

Table 7: Phase I Results from Experimenting with the Traffic Model

EB-GT	WB-GT	EB-WT	WB-WT
50	50	6,979.03	10,505.42
60	60	4,089.52	7,890.69
80	80	225.0058	3,186.14
100	100	115.173	344.3368
110	110	110.3978	192.3788
120	120	**109.3351**	154.9445
140	140	111.7418	130.1684
150	150	113.9442	129.5323
160	**160**	116.5949	**128.2344**
180	180	122.5078	132.7623
200	200	128.9449	137.8971
220	220	135.7282	143.9771
240	240	142.6467	150.9405

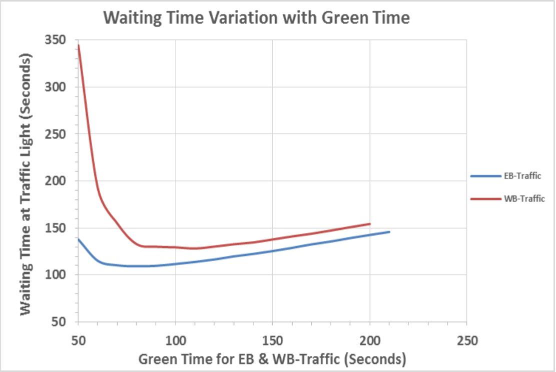


Figure 14: Waiting Time Variation with Green Time

The values obtained from phase I (highlighted in bold in Table 7.0) are used to guide phase II of the experimentation which involves determining the global minimum waiting time for all traffic. Results from this phase are summarized in Table 8.0. Optimal green times were found to be 130 and 140 seconds for east-bound and west-bound traffic respectively.

Table 8: Phase II Results from Experimenting with the Traffic Model

EB-GT	WB-GT	EB-WT	WB-WT
120	160	194.40	97.19
120	150	155.46	101.47
125	160	155.20	100.62
130	160	139.64	104.34
130	155	133.05	106.57
130	150	127.47	109.70
130	**140**	**117.84**	**117.45**

5. CONCLUSIONS

The paper presents a concise overview of simulation, the existing simulation methods, different simulation systems and studies in which simulation has been previously applied within the construction domain.

Simphony is introduced as an example of typical simulation system currently in use, its features discussed and reasons why it remains relevant in the process of defining next generation simulation tools/systems highlighted.

Two practical problems (an earth-moving problem involving shift dynamics and a traffic light problem) modeled in Simphony and experimented with to generate results that can be used to support decision making processes are described to showcase capabilities and features that exist within Simphony.

REFERENCES

AbouRizk, S. M., Ruwanpura, J. Y., Er, K. C., and Fernando, S., 1999. Special purpose simulation template for utility tunnel construction. *Proceedings of Winter Simulation Conference*, 2: 948-955. Dec. 5-8; Phoenix, AZ.

Ahn, C., Rekapalli, P. V., Martinez, J. C., and F. A. Pena-Mora., 2009. Sustainability analysis of earthmoving operations. *Proceedings of Winter Simulation Conference*, pp. 2605-2609. Dec. 13-16, Austin, TX.

Ailland, K., Bargstadt, H., and Hollermann, S., 2010. Construction process simulation in bridge building based on significant day-to-day data. *Proceedings of Winter Simulation Conference*, pp. 3250-3261. Dec. 5-8, Baltimore, MD.

Al-Bataineh, AbouRizk, S. M., and Parkis, H., 2013. Using simulation to plan tunnel construction. *Journal of Construction Engineering and Management*, 139(5): 564–571.

Alvanchi, A., Nguyen, A., and AbouRizk, S. M., 2012. Structural steel fabrication special purpose simulation. *Construction Research Congress,* pp. 1391-1399. May, 19-23, Purdue, IN.

Chehayeb, N. N. and AbouRizk S. M., 1998. Simulation-based scheduling with continuous activity relationships. *Journal of Construction Engineering & Management,* 124(2): 107-115.

Cheng, F. F., Wang, Y. W., Ling, X. Z., and Bai, Y., 2011. A Petri net simulation model for virtual construction of earthmoving operations. *Journal of Automation in Construction,* 20: 181-188.

Chui-Te, C., Tseng-Hsing, H., Ming-The, W., and Hsien-Yen, C., 2011. Simulation for steel bridge erection by using BIM tools. *Proceedings of 28th ISARC,* pp. 560-563. June 29, Seoul, Korea.

Crain, R. C., 1997. Simulation using GPSS/H. *Proceedings of Winter Simulation Conference,* pp. 567–573. Dec. 7-10, Atlanta, GA.

Dori, G., and Borrmann, A., 2011. Automatic generation of complex bridge construction animation sections by coupling constraint-based discrete-event simulation with game engines. *Proceedings of International Conference on Construction Applications of Virtual Reality*. Nov. 3-4, Weimar, Germany.

Dulcy, M. A., and Halpin, D. W., 1998. Simulation of the construction of cable-stayed bridges. *Canadian Journal of Civil Engineering,* 25: 490-499.

Fu, J., 2012. A microscopic simulation model for earthmoving operations. *World Academy of Science, Engineering and Technology,* 67: 218-223.

Hajjar, D., and AbouRizk, S. M., 1999. Simphony: An environment for building special purpose construction simulation tools. *Proceedings of Winter Simulation Conference,* 2: 998-1006. Dec. 5-8, Phoenix, AZ.

Halpin, D. W., 1973. *An investigation of the use of simulation networks for modeling construction operations.* PhD. Thesis, University of Illinois, Urbana-Champaign, III.

Halpin, D.W., and Riggs, L.S., 1992. *Planning and Analysis of Construction Operations.* John Wiley & Sons, Inc.

Hu, D., and Mohammed, Y., 2010. State-based simulation mechanism for facilitating project schedule updating. *Construction Research Congress,* pp. 369-378. May 8-10, Banff, AB.

Ioannou, P. G., and Martinez, J. C., 1996. Comparison of construction alternatives using matched simulation experiments. *Journal of Construction Engineering and Management*, 122(3): 231-241.

Liu, Y. and Mohamed, Y., 2008. Multi-agent resource allocation (Mara) for modeling construction processes. *Proceedings of Winter Simulation Conference*, pp. 2361-2369. Dec. 7-10, Miami, FL.

Liu, H., Siu, M. F., Hollerman, S., Ekyalimpa, R., Lu, M., AbouRizk, S., and Bargstadt, H., 2012. Simulation of mobile falsework utilization methods in bridge construction. *Proceedings of Winter Simulation Conference, pp. 1-13.* Dec. 9-12, Berlin, Germany.

Lu, M., 2003. Simplified discrete-event simulation approach for construction simulation. *Journal of Construction Engineering and Management*, 129(5): 537-546.

Martinez, J. C., 1996. *STROBOSCOPE – State and resource based simulation of construction processes*. PhD thesis, University of Michigan.

Marzouk, M., El-Dein, H. Z., and El-Said, M., 2007. Application of computer simulation to construction of incremental launching bridges. *Journal of Civil Engineering and Management*, 13(1): 27-36.

Marzouk, M., and Moselhi, O., 2002. Simulation optimization for earthmoving operations using genetic algorithms. *Journal of Construction Management and Economics*, 20(6): 535-543.

Mohsen, O. M., Knytl, P. J., Abdulaal, B., Olearczyk, J., and Al-Hussein, M., 2008. Simulation of modular building construction. *Proceedings of Winter Simulation Conference,* pp. 2471-2478. Dec. 7-10, Miami, FL.

Pritsker, A. A. B, O'reilly, J. J., and Laval, D. K. , 1997. *Simulation with visual SLAM and AweSim*. New York, NY: John Wiley & Sons, Inc.

Rekapalli, P. and Martinez, J., 2011. Discrete-event simulation-based virtual reality environments for construction operations: Technology introduction. *Journal of Construction Engineering and Management,* 137(3): 214–224.

Sadeghi, N. and Fayek. A. R., 2008. A framework for simulating industrial construction processes. *Proceedings of Winter Simulation Conference,* pp. 2396-2401. Dec. 7-10, Miami, FL.

Song, L., and AbouRizk, S. M., 2003. Building a virtual shop model for steel fabrication. *Proceedings of Winter Simulation Conference,* pp. 1510-1517. Dec. 7-10, New Orleans, LA.

Song, L., and AbouRizk, S. M., 2006. Virtual shop model for experimental planning of steel fabrication projects. *Journal of Computing in Civil Engineering,* 20(5): 308-316.

Sun, C.-T., Wang, D.-Y., & Chang, Y.-Y., 2013. Effects of thinking style on design strategies: Using bridge construction simulation programs. *Educational Technology & Society,* 16 (1): 309–320.

Touran, A. and Asai, T., 1987. Simulation of tunneling operations. *Journal of Construction Engineering and Management*, 113(4): 554–568.

Wang, P., Mohamed, Y., Abourizk, S., and Rawa, A., 2009. Flow production of pipe spool fabrication: Simulation to support implementation of lean technique. *Journal of Construction Engineering and Management*, 135(10): 1027–1038.

Zhang, H., Tam, G. M., and Shi, J. J., 2002. "Simulation-based methodology for project scheduling." *Journal of Construction Management and Economics,* 20(8): 667-678.

Zhou, F., AbouRizk, S. M., and Fernando, S., 2008. A simulation template for modeling tunnel shaft construction. *Proceedings of Winter Simulation Conference*, pp. 2455-2461. Dec. 7-10, Miami, FL.

ANALYSIS OF STATES IN PRODUCTION SYSTEMS MODELED BY PETRI NETS USING DIOPHANTINE EQUATIONS

Julián Gómez-Munilla[a], Emilio Jiménez-Macías [b], Juan-Ignacio Latorre-Biel [c], Mercedes Pérez de la Parte[a] , Jorge Luis García-Alcaraz[d]

[a] University of La Rioja. Industrial Engineering Technical School.
Department of Mechanical Engineering. Logroño, Spain
[b] University of La Rioja. Industrial Engineering Technical School.
Department of Electrical Engineering. Logroño, Spain
[c] Public University of Navarre. Department of Mechanical Engineering, Energetics and Materials.
Campus of Tudela, Spain
[d] Technology and Engineering Institute, Universidad Autónoma de Ciudad Juárez, Chihuahua (Mexico)

(a) julian.gomez@unirioja.es , mercedes.perez@unirioja.es , (b) emilio.jimenez@unirioja.es ,
(c) juanignacio.latorre@unavarra.es , (d) jorge.garcia@uacj.mx

ABSTRACT

This paper is devoted to study ways to determine the possible states of a production system, modeled by Petri nets (PN), making use of techniques for solving systems of linear Diophantine equations. For this porpoise, PNs have been divided into three possible types, and each provides a specific method so as to exploit its characteristics to optimize the computation time. These types are conservative systems, nonconservative systems with conservative components, and nonconservative systems without conservative components. These proposal methods have been compared with the determination of states from marking evolution, being clearly advantageous in computation time.

Keywords: Petri Net, Discrete Event Systems, Diophantine Equations, State equation

1. INTRODUCTION

Petri Nets constitutes a family of formalisms with mathematical-graph duality, which allows to efficiently representing discrete event systems, especially when they presents concurrency and shared resources. The study of the states of the systems being modeled is needed to understand its behavior and properties, so that knowing all states of the PN (the marking) is a critical task.

This paper aims to explore ways to determine the possible states of a production system (Jimenez et *al.*, 2006, 2009, 2012), modeled by PN (Latorre et *al.*, 2009, 2013a) and analyzing its states using resolution techniques (Latorre et *al.*, 2013b, 2013c) of linear Diophantine equations systems. PNs are classified into 3 types, and each provides a specific method so as to exploit its characteristics to optimize the computation time. These types are: conservative systems, but nonconservative systems with conservative components, and nonconservative systems without conservative components. These proposal methods have been compared with the determination of states from marking evolution, being clearly advantageous in computation time.

2. PETRI NETS

Petri nets are formed by a set of places (P), another of transitions (T), and arcs (F), each with a given weight (W), connecting places transitions and vice versa. The set of places, transitions, arcs and weights defines the structure of the system (Murata, 1989; Silva, 1993; Girault and Valk, 2001).

Graphically, places are represented by circles, transitions by bars or rectangles, and arcs with arrows. To indicate the weight of the arcs, when greater than one, numbers are typically placed beside the arrows. The marking of the net is represented by tokens (points) in the places, or with numbers if the number of tokens is high.

The PN structure is represented by the incidence matrix, where rows represent different places and transitions columns. The elements of the matrix indicate the number of tokens that appear (if positive) or disappear (if negative) in each place -row- when a transition-column- is fired. The PN marking is represented by a column vector that has so many components as places in the PN.

The state equation of a PN determines which is the marking (state) reached after firing a number of transitions from an initial marking.

$$m = m_0 + C \cdot \sigma \qquad (1)$$

Being:

m: Final marking vector

m_0: Initial marking vector

C: Incidence Matrix

σ: Firing verctor**.** Column vector with as many rows as transitions in the system. Their values represent the number of times that each transition is fired.

The state equation does not take into account whether the firing vector is really applicable, since it is not considered the order in the firings (possible states could be reached from the point of view of that equation, which could not be reached by a sequence firing of transitions).

The properties of the PN can be of two types: structural and dynamic. The main difference is that the first ones depend only on the structure of the system, while the latter also depends on the initial marking.

Structural properties, being only dependent on the structure, can be determined mathematically. Among these we highlight 2, which are important for this analysis:

Conservativeness: a PN is conservative if, for any marking, the sum of all their marking in the places, each multiplied by a factor, remains constant. Mathematically, if the vector of factors or weights:

$$W \cdot m = W \cdot (m_o + C \cdot \sigma) = W \cdot m_o + W \cdot C \cdot \sigma \quad (2)$$

To satisfy the condition $W \cdot m = W \cdot m_o$, it is necessary that $W \cdot C \cdot \sigma = \mathbf{0}$, and since the firing vector is indifferent because it is a structural property, observe that:

$$W \cdot C = \mathbf{0} \quad (3)$$

The vector or vectors W that satisfy the equation (3) are left cancellers of the incidence matrix, or labeled invariant or conservative components.

For the PN is conservative, all sites must be contained in (or "covered" by) any marking invariant component or conservative component. If there are places not covered, the PN is not conservative, although the covered places do present conservation propertires.

Repeatability: A PN is repetitive if there exists a sequence of firings of all transitions that returns the system to an initial marking. Mathematically, if σ_r is the firing vector that returns the PN to the initial marking:

$$m = m_o + C \cdot \sigma_r \quad (4)$$

To satisfy the condition $m = m_o$, it is necessary that:

$$C \cdot \sigma_r = \mathbf{0} \quad (5)$$

The vector or vectors σ_r satisfy the equation (5) are right cancellers of the incidence matrix, or firing invariant, or or repetitive components.

So that the PN be repetitive, all transitions must be contained in (or "covered" by) any firing invariant or repetitive component. If there are uncovered transitions, the PN is not repetitive, although the covered transitions are.

The repeatability property is not used as such in this work, although it has been considered for its close relationship with conservativeness (they are the same for the dual PN).

3. DIOPHANTINE EQUATIONS

3.1. Definition

A Diophantine equation is one that admits only as solution an integer. Specifically, diophantine linear equations have the form:

$$a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_n \cdot x_n = b \quad (6)$$

Where $a_1, a_2, \ldots, a_n$, and b are known integers and x_1, $x_2, \ldots$, and x_n are unknown integers. If $b = \mathbf{0}$, it is said that the Diophantine equation is homogeneous. A Diophantine equations linear system is a system in which equations are of the form (6).

3.2. Euclidean algorithm and Bezout identity

The resolution of a linear Diophantine equation is based on the Euclidean algorithm, which is used to calculate the greatest common divisor of two integers (Ajili and Contejean, 1995, Bradley, 1971; Havas et *al*., 1998; Hemmecke, 2011; Lazebnik, 1996). There is the property that, if d is the greatest common divisor of a and b, then there are two integers x and y such that the reste of the division $\frac{a \cdot x + b \cdot y}{d}$ is zero.

Bézout's identity says that there are two integers, m and n such that:

$$d = a \cdot m + b \cdot n \quad (7)$$

That is, the greatest common divisor of a and b can be represented as a linear combination of these two integers a and b.

3.3. Algorithms for solving Diophantine equations

There are, at present, several algorithms for solving systems of Diophantine equations. Some of them use the so-called Hermite normal form, and others are based on repeatedly test possible solutions. The algorithm used in this work to solve systems of equations is of the second type, because they do not want to find all solutions, but only those who are between zero and a limit (Clausen and Fortenbacher, 1989; Contejean and Devie, 1994), either because the system is conservative (and if so inherently limited), or because it is not but our intention is to determine progressively the states, by including limits in the not limited places and increasing these limits prograsively to build the reachability tree in an structured and organized way.

3.4. Application of Diophantine equations to PN

Diophantine equations appear many times in the mathematical resolution of PN. Both the firing vector as the marking vector are composed always by non-negative integers, and therefore any equation or system of equations in which one of the two vectors is the unknown is a diophantic equation.

4. PROPOSED METHODS

Following they are presented a series of methods that aim to solve the problem of identifying all possible state in a Petri net from an initial marking, depending on the type of PN concerned.

4.1. General methods

This case can be used in all types of PN, and is the system used to non-conservative systems without marking invariants; for those who do have conservative componentsn other methods will be proposed in the following sections.

4.1.1. Resolution of the state equation

This method is valid for conservative Petri nets and for non conservative. It consists of applying the state equation to all possible marking and selecting those for which there are non-negative integer solutions for the firing vector.

Let be a PN with n places, all of them with the marking limited between 0 and a maximum (*lim*) number of tokens. Therefore, there appear $(lim + 1)^n$ possible markings. This is equivalent to having a word of n letters, where each letter can take the values $\{0,1,\ldots,lim\}$, what is known in combinatorics as variations with repetition of $lim + 1$ elements taken from n to n or $VR(lim + 1, n)$.

For each of the possible markings, it is necessary to verify the existence of a firing vector between the initial marking and studied marking. A number of $(lim + 1)^n$ linear Diophantine equations systems should be solved, with the form:

$$m = m_o + C \cdot \sigma \Rightarrow C \cdot \sigma = m_o - m \qquad (8)$$

In all of them, C and m_o are similar. The analysed marking m belongs to the system iff there exists a vector σ, whose components are non-negative integer and such that satisfies the equation (8).

The advantage of this method is that all checked possible markings within the limits are obtained, including those that are not achievable by evolution from the initial marking by firing successive transitions (spurious states). The main drawback is the time spent, because $(lim+1)^n$ different systems of linear Diophantine equations should solved. This makes it infeasible in many cases the use of this method.

4.1.2. Evolution of successive shots marked by transitions

This method is also applicable to conservative and non-conservative PN. It consists on firing every transition from the initial marking to determine the following markings. A marked with a negative number of marks somewhere means that bthe state is not possible.

Let be a PN with n places where each marking can be between 0 and a maximum (*lim*) of tokens, and m transitions; using the state equation:

$$m = m_o + C \cdot \sigma \quad \Rightarrow \qquad (9)$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{1,0} \\ \vdots \\ x_{n,0} \end{bmatrix} + \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nm} \end{bmatrix} \cdot \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}$$

Applying only one firing from an only transition t_j, the firing vector presents this form:

$$\begin{bmatrix} t_1 \\ \vdots \\ t_j \\ \vdots \\ t_m \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \qquad (10)$$

Then, equation (9) is equivalent to (11):

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{1,0} \\ \vdots \\ x_{n,0} \end{bmatrix} + \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nm} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} x_{1,0} \\ \vdots \\ x_{n,0} \end{bmatrix} + \begin{bmatrix} c_{j1} \\ \vdots \\ x_{jn} \end{bmatrix} \qquad (11)$$

I.e., firing only once the transition t_j is equivalent to to adding the column j of the incidence matrix.

A firing of a new transition is applied to the markins so obtained, in order to obtain a new set of markings.

Naming m_1 the set of valid markings obtained after firing once every transition from m_o; m_2 the set of valid markings obtained after firing once every transition from m_1, etc.:

$$\begin{aligned} m_1 &= m_0 + C \cdot \sigma \\ m_2 &= m_1 + C \cdot \sigma \\ &\ldots \qquad (12) \\ m_n &= m_{n-1} + C \cdot \sigma \end{aligned}$$

The process stops when no new markings are obtained. This situation is caused by any of the following reasons:

1) Because it is not possible to fire any transition (there are deadlocks).
2) Because it is possible to fire any transition but would exceed the limit marks (valid states but not desired).
3) Because it is possible to fire any transition but this would already drive to previous markings.

It is possible, however, did not obtain all valid markings (with a number of tokens between 0 and *lim*). For example, you can reach invalid marking as follows:

$$m_n = \begin{bmatrix} a \\ b \\ \lim + k \\ c \end{bmatrix}; k > 0 \qquad (13)$$

There may be a column in the incidence matrix C such that:

$$c_{1:n,j} = \begin{bmatrix} d \\ e \\ -k \\ f \end{bmatrix}; |d| \le a, |e| \le b, |f| \le c \qquad (14)$$

When firing the transition j from m_n, the following marking would be obtained:

$$m_{n+1} = \begin{bmatrix} a+d \\ b+e \\ \lim \\ c+f \end{bmatrix} \qquad (15)$$

is clearly valid, as the number of tokens in all places is between 0 and, *lim* and there is no reason for having being reached previously.

A tolerance to accept provisionally markings with a number of tokens bigger than the limit can be allowed, then removing them from the final results. The advantage of this method is that it is remarkably fast: it is only necessary to sum column vectors and discard the result if an element is negative; there is no need to solve any equation.

The main drawback is that it is impossible to determine spurious states, since they, by definition, are not achievable by successive firing of transitions from the initial marking.

4.1.3. Identification of spurious markings

The application of the two methods shown so far, permits us to identify which marked are spurious and which are not.

Calling M the set of marked obtained by solving the equation of state in the $(lim + 1)^n$ possible markings (or by other means, as discussed later), and M_t the set of marked not spurious (obtained by initial developments by firing transitions) is clear that:

$$M_t \subseteq M \qquad (16)$$

Furthermore, if M_s is the set of labeled spurious:

$$M_t \cap M_s = \varnothing;\ M = M_t \cup M_s \rightarrow$$
$$M_s = \{\, M \setminus M_t\} \qquad (17)$$

That is, all the markings obtained by solving the state equation that have not been obtained by evolution by firing transitions are spurious markings.

4.2. Conservative PN

Conservative Petri Nets have marking invariants covering all places. So there are mathematical laws that apply to all possible markings, spurious or not. The conservative PN are also limited. It is convenient to take advantage of this fact to determine all possible markings.

Let be a conservative PN with n places, whose q conservative components or marking invariants are the rows of a matrix W. By definition:

$$W \cdot \mathrm{C} \cdot \sigma = 0 \iff W \cdot m = W \cdot m_{\mathrm{o}} \qquad (18)$$

Then:

$$\begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{q1} & \cdots & w_{qn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{q1} & \cdots & w_{qn} \end{bmatrix} \cdot \begin{bmatrix} x_{1,0} \\ \vdots \\ x_{n,0} \end{bmatrix} =$$
$$= \begin{bmatrix} w_{1,1:n} & \times & w_{1:n,0n} \\ \vdots & \vdots & \vdots \\ w_{q,1:n} & \times & w_{q:n,0n} \end{bmatrix} \qquad (19)$$

The only unknown in (19) is the vector $m = [x_1,\ldots x_i,\ldots,x_n]^{\mathrm{T}}$. It has to meet that $0 \le x_i \le lim\ \ \forall\ i \in \{1,\ldots,n\}$. Then, (19) is a system of linear Diophantine equations whose solutions are the possible markings of the system, including the spurious ones.

The advantages of this method are clear: after obtaining the conservative components, a system of linear Diophantine equations must be solved to determine all markings, including spurious.

The main disadvantage is precisely that it only applies to conservative systems. A modification of the method may, however, greatly simplify determining markings in nonconservative PN with marking invariants.

4.3. Nonconservative PN with marking invariants

A Petri net with marking invariants is not conservative if there are places "not covered", ie if the vectors indicating the conservative components have one or more columns whose value is null in all the elements. It can be taken advantage of the fact that the places are covered in the conservative components (and therefore the PN itself is limited in those places) to greatly simplify the search for markings.

Let be a Petri net with n places and m transitions, whose q conservative components or marking invariants are the rows of a matrix W. Let p be the number of locations covered by the conservative components, and $n\text{-}p$ the number of places not covered. Then, the set P of index of places can be divided into two subsets: $P_{n\text{-}p}$ the set of indexes of places covered, and $P_{n\text{-}p}$ with indexes of the places not covered. So:

$$P_n \cap P_{n\text{-}p} = \varnothing;\ P_p \cup P_{n\text{-}p} = P \qquad (20)$$

Then eliminating the places not covered by the conservative components (or what is the same, eliminating the columns of zeros):

$$\begin{bmatrix} w_{1,1} & \cdots & w_{1,i} & \cdots & w_{1,n} \\ \vdots & & \vdots & & \vdots \\ w_{q,1} & \cdots & w_{q,i} & \cdots & w_{q,n} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} =$$
$$= \begin{bmatrix} w_{1,1} & \cdots & w_{1,i} & \cdots & w_{1,n} \\ \vdots & & \vdots & & \vdots \\ w_{q,1} & \cdots & w_{q,i} & \cdots & w_{q,n} \end{bmatrix} \cdot \begin{bmatrix} x_{1,0} \\ \vdots \\ x_{i,0} \\ \vdots \\ x_{n,0} \end{bmatrix} \forall i \in P_p \qquad (21)$$

The only unknown in (21) is the vector $m = [x_1,\ldots x_i,\ldots,x_n]^{\mathrm{T}}\ \forall\ i \in P_p$. That must be met $0 \le x_i \le lim$ $\forall\ i \in P_p$. Then (21) is a system of linear diophantine equations whose solutions are the set of possible markings of the places covered by the conservative components.

Let be M_p the set of possible markings of places covered by the conservative components. For each of them there is a firing vector such that, applied to the initial marking in the covered places, brings the system to a final marking $m \in M_p$.

To apply the state equation to the places of P_p, it is necessary to eliminate from the incidence matrix the rows corresponding to the places not covered by conservative components (22).

The only unknown in (22) is the firing vector $\sigma = [t_1,\ldots,t_m]^{\mathrm{T}}$, whose values must be non-negative integers. Then (22) is a linear system of Diophantine equations to be applied to each marking obtained in (21).

Let be J the set of vectors $\sigma = [t_1,\ldots,t_m]^{\mathrm{T}}$ that satisfy (22) for all $m \in M_p$, ie all possible firing vectors that, applied to the initial marking, lead to a marking $m \in M_p$.

It is necessary to check that the various firing vectors J are also applicable to places of the set $P_{n\text{-}p}$, ie, the *n-p* places not covered by the conservative components.

$$\begin{bmatrix} x_1\\ \vdots\\ x_i\\ \vdots\\ x_n\end{bmatrix} = \begin{bmatrix} x_{1,0}\\ \vdots \\ x_{i,0}\\ \vdots\\ x_{n,0}\end{bmatrix} + \begin{bmatrix} c_{1,1} & \cdots & c_{1,m}\\ \vdots & & \vdots\\ c_{i,1} & \cdots & c_{i,m}\\ \vdots & & \vdots\\ c_{n,1} & \cdots & c_{n,m}\end{bmatrix} \cdot \begin{bmatrix} t_1\\ \vdots\\ t_m\end{bmatrix}$$

$$\forall m = \begin{bmatrix} x_1\\ \vdots\\ x_i\\ \vdots\\ x_n\end{bmatrix} \in M_p, \forall i \in P_p \qquad (22)$$

Simply apply the state equation, this time to all places:

$m = m_0 + \mathrm{C} \cdot \sigma \;\; \forall \;\; \sigma \in J \;\; =>$ (23)

$$\begin{bmatrix} x_1\\ \vdots\\ x_i\\ \vdots\\ x_n\end{bmatrix} = \begin{bmatrix} x_{1,0}\\ \vdots \\ x_{i,0}\\ \vdots\\ x_{n,0}\end{bmatrix} + \begin{bmatrix} c_{1,1} & \cdots & c_{1,m}\\ \vdots & & \vdots\\ c_{i,1} & \cdots & c_{i,m}\\ \vdots & & \vdots\\ c_{n,1} & \cdots & c_{n,m}\end{bmatrix} \cdot \begin{bmatrix} t_1\\ \vdots\\ t_m\end{bmatrix}; \forall \sigma = \begin{bmatrix} t_1\\ \vdots\\ t_m\end{bmatrix} \in J, \forall i \in P$$

Let M_t be the set of markings obtained after applying (23). So marked as many vectors are obtained from the above equation as firing vectors has been tested:

$$|M_t| = |J| \qquad 24)$$

It is necessary to verify that these markings are invalid. Clearly covered places will have a valid marking because the firing vectors σ ∈ *J* were obtained from them. However, it is possible that the places not covered do not have a valid marking. It should be checked for all markings $m \in M_t$ such that $0 \le x_i \le lim$ $\forall\, i \in P_{n\text{-}p}$, and discard markings that does not fulfill this condition, as well as the firing vectors that have led the system that led to them.

The non-rejected firing vectors must be applied again to the not discarded markings until any one exceeds the defined limit.

The proposed method is summarized in the following steps:

1) Regardless of the places not covered by the conservative components, find all possible markings solving the system of linear Diophantine equations (21).

2) By applying the state equation for each possible marking, again disregarding the places not covered, draw the firing vector (22).

3) Taking into account all places, apply firing vectors obtained in the previous section to the initial marking.

4) Discard the markings obtained in (23) that are not possible for having a negative number somewhere or exceeding the bound, as well as the firing vectors that led to them.

5) Continue implementing the non rejected firing vectors to the valid markings and to those obtained from them, up to reach the limits defined or even not be possible to continue applying them.

The advantages of this method with respect to the general methods is that it considerably reduces the number of equations to be solved, since it takes advantage of the conservative components (taking into account the limitation of marking locations covered). Another advantage is that it allows obtaining the spurious markings.

4.4. Recommendations on the proposed methods

In view of the described methods, the mode of operation to determine the states of a PN may be as follows:

- First, you must determine the type of PN in question: Conservative, non conservative but with marking invariants, or non conservative without marking invariants.
- In case there are invariant markings, determine all possible markings by the specific methods proposed. Apply the method of "evolution of markings by successive firing of transitions" to determine which markings are achievable. The possible markings that are not achievable, are spurious.
- In the case that the PN has no marking invariants, there is a "fast" method to determine all possible markings. It may be a reasonable option to search only those that are achievable by firing transitions. However, in PN with a limit of small tokens and few places, and always depending on the time and resources available, it may be applicable the general method of "Solving the state equation".

5. EXAMPLE OF APPLICATION

To exemplify the approach consider a production system with a robot, which takes pieces from two conveyor belts, each with a part type, and fed to two machines which make two types of products with those pieces (Fig.1).

The PN that models that process is presented in Fig2. The numbering of places and transitions, not included to simplify the drawing, is in both from top to bottom and from left to right. Thus the incidence matrix is as shown in Table 1, and the initial marking is:

m_0=[1,1,1,0,0,2,1,1,2,0,0,0,0,0,0]^T

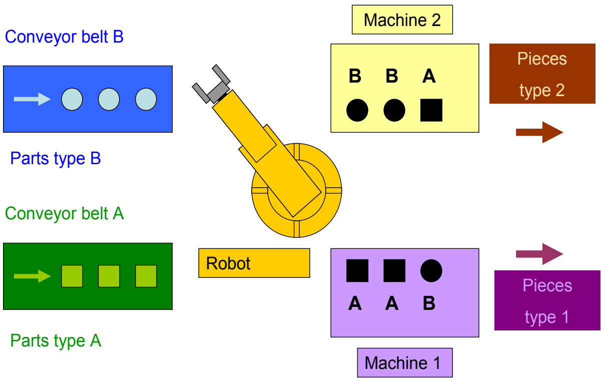


Figure 1: Sample of production system

Table 1: Incidence matrix of Pn in Fig. 2

-1	0	1	1	0	0	0	0	0	0
0	-1	0	0	1	1	0	0	0	0
-1	-1	1	1	1	1	0	0	0	0
1	0	-1	-1	0	0	0	0	0	0
0	1	0	0	-1	-1	0	0	0	0
0	0	-1	0	0	0	0	0	2	0
0	0	0	-1	0	0	0	0	0	1
0	0	0	0	-1	0	0	0	1	0
0	0	0	0	0	-1	0	0	0	2
0	0	1	0	0	0	-2	0	0	0
0	0	0	0	1	0	-1	0	0	0
0	0	0	1	0	0	0	-1	0	0
0	0	0	0	0	1	0	-2	0	0
0	0	0	0	0	0	1	0	-1	0
0	0	0	0	0	0	0	1	0	-1

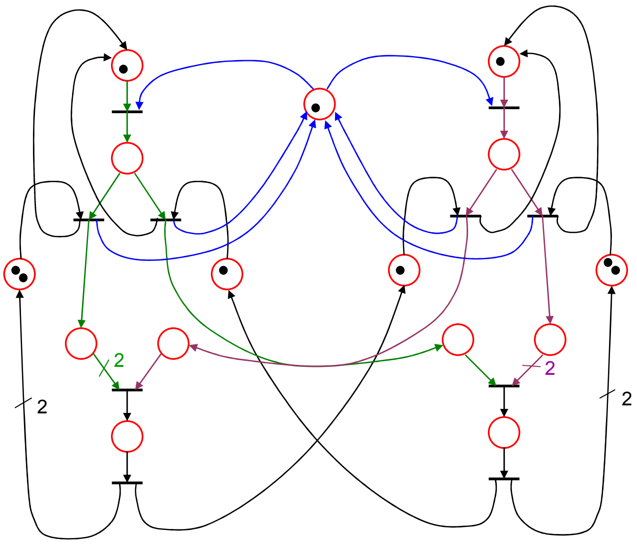


Figure 2: PN modeling system in Fig.1

From the graphic of the PN, or from the state equation, alternatively, we can study the reachability tree of the system, resulting in 147 states, simply using the method of successive firings.

Furthermore, calculating the marking invariant gives us the following seven solutions for the conservative components (table2):

Table 2: Conservative components (rows) of PN in Fig. 2

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	1	0	0	0	2	0
0	0	0	0	0	0	0	1	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	1	0	0	0	1	0	2

No column is zero, what implies that the network is conservative, and can be applied the proposed methodology for conservative PN. Therefore, the Diophantine equations are solved following the Contejean-Devie algorithm, and the 147 possible solutions are obtained, but with less computational time and effort (a summary of the states is presented in Table3, reduced to just 10 states for space cuestions).

Table 3: Summary of the 147 states of the system in Fig. 1 provided by the Contejean-Devie algorithm

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1
2	0	1	0	1	0	0	0	0	0	0	0	1	2	1	0
3	0	1	0	1	0	0	0	0	0	2	1	0	0	0	1
4	0	1	0	1	0	0	0	0	0	2	1	1	2	0	0
5	0	1	0	1	0	0	0	0	1	0	0	1	1	1	0
6	0	1	0	1	0	0	0	0	1	2	1	1	1	0	0
7	0	1	0	1	0	0	0	0	2	0	0	1	0	1	0
8	0	1	0	1	0	0	0	0	2	2	1	1	0	0	0
9	0	1	0	1	0	0	0	1	0	2	0	0	0	0	1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
147	1	1	1	0	0	2	1	1	2	0	0	0	0	0	0

6. CONCLUSIONS

The paper has presented a study on the different techniques to determine the states of a production system by analyzing the Petri net that models it. Two general methods have been discussed, one slower and depth that allows obtaining "spurious marked" and another one considerably faster but not allowing obtaining them.

The structure of Petri Nets may have different characteristics, so as to be classified into three types for those propouses: conservative to all places, nonconservative globally but for conservative for some places, and non-conservative to any place. The existence of conservation laws for all or some places considerably reduces the search for possible PN markings (ie, possible states of the system), thanks to new Diophantine equations systems that appear whose solutions are directly the potential markings. Accordingly, two methods have been described specifically for those Petri nets with these characteristics.

For the resolution of systems of Diophantine equations, the most common today algorithms have been examined, and the one which best fits the type of solutions expected in Petri Nets has been chosen.

ACKNOWLEDGMENTS

This paper has been partially supported by a grant of the University of La Rioja and Banco Santander.

REFERENCES

Ajili, F., Contejean, E., 1995. Complete solving of linear Diophantine equational and inequational systems without adding variables. *Unité de recherche INRIA* Lorraine, Metz.

Bradley, G. H. 1971. Algorithms for Hermite and Smith Normal Matrices and Linear Diophantine Equations. *Mathematics of computation* 25 (116).

Clausen, M., Fortenbacher, A., 1989. Efficient solution of linear Diophantine equations. *Journal of symbolic computation* 8, 201-216.

Contejean, E., Devie, H., 1994. An efficient incremental algorithm for solving systems of linear

Diophantine equations. *Information and computation* 113 (1), 143-172.

Girault, C., Valk, R., 2001. Petri Nets for systems engineering. A guide to modeling, verification and applications. Springer-Verlag, pp. 9-73.

Havas, G., Majewski, B. S., Matthews, K. R., 1998. Extended GCD and Hermite Normal form Algorithms via lattice basis reduction. *Experimental Mathematics* 7 (2), 125-136.

Hemmecke, R., 2011. Discrete optimization. Lecture notes SS 2011, TU Munich.

Jimenez, E., Perez, M., Latorre, J.I., 2006. Industrial applications of Petri nets: system modelling and simulation. *Proceedings of International Mediterranean Modelling Multiconference* 2006, pp. 159-164

Jimenez, E., Perez, M., Latorre, J.I., 2009. Modelling and simulation with discrete and continuous PN: semantics and delays. *Proceedings of 21st European Modeling and Simulation Symposium, Vol II*, pp. 14-19

Jimenez, E., Tejeda, A., Perez, M., Blanco, J., Martinez, E., 2012. Applicability of lean production with VSM to the Rioja wine sector. *International Journal Of Production Research*, 50 (7), 1890–1904

Latorre, J.I., Jimenez, E., Blanco, J., Sáenz-Díez, J.C., 2013a. Integrated methodology for efficient decision support in the Rioja wine production sector. *International Journal of Food Engineering*, (In press).

Latorre, J.I., Jimenez, E., Perez, M., 2009. Decision taking on the production strategy of a manufacturing facility. An integrated methodology. *Proceedings of 21st European Modeling and Simulation Symposium, Vol II*, pp. 1-7.

Latorre, J.I., Jimenez, E., Perez, M., 2013b. Simulation-based Optimisation of Discrete Event Systems by Distributed Computation. *Simulation-Transactions of the Society for Modeling and Simulation International*, (In press).

Latorre, J.I., Jimenez, E., Perez, M., 2013c. The optimization problem based on alternatives aggregation Petri nets as models for industrial discrete event systems. *Simulation-Transactions of the Society for Modeling and Simulation International*, 89 (3), 346–361.

Lazebnik, F., 1996. On systems of linear Diophantine equations. *Mathematics Magazine* 69 (4), 261-266.

Murata, T., 1989. Petri Nets: Properties, analysis and applications. Proceedings of the IEEE 77 (4), pp. 541-580.

Silva, M., 1993. Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall.

RUBIK'S CUBE AS A BENCHMARK FOR STRATEGIES OF SOLUTION SEARCH IN DISCRETE SYSTEMS PRESENTING STATE EXPLOSION PROBLEM. MODEL WITH ORDINARY AND COLORED PN

Emilio Jiménez-Macías [(a)], Francisco Javier Leiva-Lázaro[(b)],
Juan-Ignacio Latorre-Biel [(c)], Mercedes Pérez de la Parte[(d)]

[(a,b)] University of La Rioja. Industrial Engineering Technical School.
Department of Electrical Engineering. Logroño, Spain
[(c)] Public University of Navarre. Department of Mechanical Engineering, Energetics and Materials.
Campus of Tudela, Spain
[(d)] University of La Rioja. Industrial Engineering Technical School.
Department of Mechanical Engineering. Logroño, Spain

[(a)] emilio.jimenez@unirioja.es, [(b)] francisco-javier.leiva@unirioja.es,
[(c)] juanignacio.latorre@unavarra.es, [(d)] mercedes.perez@unirioja.es

ABSTRACT

This paper presents an analysis of Rubik's Cube and its methods of resolution, used to expose, in a simple and easily understandable to students way, the state explosion problem faced by discrete systems and the possibilities of dealing with the problem based on analysis, sihmulation or a combination of both. The goal is not to advance knowledge of the cube, which is used simply as a benchmark, but to show an analogy of how in discrete production systems is given that: a) you may not have a solution to evolve the system until the desired state (the desired output), b) or sometimes a solution is available, although not optimal, c) and the combination of analytical techniques and simulation often improves the solution, but still not be optimal d) and it may even known how to get the optimal solution, but it is impossible to put into practice due to the computational (or time) cost. Additionally, by modeling the system with a PN, all the developed analysis on the system is valid on the model, allowing thus advance knowledge of the PN model. The lines to develop various PN models of Rubik's cube with PN formalisms are also exposed.

Keywords: workstation design, work measurement, ergonomics, decision support system

1. INTRODUCCIÓN

One of the biggest problems faced by those who have to deal with discrete systems is the well-known state explosion problem (Ajay, 2012, Sturtevant et al., 2009). This problem occurs in many production and logistics systems whose behavior can be represented by a discrete model, and makes it very difficult to plan production to achieve the expected result, ie to bring the system to a specific state (eg, produce 20 cars of a certain model and color, 30 other ...). The use of analysis techniques allow dealing with these systems and obtained solutions to the problem, that is, allow knowing how to bring the system to the desired state (eg, how to produce exactly the desired cars, in number for any type and model, for that day). But often it is not possible, or at least it is not known, how to analytically treat these systems efficiently, and in those cases we are forced to use simulation.

But both in one case (analytical treatment) and the other one (through simulation) it is difficult to find an optimal solution of the system, ie, how to reach the desired state of the most favorable manner according to an established criteria (eg, how to produce the desired cars for the day with the fewest hours of work, or the lowest possible economic cost, or the lowest environmental impact) (Jiménez et al., 2012). Moreover, the problem of knowing the optimal solution could be solved by simulation, if infinite computational possibilities were available. Ie it is known how to find optimal solutions, but you can not get them because the computational requirements would be prohibitive with the available resources (temporary, economic, or technological). Often those cases, in which a solution is known, although it is not the optimal one, are found. Furthermore, the combination of simulation and analytical techniques can lead to improved solutions, often without being the optimal (or sometimes even being) at least it is better than the previous ones (Jimenez et al., 2006, 2009).

All of those behaviors, inherent to discrete systems, which suffer the state explosion problem, and therefore inherent to the production processes that can be modeled by discrete formalisms, can be seen in a very intuitive and graphical way, through a benchmark: the well-known Rubik cube. Thus, this system can be used as a basis for exposing students of engineering or mathematics such conduct by an easy to understand analogy.

This paper presents the analogy of solving Rubik cube with the problem to finding the solution of a production system, and discusses the types of solutions that can be found using different analytical and simulation techniques, as well as combinations between them, for different types of initial states which may exist.

2. BENCHMARK: CUBE RUBIK

2.1. The system and its states

In the original Rubik's Cube (3×3×3) we find eight vertices, with 3 possible orientations each, and twelve edges, with two possible orientations each (Rubik, 2013, 2013B), considering (without loss of generality) that the central pieces are fixed. Therefore there are 8! ways of place the vertices, with 3^8 posibilities for the orientations, and 12! ways to plave the edges, with 2^{12} orientations. That is, the system in total presents the number of $8!\cdot3^8\cdot12!\cdot2^{12}$ = 519.024.039.293.878.272.000 possibilities, or possible states of the system. But those are the states that can be obtained by decomposing the system and composing it again changing the position and orientation of the pieces (provided the central pieces are fixed). But they are not the possible states of the system that can be obtained from the initial satate by the movements of the cube (which is what we want to know). In fact, from the initial state we can obtain only 1/12 of those states, since, seven of the vertices can be oriented independently, but the eighth orientation depends on the previous seven (1/3), and also one of the orientation of the edge pieces depends on the orientation of the other eleven (1/2), and given 10 positions of the edges the other 2 edge pieces can be placed in a fixed distribution of the two rest of places (1/2). Considering the above, the real number of possible permutations is of our system is (1); ie standard 3x3x3 cube provides a quantity exceeding 43 trillion possible permutations.

$$\frac{8!\cdot3^8\cdot12!\cdot2^{12}}{12} = 43.252.003.274.489.856.000 \qquad (1)$$

2.2. Methods of resolution

The resolution algorithms used in this work are the following ones (Demaine et al., 2011; Korf, 1997):

2.2.1. Thistlethwaite method

This method was created by Morwen Thistlethwaite, a mathematician at the University of Tennessee, in 1981 (Jaapsch, 2013). The method is based on studying a problem as a group of subproblems, restricting each position in groups of positions that can be solved using a series of predetermined algorithms. Each of the sub-problems consist on fixe a certain position and see the movements that can be made free. Subgroups created are:

G0 = L, R, F, B, U, D
G1 = L, R, F, B, U2, D2
G2 = L, R, F2, B2, U2, D2
G3 = L2, R2, F2, B2, U2, D2
G4 = {I} → Cube solved

From the tables created for items in each group, he found a sequence of moves that led to another smaller group. A randomly chosen cube belongs to G0. From the G0 group element, subgroups belonging to G1will be obtained, and so on until the solved cube belonging to the group G4.

Originally this resolution algorithm solved the Rubik's Cube in 52 moves, but successive changes in the subgroups created permit solving it in fewer moves.

2.2.2. Kociemba algorithm.

Thistlethwaite algorithm was improved by Herbert Kociemba 1992, reducing the number of groups to only two (Kociemba, 2013).

G0 = L, R, F, B, U, D
G1 = L, R, F2, B2, U2, D2
G2 = {I}.

As in the Thistlethwaite method, the Kociemba method examines the elements between the groups G0 and G1, to find the optimal solution between groups. Using this method the cube can be solved with a maximum of 20 moves.

2.2.3. Layer by layer method.

This resolution system is the most used by initiates in the Cube, allowing cube solving very simple and sequential, needing no intuitive resolution system or memorizing complex solvers (Rubikaz, 2013). The disadvantage of this method of solution is its high number of moves needed to solve the cube as it generally exceeds 100 moves in most initial states.The resolution system consists of 7 stages:

S1. - Form a cross on the top face. This step creates a cross on the top, so that the colors match also in adjoining layers.

S2. - Place each of the vertices of the top face. At this stage of the placement of the upper vertices forming a T in each of the faces. Each vertex must be positioned so that each of its three colors match the colors of three adjacent faces.

S3. - Complete the central face. It is using the above process, but the central parts of the layer. These parts have only two colors, so that the resolving system is simpler than the previous one.

S4. - Form a cross on the underside, keeping the rest of unchanged faces. First, form a bar on the underside, and then continue with the cross.

S5. - Set the colors of the sides of the cross on the underside. At this stage the colors are placed at each end of the cross corresponding to each of the sides holding the cross on the underside. To do so first turn the top layer until having at least two side colors in their correct position.

S6. - Place each of the vertices of the lower face without orientation. In this step, we need to put the last 4 vertices in place, but no matter their orientation.

S7. - Align the corners of the underside. This step will guide each of the vertices that have been previously placed. Once the cube vertices are oriented, the cube is solved.

2.2.4. Optimal algorithm.

The optimal algorithm development was part of the research group of Herbert Kociemba, which managed to show that any cube can be solved in fewer than 20 moves (Kociemba, 2013).

In order to apply such a claim they had to make a thorough study of the various types of symmetries in the cube, which would reduce the number of possible states considerably.

They developed the symmetric states in the cube, since according to the arrangement of the colors, despite being in a different order, two states may be symmetrical, so that the resolution algorithm is the same for both.

The known symmetries to develop the optimal algorithm amounts to 164.604.041.664. For instance, the states that requie 20 movements are reduced from 1.091.994 to 32.625, because of the use of the simetries.

There are 48 kinds of possible symmetry elements (Table 1). All the states have at least a type of symmetry, except the initial state; however, other states present various types of symmetries. The combination of the 48 types of symmetries provide with different groups of symmetries to use.

Table 1: Types of symmetry

Type of simetry	Elements
1/2 rotation around an edge	6 elements
Reflection through a plane	6 elements
1/2 rotation around a face	3 elements
Reflection through a plane	3 elements
1/4 rotation around a face	2 x 3 elements
1/4 rotation + reflection through the center	2 x 3 elements
1/3 rotación alrededor de una arista	2 x 4 elements
Reflection through the center	1 elements
1/3 rotation + reflection through the center	2 x 4 elements
Identity (no movement)	1 elements

3. EXPERIMENTS ON THE SYSTEM

3.1. Methodology

In the study of this project, the analysis of 200 different states has been developed. Four tables have been filled with 50 different cases each, varying the number of random movements of the satates, from 1 to 50. Each state has been solved by the different methods that have been discussed in previous points.

3.2. Results

Table 2 shows the results obtained from a series of resolution with all methods with states achieved with from 1 to 50 random movements from the initial state. There is a column with the random movements that drive to the state from the initial state (so that the table is completely reproducible). The other columns presents the results of the simulation of each of the states for each resolution method, including the optimal algorithm, in which it has taken considerable time simulation for the 200 cases studied. The optimal algorithm simulation was performed using iterative deepening search by analyzing an average of 14.550.000.000 nodes at depth 18, and an average of 11.300 simulation seconds for each state (resulying in 2,91·1.012 nodes with a time of 628 simulation hours in total). The time for solving a state with the simplest methods is (depending on the state) just a few seconds.

Only 1 table is shown, for space questions, of the 4 with similar experiments developed to validate the methodology, for obvious space issues, but in the next section the figures with their results are shown.

4. ANALYSIS AND INTERPRETATION OF RESULTS

Table 2 allows reproducing the results, and presents the exact values of the simulations, but the most appropriate way to understand them is by the results showns in thw figures.

Figure 1 shows the data of the resolution for each of the methods. A first reflection is the great difference between advanced methods and method face to face. Obviously, having advanced algorithms gives a tremendous advantage in terms of results, but those advanced algoritms also present other disadvantages, specifically the time resolution, and the difficulty of implementation (face to face is so simple that a person can learn it in a few hours).

Figure 1 does not show the resolution of the face to face method starting from each of the colors, in order to simplify the drawing, and those 6 resolutions are shown in Figure 2. What Figure 1 additionally presents is, for every test, maximum value of 6 the 6 sides, the minimum value, and the one of the 6 obtained by the algorithm by default, which always select the same color (called Automatic method in the Figures).

It can be seen that choosing the right color with this method can reduce the movement cost very considerably (difference between red and green lines). In this case, even without advanced methods available, if you only have face-to-face method, the simulation of the 6 cases with little effort (6 times the effort of only one simulation) allows in average improving in a high percentage the solution.

Keep in mind that improved rarely solutions is proportional to stress (this is not an exception). For example, Figure 3 shows the advanced cases: advanced, the 2-phase, and the optimum. It also comes in red an upper bound to the minimum value of resolution, because that value will always be less than 20 (maximum demonstrated value movements required from any state), and less than or equal to the movements necessary to get the state from the initial state (as with the reverse sequence resolves sure).

Table 2: Sequence of 50 resolution tests with all methods, random states of 1 to 50 movements

n	Random movements	C1	C2	C3	C4	C5	C6	Aut.	Adv.	2Ph	Opt.
1	D	1	1	1	1	1	1	1	1	1	1
2	R2,B’	2	99	2	2	2	155	51	2	2	2
3	U2,R’,L	127	140	79	82	67	80	3	3	3	3
4	L2,B2,R,F’	123	106	148	116	107	105	146	4	4	4
5	U2,B’,L2,D’,R2	125	136	132	116	94	134	107	5	5	5
6	L2,B2,F’,R’,F’,B2	135	77	129	86	69	116	108	6	6	6
7	U,D,F2,D',R,F,L2	131	124	92	101	139	120	127	12	7	7
8	R',B2,L2,B2,R2,B',F2,U2	115	104	104	108	101	114	112	8	8	7
9	D2,L,R2,U2,D2,R,B',F2,L	154	87	104	124	110	161	111	18	9	9
10	B,D',B',L2,R',U2,L2,R,D,R'	87	117	98	112	100	105	137	22	10	10
11	L',F',L2,D2,F2,L,R',F,U2,B,L2	112	127	97	114	90	145	71	24	11	11
12	U,D2,F2,U2,L2,B,R,U,B,D',L2,U'	104	102	119	155	136	100	120	29	12	12
13	U2,L',R2,B2,U,L,R2,U',D,R2,B',F,L2	112	146	127	118	120	162	130	32	13	13
14	B',U2,B2,D2,L2,U',F',L,B,U',B2,U',D2,L	125	142	129	110	160	100	106	28	14	14
15	U,B',R2,B,L',F',L',R',B',F2,L2,R',F',D,L2	146	162	120	147	102	123	115	32	15	15
16	U2,L,R,U2,L2,B2,D,F2,L2,F2,U2,L',D,B2,L',R2	107	153	135	113	134	116	115	31	14	14
17	L2,R,U2,D,L,R',B2,U',L2,R2,F,L,D,B2,R2,B,D2	136	129	111	123	112	123	147	30	18	17
18	D,F2,R2,B,L2,R',U',L,R,B',D2,L',R,F,R',F,D',L2	106	89	148	115	116	92	96	30	19	18
19	L2,R,B',F2,L',R2,B2,R2,F2,L',R',B2,F2,U',L,R2,F,D,L	139	120	123	120	116	129	126	31	16	16
20	R2,B2,F2,U,R',B',L2,B',F2,L',R',B',F',L,B,F,R2,L2,U2,L2	119	110	129	147	133	145	114	31	17	17
21	L2,R',F',L',U2,D2,R2,B',U',D,R',B2,L2,F',R',F2,L2,R2,D2,R,L2	126	111	144	121	132	113	112	31	17	17
22	L,B2,L2,F2,D,L,R',B,U,F2,U',R,U2,L',B2,L2,R,U,B2,U,R',L2	118	131	138	147	124	147	109	35	17	17
23	D,R',F2,L,U,F2,L,R2,D,F,U2,F2,L2,F',L',R',U2,D,B',L,R2,F2,L'	110	122	114	139	116	119	121	33	19	18
24	L2,F2,U',L2,R',F,L2,R,B2,F',L2,B2,L,R',F',R2,B2,R2,B',F2,L',R',D2,R'	118	110	137	111	111	115	112	29	18	18
25	B2,U,D',R,D2,L',B,F2,D,L',B2,F,L2,R',B,L2,R2,B2,D,L',R2,F,L',U',R'	115	146	135	139	105	120	145	33	18	18
26	L2,U',F',R2,F,D,F',U',B2,D2,R2,D,B2,L,D',B',U2,L2,R2,D2,L2,F2,U,R,F',R2	134	126	163	120	153	122	105	34	19	18
27	L',D2,F,U',D2,B,L2,R2,F,L',F',D2,L,B',F',R,B2,U',B',D',L',R2,F2,L2,R',F2,R	130	144	115	109	133	154	118	32	19	18
28	R,F',L2,U2,L2,F2,R2,B2,D2,R',D,L2,U2,B2,L,F2,R,F,R2,D2,L,B2,U2,L2,U2,R',U,R	98	95	112	115	122	145	142	29	18	17
29	D,B,U,D2,R,U',L2,B,F2,U',D2,L,R',F,L,R2,F2,R2,D2,L2,U',L,D2,L2,R',B',R,B2,L2	134	107	113	115	132	154	146	29	17	17
30	L,R',B',F,U2,B2,F2,L,B2,D2,F',U2,R2,B,R2,F2,L2,D2,L',B,F2,U',D',L,D,L,R',F',D,R'	127	148	128	121	112	121	114	31	18	17
31	B',D',L2,B,L',U2,D',F,U',D',L2,D,F2,R2,U,R2,B2,D,R,U,R',B,R2,U',B2,L2,R2,D,L,U',L2	149	137	125	115	101	132	132	28	19	18
32	L,F',U2,R',U,L2,F',D',L',B2,R',B2,F2,L2,B',R',U2,B2,R',U2,L,U2,D,B',F2,R',D',R,U2,F2,R,U'	130	131	137	147	110	97	130	37	19	18
33	F2,L',B,R2,F,L',R',U',D',R2,F,U',D2,L,R',F2,L2,R',B2,F2,R',F2,L2,R2,B',U,L,R,B2,D,B2,R2,L2	112	135	137	113	136	158	112	24	17	17
34	L,U2,B',U',L2,R',B2,F,L',R2,B',F2,U2,D',L,B,L,B2,F',L,F,U2,D2,L2,D,L,R,F',L',F2,D2,R2,D2,F2	115	130	135	134	121	101	135	32	18	18
35	D',L2,F2,D',L2,R,U,F,L',R2,B2,F2,R2,B2,L2,R,D2,L2,R2,B2,U',L2,U,F2,L',B2,R,B2,F2,R2,B2,R,B',D2,L2	122	113	70	130	94	122	70	33	18	18
36	R,D2,R',U,B',L',F2,L2,R,B2,F2,R2,U',L2,R2,F,D',B2,L,R',D2,R',U,L2,B2,U',D2,B2,U2,R,F2,U,L,R',U2,L2	123	102	132	111	117	86	111	27	18	18
37	L,U2,D',R2,U2,R2,B,U2,B,L',U,L2,U',B2,L2,D',L',R2,U',L',B2,L,U2,L',U2,B2,R2,F,L2,U2,R2,D',R2,F2,L',F,L	102	101	91	134	97	98	91	32	19	18
38	R,B2,L',R,U2,D',L2,F,L,U',D2,B2,R,B',F2,R',D2,L2,R,B2,L2,R',B2,L2,U,B2,F2,U,D,R',F2,R2,B2,D2,R',B2,F',L'	149	87	106	140	130	121	140	32	19	18
39	U',B,L2,D',L2,D,F2,L,R,F',R2,D',L,U,R2,U,L',R,B',F',D2,L',R2,D2,L2,R2,F,R2,B',U2,L,B2,U',F',L,D,L,R',F2	112	124	109	120	113	147	124	30	19	18
40	B,U,D2,F2,U2,F,L2,B',U2,F2,R2,D2,F',U2,F',U2,D2,F,R2,U,L2,U2,R,U',F,R,D',L,U',D2,F2,L2,U',B2,L2,R2,U2,F2,U',L'	145	105	97	150	122	112	97	32	19	17
41	L,B',F2,U',B',D2,L,R,B',F,U,B2,L2,U,R',U',D,L,U',F2,L,R2,B2,U2,B',F,L2,R,F',U',L,B2,D,F2,L2,U,L,R2,F',R2,L	101	125	126	150	116	141	101	31	19	18
42	R,D,L',R,D,L2,U2,R',D2,F2,R,B2,D',R2,U2,F2,L',R',U',B,U2,D2,F,L',D',R,U2,L',R',D',L2,F',L2,U',D',L',R2,U,L',R',U',F2	109	129	108	134	124	129	134	24	19	16
43	D2,R2,U2,B2,F,R',U2,L',R,F',D,B2,F2,R',B,R',U2,B2,U,R2,U',L,U2,D',L,D,L,R,D,L',B2,R,U,D2,F,L',U,R2,F',L,R2,D2,F2	86	145	132	145	114	125	86	30	19	17
44	R',B2,D',B2,R2,D',R',U,L2,R',U,L,D2,F2,L2,R',B2,R',F,L,R,F2,R,B,F2,U,D,L',F,R2,B',R,B',F2,L',D,R,D',L2,R2,U2,L2,F,R2	143	130	131	133	141	90	130	30	19	18
45	R2,U',D,R',U',D2,L,F',L2,R,U',D2,L,F,L,B2,L',R2,U2,D',L',D',R',D',B',U',L2,U',L,D2,B,D2,B2,L2,B',L2,B2,U2,L2,R2,F2,D,F',L',R2	146	150	138	130	122	133	150	26	18	18
46	U',D',R',U',L',R,F',L2,B',D',F2,R',B2,U',B',D2,F,L,D2,L',U2,D',R2,F2,L',U2,L',R2,U2,D2,L2,R',F,L2,F2,D,R',B2,L2,F,L2,F,L2,B2,L2,R'	138	108	112	124	117	126	124	31	18	18
47	D',L2,B2,F2,D2,B,L2,R2,F2,D2,B2,R2,F2,R',B,L2,B2,F2,U,R',D,L',R,U2,L,R2,F2,D2,F',R',D',R',U',L2,R,F2,L,R,F2,L,U,F',L2,F',L2,R',U'	122	141	130	128	122	107	128	33	17	17
48	L2,R,B2,U2,L2,D2,L2,D,L2,D2,R2,F',L2,U2,B,U',D2,L2,U',R,U,D',R2,F',L2,F,L2,U2,L2,D,L',F2,L',R',F2,U2,L2,R',B,R2,F2,R,D',L2,F2,U,D,R2	108	117	125	107	136	104	104	25	19	18
49	R2,D2,B,F2,L2,D2,R',U,D2,L2,F,U,B',L',F2,D2,L,B,R',B2,F',L2,B2,R2,B,L2,R',U2,L2,D2,L,U,B2,F',D2,R2,U2,B',L2,B2,L2,R2,B2,F2,L,F,R',L2,U2	133	116	121	122	121	122	122	29	19	18
50	F2,D2,L,B2,L2,R2,F2,U2,L',U',B',R2,U2,L,B',U,D,L',U,R',B',R2,B',R',B2,F',U',L',R2,D2,B,R',B2,L,R,D,L2,R2,D2,L2,R2,U',D2,L,F2,L',F2,L2,R',F'	140	144	121	101	139	129	101	33	19	18

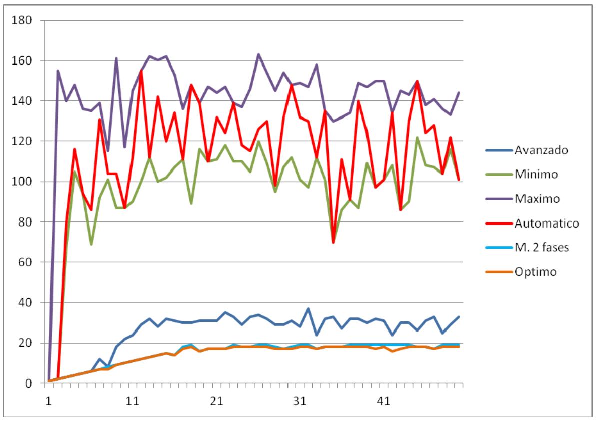


Figure 1: Resolution by all methods

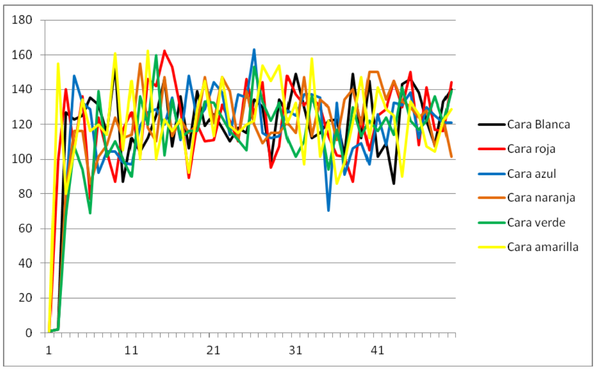


Figure 2: Resolution face to face method, implemented by the 6 possible faces

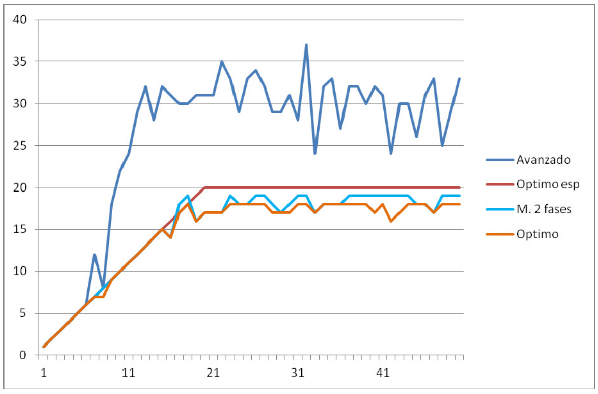


Figure 3: Resolution by the 3 advanced methods: Advanced, 2-Phases, and Optimal

We see the very advanced method substantially improves the result compared to face-to-face method, even in the best face possible. Besides its computational cost is only slightly higher (takes a few seconds on a personal computer environment), but its implementation is more difficult.

Keep in mind that solutions improvement rarely is proportional to effort (this is not an exception). For example, Figure 3 shows the advanced cases: advanced, the 2-phase, and the optimum. It also presents in red an upper bound to the minimum value of resolution,

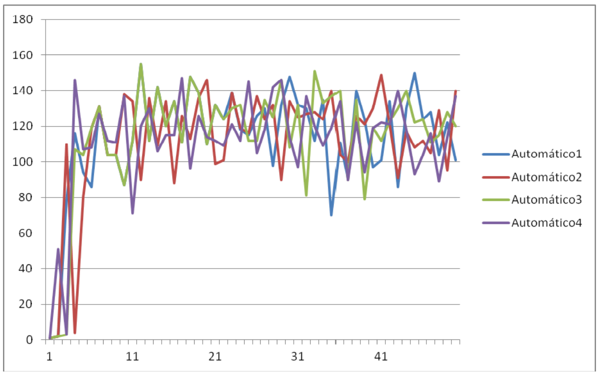


Figure 4a: Resolution by the method face to face in the 4 different sets of tests

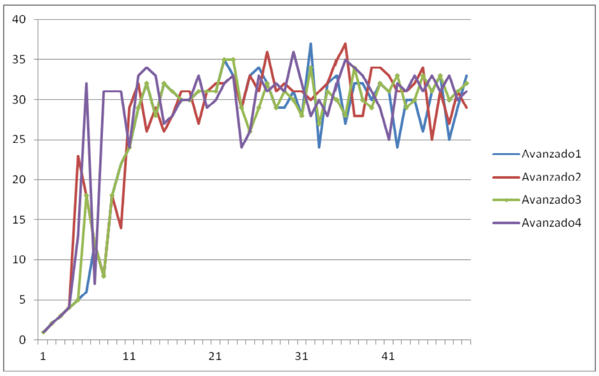


Figure 4b: Resolution by the advanced method in the 4 different sets of tests

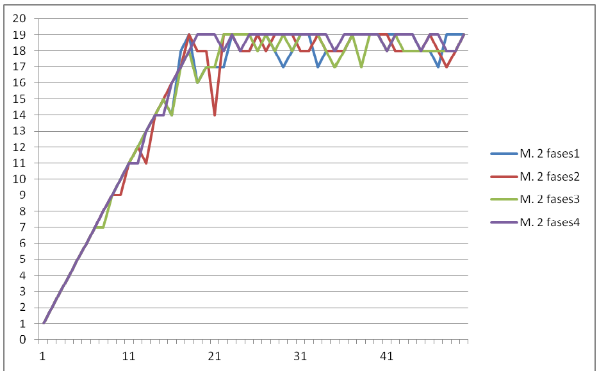


Figure 4c: Resolution method 2 phases in the 4 different sets of tests

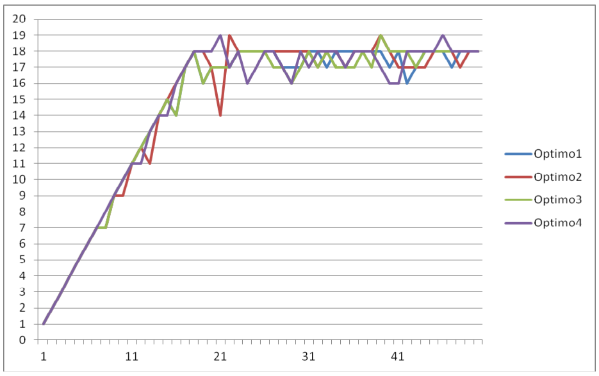


Figure 4d: Optimal Resolution method in the 4 different sets of tests

because that value will always be less than 20 (demonstrated maximum movements required from any state), and less than or equal to the movements necessary to get the state from the initial state (as the reverse sequence is a solution). The very advanced method substantially improves the result compared to face-to-face method, even in the best face possible. Besides, its computational cost is only slightly higher (takes a few seconds on a personal computer environment), but its implementation is more difficult.

The 2-phase method still further improves (less than half, in average) but computational cost is also much higher, and also the difficulty of implementation. And the optimal method improves just a little earlier, but still requires much more time to resolution (now in the order of hours).

Figures 4a to 4c show in a single graph the methods used for each of the four tests, to check that those results constitute a trend and not casuality).

5. PN APPROACH

After analyzing the system, the next step is to study the feasibility of building a model with PN. In addition, depending on the formalism and the approaches, different models could be developed (Silva 1993). The fundamental requirement is that any model represents exactly what we need to know about the system, and in this case, this is the states. Thus, all developed models will have the advantage that we know its behavious since we know a lot of information about the system (such as the number of reachable states or the ways to obtain any state). Furthermore, the analysis of models may provide information of the system; for example, any repetitive component determined from the incidence matrix of the model provides information on how to solve the system since each of the intermediate nodes (Latorre et al., 2009, 2103, 2013B).

In this section we will present different approaches to model the Rubik's Cube by PN. Such models are not developed in the paper, because each of them requires more space than an article to study it in detail, and only the Figures would also require more than one paper. So, the aim is simply to present the simplest possibilities, and the relationships between them; with this information any PN expert can build the models without problems.

5.1. Approach of labels with Colored PN

It is conceptually the simplest approach. The color lavels of the faces are considered as tokens, the places where can be each label are the PN places, and as transitions each of the 12 possible moves (spin in both directions of each of the 6 sides). This PN have these characteristics:

Tokens: 6 colors, and 8 tokens each color
Places: 48 positions
Transitions: 12 moves

In this model, the constraints derived from relative positions of the colors (such that no two of the same color on both sides of an edge) derive from the internal structure of the model and the initial marking. If anyone analises this PN model without knowing that it represents a Rubik cube, it would not be easy initially appreciate that there is a close relationship between labels for the various blocks that make up the cube.

Also noteworthy is that this model would represent the same system (would be equivalent) if were to be used 48 pieces each color, as the only possible solution in the cube presents always each label in the same position.

5.2. Approach of Physical blocks with Coloured PN

Marks are considered as tokens, and as PN places the places where the parts can be. Therefore there exist 20 places and 20 tokens. But only the corner pieces can be in the corners, and the pieces of the sides can only be in the side. Therefore, the model will have two unconnected sub-networks: une with 8 places/tokens and the other one with 12 places/tokens. But additionally must have other 20 subnetworks to indicate the orientations of each of the parts (2 possible orientations of the side pieces, and the 3 possible positions of the corners pieces). With all this, the system will have (all with 12 transitions, corresponding to the movements):

- Subnet 1
  Tokens: 8 tokens, of 1 color each
  Places: 8 positions
- Subnet 2
  Tokens: 12 tokens, of 1 color each
  Places: 12 positions
- Subnets 3 (8 subnets):
  Tokens: 1
  Places 3 (corresponding to the possible orientations)
- Subnets 4 (12 subnets):
  Tokens: 1
  Places: 2 (corresponding to possible orientations)

5.3. Approaches with ordinary PN

Any of the previous models can be made in colorless PN (generalized PN, which also are ordinary since no weights on the arcs exist in these models). It is well known that the ability of modeling the colored PN is exactly the same as ordinary PN, although obviously condensation capacity is much higher.

The way to convert the two previous models is as simple as unfolding them, ie divide each PN in as many networks as colors. For example, colored PN made in 5.1 would be now 6 ordinary PN, each with 48 places, 12 transitions and 6 tokens, or even 48 PN with 48 places, 12 transitions and one token. The PN indicated in section 5.2 could be done by 8 PN with 8 places, 12 PN with 12 places, 8 PN with 3 places, and 12 PN with 2 places, all of them with 12 transitions and one token.

6. CONCLUSIONS

The analysis developed from the tests shows an analogy between the well known system Rubik cube and real discrete production systems suffering from the state explosion problem:

- The optimum is almost always impossible to achieve, due to computational costs for its calculation and for lack of methods and algorithms to achieve it (apart from "brute force").

- With equal computational resources, research (analysis, ie, more efficient algorithms) dramatically reduces the results

- With equal computational resources and algorithms, the use of simulation (different options using the same algorithm) can lead to improvements in many cases.

- The Rubik cube benchmark is analogous to discrete production systems, because of the combinatorial explosion or state explosion inherent to discrete systems, and can be used for learning (teaching) or for deeper understanding (research).

That analogy can be enriched by modeling the system using a formalism eminently useful for modeling discrete systems with state explosion, such as Petri nets. Thus, following construction of the models (in one of the many different possibilities offered by the paradigm of Petri nets as the great family of formalisms that is), a lot of properties of these models are known (as much as are known of the system: the Rubic cube).

ACKNOWLEDGMENTS

This paper has been partially supported by e grant of the University of La Rioja and Banco Santander.

REFERENCES

Ajay, K., 2012. Search Techniques To Contain Combinatorial Explosion in Artificial Intelligence, *International Journal of Engineering Research & Technology* (IJERT), Sep 2012.

Demaine, E.D., Demaine, M.L., Eisenstat, S., Lubiw, A., Winslow, A., 2011. Algorithms for Solving Rubik's Cubes, *Cornell University*, Jun 2011.

Jaapsch http://www.jaapsch.net/puzzles/thistle.htm

Jimenez, E., Perez, M., Latorre, J.I., 2006. Industrial applications of Petri nets: system modelling and simulation. *Proceedings of International Mediterranean Modelling Multiconference* 2006, pp. 159-164

Jimenez, E., Perez, M., Latorre, J.I., 2009. Modelling and simulation with discrete and continuous PN: semantics and delays. *Proceedings of 21st European Modeling and Simulation Symposium, Vol II*, pp. 14-19

Jimenez, E., Tejeda, A., Perez, M., Blanco, J., Martinez, E., 2012. Applicability of lean production with VSM to the Rioja wine sector. *International Journal of Production Research*, 50 (7), 1890–1904

Kociemba, 2013. http://kociemba.org/cube.htm

Korf, R.E., 1997. Finding optimal solutions to Rubik's Cube using pattern databases, *Proc. Nat. Conf. on Artificial Intelligence (AAAI-97)*, Providence, Rhode Island, Jul 1997, pages 700–705.

Latorre, J.I., Jimenez, E., Blanco, J., Sáenz-Díez, J.C., 2013. Integrated methodology for efficient decision support in the Rioja wine production sector. *International Journal of Food Engineering*, (In press).

Latorre, J.I., Jimenez, E., Perez, M., 2009. Decision taking on the production strategy of a manufacturing facility. An integrated methodology. *Proceedings of 21st European Modeling and Simulation Symposium, Vol II*, pp. 1-7.

Latorre, J.I., Jimenez, E., Perez, M., 2013. Simulation-based Optimisation of Discrete Event Systems by Distributed Computation. *Simulation-Transactions of the Society for Modeling and Simulation International*, (In press).

Latorre, J.I., Jimenez, E., Perez, M., 2013. The optimization problem based on alternatives aggregation Petri nets as models for industrial discrete event systems. *Simulation-Transactions of the Society for Modeling and Simulation International*, 89 (3), 346–361.

Rubik,2013. http://en.wikipedia.org/wiki/Rubik's_Cube

Rubik, 2013b. http://www.rubiks.com/

Rubikaz, 2013. http://www.rubikaz.com/resolucion.php

Silva, M., 1993. Introducing Petri nets. In *Practice of Petri Nets in Manufacturing*, Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall.

Sturtevant, N., Felner, A., Barrer, M., Schaeffer, J., Burch, N., 2009. Memory-Based Heuristics for Explicit State Spaces, *IJCAI'09 Proceedings of the 21st international jont conference on Artifical intelligence*, Pages 609-614.

PETRI NET REDUCTION RULES THROUGH INCIDENCE MATRIX OPERATIONS

Joselito Medina-Marin[a], Juan Carlos Seck-Tuoh-Mora[b], Norberto Hernandez-Romero[c], Jose Carlos Quezada-Quezada [d]

[a,b,c]Autonomous University of Hidalgo State, Advanced Research Centre in Industrial Engineering, Pachuca de Soto, Hidalgo, México
[d]Autonomous University of Hidalgo State, Superior School of Tizayuca, Tizayuca, Hidalgo, México

[a]jmedina@uaeh.edu.mx, [b]jseck@uaeh.edu.mx, [c]nhromero@uaeh.edu.mx, [d]jcarlos@uaeh.edu.mx

ABSTRACT

A Petri net (PN) is a powerful tool that has been used to model and analyze discrete event systems. Such systems can be concurrent, asynchronous, distributed, parallel, non-deterministic, and/or stochastic. A problem in PN modelling is related to its graphical representation because it increases for each element of the system. Consequently, incidence matrix of the PN also increases the number of rows and/or columns. To verify properties in PN such as liveness, safeness, and boundedness, computer time is required, even more if we need to verify huge Petri nets. There are six simple reduction rules, which are used to produce a smaller PN preserving the properties of the original PN. In order to apply these reduction rules, we have to find the pattern and then apply the corresponding rule. In this paper, we propose to apply the reduction rules directly in the incidence matrix of the PN modelled, detecting the pattern of each rule on the incidence matrix and applying the corresponding changes on the incidence matrix.

Keywords: Petri nets, reduction method, incidence matrix

1. INTRODUCTION

Petri net (PN) is a powerful tool that gives support to theoretical and practitioners to develop models representing Discrete Event Systems (DES). It has been widely used in several fields to model and analyse flexible manufacturing systems (FMS) and information processing systems trough the application of analysis methods of PN theory, such as the coverability tree, the incidence matrix and state equation, and the reduction rules (Murata 1989).

There are several works where analysis methods of PN are applied in the study of DES.

In (Henry, Layer, and Zaret 2010) a framework that incorporates an application of the coverability analysis is presented. The coverability analysis was coupled with process failure mode analysis in order to quantify the risk induced by potential cyber attacks against network-supported operations.

The work presented in (Cabastino, Giua, and Seatzu 2006) uses the coverability graph to determine a PN system from the knowledge of its coverability graph. The authors faced the following problem: given an automaton that represents the coverability graph of a PN, determine a PN system whose coverability graph is isomorph to the automaton.

In (Latorre-Biel, and Jimenez-Macias 2011), four incidence matrix-based operations are applied to perform transformations in PN models in order to validate and verify them as models of discrete event systems. Properties of the initial PN model are preserved with these matrix operations.

A process to convert A and B contacts in the ladder diagrams into a PN model is described in (Lee, and Lee 2000). The authors construct the incidence matrix for each contact in order to obtain their corresponding state equation and perform their analysis.

In (Verbeek, et. al. 2010) the reduction method is applied to PNs with reset and inhibitor arcs. This PN extension is used to model cancellation and blocking. In (Xi-zuo, Gui-ying, and Sun-ho 2006) the reduction method of PN is applied to verify the correctness of workflow models.

Reduction rules and deadlock detection methods are proposed in (Lu, and Zhang 2010). This proposal is based on Object Oriented Petri net models and the authors take advantage of object oriented concepts to develop their methods. In (Uzam 2004) the PN reduction approach is used to set a policy for deadlock prevention in FMS.

We can notice the importance of the use of analysis methods in PN theory and applications. Nevertheless, the structures of the PN models obtained from DES have several places and transitions, which produces huge coverability trees and very big matrices. Hence, the importance to apply reduction rules in order to have smaller petri net models and perform faster analysis to DES. Therefore, the use of reduction rules and its application on the incidence matrix of the PN, instead of its graphical representation, is proposed in this work.

The remainder of the paper is organized as follows. Section 2 gives fundamental concepts of PNs and introduces the incidence matrix and reduction rules method. Section 3 describes the proposal of using incidence matrix operations to apply reduction rules. Section 4 presents two illustrative examples. Finally, section 6 shows conclusions of the work and further research.

2. PETRI NET FUNDAMENTALS

A PN is a graphical and mathematical tool that has been used to model concurrent, asynchronous, distributed, parallel, non-deterministic, and/or stochastic systems.

The graph of a PN is directed, with weights in their arcs, and bipartite, whose nodes are of two types: *places* and *transitions*. Graphically, places are depicted as circles and transition as boxes or bars. PN arcs connect places to transitions or transition to places; it is not permissible to connect nodes of the same type. The state of the system is denoted in PN by the use of *tokens*, which are assigned to place nodes.

A formal definition of a PN is presented in table 1 (Murata 1989).

Table 1: Formal definition of a PN

A Petri net is a 5-tuple, $PN = \{P, T, F, W, M_0)$ where: $P = \{p_1, p_2, \ldots, p_m\}$ is a finite set of places, $T = \{t_1, t_2, \ldots, t_n\}$ is a finite set of transitions, $F \subseteq \{P \times T\} \cup \{T \times P\}$ is a set of arcs, $W = F \rightarrow \{1, 2, 3, \ldots\}$ is a weight function, $M_0 = P \rightarrow \{0, 1, 2, 3, \ldots\}$ is the initial marking, $P \cap T = \varnothing$ and $P \cup T \neq \varnothing$.

The token movement through the PN represents the dynamical behaviour of the system. In order to change the token position, the following transition firing rule is used (Murata 1989):

1. A transition $t \in T$ is enabled if every input place $p \in P$ of t has *w(p,t)* tokens or more. *w(p,t)* is the weight of the arc from *p* to *t*.
2. An enabled transition *t* will fire if the event represented by *t* takes place.
3. When an enabled transition *t* fires, *w(p,t)* tokens are removed from every input place *p* of *t* and *w(t,p)* tokens are added to every output place *p* of *t*. *w(t,p)* is the weight of the arc from *t* to *p*.

2.1. Analysis methods

PN theory considers three groups of analysis methods: a) the coverability tree method, b) the matrix-equation approach, and 3) the reduction method. For the intention of this paper, the matrix equation approach and reduction methods are presented.

2.1.1. Incidence matrix and state equation

A *PN* with *n* transitions and *m* places can be expressed mathematically as a $n \times m$ matrix of integers $A = [a_{ij}]$. The values for each element of the matrix are given by: $a_{ij} = a_{ij}^{+} - a_{ij}^{-}$, where a_{ij}^{+} is the weight of the arc from t_i to p_j, and a_{ij}^{-} is the weight of the arc from p_j to t_i.

The state equation is used to determine the marking of a *PN* after a transition firing, and it can be written as follows:

$$M_k = M_{k-1} + A^T u_k,\ k = 1,2,\ldots \qquad (1)$$

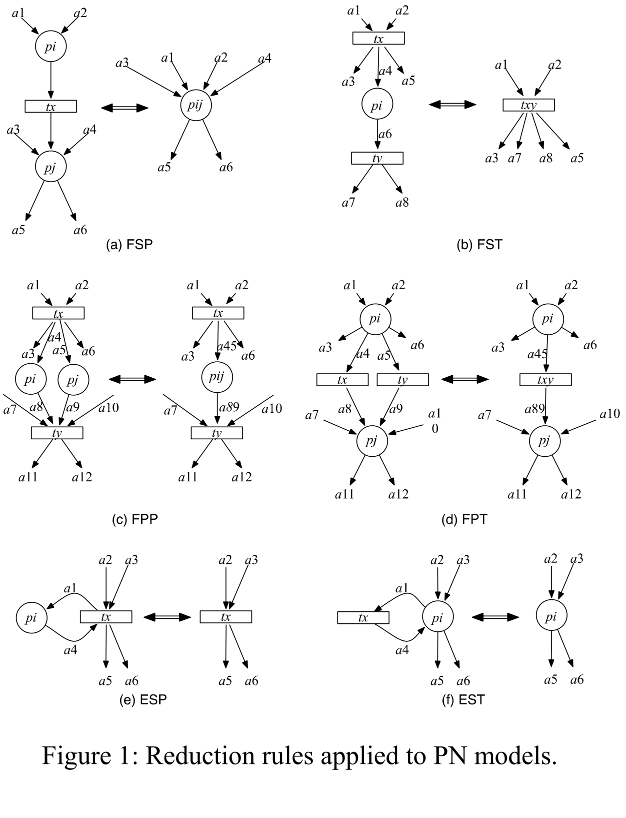
where u_k is a $n \times 1$ column vector of n - 1 zeros and one nonzero entries, which represents the transition t_j that will fire. The nonzero entry is located in the position j of u_k. A^T is the transpose of incidence matrix. M_{k-1} is the marking before the firing of t_j. And Mk is the reached marking after the firing of t_j denoted in u_k.

2.1.2. Reduction rules

In order to work with smaller PN models and analyse them in an easier way, six simple reduction rules have been proposed (Murata 1989; Zhou and Venkatesh 1999). These rules guaranty the preservation of system properties in the system modelled, such properties are safeness, liveness and boundedness.

1. Fusion of Series Places (FSP).
2. Fusion of Series Transitions (FST).
3. Fusion of Parallels Places (FPP).
4. Fusion of Parallels Transitions (FPT).
5. Elimination of Self-loop Places (ESP)
6. Elimination of Self-loop Transitions (EST).

Figure 1 shows the transformations in the PN through the application of reduction rules.



(a) FSP

(b) FST

(c) FPP

(d) FPT

(e) ESP

(f) EST

Figure 1: Reduction rules applied to PN models.

3. REDUCTION RULES ON INCIDENCE MATRIX

3.1. Fusion of Series Places (FSP) rule

In the FSP rule, the transition t_x located between the places p_i and p_j is deleted and both places are merged. Then, the result is a unique place p_{ij} with the sum of input and output arcs of places p_i and p_j less the arcs connected to t_x. (Figure 1a).

The incidence matrix for PN of figure 1a is the following.

$$A = \begin{array}{c|ccccccc} & p_1 & \cdots & p_i & \cdots & p_j & \cdots & p_m \\ \hline t_1 & \cdots & \cdots & i_1 & \cdots & j_1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ t_x & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ t_n & \cdots & \cdots & i_n & \cdots & j_n & \cdots & \cdots \end{array} \qquad (2)$$

On the incidence matrix we have to do the next steps:

1. Delete the x row from the incidence matrix.
   $A_1 = A[1 \ldots x-1, 1 \ldots m]$
   $A_2 = A[x+1 \ldots n, 1 \ldots m]$
   $A_r = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$
2. Sum columns i and j of A_r,
   $A_3 = A_r[1 \ldots n-1, i]$
   $A_4 = A_r[1 \ldots n-1, j]$
   $A_{34} = A_3 + A_4$
3. Replace column in position i by the column vector resulting A_{34}, and remove the column in position j from A_r.
   $A_5 = A_r[1 \ldots n-1, 1 \ldots i-1]$
   $A_6 = A_r[1 \ldots n-1, i+1 \ldots j-1]$
   $A_7 = A_r[1 \ldots n-1, j+1 \ldots m]$
   $A_{fsp} = A[A_5 \quad A_{34} \quad A_6 \quad A_7]$

The column vector A_{34} can be placed either on position i or j. In this case, column in position i is replaced by the column vector resulting A_{34}. At the end, the dimension of incidence matrix A_{fsp} is n-1 × m-1.

3.2. Fusion of Series Transitions (FST) rule

Now, place p_i will be deleted and transitions t_x and t_y will be merged (figure 1b). As result of this rule, we get a single transition t_{xy} whose input and output arcs are the join of input and output arcs of t_x and t_y less those arcs connecting from t_x to p_i and from p_i to t_y.

The incidence matrix of PN depicted in figure 1b is the next.

$$A = \begin{array}{c|ccccc} & p_1 & \cdots & p_i & \cdots & p_m \\ \hline t_1 & \cdots & \cdots & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 & \cdots & \cdots \\ t_x & x_1 & \cdots & 1 & \cdots & x_m \\ \cdots & \cdots & \cdots & 0 & \cdots & \cdots \\ t_y & y_1 & \cdots & -1 & \cdots & y_m \\ \cdots & \cdots & \cdots & 0 & \cdots & \cdots \\ t_n & \cdots & \cdots & 0 & \cdots & \cdots \end{array} \qquad (3)$$

For the application of the FST rule on the incidence matrix, we have to perform the following steps.

1. Delete the i–th column from the incidence matrix.
   $A_1 = A[1 \ldots n, 1 \ldots i-1]$
   $A_2 = A[1 \ldots n, i+1 \ldots m]$
   $A_r = [A_1 \quad A_2]$
2. Sum rows x and y of A_r,
   $A_3 = A_r[x, 1 \ldots m-1]$
   $A_4 = A_r[y, 1 \ldots m-1]$
   $A_{34} = A_3 + A_4$
3. Replace row in position x by the row vector resulting A_{34}, and remove row in position y from A_r.
   $A_5 = A_r[1 \ldots x-1, 1 \ldots m-1]$
   $A_6 = A_r[x+1 \ldots y-1, 1 \ldots m-1]$
   $A_7 = A_r[y+1 \ldots n, 1 \ldots m-1]$
   $A_{fst} = A \begin{bmatrix} A_5 \\ A_{43} \\ A_6 \\ A_7 \end{bmatrix}$

A_{fst} is a n-1 × m-1 matrix, after the elimination of p_i and the fusion of t_x and t_y.

3.3. Fusion of Parallel Places (FPP) rule

The aim of this rule is to fuse places with the same input transition, same output transition, and only with one input arc and one output arc.

In order to apply FPP rule on the incidence matrix, parallel places involved are deleted except one of them. The incidence matrix related to PN of figure 1c is the following.

$$A = \begin{array}{c|ccccccc} & p_1 & \cdots & p_i & \cdots & p_j & \cdots & p_m \\ \hline t_1 & \cdots & \cdots & 0 & \cdots & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 & \cdots & 0 & \cdots & \cdots \\ t_x & x_1 & \cdots & 1 & \cdots & 1 & \cdots & x_m \\ \cdots & \cdots & \cdots & 0 & \cdots & 0 & \cdots & \cdots \\ t_y & y_1 & \cdots & -1 & \cdots & -1 & \cdots & y_m \\ \cdots & \cdots & \cdots & 0 & \cdots & 0 & \cdots & \cdots \\ t_n & \cdots & \cdots & 0 & \cdots & 0 & \cdots & \cdots \end{array} \qquad (4)$$

In this case, there is only one operation to perform on the incidence matrix.

1. Delete the j–th column from the incidence matrix.
   $A_1 = A[1 \ldots n, 1 \ldots j-1]$
   $A_2 = A[1 \ldots n, j+1 \ldots m]$
   $A_{fpp} = [A_1 \quad A_2]$

The fusion of places p_i and p_j is denoted on the incidence matrix with the elimination of either p_i or p_j. Both places have the same arcs, then $A[1 \ldots n, i] = A[1 \ldots n, j]$. The dimension of A_{fpp} is n rows by m-1 columns.

3.4. Fusion of Parallel Transitions (FPT) rule

For this rule, the rows of merged transitions are deleted except one of them. The transitions that will be fused have the same input place and output place, so the rows in incidence matrix corresponding to these transitions are similar.

The incidence matrix representing the PN with parallel transitions (figure 1d) is the following.

$$A = \begin{array}{cc} & \begin{array}{ccccccc} p_1 & \cdots & p_i & \cdots & p_j & \cdots & p_m \end{array} \\ \begin{array}{c} t_1 \\ \cdots \\ t_x \\ \cdots \\ t_y \\ \cdots \\ t_n \end{array} & \begin{bmatrix} \ldots & \ldots & i_1 & \ldots & j_1 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & i_n & \ldots & j_n & \ldots & \ldots \end{bmatrix} \end{array} \tag{5}$$

To apply the FPT rule on incidence matrix, either t_x or t_y must be deleted, and the other one must be kept.

1. Delete the y–th row from the incidence matrix.
   $A_1 = A[1 \ldots y-1, 1 \ldots m]$
   $A_2 = A[1 \ldots y+1, 1 \ldots m]$
   $A_{fpt} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$

A_{fpt} is a matrix of n-1 rows by m columns.

3.5. Elimination of Self-loop Places (ESP) rule

Self-loop places can be seen in the PN graph, moreover on the incidence matrix this kind of places have only zero entries in their corresponding column. However, isolated places also present only zero entries on the incidence matrix, but in the intention of reduction method isolated places can also be deleted.

Incidence matrix for PN depicted in figure 1e is defined as follows.

$$A = \begin{array}{cc} & \begin{array}{ccccc} p_1 & \cdots & p_i & \cdots & p_m \end{array} \\ \begin{array}{c} t_1 \\ \cdots \\ t_x \\ \cdots \\ t_n \end{array} & \begin{bmatrix} \ldots & \ldots & 0 & \ldots & \ldots \\ \ldots & \ldots & 0 & \ldots & \ldots \\ x_1 & \ldots & 0 & \ldots & x_m \\ \ldots & \ldots & 0 & \ldots & \ldots \\ \ldots & \ldots & 0 & \ldots & \ldots \end{bmatrix} \end{array} \tag{6}$$

The elimination of the self-loop place on the incidence matrix is done through the following step:

1. Delete the i–th column from the incidence matrix.
   $A_1 = A[1 \ldots n, 1 \ldots i-1]$
   $A_2 = A[1 \ldots n, i+1 \ldots m]$
   $A_{esp} = [A_1 \quad A_2]$

After the elimination of i-th column, the A_{esp} matrix has n rows by m-1 columns.

3.6. Elimination of Self-loop Transitions (EST) rule

EST rule indicates that transitions with only an input arc and an output arc to the same place have to be deleted. On the incidence matrix, rows with zero entries in all values denote a self-loop transition or even an isolated one. In both cases the row must be removed from the matrix.

The PN depicted in figure 1f has the following incidence matrix.

$$A = \begin{array}{cc} & \begin{array}{ccccc} p_1 & \cdots & p_i & \cdots & p_m \end{array} \\ \begin{array}{c} t_1 \\ \cdots \\ t_x \\ \cdots \\ t_n \end{array} & \begin{bmatrix} \ldots & \ldots & i_1 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & i_n & \ldots & \ldots \end{bmatrix} \end{array} \tag{7}$$

t_x row is a zero row vector because p_i is an input an output place of t_x, i.e. $A(t_x,p_i) = 1^+ - 1^- = 0$.

The elimination of t_x on the incidence matrix can be done with the following instruction.

1. Delete the x–th row from the incidence matrix.
   $A_1 = A[1 \ldots x-1, 1 \ldots m]$
   $A_2 = A[x+1, \ldots n, 1 \ldots m]$
   $A_{est} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$

Incidence matrix A_{est} has one row less that A, and the dimension of the matrix now is n-1 rows by m columns.

4. ILLUSTRATIVE EXAMPLES

In order to show the applicability of reduction rules on incidence matrix of PN, two examples taken from the literature are presented.

4.1. Example 1

The PN used in this example was presented in (Murata 1989) and it is shown in figure 2.

The incidence matrix is the following.

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Places p_3 and p_4 are serial places. To reduce the PN firstly we apply the steps described for FSP rule in section 3.1.

1. Delete the 4th row from the incidence matrix.
2. Sum column vectors 3 and 4.
3. Replace the 3rd column with the values obtained in the sum, and remove the 4th column.

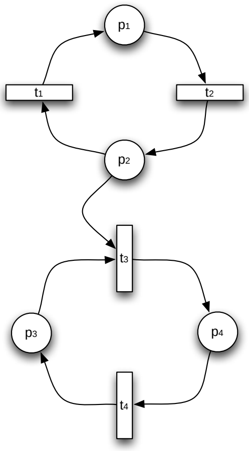


Figure 2: PN model for example 1.

After these operations, the resulting incidence matrix is the following.

$$A_{fsp} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

Now, the rule that can be applied is the FST, because transitions *t_1* and *t_2* are series transitions.

1. Delete the 1st column from the incidence matrix (place p_1).
2. Sum row vectors 1 and 2.
3. Replace the first row by the result of the sum and delete row 2.

The incidence matrix after these operations is the following.

$$A_{fst} = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$$

In this phase, the ESP rule can now be applied to this incidence matrix, because the place of second column of A_{fst} is a self-loop place. The instruction for ESP rule is applied.

1. Delete the second column from the incidence matrix.

The incidence matrix is the following.

$$A_{esp} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Finally, EST rule can be applied on incidence matrix A_{esp} because the transition of first row represents a self-loop transition.

1. Delete the first row from the incidence matrix A_{esp}.

Applying the EST instruction the last incidence matrix is as follows.

$$A_{esp} = [-1]$$

Figure 3 shows the evolution in the PN when the reduction rules are applied.

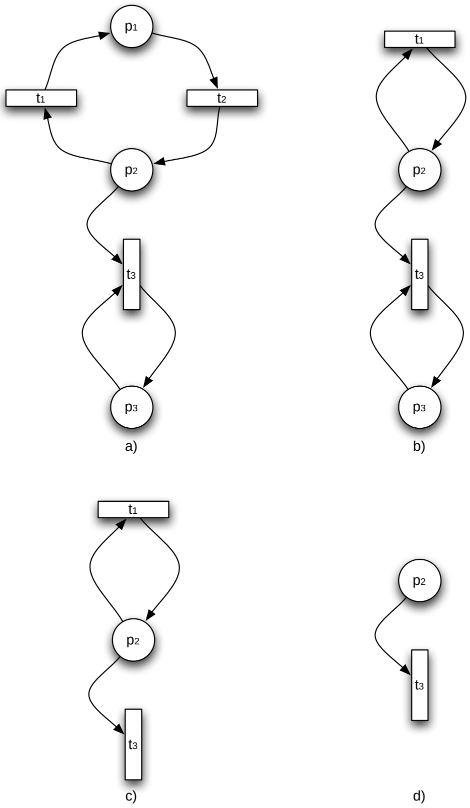


Figure 3. PN changes after the application of some reduction rules.

4.2. Example 2

The second example is a PN model that was reduced applying the reduction method by (Zhou and Venkatesh 1999). Figure 4 shows the initial PN model and its incidence matrix is the following.

$$A = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{array} \begin{array}{c} \begin{array}{cccccccc} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 \end{array} \\ \begin{bmatrix} -1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

FSP rule is applied to fuse places p_2 and p_3, where transition t_2 is between these places.

1. Delete the *2nd* row from the incidence matrix (t_2).

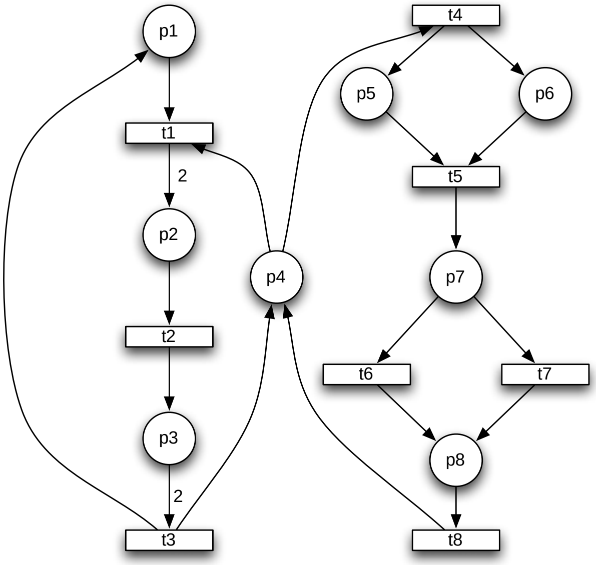


Figure 4. PN model used in example 2.

$$
A_{fsp} = \begin{array}{c} \\ t_1 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{array}
\begin{array}{c} \begin{array}{cccccccc} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 \end{array} \\
\begin{bmatrix}
-1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\
1 & 0 & -2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix} \end{array}
$$

2. Sum columns *2* and *3* of A_{esp},

$$
\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

3. Replace the column in position *2* by the result of the sum, and remove the column in position *3* from A_{esp}.

$$
A_{fsp} = \begin{array}{c} \\ t_1 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{array}
\begin{array}{c} \begin{array}{ccccccc} p_1 & p_2 & p_4 & p_5 & p_6 & p_7 & p_8 \end{array} \\
\begin{bmatrix}
-1 & 2 & -1 & 0 & 0 & 0 & 0 \\
1 & -2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & -1 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix} \end{array}
$$

Next, rule FPP is applied to parallel places p_5 and p_6.

1. Delete the column corresponding to p_6 from the incidence matrix.

$$
A_{fpp} = \begin{array}{c} \\ t_1 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{array}
\begin{array}{c} \begin{array}{cccccc} p_1 & p_2 & p_4 & p_5 & p_7 & p_8 \end{array} \\
\begin{bmatrix}
-1 & 2 & -1 & 0 & 0 & 0 \\
1 & -2 & 1 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & 0 & 0 & -1
\end{bmatrix} \end{array}
$$

Parallel transitions t_6 and t_7 are reduced with the rule FPT.

1. Delete the t_7 row from the incidence matrix.

$$
A_{fpt} = \begin{array}{c} \\ t_1 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_8 \end{array}
\begin{array}{c} \begin{array}{cccccc} p_1 & p_2 & p_4 & p_5 & p_7 & p_8 \end{array} \\
\begin{bmatrix}
-1 & 2 & -1 & 0 & 0 & 0 \\
1 & -2 & 1 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & 0 & 0 & -1
\end{bmatrix} \end{array}
$$

Serial transitions t_1 and t_3 are fused through FST rule. Place p_2 is the output and input place from t_1 and to t_3, respectively.

1. Delete the p_2 column from the incidence matrix.

$$
A_{fst} = \begin{array}{c} \\ t_1 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_8 \end{array}
\begin{array}{c} \begin{array}{ccccc} p_1 & p_4 & p_5 & p_7 & p_8 \end{array} \\
\begin{bmatrix}
-1 & -1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & -1 & 1 \\
0 & 1 & 0 & 0 & -1
\end{bmatrix} \end{array}
$$

2. Sum rows t_1 and t_3 from A_{fst},

$$
\begin{array}{l}
[-1 \quad -1 \quad 0 \quad 0 \quad 0] + \\
[\ 1 \quad\ \ 1 \quad 0 \quad 0 \quad 0] = \\
[\ 0 \quad\ \ 0 \quad 0 \quad 0 \quad 0]
\end{array}
$$

3. Replace the row in t_1 position by the result of the sum, and remove the row of t_3 from A_{fst}.

$$
A_{fst} = \begin{array}{c} \\ t_1 \\ t_4 \\ t_5 \\ t_6 \\ t_8 \end{array}
\begin{array}{c} \begin{array}{ccccc} p_1 & p_4 & p_5 & p_7 & p_8 \end{array} \\
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & -1 & 1 \\
0 & 1 & 0 & 0 & -1
\end{bmatrix} \end{array}
$$

Next, FST rule is used to fuse serial transitions t_4 and t_5. Place p_5 is between t_4 and t_5.

1. Delete the p_5 column from the incidence matrix.

$$A_{fst'} = \begin{array}{c|cccc} & p_1 & p_4 & p_7 & p_8 \\ t_1 & 0 & 0 & 0 & 0 \\ t_4 & 0 & -1 & 0 & 0 \\ t_5 & 0 & 0 & 1 & 0 \\ t_6 & 0 & 0 & -1 & 1 \\ t_8 & 0 & 1 & 0 & -1 \end{array}$$

2. Sum rows t_4 and t_5 from $A_{fst'}$,

$$\begin{aligned} &[0 \quad -1 \quad 0 \quad 0] + \\ &[0 \quad\ \ 0 \quad 1 \quad 0] = \\ &[0 \quad -1 \quad 1 \quad 0] \end{aligned}$$

3. Replace the row in t_4 position by the result of the sum, and remove the row of t_5 from $A_{fst'}$.

$$A_{fst'} = \begin{array}{c|cccc} & p_1 & p_4 & p_7 & p_8 \\ t_1 & 0 & 0 & 0 & 0 \\ t_4 & 0 & -1 & 1 & 0 \\ t_6 & 0 & 0 & -1 & 1 \\ t_8 & 0 & 1 & 0 & -1 \end{array}$$

Then, serial transitions t_6 and t_8 are fused through the application of FST rule. Place p_8 is the output place from t_6 and the input place to t_8.

1. Delete the p_8 column from the incidence matrix.

$$A_{fst''} = \begin{array}{c|ccc} & p_1 & p_4 & p_7 \\ t_1 & 0 & 0 & 0 \\ t_4 & 0 & -1 & 1 \\ t_6 & 0 & 0 & -1 \\ t_8 & 0 & 1 & 0 \end{array}$$

2. Sum rows t_6 and t_8 from $A_{fst''}$,

$$\begin{aligned} &[0 \quad 0 \quad -1] + \\ &[0 \quad 1 \quad\ \ 0] = \\ &[0 \quad 1 \quad -1] \end{aligned}$$

3. Replace the row in the position of t_6 by the result of the sum, and remove the row of t_8 from $A_{fst''}$.

$$A_{fst''} = \begin{array}{c|ccc} & p_1 & p_4 & p_7 \\ t_1 & 0 & 0 & 0 \\ t_4 & 0 & -1 & 1 \\ t_6 & 0 & 1 & -1 \end{array}$$

Serial transitions t_4 and t_6 are merged with the FST rule, where p_7 is the output place from t_4 and the input place to t_6.

1. Delete the p_7 column from the incidence matrix.

$$A_{fst'''} = \begin{array}{c|cc} & p_1 & p_4 \\ t_1 & 0 & 0 \\ t_4 & 0 & -1 \\ t_6 & 0 & 1 \end{array}$$

2. Sum rows t_4 and t_6 from $A_{fst'''}$,

$$\begin{aligned} &[0 \quad -1] + \\ &[0 \quad\ \ 1] = \\ &[0 \quad\ \ 0] \end{aligned}$$

3. Replace the row in the position of t_4 by the result of the sum, and remove the row of t_6 from $A_{fst'''}$.

$$A_{fst'''} = \begin{array}{c|cc} & p_1 & p_4 \\ t_1 & 0 & 0 \\ t_4 & 0 & 0 \end{array}$$

Finally, place p1 is removed because it is a self-loop place.

1. Delete the column of p_1 from the incidence matrix $A_{fst'''}$.

$$A_{esp} = \begin{array}{c|c} & p_4 \\ t_1 & 0 \\ t_4 & 0 \end{array}$$

A_{esp} is the incidence matrix of the PN model obtained in (Zhou and Venkatesh 1999). Figure 5 shows the reduced PN after the application of the reduction rules.

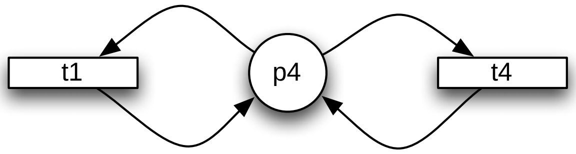


Figure 5. Reduced PN after the application of reduction rules.

5. CONCLUSIONS AND FUTURE WORK

The reduction method in PN is used to generate smaller PNs that preserve structural properties from the initial model.

Other analysis method in PN is the state equation, which uses the concept of incidence matrix. The incidence matrix is the mathematical representation of PNs, and denotes the relationship between places and transitions of the PN.

Reduction rules have been applied on the PN graphical representation; however, this work shows that reduction method can be applied on the incidence matrix, with the same results. This result is important because now the reduction method, based on matrix operations, can be inserted in a computational algorithm.

Two examples were developed to show the feasibility of matrix operations on PN reduction method.

As further research, it is possible to develop an algorithm taking into account the matrix operations proposed, and reduce very big PNs into smaller ones.

REFERENCES

Cabastino, M.P., Giua, A., and Seatzu, C., 2006. Identification of unbounded Petri nets from their coverability graph. *Proceedings of the 45th IEEE Conference on Decision & Control*, pp. 434–440. December 13-15, San Diego, CA, USA.

Henry, M.H, Layer, R.M., and Zaret, D.R., 2010. Coupled Petri nets for computer network risk analysis. *International Journal of Critical Infrastructure Protection*, 3(2), 67 – 75.

Latorre-Biel, J., and Jiménez-Macías, E., 2011. Matrix-based operations and equivalent classes in alternative Petri nets. *Proceedings of the 23rd European Modeling & Simulation Symposium*, pp. 587 – 592. September 12-14, Rome, Italy.

Lee, G.B., and Lee, J.S., 2000. The state equation of Petri Net for LD program. Systems, Man, and Cybernetics, 2000 IEEE International Conference on, pp. 3051 – 3056 vol.4. October 8-11, Nashville, TN, USA.

Murata, T., 1989. Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE*, 77(4), 541 – 580.

Uzam, M., 2004. The use of the Petri net reduction approach for an optimal deadlock prevention policy for flexible manufacturing systems. *The International Journal of Advanced Manufacturing Technology*, 23(3-4), 204 – 219.

Verbeek, H.M.W, Wynn, M.T., van der Aalst, W.M.P., and ter Hofstede, A.H.M., 2010. Reduction rules for reset/inhibitor nets. *Journal of Computer and System Sciences*, 76(2), 125 – 143.

Xi-zuo, L., Gui-ying, H., and Sun-ho, K., 2006. Applying Petri-Net-Based Reduction Approach for Verifying the Correctness of Workflow Models. *Wuhan University Journal of Natural Sciences*, 11(1), 203 – 210.

Zhou, M.C., and Venkatesh, K., 1999. *Modeling, Simulation, and Control of Flexible Manufacturing Systems*. New York: World Scientific.

AUTHORS BIOGRAPHY

Joselito Medina-Marin. He received the M.S. and Ph.D. degrees in electrical engineering from the Research and Advanced Studies Centre of the National Polytechnic Institute at Mexico, in 2002 and 2005, respectively. Currently, he is a Professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include Petri net theory and its applications, active databases, simulation, and programming languages.

Juan Carlos Seck-Tuoh-Mora. He received the M.S. and Ph.D. degrees in electrical engineering (option: Computing) from the Research and Advanced Studies Centre of the National Polytechnic Institute at Mexico, in 1999 and 2002, respectively. Currently, he is a Professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include cellular automata theory and its applications, evolutionary computing and simulation.

Norberto Hernandez-Romero. He received the M.S. degree from Department of Electrical Engineering, Laguna Technological Institute at México, in 2001 and the Ph. D. from Autonomous University of Hidalgo State at México in 2009. Currently, he is a professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include system identification, feedback control design, genetic algorithms, fuzzy logic, neural network and its applications.

Jose Carlos Quezada-Quezada. He received the M. S. degree in Mechatronics Engineering from Tecnológico de Estudios Superiores de Ecatepec, Estado de México, Mexico, in 2008, and the Bachelor degree in Electronic Engineering at the Technological Institute of Lázaro Cárdenas, Michoacán, Mexico, in 1992. He has worked in Fertilizantes Mexicanos S.A. de C.V., CFE, SICARTSA and Fertinal. Currently, he is a Research Professor of the Superior School of Tizayuca, Hidalgo, México. His main research interest is on process automation by means of PLC, PAC and HMI.

STATE SPACE ANALYSIS FOR MODEL PLAUSABILITY VALIDATION IN MULTI AGENT SYSTEM SIMULATION OF URBAN POLICIES

Miquel Angel Piera [(a)], Roman Buil [b)], Egils Ginters [(c)]

[(a)(b)] Universitat Autònoma de Barcelona, Department of Telecommunications and Systems Engineering, Unit of Logistics and Aeronautics, 08193, Bellaterra, Barcelona
[(c)] Sociotechnical Systems Engineering Institute, Vidzeme University of Applied Sciences, Latvia

[(a)]miquelangel.piera@uab.cat, [(b)]roman.buil@uab.cat, [(c)]egils.ginters@va.lv (please use "hyperlink" style)

ABSTRACT

Multi-agent models have been increasingly applied to the simulation of complex phenomena in different areas, providing successfully and credible results, however, model validation is still an open problem. The complexity of the stochastic interaction between agents together with a large numbers of parameters can make validation procedures intractable.

Particular validation difficulties appear in social science using multi-agent models, when agents are defined as spatial objects to computationally represent the behavior of individuals in order to study emergent patterns arising from micro-level interactions.

This paper considers some of the difficulties in establishing verification and validation of agent based models, and proposes the use of colored petri net formalism to specify agent behavior in order to check if the model looks logical and the model behaves logical. Model plausibility is used to express the conformity of the model with a priori knowledge about the process. A proof-of-concept is presented by means of a case study for testing the robustness of emergent patterns through sensitivity analyses and can be used for model calibration.

The proposed methodology has been applied in the European Future Policy Modeling project (www.fupol.eu) to create trust and increase the credibility of the agent based models developed to foster e-participation in the design of urban policies by means of simulation techniques.

Keywords: model validation, state space, petri net, agents

1. INTRODUCTION

Urban policies can be understood as the different mechanisms arranged to reach the goals set by the urban authorities in which political, management, financial, and administrative aspects are involved. Active participation in the design of urban policies can be seen as the involvement either by an individual or a group of individuals in their own governance or other activities, with the purpose of exerting influence.

Unfortunately, despite a city should be driven by their inhabitants, considering the complexity in the interrelationship between the different urban policy domains, citizens use to participate scarcely in the decision making process, since citizen's knowledge has been often considered to be ''opinion'' or ''belief'' (ie. influenced by subjective elements) and thereby dismissed during the planning of urban policies, relying mostly on ''hard'' technical knowledge and professional expertise (Albeveiro et all 2004).

Digital simulation techniques could contribute to increase the citizen's participation in the urban policy design by transforming their opinion's in valuable knowledge by means of evaluating certain decisions in digital urban scenario which would allow to understand the impact of different choices not only in the urban problem under study, but also on different hidden indicators (i.e. dynamics not considered by a citizen due to its limited cognitive horizons). Furthermore, the try-and-error experimental approach inherent to simulation models could act as an Enabler of open deliberations between citizens to foster mutual learning process of the complex urban dynamics

To foster citizens' involvement in the design of urban policies by means of scenarios simulation, there are 2 key factors that a new urban modeling approach should consider in order to place citizens at the center of the decision making process:

1. Model Transparency: The use of overly ambitious computer simulation models has been up to now very limited and restricted to certain types of planning, e.g. delivering regularly forecasts for transportation capacity planning, in which scientific knowledge has been used to legitimize political arguments. Models based on complex mathematical computations are oriented to specialist and require trained users.
   The use of a causal formalism to specify well accepted social urban behavior together with a visualization tool to understand the cause-effect relationships described as rules in the simulation model, would contribute to embed local knowledge into urban planning processes

in such a way that a more comprehensive knowledge base for decision-making can be attained.

2. Scenario Acceptability: A formal and easy to understand scenario specification in which citizens will experiment their thoughts is a critical aspects to create trust and increase the credibility of the model and the results delivered during the simulation. Usually, scenarios are specified by means of hypothesis which constraints the numerical results to certain regions of bounded values. Citizen's should easily understand the scenario components and to perform the required changes to deal with new boundary conditions.

It must be considered that most urban policy simulation models do not take into account the citizen participation in democratic decision-making processes, instead there exist fears, expectations and prejudices among the practitioners against the models: quantitative models are monsters and do not capture the social aspects in an adequate way. A lack of credibility in the models is reported due to non-scientific actors being not aware of the uncertainty inherent in such models.

To create trust and increase the credibility of the model and the simulation results delivered, it is essential to deal with a validation approach in which non-simulation-trained end-users (i.e. practitioners) could feel comfortable and trust the simulation model.

By means of a Multi Agent System simulation platform, it has been developed in the FP7 project FUPOL, a library of causal models to allow citizens testing the benefits and shortages of different proposed urban policies and check new policies according to their own beliefs.

Model validation of urban policy models using Multi Agent Simulation is a complex task that plays an important role for model acceptability. Consider for example those models, which assign high levels of cognition to their agents, so that agents can act intuitively rather than rationally. In these contexts, model validation becomes a challenge.

In this paper it is introduced in section II present shortages in the validation of Agent Based Models, while in section III it is proposed a model plausibility approach to improve the acceptability of the model. Section IV provides a short background on the use of colored petri net modeling formalism which is used to open the state space of the system to better understand the model causality and evaluate the reachability of some system states. Finally, section V illustrates the proposed validation approach by means of a real case study in Zagreb.

2. ABM VALIDATION SHORTAGES

Assessment in modeling and simulation confidence is considered a critical issue. One of the primary methods for building and quantifying this confidence are Verification and validation (V&V) of computational simulations. Verification can bee seen as the assessment of the accuracy of the solution to a computational model by comparison with known solutions, while validation can be seen as the assessment of the accuracy of a computational simulation by comparison with experimental data.

According to (Ormerod and Rosewell 2009) in social science, no firm conclusions have been reached on the appropriate way to verify or validate Agent Based Models, due to several aspects, such as agent capacity to take decisions autonomously.

In the field of Multi-agent Systems (MAS) an Agent can be seen as actor/entity with sensing capabilities (representing the means of collecting information), decision making capabilities (representing the means of transforming the sensed information/inputs into a meaningful action), and actuation capabilities (representing the means by which the agent can execute the selected action).

For software engineers the agent concept is related to the concept of objects in Object Oriented Programming (OOP), since agents has properties, functions, and possesses the ability to be abstracted and to integrate with other objects, and so forth. However there is an important difference for model validation is that agents are "autonomous" while objects are "obedient" (Grimshaw 2001).

One of the main problems of ABMs verification process is its sensitivity to replicate statistical results in a multirun approach:

- Numerical Identity: the original and replicated model should produce exactly the same results. This is the case in discrete event system simulation and also in continuous process simulation, however, behavioral rules in many ABMs typically contain stochastic elements designed to provide a wider agents behavior scope.
- Distributional equivalence: the properties of the original and replicated model should be statistically indistinguishable from each other. Note that when scaling the amount of agents, using the same distribution rate, the results obtained usually differs since one of the properties of the agents is their learning capacity which depends somehow on the amount of interactions (ie. Critical mass).
- Relational alignment: if input variable x is increased in both models by a given amount, the distribution observed in the changes in output variable y should be statistically indistinguishable. In ABM, the different affinities that can be generated randomly between the agents, do not guarantee relational alignment.

The process of validation requires a clear view of what the model is attempting to explain and for what purpose. Even when the modeler has a clear mind of what are the key facts that the model needs to explain, it is a hard task to clearly formalize how ABM can do it. An aspect of problem description in ABMs which is much more critical than with other modeling strategies is the time representation. ABMs are typically solved in steps, but there is no clear equivalent in real time with the step in any given ABM. The difficulty of translating steps in a model into real time sometimes is seen as a weakness of ABMs, however if provides an important strength: mapping a step into a real time equivalent can be a useful part of the model calibration process.

3. ABM VALIDATION IN URBAN POLICY DESIGN

Generally speaking, the quality of an urban policy model can be judged with respect to several features. One of the most widely used features in some modeling domains is related to direct comparison of input–output data from the model and from the real system. Unfortunately, in urban policy design simulation domain, the main problem of this approach is that real system data are, by definition in complex systems, numerous, and their units and types (e.g. qualitative or quantitative) are different. In this context, the proximity determination of a simulation from a reality can be difficult, particularly because the quality of their information may be very imperfect (i.e. uncertain, imprecise).

Urban policy complexity inevitably leads to a difficult access to parameter values, e.g. it seems rather difficult to observe and know certain policy acceptability by each citizen at a given time. It is well known that, citizens' affinities and priorities are influenced by several aspects that range from urban context changes (i.e. worsening economy), citizens interactions (i.e. opinion leaders, lobbies, etc.), or media information. Real parameters can only be observed occasionally and usually at so called ''simple'' or ''obvious'' moments. Within the framework of the agent based simulation validation, the lack of citizens status knowledge with respect to a certain urban policy, makes difficult to use classical comparison of input-output data in which agents status could be compared with citizens status.

A different feature to validate a model which better fits the urban policy modeling domain using ABM is Plausibility, also referred to as "conceptual validity" or "face validity", which expresses the conformity of the model with a priori knowledge about the process.

Assessment of model plausibility is tightly related to expert judgment of whether the model is good or not. The level of plausibility, or better said the expert opinion about it, is basically related to two features of the model:

- The first one considers the question whether the model "looks logical". This question concerns characteristics of the model structure (type of equations/rules, connections between equations/rules, etc.) and its parameters, and is relevant when the model is derived from first principles or well accepted hypothesis. If the structure and the parameters are feasible, which means comparable to what experts know about the real process, then the confidence into the model is greater.
- The second one is related to the question whether the model "behaves logically". This part concerns assessment of the reaction of the model outputs (dynamics, shape, etc.) to typical events (scenarios) on the inputs. If the model in different situation reacts in accordance with expectations of the experts, then again our confidence about its validity is increased.

3.1. Model Structure: Simple rules.

Simplicity of behavior is an important criterion that can contribute drastically to answer whether a model looks logical. If simple agent rules can produce a good description, this is better than having complicated ones. By ensuring that agents are only required to have the minimum necessary ability to process information or to learn, ABM model plausibility is facilitated. Indeed one way of testing an ABM in the social sciences is to assign increasing levels of cognition to agents to see at what point the model ceases to provide a description of reality. Thus, it is important to design agents with low, or even zero, cognition capacity and minimize the amount of agents with a high level of cognition to those which need special justification rather than those which do not.

In Section IV it is illustrated how to obtain simple rules to describe agents behavior by formalizing the system dynamics using colored petri net formalism.

3.2. Model Outputs: Expected Behavior

The different states that could be reached in an ABM describes the scope of the agent dynamics. Thus, it is proposed to compute the state space of the ABM without considering particular time constraints (time events), neither particular stochastic factor constraints. The full state space of the system can be computed providing all the rules specified in the different agents, considering different sequence rule combinations together with the evolution of the state variables defined in each agent.

State space analysis tools allows the evaluation of the different states that can be reached, and in case a feasible final state is never reached (i.e. for example the swings zone in a green park is never used), it is possible to check why the agents rules defined has not been executed, and add or modify the rules in order to achieve an acceptable representation of the system.

4. COLORED PETRI NET FORMALISM: AGENT RULES

The causal models developed for FUPOL are based in a set of rules generated using the information obtained from the cities. These rules are defined in such a way that models are more understandable for people without modelling background, as could be citizens without modelling skills. A proper visualization will also contribute to achieve a level of transparency good enough to allow a better understanding of the models for any kind of user.

The specification of a system dynamic by a set of rules would lead to a poor modelling approach lacking of the most essential modelling analysis tools that would lead to unpredictable simulation results. Colored Petri Net formalism (Jensen and Kristensen 2009) allows the specification rule based system dynamics as a formal language in which it will be possible to determine if the rules are consistent with the observed system dynamics, which dynamics has been properly formulated, which system states can be reached using the rules and check which rules should be added to reach certain final system states.

Rules can be seen as a relationship between precedent conditions and a consequent body. This form of rules can be interpreted in CPN formalism as a set of pre-conditions, which must be satisfied in order to fire an event, and a set of post-conditions, which represent the new state of the system reached after firing the event. Each rule can be formulated as a transition, in which the pre-conditions will be formulated by means of input arc expressions of the place nodes connected to the transition, and the post-conditions will be computed by means of output arc expressions connected at the output place nodes.

This one-to-one representation between rules and CPN transitions is a positive feature to improve simulation transparency.

The coloured Petri net formalism (CPN) has been widely used by the simulation community for different purposes. It has characteristics that allow modelling true concurrency, parallelism or conflicting situations present in dynamic systems. The formalism allows not only developing dynamic discrete-event oriented models without ambiguity and in a formal way but also it allows modelling the information flow, which is an important characteristic and very useful in systems modelling and decision making.

Traditionally in CPN, the place nodes are used to model resource availability or logic conditions that need to be satisfied. The transition nodes can be associated to activities of the social system or social actors.

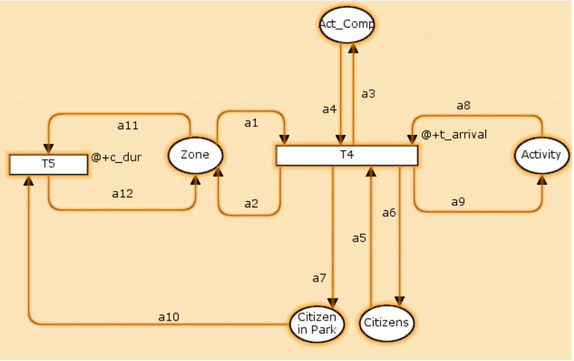


Figure 1: Citizens Activities in a Green Park

In the case of social systems these transition nodes could represent activities or decisions to be taken by each agent. Thus, the information attached to each transition is used as a base to define the agent behaviour rules. These rules will depend on the tokens attributes, and they will define the activities to be performed, the time frame and the companion when performing the activity. For example, Figure 1 shows the events related to the use of a specific zone in a green park.

There are 5 different entities (agents) that take part when evaluating this driving force:

- The activity to be performed (Place node **Activity**)
- Citizens profile: potential users of the park with preferences for certain activities and time scheduling (Place node **Citizens** and **CitizensinPark**)
- A description of the zones in the green park where different activities can be performed, considering also the incompatibilities between simultaneous activities (Place node **Act_Comp**).
- A description of the surface and the capacity of the different zones (Place node **Zone**)

As a consequence of this driving force, when a citizen agent try to perform an certain activity in the green park under study according to the citizen time scheduling availability, their preferences, if the zone occupancy is above a certain threshold value or present activities in the zone are incompatible with the one the agent would like to perform, a conflict is generated.

Well-used (and well-maintained) city parks are likely to be perceived as safe places to visit, sit on the grass, etc., but this may not be true for emptier or poorly maintained spaces, or where there is no surrounding land use that provides informal policing of the area. A CPN based agent model would contribute to a better understanding of the conflicts that could be generated in a green park and design policies for their mitigation.

4.1. ABM State Space Analysis

One of the most powerful quantitative analysis tools of PN and CPN is the coverability tree (Mujica et all 2010). The goal of the coverability tree is to find all the markings, which can be reached from a certain initial system state, representing a new system state in each tree node and representing a transition firing in each arc. The coverability tree allows:

- All the urban policy ABM states (markings) that can be reached starting from certain initial system conditions M0.
- The transition sequence to be fired to drive the system from a certain initial state to a desired end state.

Figure 2 illustrates the first two levels of a coverability tree, the state vector of the CPN model with 8 Places is represented. In each position of the vector, the tokens and its colours stored in each place node are represented. Given this initial marking, the only enabled events are those represented by transition T1 and transition T2. It should be noted that transition T2 could be fired using three different combinations of tokens (i.e. different entities). Once a transition has been fired, a new state vector is generated. Thus, a proper implementation of a CPN model in a simulation environment should allow automatic analysis of the whole search space of the system by firing the different sequences of events without requiring any change in the simulation model.

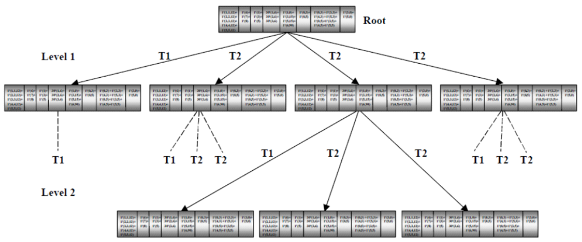


Figure 2: First Two Levels of the Reachability Tree

The coverability tree of the green park CPN model has been analysed for a reduced amount of entities, obtaining as main results:

- The sequence of actions obeys to the logic reaction of the model outputs to the different events activated by the context. Thus, for example, when all the zones are above their capacity the new citizens arrivals are not allowed to perform no one of their preferred activities.
- The state change behaves as it was reported by the fieldwork. The different real context specifications (i.e. early morning no kind in the play a round zones, only people walking with a dog and walking through the park pathway) have been obtained in the coverability tree.

This validation provides the confidence that the ABM model will be complete. Thus, by the proper tuning of the parameters, it should be possible to obtain a good prediction of the reality (i.e. validation for model falseness).

5. CASE STUDY: GREEN PARK DESIGN

The simulation objective for a green park design in Zagreb is to provide the best solution for the facilities that would be included in the 30000 m2 of green area situated near the Autism Centre. The design must satisfy most of the potential users demands and must encourage interactions between autism people and non-autism users, while avoiding possible conflicts between them. At the same time, possible conflicts between all kinds of users must be avoided. For example, nobody likes to have dogs around kids while these are playing in a playground. In Table 1 some modelled activities are summarized.

Table 1: Some Green Park Activities

ID	Activity	ID	Activity
A1	Walking through the park pathway	A21	Walking with a dog
A2	Sitting on a bench	A22	Touching objects and surfaces
A5	Playing in the Sandbox	A24	Playing with a dog
A7	Sitting and playing on the grass	A25	Playing bocce ball
A8	Swinging	A27	Playing in the labyrinth
A9	Sliding down the toboggan (slide)	A29	Standing with a parm
A10	Spinning on roundabouts	A30	Watching water movement and/or listening the sound of the water
A11	Climbing on monkey bars	A31	Following the paths with motoric tasks
A13	Playing ball on the grass	A32	Walking through the labyrinth
A14	Playing rackets on the grass	A33	Sitting in the aromatic herbs garden
A15	Riding a bicycle	A34	Walking through the aromatic herbs garden
A16	Playing football	A35	Listening bells sounds
A17	Playing basketball	A38	Playing theatre games
A19	Playing social games sitting around a table	A39	Playing in the Sensory Park

Some of the activities can be grouped depending, for example, on the age of people doing them. For each activity, the main characteristics of the citizens that could be candidates at different affinity levels together with the most expected time-frame are formalized in CPN. In Table 2, some agent attributes formalized in CPN are summarized.

Specific citizen behavioral rules that has been formalized in CPN and later on codified in Repast are:

S1: Kids between 3 and 5 years of age are mostly interested on playing (i.e. swings, sand area); however, sometimes they could prefer to sit on a bench or around a table to have some snack or even to have lunch.

S2: Kids between 6 and 12 years of age are also mostly interested on playing (i.e. playing ball) ; however, their plays are quite different than kids between 3 and 5.

S3: Young people between 13 and 19 years should have other concerns and responsibilities. Some of them can be really interested in sports (individual and collective), and they can be responsible of a dog. And preferences can vary between males and females. However, if they don't like sports, males and females share preferences (it will implies to include them in the same rule), and these are sitting (could be on the grass, around a table, on a bench,...) or some pair situations like walking or sitting together. Due to the fact they can move around by themself, walking pass through the park can be also one of their activities. Regarding sports, they have different preferences that can be summarized using statistics.

Table 2: Some Agent's Attributes

Data (attribute)	Meaning and Ranges
Age	Age of the citizen. Range: 0 - 120
Gender	Gender of the citizen. Range: M (male) or F (female)
Dog	Having or not having a dog. Range: 0 (no dog) or 1 (dog)
CultLevel	Cultural level. Range: from 1 (no studies) to 5 (Master or PhD level). This information is useful to describe how the affinities can affect or can be influenced by their neighbourhood´s opinions.
HouseType	Type of house. It is different to live in an apartment than in a house. Range: 1 (Apartment), 2 (Town house without too much garden) or 3 (House with garden)
CitizenOrigin	Type of citizens depending on where they live. Range: 1 (Autism centre), 2 (Surrounding neighbourhood), 3 (Closest districts), 4 (Other town parts) or 5 (Outside the town)
Personality	Level of personality of the agent. High personality indicates that it can influence the other agents. Low personality indicates that other agents with higher personality can influence it. Range: 0 to 100.

By means of a multirun approach, the agent observer computes the best surface distribution between different zones that should allocated the different described activities, while minimizing the potential conflicts between activities in the same zone.

Different output parameters have been used for the simulation model acceptability. Thus, in Figure 3, it has been represented the occupancy of the different zones considering a reduced size scenario with a particular neighborhood profile. By introducing extreme scenario conditions (i.e. Only elderly citizens or only young couples without kids, or only teenagers, …) it has been possible to involve end users in the model acceptability.

6. CONCLUSIONS

Under the framework of the FUPOL project one challenging task is the policy modelling and analysis. The proposed methodology has been performed through a novel approach which models the different actors in a policy process as agents whose behaviour is governed by a causal modelling developed in coloured Petri nets. The translation of the CPN models into the Repast environment allows a novel way of understanding the causal relationships that are behind decision making in society. With the use of CPN it is possible to implement the causal relationships that govern the agent behaviour in such a way that more transparency is achieved during the evaluation of a particular policy.

Inherent difficulties of ABM verification and validation has been tackled by formalizing agents behavior using CPN and analyzing the state space to determine model plausibility. In general, we consider that simpler agents with simpler rules are to be preferred. The simpler the rule, the easier it becomes to test the model and discover its implications.

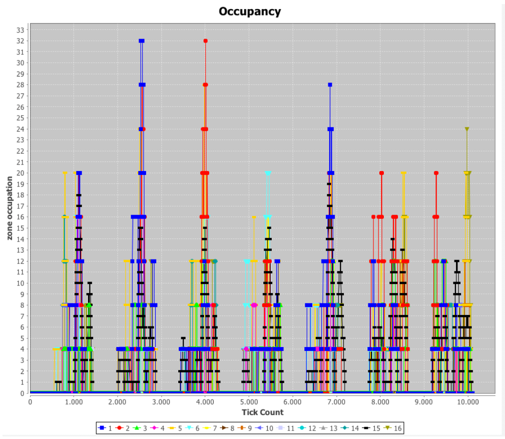


Figure 3: Occupancy of Green Park Zones

ACKNOWLEDGMENT

This work is funded by Future Policy Modeling Project (FUPOL), FP7-ICT-2011-7, ref. 287119 (www.fupol.eu).

REFERENCES

Grimshaw, D., 2001. CPS 720 Artificial Intelligence Programming Course. Ryerson University. Available from: http://www.ryerson.ca/~dgrimsha /courses/cps720/agentdef3.html [accessed June 2012]

Ormerod, P. and Rosewell, B., 2009. Validation and Verification of Agent-Based Models in the Social Sciences. *Lecture Notes on Artificial Intelligence*, V. 5466. Springer.

Jensen K; Kristensen L.M.; 2009."Coloured Petri Nets: Modelling and Validation of Concurrent Systems", Springer,2009.

Mujica, M., Piera, M.A., Narciso M., 2010. Revisiting state space exploration of timed coloured petri net models to optimize manufacturing system's performance. *Simulation Modelling Practice and Theory*, vol.8(9), pp. 1225-1241.

Albeverio, S., Andrey D., Giordano P., Vancheri A., 2004. *The Dynamics of Complex Urban Systems.* Springer, Physica-Verlag.

A MAS APPROACH BASED ON COLOURED PETRI NET FORMALISM FOR URBAN POLICY DESIGN

Roman Buil [(a)], **Miquel Angel Piera** [(b)]

Universitat Autònoma de Barcelona, Department of Telecommunications and Systems Engineering, Unit of Logistics and Aeronautics, 08193, Bellaterra, Barcelona

[(a)]miquelangel.piera@uab.cat, [(b)]roman.buil@uab.cat

ABSTRACT

Simulation transparency is becoming more crucial in the decision making process when quantitative computer tools are used to justify some strategies. E-governance is one of these areas in which the use of Multi Agent System (MAS) simulation systems could be used to foster e-participation in which citizens could be involved in the design of urban policies that affects their habitat environment.

The Colored Petri Net (CPN) formalism is a promising modelling approach to foster simulation transparency by means of state space traceability tools and it has been proven to be useful for modelling system dynamics with concurrent and conflict patterns in more efficient ways.

We propose a modeling methodology to represent and analyze a context-aware multi agent-based system, which tends to be highly complex. We introduce CPNs as a method of capturing the dynamics of this contextual change. We define CPNs and a way to apply them in context-aware agent-based systems. We also describe a prototype system that we have developed which translates CPN specification into Repast Simphony agents' behaviour.

Keywords: Urban policy, state space, petri net, MAS simulation, e-participation

1. INTRODUCTION

The European Commission launched a call under its framework 7 program dealing with ICT solutions for governance and policy modelling (Objective ICT-2011.5.6).

Target Outcomes are ICT solutions for governance and policy modelling. The research is focused in the development of advanced ICT tools for policy modelling, prediction of policy impacts, development of new governance models and collaborative solving of complex societal problems.

The main objective of FUPOL is to demonstrate that, with ICT support the whole policy development lifecycle of policy formulation, collaborative stakeholder involvement, policy modelling, scenario generation, visualization of results and feedback is feasible and a core element of future policy development at local, regional, national as well as global level (Mujica and Piera 2012).

Most urban simulation models do not take into account the demands and skills of the multiple and heterogeneous users of a model in practice. There exist fears, expectations and prejudices among the practitioners against the models: quantitative models are monsters and do not capture the social aspects in an adequate way. A lack of credibility in the models is reported due to non-scientific actors being not aware of the uncertainty inherent in such models. As a result, mistrust or communication problems can be found generally between scientists and actors.

In contrast to traditional urban modelling methodologies, the proposed causal modelling approach in FUPOL takes into consideration that actors have different skills and pre-knowledge of the complex urban system. By incorporating users' heterogeneity in non-scientific knowledge (subjective values), interests and preferences, FUPOL modelling approach contributes to overcome one of the main shortages of present policy models: citizen e-participation.

The term e-Participation is quite new and it is widely used in e-Governance and e-Democracy programs. E-participation in urban decision-making means the use of ICT for enabling and strengthening citizen participation in democratic decision-making processes. The use of ICT in e-Participation process consists on motivation and engagement of a large number of citizens through diverse modes of technical and communicative skills to ensure broader participation in the policy process, real-time qualitative and accessible information, and transparent and accountable governance (Islam 2008).

By means of a Multi Agent System (MAS) simulation platform, FUPOL models allow citizens to test the benefits and shortages of different proposed urban policies and check new policies according to their own beliefs.

Figure 1 illustrates the selected tools to model the different urban policy domains by properly integrating fuzzy cognitive maps (FCM) and coloured Petri nets (Moore and Gupta 1996, Jensen 1997, Christensen et al. 2001, Mujica and Piera 2011) under a MAS environment, which presents great advantages in order to analyse social systems (North and Macal 2007).

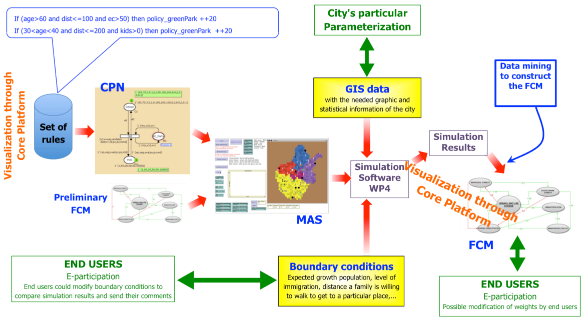

If (age>60 and dist<=100 and ec>50) then policy_greenPark ++20
If (30<age<40 and dist<=200 and kids>0) then policy_greenPark ++20
City's particular Parameterization
Visualization through Core Platform
Set of rules
CPN
GIS data
with the needed graphic and statistical information of the city
Data mining to construct the FCM
Simulation Results
Simulation Software WP4
Visualization through Core Platform
Preliminary FCM
MAS
FCM
END USERS
E-participation
End users could modify boundary conditions to compare simulation results and send their comments
Boundary conditions
Expected growth population, level of immigration, distance a family is willing to walk to get to a particular place,...
END USERS
E-participation
Possible modification of weights by end users

Figure 1: Main FUPOL Simulation Tools

The urban policy models consist of a library with citizens' affinities together with urban areas defined by social activities (constrains and facilities). Citizen affinities can be categorized considering the particular attributes of each person, such as their age, the amount of kids, the incomes and sex among others. Thus, the different citizen behavioural rules are defined by a set of acceptable attribute interval values. All these citizen affinities are described as a set of informal rules that can be easily checked/maintained/modified by different end-users.

The library of rules (citizen affinities together with social activities) is automatically translated to CPN formalism for a causal-based state space analysis. A set of rules is obtained from CPN and used to describe the agents' behaviour, which are modelled using Repast Simphony.

As a result of the simulation, numerical data generated is processed under a data mining approach (ANFIS) to automatically generate the main weights for the causal relationships between Fuzzy Cognitive Maps items. End users can interact with the Repast Simphony Model and the final FCM.

First model and results for Community Facilities: Area Design domain, developed for the City of Zagreb, will be presented in this paper. The main target of this model is the design (or configuration) of an area, in this particular case, a green park near the new national centre of autism, in order to overcome conflicts between neighbours and autistics, and also to facilitate the autistics integration in the society.

Section 2, 3 and 4 present the three modelling approaches combined in the methodology: CPN, MAS and FCM. The methodology phases are described in Section 5, and the Zagreb use case in Section 6. Finally, conclusions are presented in Section 7.

2. COLOURED PETRI NETS

Coloured Petri Nets (CPN) is a simple yet powerful modelling formalism, which allows to properly modelling discrete-event dynamic systems that present a concurrent, asynchronous and parallel behaviour (Moore et al. 1996, Jensen 1997, Christensen et al. 2001). CPN can be graphically represented as a bipartite graph, which is composed of two types of nodes: the place nodes and the transition nodes. Place nodes are commonly used to model system resources or logic conditions, and transition nodes are associated to activities of the real system. The entities that flow in the model are known as tokens and they have attributes known as colours. The use of colours allows modelling not only the dynamic behaviour of systems but also the information flow, which is a key attribute in decision making (Mujica and Piera 2011).

The formalism can be graphically represented by a bipartite graph where the place nodes are represented by circles and the transition nodes by rectangles or solid lines. Figure 2 illustrates a graphical representation of a CPN model.

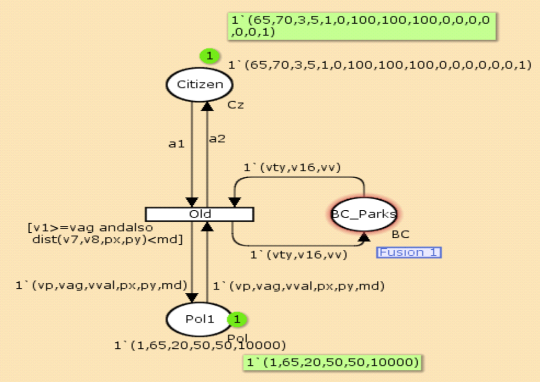


Figure 2: CPN Model Example

3. MULTI AGENT SYSTEMS

Multi-agent systems (MAS) have been applied in various fields related to Human Sciences, such as Political Sciences, Economics, and Social Sciences (Segal-Halevi (2012), Wilensky, 1999), in which an agent can be seen as an actor that has the power of "Agency". The concept of agency is paramount since it extends beyond a single human being to organizations and social systems, where some have "agency" power with the ability to make decisions, while others do not have such power. These agents communicate, collaborate and negotiate among each other in order to meet their design objectives.

During the developing phase of MAS, it is necessary to simulate the system feasibility before the formal implementation. One way to achieve this objective is to develop a model to represent the behaviours of the MAS and then to simulate the model performance. In practice, a lot of time might be spent on modelling MAS behaviours while the resulting process models are typically still not compatible with the original system due to the lack of information about actual behaviours of the MAS. To avoid this difficulty, CPN models can be used to describe agents' behaviour and predict the performance by means of state space analysis tools.

In FUPOL several MAS platforms has been analysed and, among them, Repast Symphony has been chosen as a simulation environment.

Repast Simphony is a widely used, free, and open source environment for agent-based modelling of complex adaptive system. The most recent version, 2.0, was released on March 5, 2012. Repast Simphony is a second-generation environment that builds upon the previous Repast 3 library (North et al. 2013).

Due to its flexibility it is suited to fully integrate the semantic rules present in CPN, and their implementation allow to govern the agents that interact within the environment in a more transparent way which is useful to understand the emergent dynamics caused by the agent interaction.

4. FUZZY COGNITIVE MAPS

Cognitive maps (CMs) are qualitative models of a system, consisting of variables and the causal relationships between those variables. Causal relationships among the variables in the models are specified and tested with parameter estimation procedures, usually maximum likelihood. These variables can be physical quantities that can be measured. The decision-makers' maps can be examined, compared as to their similarities and differences, and discussed (Özesmi,2003). Stakeholders can be compared to see which groups have more relationships among variables. If some groups perceive more relationships, they will have more options available to change things. The person who develops the cognitive map decides what the important concepts that affect a system are, and then draws causal relationships among these variables indicating the relative strength of the relationships with a positive/negative/none sign between concepts. The directions of the causal relationships are indicated with arrowheads. Also, he decides on the strengths that can be changed easily. More simulations can be done in order to see how the model changes with changing strengths of relationships

Strictly speaking, a FCM is a figure composed of nodes and edges, the former introducing the qualitative concepts of the analysis while the latter are indicating the various causal relationships. Each concept node possesses a numeric state, which denotes the qualitative measure of its presence in the conceptual domain. Thus, a positive numeric value indicates that the concept is strongly present in the analysis, while a negative or zero value indicates that the concept is not currently active or relevant to the conceptual domain. A FCM works in discrete steps. When a strong positive correlation exists between the current state of a concept and that of another concept in a preceding period, we say that the former positively influences the latter indicating this by a positively weighted arrow directed from the cause to the effect concept. By contrast, when a strong negative correlation exists, it reveals the existence of a negative causal relationship indicated by an arrow charged with a negative weight. Two conceptual nodes without a direct link are, obviously, independent.

5. MODELLING METHODOLOGY PHASES

There are several steps to be followed in order to ensure a good performance of the final simulation models. These steps are:

1. Rules generation using the information obtained from the field work make in pilot cities to collect data regarding the corresponding domain. In the case of Zagreb, to collect data regarding the use of the parks in the city
2. These rules are internally verified and validated using CPN models and then, they are used to formalise agent's behaviour in Repast Symphony (MAS)
3. Numerical data obtained by running several MAS simulations is processed under a data mining approach to automatically generate the main weights of the FCM.

After these steps, the MAS model and the FCM are ready to be used by end users, which will be able to modify the boundary conditions in order to test what happen if the initially considered ones are not really satisfied.

6. CASE STUDY: GREEN PARK AREA DESIGN

6.1. Description of the System

The simulation objective for a green park design in Zagreb is to provide the best solution for the facilities that would be included in the green park situated near the Autism Centre. The surface of the park is around 30.000m^2, and the design must satisfy most of the potential users demands and must encourage interactions between autistic people and non-autistic users, while avoiding possible conflicts between them. At the same time, possible conflicts between all kinds of users must be avoided. For example, nobody likes to have dogs around kids while these are playing in a playground. Or nobody is going to sit on the grass near 20 young guys playing football.

6.2. Model Specifications

The officers of the city of Zagreb together with experts on autism disorder have generated the data used in this paper in order to test the simulation model while they perform the field work to collect the real data that they will finally use. They have considered their own experience in green parks around the city. Table 1 presents the considered activities.

Table 1: Green Park Activities

ID	Activity	ID	Activity
A1	Walking through the park pathway	A21	Walking with a dog
A2	Sitting on a bench	A22	Touching objects and surfaces
A3	Sitting around a table	A23	Playing frisbee
A4	Feeding children	A24	Playing with a dog
A5	Playing in the Sandbox	A25	Playing bocce ball
A6	Sitting on the grass	A26	Resting
A7	Sitting and playing on the	A27	Playing in the labyrinth

	grass		
A8	Swinging	A28	Standing
A9	Sliding down the toboggan (slide)	A29	Standing with a parm
A10	Spinning on roundabouts	A30	Watching water movement and/or listening the sound of the water
A11	Climbing on monkey bars	A31	Following the paths with motoric tasks
A12	Roller skating	A32	Walking through the labyrinth
A13	Playing ball on the grass	A33	Sitting in the aromatic herbs garden
A14	Playing rackets on the grass	A34	Walking through the aromatic herbs garden
A15	Riding a bicycle	A35	Listening bells sounds
A16	Playing football	A36	Playing passing obstacles (climbing, crawling, going in and out, etc.)
A17	Playing basketball	A37	Having a party in the amphitheatre
A18	Skateboarding	A38	Playing theatre games
A19	Playing social games sitting around a table	A39	Playing in the Sensory Park
A20	-		

These activities can be performed in specific zones, and these zones are summarized in Table 2.

Table 2: Green Park Zones

ID	Description
Z1	Playground with games 1, for little children
Z2	Playground with games 2, for younger school age
Z3	Playground with games 3, for adolescents
Z4	Sandbox
Z5	Grass zone (with some trees to have some shade)
Z6	Picnic zone (with tables and benches)
Z7	Bike training ground (with basic traffic signs and rules)
Z8	Sensory Park: Labyrinth
Z9	Sensory Park: Music bells corner
Z10	Sensory Park: aromatic herbs garden
Z11	Sensory Park: building in the nature zone
Z12	Sensory Park: amphitheatre
Z13	Sensory Park: paths with different colours, textures and training ground for motoric tasks (jump, skip over, scrape through, reach and pull, etc.)
Z14	Sensory Park: Fountain
Z15	Pathways
Z16	Bocce ball fields

For each activity, the main characteristics of the citizens that could be candidates at different affinity levels together with the most expected time-frame are formalized in CPN. Some agent attributes are summarized in Table 3.

Table 3: Some Agent's Attributes

Data (attribute)	**Meaning and Ranges**
Age	Age of the citizen. Range: 0 - 120
Gender	Gender of the citizen. Range: M (male) or F (female)
Dog	Having or not having a dog. Range: 0 (no dog) or 1 (dog)
CultLevel	Cultural level. Range: from 1 (no studies) to 5 (Master or PhD level). This information is useful to describe how the affinities can affect or can be influenced by their neighbourhood´s opinions.
HouseType	Type of house. It is different to live in an apartment than in a house. Range: 1 (Apartment), 2 (Town house without too much garden) or 3 (House with garden)
CitizenOrigin	Type of citizens depending on where they live. Range: 1 (Autism centre), 2 (Surrounding neighbourhood), 3 (Closest districts), 4 (Other town parts) or 5 (Outside the town)
Personality	Level of personality of the agent. High personality indicates that it can influence the other agents. Low personality indicates that other agents with higher personality can influence it. Range: 0 to 100.

Specific citizen behavioural rules that has been formalized in CPN and later on codified in Repast Simphony are:

S1: Kids between 3 and 5 years of age are mostly interested on playing (ie. swings, sand area); however, sometimes they could prefer to sit on a bench or around a table to have some snack or even to have lunch.

S2: Kids between 6 and 12 years of age are also mostly interested on playing (i.e. playing ball); however, their plays are quite different than kids between 3 and 5.

S3: Young people between 13 and 19 years should have other concerns and responsibilities. Some of them can be really interested in sports (individual and collective), and they can be responsible of a dog. And preferences can vary between males and females. However, if they don't like sports, males and females share preferences (it will implies to include them in the same rule), and these are sitting (could be on the grass, around a table, on a bench,...) or some pair situations like walking or sitting together. Due to the fact they can move around by themself, walking pass through the park can be also one of their activities. Regarding sports, they have different preferences that can be summarized using statistics.

6.3. Model Representation

Figure 3 presents the preliminary MAS model representation used by modellers; final interface is still under development. Circles in purple are non-autistics, they are located anywhere, and each point can represent more than one agent because the members of the same family unit are located at the same place, their residence (home); triangles in purple are autistics, and in the current stage they are all located in the cell representing the centre of autism. Red circles are monitors in charge of escorting the autistics, and they are located in the centre of autism also (there are two cells representing the centre). Big circles on the left side represent the different zones considered in the park. Their colour change depending on its occupancy: Green means there is a lot of space (low occupancy rate); orange means that the zone is almost full of people (elevate occupancy rate); and red means that the occupancy rate is at 100% or more, indicating that more space for that zone is needed to fit all the people interested in it.

Figure 4 present the map of the area where the autism centre and the park will be placed. The red area marked with "D" is the area for the autism centre, and the green area with "Z1" is the area for the green park.

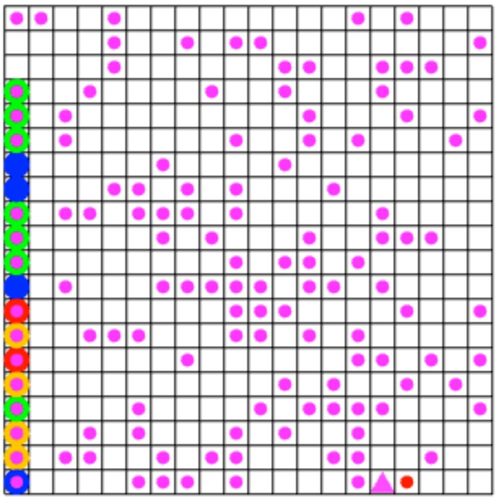


Figure 3: Preliminary MAS Model Representation

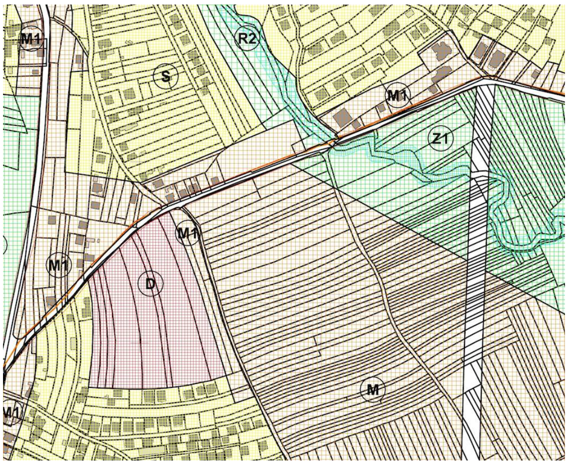


Figure 4: Autism Center and Green Park Area

By means of a multirun approach, the agent observer computes the best surface distribution between different zones that should allocated the different described activities, while minimizing the potential conflicts between activities in the same zone.

6.4. Test Results

Reduced size scenario has been considered to test the model (All zones of 20m2, but zone 15, which are the pathways, with a fixed surface of 80m2), and end users model acceptability has ben reached by introducing extreme scenario conditions, one of which is "Mostly elders and 20% with dog" and it is presented in this paper.

Tests consist on simulations of one week, from Monday to Sunday, and different time frames during the day (9:00 to 12:00, 12:00 to 15:00. 15:00 to 17:00, 17:00 to 20:00, and 20:00 to 00:00). Outputs graphically presented in this paper are:

1) Zones Occupancy presented in Figure 5: it can be observed that zones 6 (picnic zones where elders play table games) and 16 (bocce ball fields) are more occupied than other zones.
2) Zones Surface in Figure 6: due to zones 6 and 16 are more occupied; they need more square meters (surface).
3) Number of conflicts in Figure 7: initial conflicts are in zones 6 and 16 before more square meters are assigned to them. Latter on, due to there are not just elders and zone 1 surface has been reduced, more conflicts appear in zone 1.

Notice that the convergence of the needed surface will be stabilised as many days are simulated.

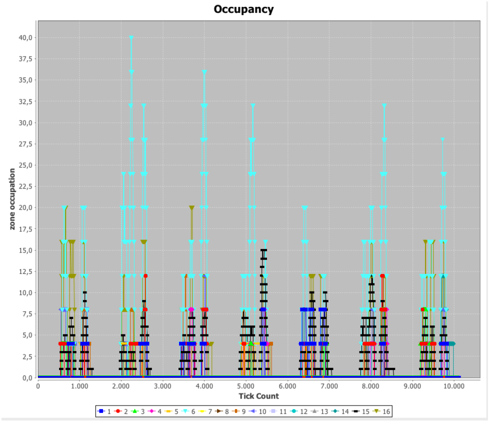


Figure 5: Occupancy of Zones

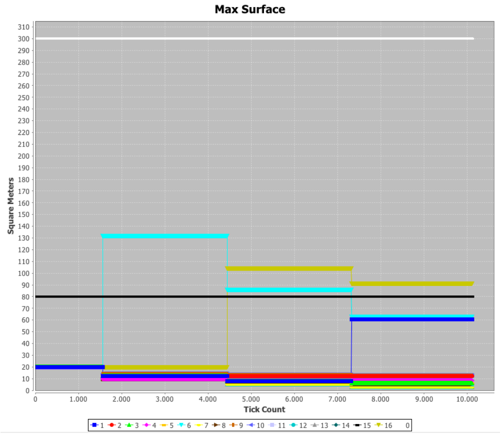


Figure 6: Zones Surface

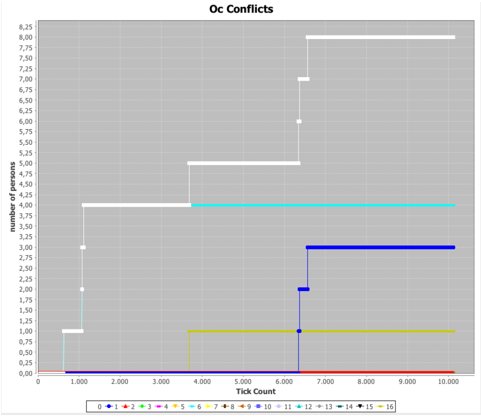


Figure 7: Occupancy conflicts

7. CONCLUSIONS

Under the framework of the FUPOL project one challenging task is the policy modelling and analysis. The proposed methodology has been performed through a novel approach which models the different actors in a policy process as agents whose behaviour is governed by a causal modelling developed in coloured Petri nets. The translation of the CPN models into the Repast Simphony environment allows a novel way of understanding the causal relationships that are behind decision making in society. With the use of CPN it is possible to implement the causal relationships that govern the agent behaviour in such a way that more transparency is achieved during the evaluation of a particular policy.

Results of the green park area design preliminary model show that the developed model can be used to estimate the number of persons performing certain activities in certain zones of the park, which can be used to calculate the surface needed for each zone in order to do not have occupancy conflicts in the park. Final simulation software allowing much more agents will be used to initially design the park and also to predict park maintenance activities.

ACKNOWLEDGMENTS

This work is funded by Future Policy Modelling Project (FUPOL), FP7-ICT-2011-7, ref. 287119 (www.fupol.eu).

REFERENCES

Christensen, S., Jensen, K., Mailund, T., Kristensen, L.M., 2001. State Space Methods for Timed Coloured Petri Nets. Proc. of 2nd International Colloquium on Petri Net Technologies for Modelling Communication Based Systems, 33-42, Berlin.

Islam M. S.,2008, Towards a sustainable e-Participation implementation model, European Journal of ePractice ,www.epracticejournal.eu 1 Nº 5, October 2008, ISSN: 1988-625X

Jensen, K., 1997. Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use. 1 Springer-Verlag, Berlin.

Moore, K.E., Gupta, S.M., 1996. Petri Net Models of Flexible and Automated Manufacturing Systems: A Survey. International Journal of Production Research, 34(11), 3001-3035.

Mujica, M.A., Piera M.A., 2012. The translation of CPN into NetLog environment for the modelling of political issues: FUPOL project. *24th European Modeling and Simulation Symposium, EMSS 2012*, pp. 555.

Mujica, M.A., Piera, M.A., 2011. A Compact Timed State Approach for the Analysis of Manufacturing Systems: Key Algorithmic Improvements, *International Journal of Computer Integrated Manufacturing, Vol.24 (2),* February 2011.

North M. J., Macal C. M., 2007, Managing Business Complexity Discovering Strategic Solutions with Agent-Based Modelling and Simulation, *OUP*.

North, M.J., N.T. Collier, J. Ozik, E. Tatara, M. Altaweel, C.M. Macal, M. Bragen, and P. Sydelko, 2013. Complex Adaptive Systems Modeling with Repast Simphony, *Complex Adaptive Systems Modeling, Springer, Heidelberg*, FRG (2013).

Predictive monitoring of production line efficiency

Tomi Zebič[(a)], Boštjan Hauptman[(b)], Peter Rogelj[(b)], Gašper Mušič[(c)]

[(a)]Metronik d.o.o, Ljubljana, Slovenia
[(b)]Trimo d.o.o, Trebnje, Slovenia
[(c)]University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

[(c)]gasper.music@fe.uni-lj.si

ABSTRACT

The paper deals with analysis of production line performance efficiency and builds a prediction model that enables a short-term prediction of the expected performance based on the scheduled product mix. The actual schedule can be used as the model input to assist the production operators at the on-line production management. The same model can be used to experimentally evaluate the effect of different scheduling strategies by linking the performance model to discrete event simulator.

Keywords: Production process efficiency, modelling, simulation, optimization

1. INTRODUCTION

Manufacturing companies are facing increasing competitive pressures characterized by requirements on fast responsiveness while maintaining high productivity at high quality. Effective production management is one of the fundamental operational activities that has to be carefully designed and integrated into the overall management structure in order to meet the given requirements.

The integration of production management is a process aiming to upgrade and strengthen the links among the existing management activities. Various information technology products are used to support and improve the efficiency of production management. This way large volume of data is collected that contain useful information about production process performance. But the quality operational decision making still remains one of the most critical challenges for present manufacturing companies. The sole collection of production data is inadequate and a more tangible decision-making support is needed.

Within the third part of IEC 62264 standard - Enterprise-control system integration - entitled Activity models of manufacturing operations management (IEC 2007) the production management activities are decomposed in details. The production control is included among activities of the third level, and is further decomposed. Among others, the Production performance analysis is listed as a part of Production operations management. It is defined as a collection of activities that analyze information on the effectiveness and report the outcomes to the business level. This includes analysis of information on cycle times, resource utilization, equipment utilization, equipment and procedures' efficiency, and product variability. Relationship between these and other analyses can be the basis for the preparation of Key Performance Indicators (KPI) reports. This information can be used to optimize the production and the use of resources.

One way is to develop useful production performance models and integrate them in appropriate software tools to provide a better insight into performance mechanisms. The term model is referred to herein as a relationship between the relevant influence quantities, and one or more selected output variables. Such a model can be used either to analyze the relations between quantities, either for the analysis of scenarios, but it can also be used for short-term prediction of production performance. The models can be used to simulate production efficiency. In this way the effect of changes in the process can be tested experimentally, either changes in the operational settings, changes of production procedures or changes in the production path. Also the impact of external influences can be tested, such as the structure of the work orders, changes of material inputs etc. This enables an advanced evaluation of control measures and improves the quality of the operations management decisions. The models can be built on the basis of historical production data and may reflect different aspects of the production based on the intended purpose of their use.

This way the models are used in the context of production efficiency management, which is a set of activities that systematically record, manage and present information about the performance in a consistent manner, including corrective actions to affect the operational improvements. One of the main activities of the production efficiency management is the transformation of large amounts of raw data into information that can be used as a decision support on the management of production.

As discussed above, the ability of KPIs' prediction is an important aspect of the production efficiency management. The traditional implementation of this prediction is within the production planning/scheduling. Plans and schedules contain information that indicates the future production activities and can be used to estimate the future KPI values. An advanced implementation of KPI predic-

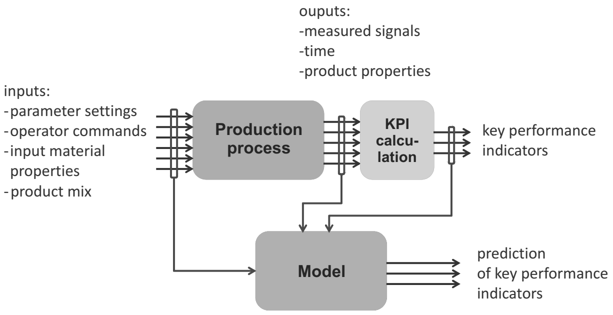


Figure 1: KPI prediction model

tion is based on the use of statistical techniques and experimental modeling methods (identification) on the existing KPI data values, and development of prediction models for the future values (Figure 1).

The paper investigates the applicability of modelling methodology in the context of production performance prediction and optimization. This is one of the fields where information technology has an immediate and considerable impact on the efficiency and quality of production control and related manufacturing processes. A case study is presented, where a production line for making building construction panels is analysed and various uses of derived model are proposed to improve the production management and manufacturing execution.

2. PRODUCTION PERFORMANCE MEASUREMENT

Measurement and evaluation of production performance is a key component of production efficiency management. One of the main challenges it to orchestrate various performance metrics in view of the changing list of production objectives. The purpose of performance indicators is on the one hand to provide information on the achievement of a set of objectives and on the other hand to connect the observed values with the improvement measures that should be taken. In this sense, the Performance Measurement Systems (PMS) are tools for decision support (Kaplan and Norton 1992, Neely et al. 1995, De Toni and Tonchia 2001) in a process of continuous improvement.

The importance of performance measurement drives a rich variety of method proposals and approaches that can be found in research literature. Nudurupati et al. (2011) give a recent survey of PMS developments in relation to management information systems and change management. From a global point of view, PMS can be treated as a multi-criteria instrument, which consists of a set of performance expressions (also called metrics) that are consistently organized according to the objectives of the company (Berrah and Cliville 2008). In doing so, the metrics can be based on actual measurements as well as on the other types of effects evaluations. PMS is always defined in relation to the global objective, and gives as a result one or more efficiency measures with the purpose of quantitative evaluation of the fulfillment of this objective.

In general the considered global objective is decomposed to a more elementary objectives along organizational levels (strategic, tactical, operational), while the elementary performance expressions associated with the decomposed objectives are aggregated to provide information on the achievement of the global objective. Various quantitative decomposition/aggregation performance measuring models have been proposed in order to control and manage the process of improving efficiency, thus supporting decision making in this process (Ghalayini et al. 1997, Suwignjo et al. 2000, Cliville et al. 2007, Berrah and Foulloy 2013).

2.1. Overall Equipment Effectiveness

The top-down oriented PMS are often combined with specific performance measures, such as Overall Equipment Effectiveness indicator (OEE) (Nakajima 1988, Muchiri and Pintelon 2008). Although OEE measure is not a complete PMS, it is an important complement to the traditional PMS when applied by autonomous small groups on the shop-floor together with quality control tools (Jonsson and Lesshammar 1999). As such the OEE measure is one of the standard indicators of technological performance.

In the foreground of the OEE assessment is the treatment of losses due to various interruptions in production process. They have a variety of causes, but commonly result in activities that consume resources without creating new value. To control the efficiency of discrete production in the context of TPM (Total Productive Maintenance), Nakajima (1988) developed a model of quantitative evaluation of the overall effectiveness of equipment that identifies each type of loss. The model includes the following losses:

- in terms of the availability:
  - time loss due to equipment failure,
  - level of use of the equipment or time wasted due to the preparation, set-up and adjustment of equipment;
- in terms of effectiveness:
  - production speed reduction due to minor stoppages, e.g. abnormal operation of machinery, unexpected shutdown, etc.,
  - production speed deterioration due to the operation of equipment at a speed below the nominal;
- in terms of quality:
  - level of production losses, measured by the volume of low-quality production due to scrap and re-work,
  - the level of other losses, representing lower production yields due to machine start-up runs before the establishment of stable operation of the equipment.

The model is partly based on the SEMI E10 standard that defines the operational status of equipment for semiconductors manufacturing and classifies the associated time intervals. Through the comparison of the time intervals the OEE provides a comprehensive insight into the utilization of available resources. Production is most effective according to the OEE indicator when the production system is operating at full capacity, producing the required product quality while production process is working without interruption.

There are several slightly different OEE definitions. E.g., in determining the availability, the planned production time can include the time that is used for preventive maintenance or not. In the first case we get a higher level of availability while planning more frequent maintenance tasks will not decrease it. But this can lead to poor planning and excessive maintenance time. On the other hand, the inclusion of this time in the planned production time lowers the availability indicator, but it reflects the actual availability and at the same time motivates more effective maintenance planning (Jonsson and Lesshammar 1999).

Similarly, slightly different definitions of the performance and quality indicators are used. At this point, the definition of OEE is used, as defined by the ISO 22400-2 standard. OEE is composed as a product of three independent components

$$OEE = A \cdot P \cdot Q \tag{1}$$

In equation (1) the three components have the following meanings:

- A - Availability (in terms of performance availability)
- P - Performance (efficiency in terms of capacity)
- Q - Quality (effectiveness in terms of quality)

Such a structure of the OEE indicator means that a disturbance in the manufacturing process is reflected on one of the components, enabling the identification of the cause of loss in efficiency. At the same time such a structure allows easy detection of weaknesses in the organization or operation of the production process.

3. PRODUCTION PERFORMANCE MODELLING

Production models are often developed to cope with technological efficiency. In particular, the model is used to obtain an insight into the future behaviour of key variables. This information can facilitate the effective ergonomic (re)design and its optimization (del Rio Vilas et al. 2013, Latorre et al. 2013) but can also be used in the context of real-time operational management (Curcio et al. 2007, Mujica et al. 2010).

The primary purpose of the modelling of technological efficiency, therefore, is the prediction of technological efficiency-related indicators. This allows for better corrective actions in the context of production efficiency management in order to increase technological efficiency. Particularly, modelling can be seen as an aid in deciding on production management measures.

In terms of the standard performance indicators it is particularly important to enable the prediction of indicators that are linked to productivity (efficiency, availability, ...) and quality. The impact on these indicators is highly dependent on the actual production process, but some general dependencies can also be extracted.

As a starting point for performance modeling standard dynamic systems modelling approaches can be applied: theoretical modeling based on the known physical and other relationships among the considered quantities, experimental modelling based on the measured signals and archived data, and a combination of the two approaches.

3.1. Theoretical modelling

The theoretical modeling is based on assumed a-priori known relationships. For example in the process industry the duration of the manufacturing operations can be determined if the dynamics of processes within a specific operation is modelled. Such theoretical modelling works well for smaller problems where we have a good insight into the system.

Another aspect of theoretical modeling, which is particularly important in the discrete production, is the structural aspect. Production processes are very diverse, so we use the analysis of the general properties and characteristics of the production processes during the modelling.

With the model of the structure of the manufacturing process, we wish to create a universal presentation for any real production process, and on the basis thereof design a performance model of the specific production process. The structure model should include basic building blocks and the links between them that can describe the nature of most manufacturing processes, with an emphasis on the presentation of the flow of material among operations. In literature, the basic building blocks of the production process model are production equipment devices (machines), and these can be then connected together in different ways. In general, in any manufacturing process we can distinguish:

- equipment,
- connections among pieces of equipment.

3.2. Manufacturing process effectiveness

The OEE indicator is calculated for each peace of equipment, e.g. each machine, but for a comprehensive evaluation of the production efficiency the metrics for individual machines must be combined into a single performance metric of the manufacturing process. The problem here is that not only the integration of various indicators has to be considered, as is the case with PMS systems, but it is necessary to take into account the structure model of the process.

Only the evaluation of the effectiveness of the production process as a whole provides a link with cost efficiency. For performance measures of the entire production process the abbreviation OFE - Overall Effectiveness Factory is used in the literature.

Metrics at the level of the production process should summarize the situation in individual machines and at the same time add a holistic perspective - the aspect of machines coordination. Therefore, the way of machine level metrics integration to the level of the production process is significantly affected by the interconnection of equipment.

3.3. Overall throughput effectiveness

In the literature several attempts can be found to design procedures for determining metrics based on the model of the structure of the production process. Huang and coauthors designed an OEE type metric to measure the efficiency level of the plant or factory. The measure is called Overall Throughput Effectiveness (OTE) (Muthiah and Huang 2007).

This measure was derived similarly to the overall effectiveness of equipment (OEE), on the basis of the finding that the efficiency of the entire production process, is the ratio between the actual quality product and the normalized product (quantities produced) in a selected time interval. This relationship is given by:

$$OTE = \left. \frac{P_{act}}{P_{norm}} \right|_{t_{observation}} \tag{2}$$

where P_{act} is the actual amount of quality (appropriate) product after completion of the production procedure, P_{norm} is the normalized (expected) quantity of product at the end of the production process.

In determining the two variables that appear in this equation the machine level quantities should be acquired and mapped to the production level, whereby it is necessary to take into account the connection between the machines. For the assumed model structure the OTE calculation formula can be written.

Serial connection of machines

In the calculation of this type of structure we assume that the machines are rigidly connected, so the slowest machine dictates the operation of the entire line. The basic equation is adapted and taking into account the structure of the OEE metric we obtain

$$\begin{aligned} &OTE_{ser} \\ &= \frac{1}{\min\limits_{i=1,2,\ldots,n} \left\{R_{th(i)}\right\}} \min \left\{ \min_{i=1,2,\ldots,n-1} \left\{ OEE_{(i)} \right.\right. \\ &\qquad \left.\left. \times R_{th(i)} \times \prod_{j=i+1}^{n} Q_{eff(j)} \right\}, OEE_{(n)} \times R_{th(n)} \right\} \end{aligned} \tag{3}$$

where $Q_{eff(i)}$ is quality component for machine i (quality efficiency); $R_{th(i)}$ is a capacity component for the machine i (theoretical processing rate), $OEE_{(i)}$ is the OEE for machine i, and n is the number of machines in the chain.

The equation can be illustrated with the following example (Muthiah and Huang 2007): suppose that the line consists of machine A and machine B, processing begins on machine A, and is continued on machine B. So the amount of good product of the machine B is limited to that of machine A. If the performance of machine B is not limited by the quantity of product from the machine A, then the quantity of product from the machine B, and thus of the entire line is dependent on the efficiency of the machine B. Otherwise, the amount of good product is limited by the efficiency of machine A and machine B quality factor.

Similarly the formulas for parallel connection of machines and assembly/distribution connection of machines can be derived.

OTE metric restrictions

Weaknesses of the metric for serial connection of machines originates from the assumption that the connection is rigid, so a fixed line structure is assumed. Not every serial link between the machines is rigid, and in such cases the stoppage of certain machines does not block the entire production. Other machines can continue working and generating in-process inventory.

Another type of the disadvantages of the OTE is that it does not have explicit components and only the overall efficiency of the production process is evaluated. It does not give information on whether the source of the problem is the machine operation or there is a problem at the level of machines coordination. As a result of this it is not possible to determine the reasons for the loss at the level of the entire production process.

The latter disadvantage is partially eliminated by some of the other metrics, which can be found in the literature, such as OLE - Overall Line Effectiveness and OEEML - Overall Equipment Effectiveness of a Manufacturing Line (Mathur et al. 2011).

4. Efficiency monitoring case study

Presented methods of determining the efficiency of production represent a static model of technological efficiency. Such a model is useful for the evaluation and analysis of specific situations in the manufacturing process, but it is not useful for control in terms of an integrated production management. The prediction of the indicators is needed, which can be used for making decision on the management measures.

In the discrete production there is often a strong dependence of the reachable production speed on the type of product. Knowing the structure of already dispatched work orders, the planned schedule can be considered as a model of future conditions in production. Inputs to this model are: production plan, production procedures, production resources, production times and other parameters, the model outputs are sequences and durations of tasks. Prediction of certain performance indicators can be obtained through the proper evaluation of a given schedule.

4.1. Production line performance modelling

As a practical example the presented approach was tested within a demonstration project in a Slovene company. A building construction panels production line is considered. The panels are produced by gluing the appropriately processed thin plates with a layer of the mineral wool. This way a stripe shape sandwich structure of a fixed width is produced, which is then cut into panels of desired lengths. The basic layout of the line is shown in Figure 2.

In modelling of the production line we start from the assumption of the serial connection among machines that has already been discussed in the presentation of the OTE metric (equation (3)).

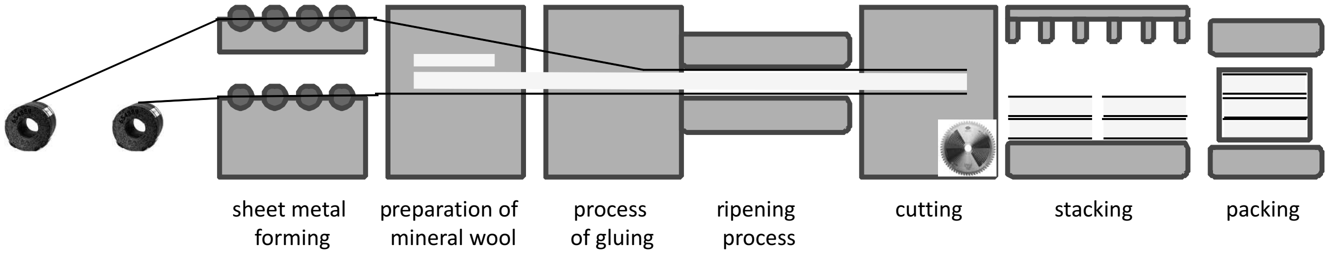


Figure 2: Production line

To better understand equation (3) the equation to calculate OEE for machine i is rewritten by expressing P factor slightly differently, namely as the ratio of the actual (R_{act}) and the theoretical processing rate (R_{th}).

$$OEE_{(i)} = A_{(i)} P_{(i)} Q_{eff(i)} = A_{(i)} \frac{R_{act(i)}}{R_{th(i)}} Q_{eff(i)} \tag{4}$$

It can be seen that the product $OEE_{(i)} R_{th(i)}$ in (3) actually reflects the product of availability factor, the actual performance and quality factor of the machine i. Equation (3) to calculate OTE in the case of serial machine setup can therefore also be written as

$$\begin{aligned} OTE_{ser} = &\left(\min_{i=1,2,\ldots,n} \left\{ R_{th(i)} \right\} \right)^{-1} \cdot \\ &\min \left\{ \min_{i=1,2,\ldots,n-1} \left\{ A_{(i)} \times R_{act(i)} \times Q_{eff(i)} \right. \right. \\ &\left. \left. \times \prod_{j=i+1}^{n} Q_{eff(j)} \right\}, A_{(n)} \times R_{act(n)} \times Q_{eff(n)} \right\} \end{aligned} \tag{5}$$

Next, a new label $Q_{(i..n)}$ is introduced to denote a factor of quality of the rest of the line from the i-th machine (included) to the end of the line:

$$Q_{(i..n)} = \begin{cases} Q_{eff(i)} \times \prod_{j=i+1}^{n} Q_{eff(j)}, & i < n \\ Q_{eff(n)}, & i = n \end{cases} \tag{6}$$

which simplifies Equation (5) into

$$OTE_{ser} = \frac{\min\limits_{i=1,2,\ldots,n} \left\{ A_{(i)} \times R_{act(i)} \times Q_{eff(i..n)} \right\}}{\min\limits_{i=1,2,\ldots,n} \left\{ R_{th(i)} \right\}} \tag{7}$$

In the following we focus on the type of production lines, such as discussed in the context of the decribed project. It is closely related to (rigid) production line with two specific features:

1. Failure of any machine will result in failure of the entire line.

2. Quality is not measured in individual machines, but only at the end of the line.

In calculating OTE, this means that the availability factor $A_{(i)} = A$ is the the same for all of the machines, and the quality factor of the remainder of the line can be replaced by the total quality factor: $Q_{eff(i..n)} = Q_{eff}$. The equation for calculating OTE thus simplifies to

$$OTE_{ser} = A \times \frac{\min\limits_{i=1,2,\ldots,n} \left\{ R_{act(i)} \right\}}{\min\limits_{i=1,2,\ldots,n} \left\{ R_{th(i)} \right\}} \times Q_{eff} \tag{8}$$

We can see that the availability of machines on the line affects the efficiency of the line, but since the loss of any equipment will result in failure of the entire line, we can only discuss the availability of the line as a whole. Availability is therefore important especially in terms of monitoring and analysis, but it is less useful in the on-line production management. The same goes for quality. Although it is an important aspect we can only discuss the quality of the entire line. From the perspective of the individual machines we therefore focus only on performance.

A simplified model describes the performance of the line depending on the performance of machines on the line. If we compare equations (4) and (8), we see that the result is expected. The described derivation actually only shows that OTE of a rigid production line can be seen as OEE of a single machine. But the calculation in accordance to the equation (8) has a significant advantage in terms of operational management of production - because we deal with the capacity of individual machines a detailed analysis of bottlenecks is possible, and in particular we can predict the performance depending on the flow of products through the line.

The key variable that determines performance is the speed of the line. Since we deal with a closely linked production line with no intermediate storage buffers, the capacity is proportional to the line speed.

If the processing rate is observed by the number of produced panels, $R_{(i)} = V_{(i)}/L$, where $V_{(i)}$ is the speed of the line at the point of the i-th machine and L is the length of the panel. Assuming a strip shape of the intermediate product, which is cut to panels of variable lengths, a more appropriate expression of the capacity is based on the produced panel surface. In this case, the performance is expressed as $R_{(i)} = V_{(i)} \cdot D$, where D is the lateral dimension (width) of the strip. In both cases, the ratio of actual and theoretical performance is expressed by the ratio of actual and theoretical speeds, which is in the case of closely linked machines determined by the minimum of the maximum attainable speed for each machine:

$$OTE_{ser} = A \times \frac{\min\limits_{i=1,2,\ldots,n}\left\{V_{act(i)}\right\}}{\min\limits_{i=1,2,\ldots,n}\left\{V_{max(i)}\right\}} \times Q_{eff} \qquad (9)$$
$$= A \times \frac{V_{act}(t)}{V_{max}} \times Q_{eff}$$

The theoretical maximum line speed V_{max} is a technological parameter and is constant at the given operating conditions. For short-term forecasting of overall production efficiency changes it is necessary to build a prediction model of the actual line speed $V_{act}(t)$.

Model of the actual line speed is built from submodels of attainable speed for each machine, by considering the dependence of the speed of the product on the machine:

$$V_{act}(t) = \min_{i=1,\ldots,n}\left\{V_{act(i)}(I_{(i)}(t))\right\} \qquad (10)$$

Here V_{act} is the actual achievable line speed, $V_{act(i)}$ is attainable speed for the machine i, $I_{(i)}(t)$ is a set of parameters linked to the product, which is produced at machine i at time t, and n is the number of machines in the line.

Due to flow of products, which have different parameters, the actual speed varies. Its prediction may be based on products that are already on the line and the known schedule of products that will be produced within the prediction horizon.

For easier computation both the products that are already under processing on the line, as well as the planned items for a chosen horizon, are put in a common queue, called stack S. Product is in line until the last technological operation on the line is carried out, then it is removed from the stack. Products in the stack are indexed backwards from the end of the line. The product is presented in accordance with a set of parameters which affect the speed. The first record in the stack $S(1)$ thus represents the parameters of the product at the last machine of the line, $I_{(n)} = S(1)$.

Also for other products in the stack the mapping

$$F : i \mapsto s; \quad i = 1, \ldots, n, \quad s = 1, \ldots, k \qquad (11)$$

can be defined connecting the machine i and the index s of the product in the stack, k is the number of products in the stack. The searched parameters of the product at the machine are obtained as $I_{(i)} = S(F(i))$. Here $F(n) = 1$.

Since the dimensions of the product, as well as other parameters may have an impact on the number of products between the machines, the mapping depends on all the products, which are in the processing between the machine and the end of the line at time t. In addition, the mapping F is time-dependent, because the composition of the product mix between the machines changes.

Specifically, the mapping has to be determined for each specific type of production line. If for example the final products are obtained by cutting the intermediate product stripe into panels, then the relative positions of machines $X_{(i)}$ to the end of the line can be determined. Mapping F is then determined by the sum of lengths of the products in in the stack:

$$L(m) = \sum_{j=1}^{m} L_j \qquad (12)$$

$$F(i) = \underset{m}{\arg\min}\ L(m) \qquad (13)$$
$$s.t. L(m) > X_{(i)}$$

At the known product mix in the stack for the selected time t the actually achievable speed of the line at the moment can be calculated as:

$$V_{act}(t) = \min_{i=1,\ldots,n}\left\{V_{act(i)}(S(m))\right\} \qquad (14)$$

Consequently, for a period of planned production, which corresponds to the length of the stack, the prediction of overall production efficiency indicator OTE can be calculated by inserting (14) into (9).

$$OTE_{ser} = A \times \frac{\min\limits_{i=1,\ldots,n}\left\{V_{act(i)}(S(m))\right\}}{V_{max}} \times Q_{eff} \qquad (15)$$

Dependencies of the attainable speed $V_{act}(i)$ on the currently processed item parameters have to be empirically determined for every machine i.

5. IMPLEMENTATION AND RESULTS

To asses the usability of the proposed model the speed dependencies were first identified through the interviews with production operators and critical evaluation of results. An example of the determined dependency for the wool preparation machine is shown in Figure 3. The speed depends on the product type, the thickness and the lateral dimension of the product.

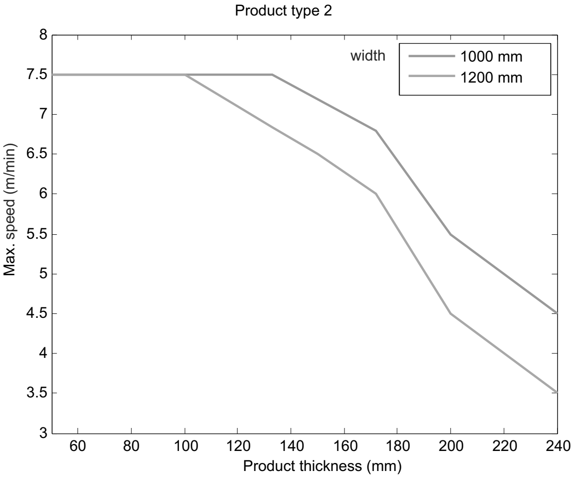


Figure 3: Speed of mineral wool preparation

Next the dependencies were coded into SQL procedures, which collect the data and perform all the necessary calculations. The overall structure of the solution is shown in Figure 4.

The relation of the actual speed to the maximum attainable speed, which is a direct measure of the line performance, is shown through a Web interface. An example

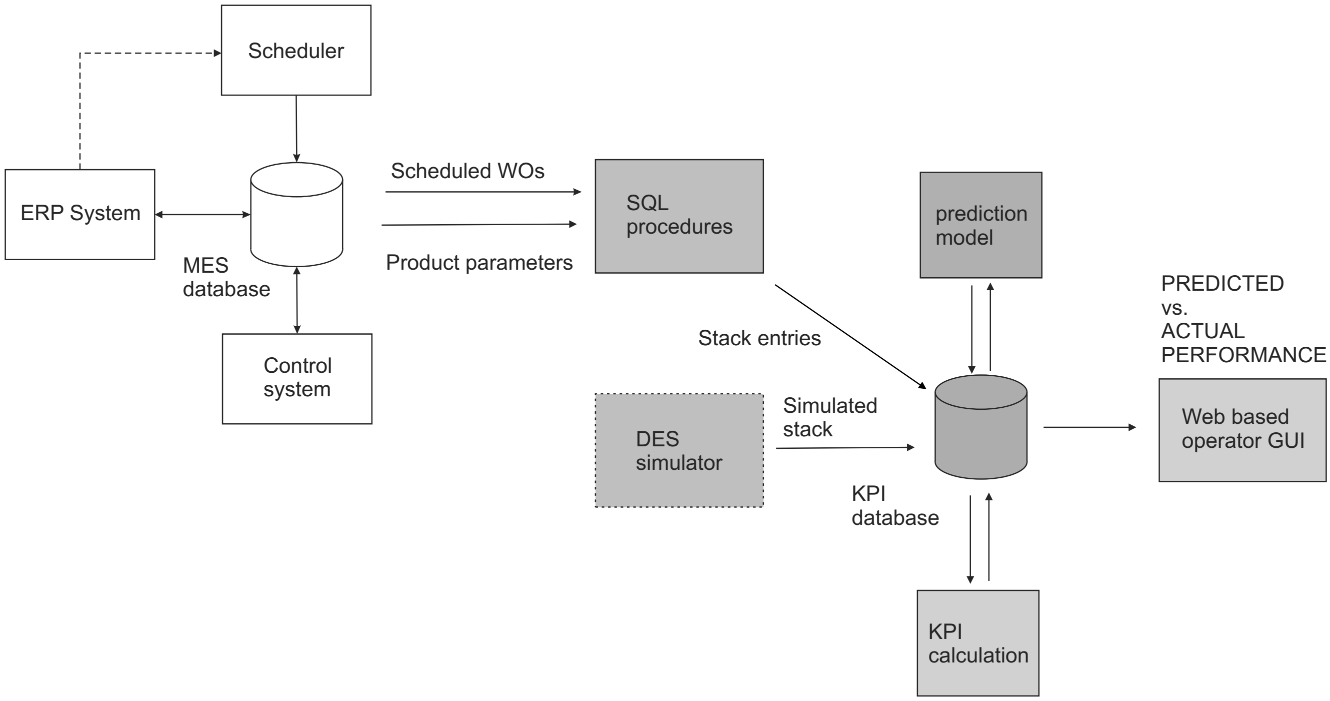
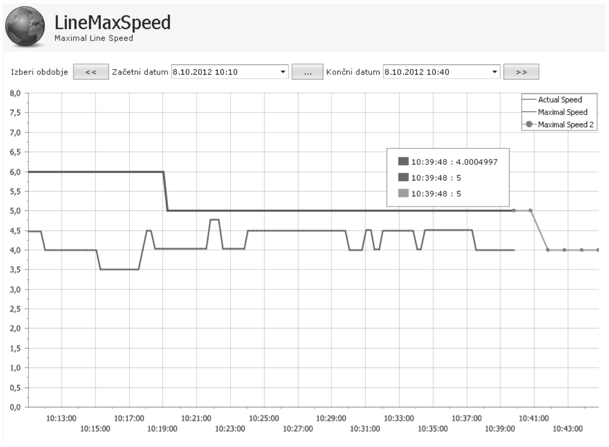


Figure 4: Solution structure


LineMaxSpeed
Maximal Line Speed

Izberi obdobje << Začetni datum 8.10.2012 10:10 ... Končni datum 8.10.2012 10:40 >>

10:39:48 : 4.0004997
10:39:48 : 5
10:39:48 : 5

Figure 5: Actual vs. predicted line speed in the operator interface

is shown in Figure 5, where also a few minute prediction of the expected speed is shown.

The actual and predicted line speeds are archived for analysis purposes. Historical values for a day of production are shown in Figure 6. Note that production started at 6 a.m. while the line was stopped next day at 2 a.m. It can be observed that we have a relatively good matching of the actual and predicted line performance except at downtime. The model can not predict unplanned interruptions, it only predicts temporarily stops when the number of panels at stacking machine exceeds certain threshold. The matching of the two speeds is better shown in Figure 7 where only non-zero speeds are shown.

Its is also interesting to analyze the performance of individual machines on the line, in particular to determine the machine that limits the performance of the line at the given moment, i.e., the bottleneck. This can be analyzed employing the influential variable selection methods that are used in data driven model building. Application of the software tool build to support the Holistic Production Control (HPC) concept (Glavan et al. 2013) results in Figures 8 and 9. Various variable selection methods were used: Linear Correlation, Partial correlation (with forward selection approach), PLS (variable importance in projection – PLS VIP), Non-Negative Garrote, LASSO, DMS search (pareto search of minimum error of linear model as objective function) (Glavan et al. 2013).

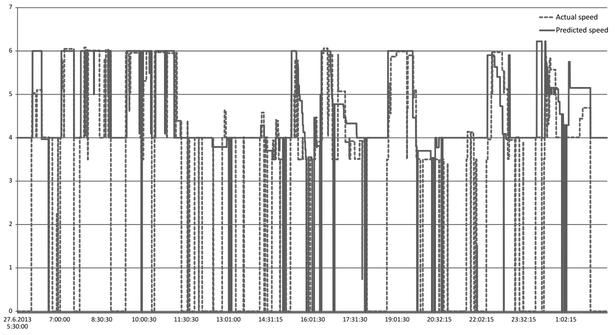


Figure 6: Actual vs. predicted speed for a day of production

It is clear that gluing process represents the bottleneck that dominantly restricts the production speed. Since the gluing process directly affects the quality of the product, operators are prone to further decrease the speed to avoid the quality drop. At the same time the attainable speed model of the gluing process is rather coarse, because no measurements of the glue distribution are available. The

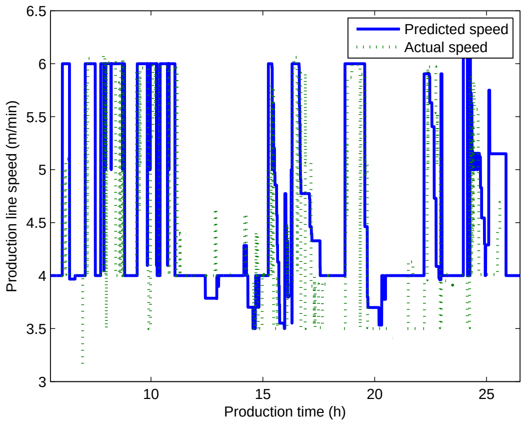
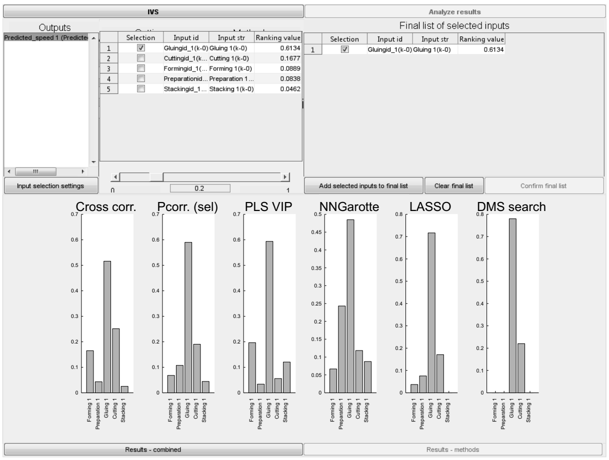


Figure 7: Actual vs. predicted speed without downtime



	Selection	Input id	Input str	Ranking value
1	☑	Gluingid_1(k-0)	Gluing 1(k-0)	0.6134
2	□	Cuttingid_1(k...	Cutting 1(k-0)	0.1677
3	□	Formingid_1(...	Forming 1(k-0)	0.0889
4	□	Preparationid...	Preparation 1 ...	0.0838
5	□	Stackingid_1 ...	Stacking 1(k-0)	0.0462

Final list of selected inputs

	Selection	Input id	Input str	Ranking value
1	☑	Gluingid_1(k-0)	Gluing 1(k-0)	0.6134

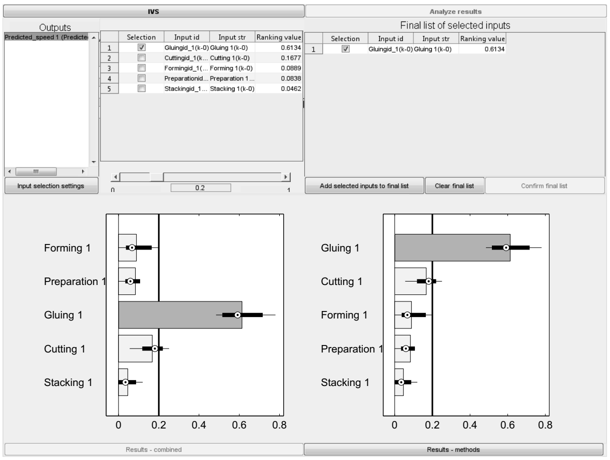
Figure 8: Analysis of the bottleneck by input variable selection methods

improvement of the gluing model therefore remains one of the main open challenges.

Another open issue is rising the operator confidence into the model. E.g., Figure 10 shows another part of the speed archive, where the operator obviously followed the model recommendation with a large safety margin. This should be overcome by the gradual improvements of the model and positive operator experience.

6. CONCLUSIONS

The presented results indicate that the derived model can be used to predict the production line performance. The operators can use the prediction to adjust the actual conditions on the line as close as possible to the optimal ones. This is particularly important for new operators, which can faster gain the necessary experience. Nevertheless, the analysis of history logs show that also experienced operators often drive the line at a lower speed, which decreases the line performance. The main reason is the potential drop of quality if the product moves through gluing machine too fast. This way the obtained information from the prediction model can be used to improve the production operation and manufacturing execution performance while maintaining product quality. The decision-supporting functionality can be even increased by implementing a link to discrete-event simulator, which is one of the issues for the future work.



	Selection	Input id	Input str	Ranking value
1	☑	Gluingid_1(k-0)	Gluing 1(k-0)	0.6134
2	□	Cuttingid_1(k...	Cutting 1(k-0)	0.1677
3	□	Formingid_1(...	Forming 1(k-0)	0.0889
4	□	Preparationid...	Preparation 1 ...	0.0838
5	□	Stackingid_1 ...	Stacking 1(k-0)	0.0462

Final list of selected inputs

	Selection	Input id	Input str	Ranking value
1	☑	Gluingid_1(k-0)	Gluing 1(k-0)	0.6134

Figure 9: Results of the bottleneck analysis

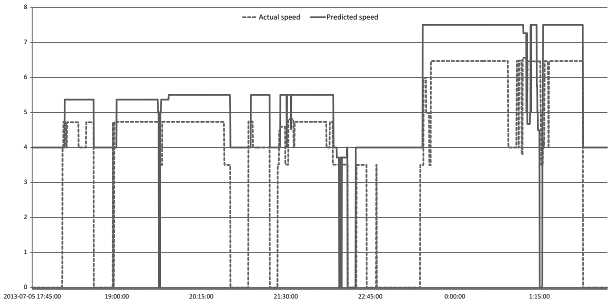


Figure 10: Actual vs. predicted speed with a conservative operator

ACKNOWLEDGMENTS

The presented work has been partially performed within the Competence Centre for Advanced Control Technologies, which is partly financed by the Republic of Slovenia, Ministry of Education, Science and Sport, and European Union, European Regional Development Fund. The support is gratefully acknowledged.

REFERENCES

Berrah, L. and Clivillé, V., 2008. Towards a quantitative performance measurement model in a buyer-supplier relationship context, *in* V. Kordić (ed.), *Supply Chain, Theory and Applications*, I-Tech Education and Publishing, Vienna, Austria.

Berrah, L. and Foulloy, L., 2013. Towards a unified descriptive framework for industrial objective declaration and performance measurement, *Computers in Industry*, 64 (6), 650 – 662.

Clivillé, V., Berrah, L. and Mauris, G., 2007. Quantitative expression and aggregation of performance measurements based on the MACBETH multi-criteria method, *International Journal of Production Economics*, 105 (1), 171–189.

Curcio, D., Longo, F. and Mirabelli, G., 2007. Manufacturing process management using a flexible modeling and simulation approach, *Proceedings of the 39th conference on Winter simulation*, WSC '07, IEEE Press, Piscataway, NJ, USA, pp. 1594–1600.

De Toni, A. and Tonchia, S., 2001. Performance measurement systems - Models, characteristics and measures, *International Journal of Operations & Production Management*, 21 (1/2), 46 – 71.

del Rio Vilas, D., Longo, F. and Monteil, N. R., 2013. A general framework for the manufacturing workstation design optimization: a combined ergonomic and operational approach, *Simulation*, 89 (3), 306–329.

Ghalayini, A. M., Noble, J. S. and Crowe, T. J., 1997. An integrated dynamic performance measurement system for improving manufacturing competitiveness, *Int. J. Production Economics*, 48 (3), 207–225.

Glavan, M., Gradišar, D., Atanasijević-Kunc, M., Strmčnik, S. and Mušič, G., 2013. Input variable selection for model-based production control and optimisation, *International Journal of Advanced Manufacturing Technology*, in press.

IEC, 2007. *Enterprise-Control System Integration – Part 3: Activity models of manufacturing operations management*, IEC 62264-3, International Electrotechnical Commission, Geneva.

Jonsson, P. and Lesshammar, M., 1999. Evaluation and improvement of manufacturing performance measurement systems - the role of OEE, *International Journal of Operations & Production Management*, 19 (1), 55–78.

Kaplan, R. S. and Norton, D. P., 1992. The balanced scorecard - measures that drive performance, *Harvard Business Review*, 70 (1), 71–79.

Latorre, J. I., Jiménez, E. and Pérez, M., 2013. The optimization problem based on alternatives aggregation Petri nets as models for industrial discrete event systems, *Simulation*, 89 (3), 346–361.

Mathur, A., Dangayach, G. S., Mittal, M. L. and Sharma, M. K., 2011. Performance measurement in automated manufacturing, *Measuring Business Excellence*, 15 (1), 77–91.

Muchiri, P. and Pintelon, L., 2008. Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion, *International Journal of Production Research*, 46 (13), 3517–3535.

Mujica, M., Piera, M. A. and Narciso, M., 2010. Revisiting state space exploration of timed coloured Petri net models to optimize manufacturing system's performance, *Simulation Modelling Practice and Theory*, 18 (9), 1225–1241.

Muthiah, K. M. N. and Huang, S. H., 2007. Overall throughput effectiveness (OTE) metric for factory-level performance monitoring and bottleneck detection, *International Journal of Production Research*, 45 (20), 4753–4769.

Nakajima, S., 1988. *Introduction to TPM: total productive maintenance*, Productivity Press, Cambridge.

Neely, A., Gregory, M. and Platts, K., 1995. Performance measurement system design - a literature review and research agenda, *International Journal of Operations & Production Management*, 15 (4), 80–116.

Nudurupati, S. S., Bititci, U. S., Kumar, V. and Chan, F. T. S., 2011. State of the art literature review on performance measurement, *Computers & Industrial Engineering*, 60 (2), 279–290.

Suwignjo, P., Bititci, U. S. and Carrie, A. S., 2000. Quantitative models for performance measurement system, *Int. J. Production Economics*, 64 (1-3), 231–241.

AUTHORS' BIOGRAPHIES

TOMI ZEBIČ received B.Sc. degree in electrical engineering and M.Sc. degree in computer science from the University of Ljubljana, Slovenia in 1985 and 1990, respectively. He is high school and college professor. As a project manager he works for Metronik engineering company. His research interests are in planning, scheduling, and manufacturing execution systems.

BOŠTJAN HAUPTMAN received B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Slovenia in 1996, 2000, and 2004, respectively. He works at the Trimo Trebnje company as IT project manager. His research interests are in planning, scheduling, SAP (PP, PS, MM and QM module), and MES systems in production and automated warehouses. His company Web page can be found at `http://www.trimo.eu`.

PETER ROGELJ received B.Sc. degree in electrical engineering from the University of Ljubljana, Slovenia in 2006. He is technology developer at Trimo Trebnje company in Slovenia. His research interests are in developing MES and Scada HMI systems for automatic productions and warehouses. His company Web page can be found at `http://www.trimo.eu`.

GAŠPER MUŠIČ received B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Slovenia in 1992, 1995, and 1998, respectively. He is Associate Professor at the Faculty of Electrical Engineering, University of Ljubljana. His research interests are in discrete event and hybrid dynamical systems, supervisory control, planning, scheduling, and industrial informatics. His Web page can be found at `http://msc.fe.uni-lj.si/Staff.asp`.

ADVANCED RIVER FLOOD FORECASTING AND SIMULATION

Galina V. Merkuryeva[(a)], Yuri A. Merkuryev [(b)]

[(a) (b)]Riga Technical University, Kalku Street 1, LV-1658, Riga, Latvia
[(a)]Galina.Merkurjeva@rtu.lv, [(b)] Jurijs.Merkurjevs@rtu.lv

ABSTRACT

The paper presents the state-of-the-art in flood forecasting and simulation applied to river flood analysis and risk prediction. Different water flow forecasting and river simulation models are analysed. An advanced river flood forecasting and modelling approach developed within the ongoing project INFROM is presented. It provides an integrated procedure for river flow forecasting and simulation advanced by integrating different models for improving flood risk outputs prediction. Demonstration cases in river flow forecasting and floods modelling are given.

Keywords: water flow forecasting, river simulation, integrated forecasting and simulation, flood risk outputs analysis

1. INTRODUCTION

Flooding is one of the natural disasters which often cause significant economic losses, human and social tragedies. Due to this, flood forecasting and its effective control is always a huge challenge for governments and local authorities (Chiang et al. 2010). Forecasts of river flow may be developed in a short-term, over periods of few hours or a few days, and in a long-term, up to nine months (Georgakakos and Krzysztofowicz 2001). An efficient flood alarm system based on a short-term flow forecasting may significantly improve public safety, mitigate social damages and reduce floods economical loss.

Flooding may be caused by several reasons such as snow and ice melting in rivers in the spring causing freshet; heavy raining in the neighbouring areas, and wind-generated waves in the areas along the coast and river estuaries. In Latvia, springtime ice drifting and congestion can cause a rapid rise in water levels of Daugava, Gauja, Venta, Dubna, Lielupe, Ogre and Barta rivers. The risk of flooding along the Daugava River is relatively high, and in most flood sensitive areas (e.g., in Daugavpils district) it may occur even twice a year. Floods in Riga and Jurmala districts located in the deltas of Daugava and Lielupe rivers and on the Gulf of Riga coast may be caused by the west wind during 2-3 days with speed greater than 20 m/s following by winds in the north-west direction. As a result, the reverse water flow from the Gulf of Riga into Daugava and Lielupe rivers may significantly rise to floods levels in these areas.

Flood forecasting and modelling is undoubtedly a challenging field of operational hydrology, and a huge literature has been written in that area in recent years. A flow forecast is an asset for flood risk management, reducing damage and protecting environment (Tucci and Collischonn 2010). Reliable flow forecasting may present an important basis for efficient real-time flood management including floods monitoring, control and warning. The integration of monitoring, modelling and management becomes important in construction of alert systems. Nowadays, application of remote sensing and GIS software that integrates data management with forecast modelling tools becomes a good practice (Pradhan 2009, Irimescu et al. 2010, Skotner et al. 2013). Additionally, different flooding scenario may be simulated based on the results of forecasting models to allow analysing river flood dynamics and evaluating their potential effects in the near future.

This paper provides the state-of-the-art in river flood forecasting and modelling as well as describes advanced river flood forecasting and simulation models developed within the ongoing research project INFROM “Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems”. Different water flow and flood forecasting techniques have been used and compared - traditional regression-based forecasting techniques, symbolic regression, cluster analysis of dynamic data, and identification of typical dynamic patterns. Among flood monitoring models, hydrodynamic and hydrological models were reviewed and compared. A procedure for integrated river flow forecasting and simulation has been developed and advanced by integrating different models and metamodels for improving flood risk outputs analysis.

The project itself addresses (Merkuryev at al. 2012) the problem of integrated monitoring and control of natural-technological systems based on analysis of heterogeneous data both from space and ground-based facilities and integration of different types of models (i.e., analytical, algorithmic, mixed) used to model behaviour of these systems.

2. STATE-OF-THE-ART

There are several models and systems that allow predicting flood risk outputs by remote sensing, GIS, hydraulic and hydrology modelling. In this paper, flood forecasting and simulation models and techniques which are used for river flow prediction and flood risk outputs generation are reviewed.

River flood monitoring and control requires measurement and notification of the water level, velocity, and precipitation. Input data for precipitation forecast are meteorological data and weather forecasts as the most important components of a flood forecasting and early warning system (Badila 2008, Crooks 2011). In practice, river flood forecasting is based on mining historical data and specific domain knowledge to deliver more accurate floods forecasts. Effective flood monitoring and control use space and ground-observed data received from satellites and terrestrial (meteorological, automatic raingauge, climatological) stations. These data may be represented as images, terrain information, and environmental information, i.e., soil type, drainage network, catchment area, rainfall, hydrology data, etc. Data representation and processing proven technologies and expertise are offered in (Astrium web site).

Besides, expert knowledge may be integrated into the flood risk assessment procedure, producing river flood scenarios to be simulated and measures for flood damage prevention or reduction. When risk outputs are calculated, decisions for preventive actions can be made based on flood risk maps, flood forecast maps, flood emergency response maps, and based on detection and monitoring for early warning mitigation and relief.

Hydrodynamic river flow processes might be represented by a variety of different models based on geological surroundings, for example, the conceptual HBR model (Irits 2005), ANN-based runoff predictors with a fuzzy classifier of the basin states (Corani and Guariso 2005), hydrodynamic deterministic models improved by uncertainty coping to produce the probabilistic hydrological forecast (ICPDR 2010), etc.

A conceptual model of the river may be described in different ways due to different scope of the model (Dharmasena 1997, Badilla 2009, Chiang et al. 2010). One of common simplifications of the hydrodynamic river flow processes is achieved by lumping of the processes in space and limiting the study area to the region affected by the flood control. Lumping of the processes in space is done by simulation of the water levels only at the relevant locations. These locations are required to be selected in upstream and downstream points of each hydraulic regulation structure and places along the river (Chiang et al. 2010).

Floods monitoring models may be classified as hydrodynamic and hydrological models. *Hydrodynamic models* describe and represent the motion of water flow using so called Navier-Stokes equations which describe the motion of fluid substances in physics.

Hydrological models are simplified conceptual representations of a part of the hydrologic cycle. Hence, they are considered as more suitable for water flow modelling in flood monitoring. Hydrological models used in the forecasts can be grouped as follows (Dharmasena 1996): 1) *stochastic hydrological black-box models* that define input-output relations based on stochastic data and use mathematical and statistical concepts to link a certain input to the model output; and 2) *conceptual or process-based models* that represent the physical processes observed in the real world. While black-box models are empirical models and use mathematical equations with no regards to the system physics, conceptual models apply hydrological concepts to simulate the basin or river behaviour.

Stochastic hydrological models are more popular in literature due to their simplicity. Among them, linear perturbation models, HEC models and neural networks-based flood forecasts systems are considered to be the most efficient tools in practice (Dharmasena 1997). In particular, linear perturbation models assume that the perturbation from the smoothened seasonal input rainfall and that of discharge are linearly related. However, the rainfall-runoff relationship has been recognized to be nonlinear, and coupling fuzzy modelling and neural networks for flood forecasting that do not assume input-output model relationship to be linear was suggested in (Corani and Guariso 2005). In The Hydrologic Engineering Center (HEC) models are numerical models for simulation of hydrologic and hydraulic process. HEC models solve the Saint-Venant equations using the finite-element method. The primary surface water hydrology model is HEC-1 Flood Hydrograph Package which can simulate precipitation-runoff process in a wide variety of river basins. The predictive power of HEC models is also discussed in (Horritt and Bates 2002; Chiang 2010)**.**

Conceptual models usually have two components (Tucci 2006), i.e. a rainfall-runoff module which transfer rainfall into runoff through water balance in the river hydrological components, and a routing module which simulates the river flow. Conceptual models such as Soil Moisture Accounting and Routing (SMAAR) model, NAM and Xinanjiang models which have a number of parameters 5, 13, 15, correspondingly, were applied to seven river basins in Sri Lanka (Dharmasena 1997). Data requirements for modelling were formulated, and calibration and validation of models was done. The results obtained demonstrated applicability of all models, but NAM and Xinanjiang models were found more appropriate as flood peaks were represented by separate parameters in these models.

There are several major river modelling software tools such as HEC-RAS, LISFLOOD-FP and TELEMAC-2D. HEC River Analysis System (HEC-RAS) allows performing one-dimensional steady flow, unsteady flow, and water temperature modelling. The HEC-RAS model solves the full 1D Saint Venant equations for unsteady open channel flow. LISFLOOD-FP is a raster-based inundation model specifically developed to take advantage of high resolution topographic data sets (Bates and De Roo 2000) and adopted to 2D approach. TELEMAC-2D is a powerful and open environment used to simulate free-surface flows in two dimensions of a horizontal space. At each point of the mesh, the program calculates the depth of water and the two velocity components. The model solved 2D shallow water (also known as Saint-Venant equations or depth average) equations for free surface flow using the finite-element or finite-volume method and a computation mesh of triangular elements (see http://www.opentelemac.org).

The predictive performance of three models is analysed in (Horritt and Bates 2002). The different predictive performances of the models stem from their different responses to changes in friction parameterisation. Also, the performance of the LISFLOOD-FP model is dependent on the calibration data used. Nevertheless, performance of 1D HEC-RAS model gives good results which are comparable with ones received from more sophisticated 2D approaches adopted by LISFLOOD-FP and TELEMAC-2D. Also, HEC-RAS models allow building long-term flood forecasts, but require large input datasets. Finally, these models reflect moving in recent years from a 1D approach (represented by the US Army Corps of Engineers HEC-RAS model) towards 2D finite element (TELEMAC-2D developed by Electricite´ de France) and raster-based (LISFLOOD-FP) models.

3. ADVANCED APPROACH

River flow forecasting and simulation is advanced by integrating different models for improving flood risk outputs prediction including input data clustering, digital maps of the relief, data crowd sourcing technology, symbolic regression-based short-term forecasting models, different hydrological models for modelling water flows in short-term, mid-term and long-term forecasts, computer simulation models for simulating behaviour of the river and its visualisation, techniques for flooding scenario generation and comparison. Real-time food forecasting and monitoring is based on processing data received from both space and ground based information sources.

Clustering of dynamic historical data is introduced which allows identifying typical dynamic flooding patterns in the real-life situations. A symbolic regression-based forecasting model is integrated for river flow short-term forecasting and monitoring in a specific real-life situation. Here, main challenges are a small number of input factors and a small set of flow measurements. For developing a symbolic regression-based forecasting model, genetic programming within HeuristicLab (Affenzeller et al. 2009, Wagner 2009) is used.

Hydrological models are advanced by realistic physical models that are derived from topological maps and represent geo information of the river and neighbouring areas. Additionally, different regression-based metamodels using river simulation results are introduced which allow performing sensitivity analysis of input factors influencing river water levels and flooding risk as well as improving output results received from the forecasting models. In the future, this approach will be extended by automatic generation and analysis of flooding scenarios for medium and long-term flood management operations.

4. DEMONSTRATION CASES

Two cases below were developed for river flood monitoring and forecasting in two Latvian districts and demonstrate applicability of the proposed approach.

The first demonstration case is developed for the Dubna River water flow modelling and flood areas modelling and simulation. Hydrological data from three hydrologic stations (water levels and flow direction) and topographic data from topographic maps are used as inputs to water flow simulation and developing a river physical model. The water level is measured as water height from the bottom of the river in millimetres, and flow direction in degrees, considering north direction as a zero degree. Geographic information is used to develop a realistic model of the river basin using information on depth of the river and specifying a sufficient amount of the river cross sections.

A simulation model prototype using HEC-RAS River modelling system software was built that models geometry of the Dubna River and simulates its flows. The graphical model of the river is shown in Figure 1 and contains information about 8 cross sections that defines all information required for calculations. The model is capable to simulate both steady and unsteady flows. Here, the flow is assumed to be unsteady as typically for areas with flooding chance.

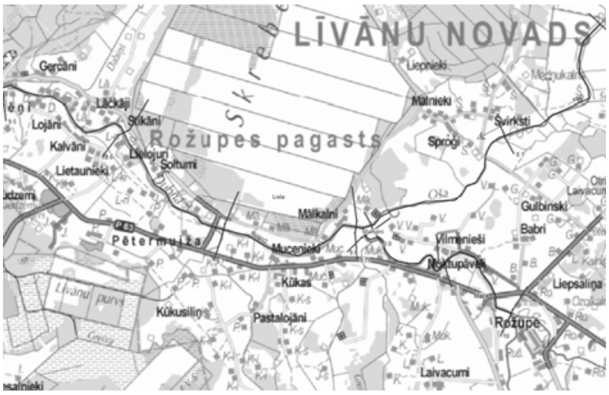
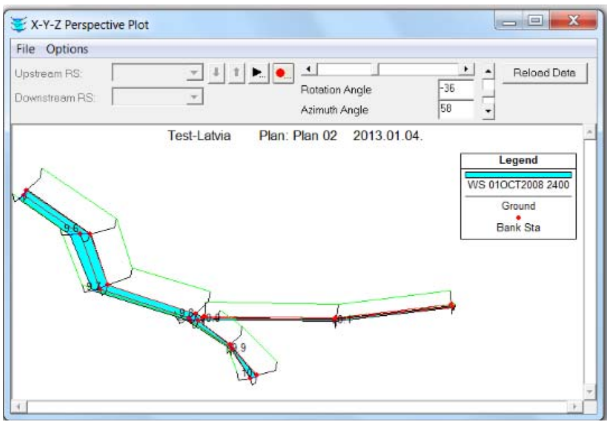


Figure 1: HEC-RAS based model of Dubna river



Rotation Angle -36
Azimuth Angle 58

Test-Latvia Plan: Plan 02 2013.01.04.

WS 01OCT2008 2400

Figure 2: River simulation model output visualisation

Numerical results of modelling are observed as a data set or visualised means (see Figure 2).

The results of modelling demonstrate possibilities of an intense river flow in late Autumn and Winter, however the level of water does not rise higher than river banks at the observed section of the river.

The second case (see Potryasaev et al. 2013) was developed for a short-term flood forecast using space-ground monitoring data of the Daugava River near Daugavpils city in Spring 2013. The forecast horizon was defined by a period of up to 12 hours. A digital map of the relief of the specified area and hydrological river characteristics were received and integrated into the models. To train the forecasting model, historical data from the Daugava River monitoring station near Daugavpils city were used. Several forecasting scenarios – by using linear and nonlinear regression models, and symbolic regression - were tested. For operational forecasting in a time step of an hour and predicting related flood territory, real-time data received from the hydrological station were integrated into the input dataset. Figure 3 illustrates applicability of the developed symbolic regression-based models for predicting the Daugava River flow and flood forecast. The forecasting accuracy of the river water flow was within 95 %.

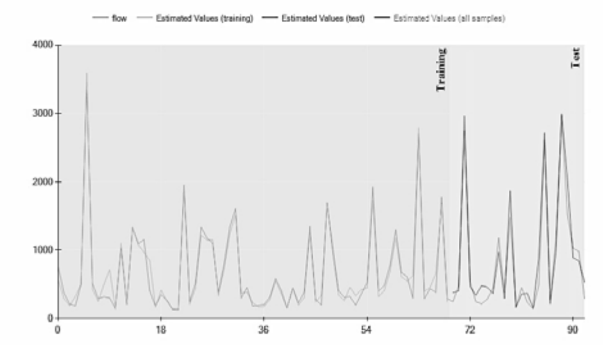


Figure 3: Empirical data versus model-based forecasting results

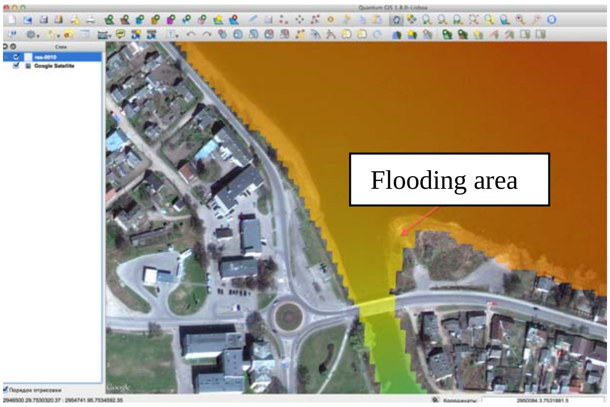


Figure 4: LISFLOOD model-based flooding area visualisation screen-short

A LISFLOOD model was developed to simulate water flows in the Daugava River and its floating routes. Calibration of the model has been performed based on satellite images and using data crowd sourcing through the geo-portal. As a result, the coincidence of flooding of significant objects (Fig. 4) has been received within 90%.

5. CONCLUSIONS

The review of the state-of-the-art in river flood flow forecasting and simulation allows defining the most efficient models and tools for water flows forecasting and river simulation. The river flood forecasting and simulation procedure proposed in the paper allows integrating capabilities of both forecasting and simulation techniques for advancing risk analysis of river floods.

ACKNOWLEDGMENTS

This presented research is supported by the project 2.1/ELRI -184/2011/14 “Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems» as a part of «Estonia-Latvia-Russia cross border cooperation Programme within European Neighborhood and Partnership instrument 2007-2013”.

The authors would like to acknowledge valuable support by Dr. Valentina Strauitmane and Dr. Guntis Strautmanis from RTU Daugavpils branch. RTU young researchers V. Bolshakov and M. Kornev are acknowledged for their contribution to the programming tasks.

REFERENCES

Affenzeller, M., Winkler, S., Wagner, S., Beham, A., 2009. Genetic *Algorithms and Genetic Programming: Modern Concepts and Practical Applications*. Chapman & Hall/CRC.

Badilla, Roy A., 2008. *Flood Modelling in Pasig-Marikina River Basin*, Master Thesis, International Institute for Geo-information science and Earth Observation Enschede, Netherlands, 73 p.

Bates, P.D., De Roo, A.P.J., 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, 236, 54–77.

Chiang, P.-K., Willems, P., Berlamont, J., 2010. A conceptual river model to support real-time flood control. *In: Demer,' River Flow 2010 - Dittrich, Koll, Aberle & Geisenhainer*, pp. 1407-1414.

Corani, G., Guariso, G. 2005. Coupling fuzzy modelling and neural networks for river flood prediction. *IEEE Transactions on Men, Systems and Cybernetics*, 3 (35), pp. 382-391.

Crooks, S. M., 2011 Flood Studies at the River Basin Scale: Case Study of the Thames at Kingston (UK), Centre for Ecology and Hydrology / Technical Report No. 53

Dharmasena, G.T., 1997. Application of mathematical models for flood. *Proc. Of the conference on Destructive Water: Water-Caused Natural*

Disasters, their Abatement and Control, IAHS Publ., No 239, pp. 225-235.

Galland, J., Goutal, N., Hervouet, J.-M., 1991. TELEMAC—a new numerical-model for solving shallow-water equations. *Advances in Water Resources,* 3 (14), pp. 138-148.

Georgakakos, K. P., Krzysztofowicz, R., 2001. Probabilistic and Ensemble Forecasting (Editorial). *Journal of Hydrology*, 249(1), pp.1-4.

Gouweleeuwa, B., Ticehurst, C., Dycea, P., Guerschmana, J., Van Dijka, Thew, P. An experimental satellite-based flood monitoring system for Sourthen Queensland, Australia

Horritt, M., Bates, P., 2002. Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology,* no. 268, pp. 87–99.

ICPDR Report, 2010, Assessment of Flood Monitoring and Forecasting in the Danube River Basin, International commission for protection of the Danube River, Flood Protection Expert Group, P. 19.

Irimescu, A., Craciunescu, V., Gheorghe, . S., Nertan, A., 2010. Remote Sensing and GIS Techniques for Flood Monitoring and Damage Assessment. Study Case in Romania, Proceedings of BALWOIS 2010, opp.1-10. Ohrid, Republic of Macedonia, May 25-29.

Iritz, L., Conceptual Modelling of Flood Flowin Central Vietnam, Department of Water Resources Engineering, School of Engineers, Lund University, Sweden, P. 10

Merkuryev, Y.A., Sokolov, B.V, Merkuryeva, G.V., 2012. Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems. *Proceedings of HMS 2012,* pp. 7-10, September 19-21, 2012, Vienna, Austria.

Pradhan, B., 2009. Effective Flood Monitoring System Using GIT Tools and Remote Sensing Data, Applied. *Geoinformatics for Society and Environment,* Stuttgart University of Applied Sciences, pp. 63-71.

Skotner, C., Klinting, A., Ammentorp, H.C., Hansen, F., Høst-Madsen, J., Lu, Q.M., Junshan, H., 2013. A tailored GIS-based forecasting system of Songhua river basin, China. Proceedings of Esri International user conference, San Diego, 6 p.

Tucci, C., Collischonn, W., 2006. Flood forecasting, WMO Bulletin 55 (3), pp. 179-184.

Wagner, S., 2009. *Heuristic Optimization Software Systems - Modeling of Heuristic Optimization Algorithms in the HeuristicLab Software Environment*. PhD Thesis. Institute for Formal Models and Verification, Johannes Kepler University Linz, Austria.

Zelentsov, V.A., Petuhova J.J., Potryasaev S.A., Rogachev S.A., 2013. Technology of operative automated prediction of flood during the spring floods. Proceedings of SPIIRAS, St-Petersburg, pp.

Astrium web site: www.astrium-geo.com/sg/3271-data-processing. Accessed 31.07.2013.

AUTHORS BIOGRAPHIES

GALINA V. MERKURYEVA is professor at the Institute of Information Technology, Department of Modelling and Simulation of Riga Technical University. Her professional interests and experiences are in the areas of discrete-event simulation, simulation metamodelling, simulation-based optimisation, decision support systems, logistics, production planning and control, supply chain simulation and management, and simulation-based training. She authors more than 170 publications, including 5 books.

YURI A. MERKURYEV is professor, head of the Department of Modelling and Simulation of Riga Technical University. He earned the Dr.sc.ing. degree in 1982 in systems identification, and Dr.habil.sc.ing. degree in 1997 in systems simulation, both from Riga Technical University, Latvia. His professional interests include methodology of discrete-event simulation, supply chain simulation and management, as well as education in the areas of simulation and logistics management. Professor Merkuryev is a corresponding member of the Latvian Academy of Sciences, president of Latvian Simulation Society, Board member of the Federation of European Simulation Societies (EUROSIM), senior member of the Society for Modelling and Simulation International (SCS), and Chartered Fellow of British Computer Society. He authors more than 300 publications, including 6 books and 6 textbooks. Professor Merkuryev is manager of the INFROM project.

SIMULATION AND OPTIMAL CONTROL OF HYBRID ELECTRIC VEHICLES

Guillermo Becerra[(a)], Alfonso Pantoja-Vazquez[(b)], Luis Alvarez-Icaza[(c)], Idalia Flores[(d)]

[(a),(b),(c)] Instituto de Ingenieria-Universidad Nacional Autonoma de Mexico
[(d)]Posgrado de Ingenierıa-Universidad Nacional Autonoma de Mexico
Coyoacan D. F. 04510, Mexico

[(a)]guillermobec@gmail.com, [(b)]apantojav@iingen.unam.mx, [(c)]alvar@pumas.iingen.unam.mx,
[(d)]idalia@unam.mx

ABSTRACT

New strategies for controlling the power split in hybrid electric vehicles (HEV) are described. The strategies focus in a planetary gear system, where kinematic and dynamic constraints must be fulfillied. The aim is to satisfy driver demands and to reduce fuel consumption. Two strategies are presented, one inspired on optimal control and the other derived from Pontryagin's Minimum Principle. It is shown that, under appropriate choice of weighting parameters in the cost function of the Hamiltonian, both strategies are similar. The resultant power flow control is continuous and uses the internal combustion engine with the maximum efficiency possible. The main advantages are the low computational cost, when compared to other optimization based approaches, and the easiness to tune. The strategy is tested by simulations using a mathematical model of a power train of a hybrid diesel-electric bus subject to the power demands of representative urban area driving cycles. The main elements of the vehicle, internal combustion engine (ICE), battery state of charge (soc), electric machine (EM) and vehicle inertia are simulated with high order models. Simulation results indicate that both strategies achieves small speed tracking errors and attain good fuel consumption reduction levels.

Keywords: Optimal control, Pontryagin's minimum principle, simulation, hybrid electric vehicles, internal combustion engine, electric machine, fuel consumption.

1. INTRODUCTION

Optimal power control on hybrid electric vehicles (HEV) is an important topic for power management. HEV may have different architectures that require the use of diverse energy management strategies. The main architectures, as presented in (John. M. Miller 2006), are series, parallel or series-parallel. A comparison of the architectures is presented in (Ehsani, Gao, and Miller 2007).

Power distribution in HEV can be performed by the use of different controllers, as described in the comparative study of supervisory control strategies for HEV presented in (Pisu and Rizzoni 2007). Rule based approaches can use heuristic, fuzzy logic, neural networks, etc. Examples of these techniques are (Xiong, Zhang, and Yin 2009), that proposes a fuzzy logic control for energy management and (Xiong and Yin 2009), that presents a fuzzy logic controller for energy management of a series-parallel hybrid electric bus with ISG.

There are also power flow control strategies based on optimization, like those revised in (Pisu and Rizzoni 2007). They are normally not implemented in real time, only proved in simulation and their off-line optimization results are used with a look-up table.

(Delprat, Lauber, Guerra, and Rimaux 2004) propose the control of parallel hybrid power train that splits the power between the engine and electric motor in order to minimize the fuel consumption. This strategy optimize the fuel consumption considering the torque engine and the gear ratio.

(Musardo, Rizzoni, and Sataccia 2005) present the Adaptive Equivalent Consumption Minimization Strategy (A-ECMS), which is an algorithm for hybrid electric vehicles that attempts to minimize the vehicle fuel consumption using an equivalence between fuel energy and electric energy. To prove its effectiveness, A-ECMS strategy is compared with Dynamic Programming (DP) and a non adaptive ECMS (Paganelli, Guerra, Delprat, Santin, Delhom, and Combes 2000). (Koot, Kessels, de Jager, Heemels, van den Bosch, and Steinbuch 2005) establish energy management strategies for HEV using dynamic programming and quadratic programming with Model Predictive Control (MPC), to minimize fuel consumption.

(Kim, Cha, and Peng 2011) reported a optimal control of parallel-HEV based on Pontryagin's minimum principle (PMP) that takes into account the state of charge and fuel consumption.They compare the strategy with dynamic programming and ECMS. In (Zou, Teng, Fengchun, and Peng 2013) they compared Pontryagin's minimum principle (PMP) with dynamic programming (DP), finding that the simulation time is significantly lower in PMP than DP. This is important for real time for implementation.

In this paper new strategies to control the power flow in a parallel power train HEV are presented. In this configuration, shown in Fig. 1, the internal combustion engine (ICE) and the electric machine (EM) can directly supply their torque to the driven wheels through a planetary gear system[1].

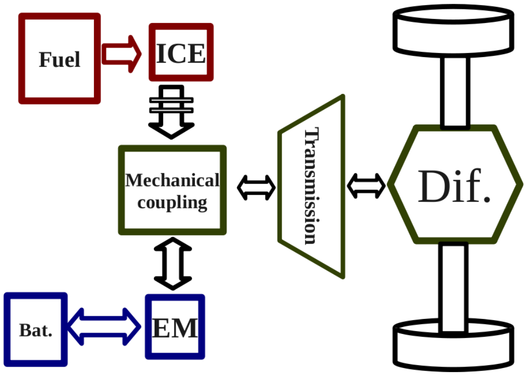


Fig. 1. Parallel hybrid electric vehicle power-train.

The design of the strategies recognizes, as is also pointed out in (Musardo, Rizzoni, and Sataccia 2005), that optimization based solutions to HEV power flow control are very difficult to implement in real-time. Moreover, their results can not be as effective when real driving conditions differ from those used in the optimization problem solution. A similar problem occurs when the uncertainty in the models is considered.

In the first strategy presented this paper, a local criteria is used based on the kinematic and dynamic constrains at the planetary gear system, that must be satisfied when distributing the power demanded by the vehicle between the ICE and EM, and in maximum efficiency curves for the ICE. There are only two pairs of parameters to tune that, as it will be shown later, have a similar behavior for all driving cycles employed in the simulations.

The second strategy is based on PMP and use the errors in the state of charge soc, fuel consumption m_f and power demanded by the user P_d, with respect to some reference values, as system states. Electric power P_{bat} and engine power P_{mci} are system inputs.

To test the developed strategy, simulations of a mathematical model of the main components of the hybrid power-train which includes the ICE and EM, clutch, planetary gear system and battery were performed. The strategy was tested using three standard driving cycles for a bus in Mexico City.

The rest of paper is organized as follows. Section 2 presents the models for simulation of the vehicle subsystems, section 3 describes details the strategies for power flow control. Simulation results are presented in the section 4 while section 5 presents the conclusions and directions for future work.

2. HEV MODELING

2.1. Internal combustion engine model

The model is taken from (Outbib, Dovifaaz, Rachid, and Ouladsine 2002). It is assumed that the air entering the intake manifold follows the ideal gas law and that the manifold temperature varies slowly with respect to pressure and engine speed. The model is described by

$$\begin{aligned}\frac{d\omega_{ice}}{dt} &= \frac{h_1}{\omega_{ice}}\dot{m}_f + h_2 p_a + \frac{h_3}{\omega_{ice}}P_b + \frac{h_4}{\dot{m}_f} \quad (1)\\ \frac{dp_a}{dt} &= h_5\dot{m}_{ai} - h_6\omega_{ice}p_a\end{aligned}$$

and the efficiency is

$$\eta_{ice} = \frac{P_{ice}}{\dot{m}_f p_{th}} \quad (2)$$

where ω_{ice} the engine speed, $\dot{m}_f$ the fuel flow rate used as control signal, p_a the intake manifold pressure, $\dot{m}_{ai}$ air flow entering the manifold, P_b the total brake power, P_{ice} the output power and p_{th} the fuel heating value. Terms h_j are constants determined in the model of (Outbib, Dovifaaz, Rachid, and Ouladsine 2002).

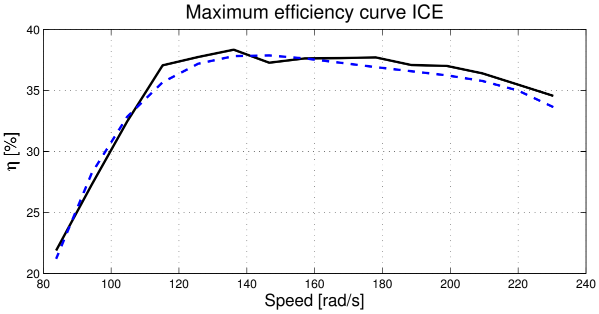


Fig. 2. Efficiency curve of the ICE Diesel

Fig. 2 shown an example of maximum efficiency curve in terms of the engine velocity. One key assumption in Eq. (2) is that air-fuel ratio can be controlled independently of ICE velocity..

2.2. Battery model

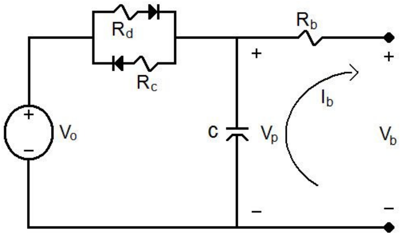


Fig. 3. Battery circuit

In the HEV, batteries are used as a temporary energy storage that helps saving fuel and reducing emissions. The state of charge of the battery (soc) is defined as a measure of the amount of electrical energy stored in it. It is analogous to the fuel gauge in the tank. Its dynamics is given by

$$\dot{soc}(t) = -\frac{P_b}{V_b Q_t} \quad (3)$$

[1] More details about HEV architectures can be found, for example, in (Ehsani, Gao, and Miller 2007)

where $P_b(t)$ is the power, V_b the voltage and Q_t denoting the total charge the battery can store.

The circuit model shown in Fig. 3, contains elements for discharging and charging mode.

2.3. Electrical machine model

The EM is an induction motor that can operate as motor or generator. When operating as motor, it draws power from the battery and the output torque drives the wheels, in possible combination with the ICE torque. Functioning as generator, it can recover kinetic energy from regenerative braking, or take energy from the ICE for battery recharging. Although the model obtained from (Peresada, Tilli, and Tonielli 2004) and used in simulations is fifth order, for the power split strategies it suffices with the relation between output power P_{em} and input power P_{bat} given by

$$P_{em} = \eta_{bm} P_{bat} \tag{4}$$

where η_{bm} is battery and EM efficiency.

2.4. Planetary gear system

The coupling of the power sources to traction is accomplished through a planetary gear system (PGS). Fig. 4 shows a schematic of this mechanical device. The ICE is connected through a clutch-brake to the sun gear of the PGS, the EM is connected to the ring gear and the wheels are connected to the carrier gear (Ambarisha and Parcker 2007),(Szumanowski, Yuhua, and Pi´orkowski 2005).

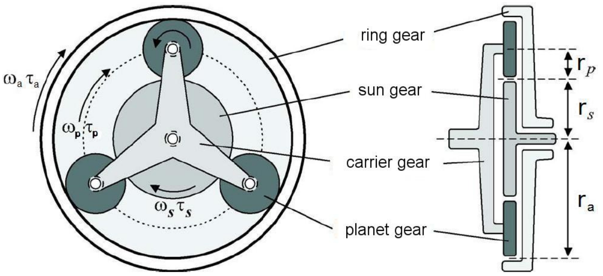


Fig. 4. Planetary gear system

The gear ratio is $k = \frac{r_a}{r_s}$, where r_a is the ring gear radius, r_s the sun gear radius and the angular velocities in the PGS satisfy

$$\omega_p = \frac{1}{(k+1)}\omega_s + \frac{k}{(k+1)}\omega_a \tag{5}$$

where ω_p, ω_s and ω_a are the angular velocities of the planet carrier, ICE and EM.

The balance of power in the PGS satisfies

$$P_p = T_s\omega_s + T_a\omega_a \tag{6}$$

Eqs. (5)-(6) are the kinematic and dynamic constraints, respectively, that any power flow strategy that employs a PGS must satisfy at all times.

The PGS is equipped with appropriate brakes to allow only one power source when convenient.

2.5. Clutch system

To disengage the ICE from the sun gear of the PGS a clutch is included (see Fig. 5). Three modes of operation for the clutch are modeled: when the ICE is disengaged, sliding and engaged (James and Narasimhamurthi 2005).

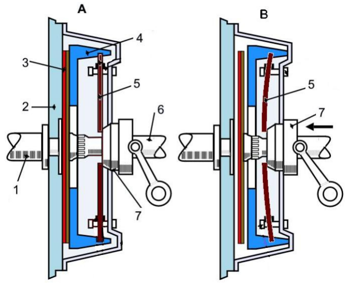


Fig. 5. Clutch system.

The clutch is modeled by

$$(J_{ice} + J_{clu})\dot{\omega}_{ice} = T_{ice} - T_{clu} - T_f \tag{7}$$

where J is the inertia, ω the velocity, T the torque, subscripts ice, clu and f are for ICE, clutch and friction, respectively. When the clutch is disengaged, $T_{clu} = 0$. When it is slipping,

$$\begin{aligned} T_{clu} = \left[k_{e1}\int(\mid \omega_{ice} - \omega_{clu} \mid)dt\right] \times \\ [\mid (\omega_{ice} - \omega_{clu}) \mid (-0.0005) + 1] \times \\ f(\mid \omega_{ice} - \omega_{clu} \mid) \end{aligned} \tag{8}$$

where, k_{e1} is the stiffness coefficient of sliding.

Finally, when the clutch is engaged $\omega_{ice} = \omega_{clu}$ and

$$\begin{aligned} T_{clu} = k_{e2}(\textstyle\int(\omega_{ice} - \omega_{clu})dt) + \\ f_{es}(\omega_{ice} - \omega_{clu}) \end{aligned} \tag{9}$$

where k_{e2} is a stiffness coefficient, f_{es} an absorption coefficient.

2.6. Vehicle model

Vehicle is modeled like a moving mass subjected to a traction force $F_{tr}(t)$. The forces at the power-train also include the aerodynamic drag force $F_a(t)$, the rolling resistance $F_r(t)$ of the tires and the gravitational force $F_g(t)$ induced by the slope in the road, that are given by (Xiong, Zhang, and Yin 2009), (Kessels, Koot, van den Bosch, and Kok 2008)

$$\begin{aligned} F_a(t) &= 0.5\rho_a v(t)^2 C_d A_d \\ F_r(t) &= mgC_r\cos\gamma(t) \\ F_g(t) &= mg\sin\gamma(t) \end{aligned} \tag{10}$$

where ρ_a is the air density, $v(t)$ the vehicle speed, C_d the aerodynamic drag coefficient, A_d the vehicle frontal area, m the vehicle mass, g the gravity acceleration constant, C_r the tire rolling resistance coefficient and γ the road slope.

The vehicle velocity $v(t)$ dynamics is given by

$$m\frac{dv(t)}{dt} = F_{tr} - F_a(t) - F_r(t) - F_g(t) \tag{11}$$

3. POWER FLOW CONTROL STRATEGIES

It is assumed that the ICE and EM are controlled with two independent controllers, whose set points must be determined by power flow control strategy.

The approach developed in this paper is based in the following observations:

1) The most important requirement in HEV power flow control is the ability to satisfy driver requirements.
2) All optimal solutions to power flow control preserve the state of charge of batteries, averaged over a long enough time period.
3) To minimize fuel consumption, ICE must be operated at high efficiency regions.

Observation 2, key in this paper strategy, is easily confirmed by noticing that all optimal solutions based on driving cycles must preserve the initial state of charge on the batteries at the end of cycle, otherwise the vehicle can not sustain repetitions of the same cycle. A similar observation is also made in Musardo et al (Musardo, Rizzoni, and Sataccia 2005), when discussing the tuning of A-ECMS. Observation 3 can be verified, for example, in (John. M. Miller 2006) or (Ehsani, Gao, and Miller 2007), and it is one of the main reasons HEV are overall more efficient that normal vehicles.

3.1. *Optimal strategy*

This strategy is based on PMP. The problem is to find an admissible control $u^* \in U$ that causes the system

$$\dot{x}(t) = a(x(t), u(t), t) \tag{12}$$

to follow an admissible trajectory $x^* \in X$that minimizes the performance cost

$$J(u) = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u(t), t)dt \tag{13}$$

Using as system state the errors in the demanded power $e_{P_p} = P_p - P_d$, state of charge $e_{soc} = soc - soc_{ref}$ and fuel flow $e_{m_f} = m_f - m_{f_{ref}}$, and the battery power and engine power $u^T = [P_{bat} \quad P_{ice}]$ as inputs. It follows that

$$\begin{aligned} e_{P_p} &= \eta_{me}P_{bat} + P_{ice} - P_d \\ \dot{e}_{soc} &= -\frac{1}{V_{bat}Q_{nom}}P_{bat} - \dot{soc}_{ref} \\ \dot{e}_{m_f} &= \frac{P_{ice}}{\eta_{ice}p_{th}} - \dot{m}_{f_{ref}} \end{aligned} \tag{14}$$

where P_d is the demanded power and P_p the power supplied by the engines.

The performance cost can be expressed as

$$\min J = \int e^T G_1 e + u^T G_2 u \quad dt \tag{15}$$

where G_1 and G_2 are appropriate weighting matrices.

The Hamiltonian is defined as

$$H = e^T G_1 e + u^T G_2 u + p^T[a(e, u, t)] \tag{16}$$

Using this notation, the necessary conditions to find the optimal control $u^* \in U$ that causes an admissible trajectory $e^* \in X$ and minimizes the performance cost as follows:

$$\begin{aligned} e_{P_p} &= \eta_{me}u_1 + u_2 - P_d \\ \dot{e}_{soc} &= -\frac{1}{V_{bat}Q_{nom}}u_1 \\ \dot{e}_{m_f} &= \frac{u_2}{\eta_{mci}p_{th}} \end{aligned} \tag{17}$$

where the soc and m_f are assumed constant references with zero time derivatives.

The costate equations are

$$\begin{aligned} \dot{p}_1 &= g_{11}e_{P_p} \\ \dot{p}_2 &= g_{12}e_{soc} \\ \dot{p}_3 &= g_{13}e_{m_f} \end{aligned} \tag{18}$$

and the restriction for the inputs are as follows

$$\begin{aligned} 0 &= g_{21}u_1^* + p_1\eta_{me} - p_2\frac{1}{V_{bat}Q_{nom}} \\ 0 &= g_{22}u_2^* + p_1 + p_3\frac{1}{\eta_{mci}p_{th}} \end{aligned} \tag{19}$$

If Eq. (19) is solved for u^* and substituted into the state Eqs. (17), three equations for the state and three for the costates are obtained

$$\begin{aligned} e_1^* &= \eta_{me}u_1^*(p_1, p_2) + u_2^*(p_1, p_3) - P_d \\ \dot{e}_2^* &= -\frac{1}{V_{bat}Q_{nom}}u_1^*(p_1, p_2) \\ \dot{e}_3^* &= \frac{1}{\eta_{mci}p_{th}}u_2^*(p_1, p_3) \\ \dot{p}_1^* &= g_{11}e_1^* \\ \dot{p}_2^* &= g_{12}e_2^* \\ \dot{p}_3^* &= g_{13}e_3^* \end{aligned} \tag{20}$$

Eqs. (20), the state and costate equations, are a set of linear first order, homogeneous algebraic-differential equations, that distribute power optimally between EM and ICE, given parameters G_i.

3.2. *PGS Strategy*

The second strategy, named PGS strategy, is inspired in optimal control and tries to reduce fuel consumption by using the EM as much as possible, that is, by maximizing electrical energy use. Assuming that the state of charge in the batteries must be kept at a reference value, for the traction case, $P_p > 0$,

$$J_1 = max(\int_0^{T_c} (sign(P_p)sign(soc - soc_{ref}))P_{me}dt \tag{21}$$

where T_c is the duration of the driving cycle, soc_{ref} is a reference value for soc. This expression is useful for the cases of traction and traction-recharging batteries.

For the braking case, $P_p < 0$, the criteria is

$$J_2 = \max \int_0^{T_c} (sign(P_p)P_{em}) \; dt \tag{22}$$

The value of Eqs. (21)-(22) is maximized when $P_{me} = \min\{sign(P_p)P_p, sign(P_p)P_{em}^{max}\}$, with P_{em}^{max} the maximum power attainable by the EM (assumed equal for the motor and generator cases). To avoid the switching induced by $sign(P_p)$ a smooth function of the soc is used. Therefore

$$P_{em} = P_{em}(soc) = \alpha_i(soc)P_{em}^{max} \tag{23}$$

where subindex in Eq. (23) is 1 when $P_p > 0$ and 2 when $P_p < 0$, $\alpha_i \in [-1, 1]$.

Assuming that P_p and ω_p are known, the proposed solution to the power flow control starts by substituting Eq. (23) in Eq. (6) leads to

$$P_p = \alpha_i P_{em}^{max} + P_{ice} \tag{24}$$

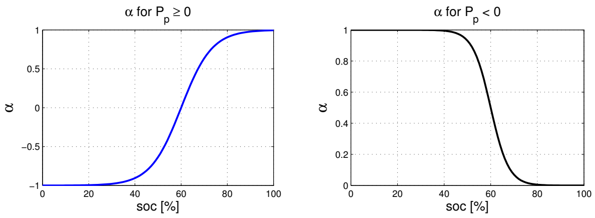


Fig. 6. α for $P_p \geq 0$ and $P_p < 0$

The shape of $\alpha_i(soc)$ determines how much electric power is taken or delivered at a given point. One possible choice for $\alpha_i(soc)$ is shown in Fig. 6, that is described by

$$\alpha_1 = \tanh(A_1(soc - soc_{ref})) \qquad Pp \geq 0 \tag{25}$$

$$\alpha_2 = 0.5 - 0.5(\tanh(A_2(soc - soc_{full}))) \qquad Pp < 0 \tag{26}$$

where soc_{ref} is a reference value for the batteries if the EM acts as a motor and soc_{full} is a reference value to avoid battery overcharging in the generator case.

If α_1 is positive the EM operates as motor, otherwise it operates as generator. Fig. 6 reveals that when $P_p \geq 0$, $\alpha_1 \in [-1, 1]$ depending on the state charge of the battery. When $P_p < 0$, $\alpha_2 \in [0, 1]$, regenerative braking is possible and the EM can work only as generator. This choice allows to make maximum use of electric power for traction or recharging of the batteries.

With α_i chosen, electric power in Ec.(24) is fixed. P_{ice} is determined as follows:

$$P_{ice} = min(P_p - P_{em}, P_{ice}^{max}); \quad P_p \geq 0$$

that guarantees that the ICE provides power below its maximum available power P_{ice}^{max}.

3.3. *Assigning speed and torque*

Given P_{ice}, the angular velocity at which the ICE must operate, ω_{ice}, is obtained from the maximum efficiency curve in the power vs. angular velocity curve. This curve has a shape similar to that shown in Fig. 7 and is approximated by a polynomial with P_{ice} as independent variable. With this choice, it is assured that the ICE is always used with maximum efficiency.

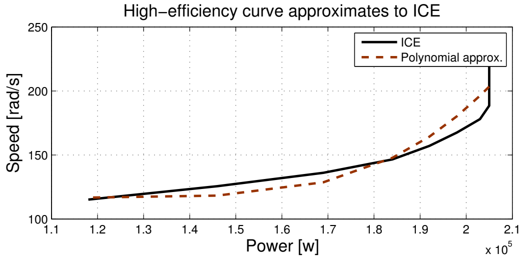


Fig. 7. Power vs. Speed high efficiency curve of the ICE.

Once ω_{ice} is obtained from the maximum efficiency curve, the required torque is

$$T_{mci} = \frac{P_{mci}}{\omega_{mci}} \qquad for \qquad \omega > 0 \tag{27}$$

$$T_{mci} = 0 \qquad for \qquad \omega = 0 \tag{28}$$

The final step is to determine the angular velocity and torque for the EM. From Eq. (5) ω_{em} is

$$\omega_{me} = \frac{(k+1)}{k}(\omega_p - \frac{1}{(k+1)}\omega_{mci}) \tag{29}$$

and the torque T_{em} is derived from

$$T_{me} = \frac{P_{me}}{\omega_{me}} \qquad for \qquad \omega > 0 \tag{30}$$

$$T_{me} = 0 \qquad for \qquad \omega = 0 \tag{31}$$

4. SIMULATION RESULTS

Simulations were carried out on SIMULINK MATLAB software for a bus with mass of 15,000 $[kg]$, a diesel ICE of 205 $[kw]$, a clutch between the ICE and a PGS with $k = 5$. The electric machine is a induction motor of 93 $[kw]$ and the batteries are $25[Ah]$ at $288[V]$. ll the components were simulated and tested separately. The bus is commanded to follow the three standard driving cycles: low velocity (1) c_1, medium velocity (2) c_2 and high velocity (3) c_3. On example of a driving cycle is shown in Fig. 8.

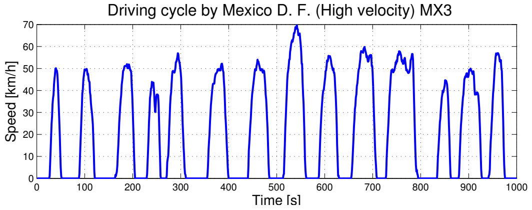


Fig. 8. High velocity driving cycle, c3

As mentioned before, the most important feature of any power flow control strategy in HEV is the ability to track driver power demands. Typical examples of speed tracking capability are shown in Fig. 9, that illustrate, for driving cycles 2 and 3, the velocity of the HEV obtained by the PMP strategy and velocity obtained by the PGS strategy. Velocity tracking is very good for both strategies. The first observation that inspired the strategies is satisfied.

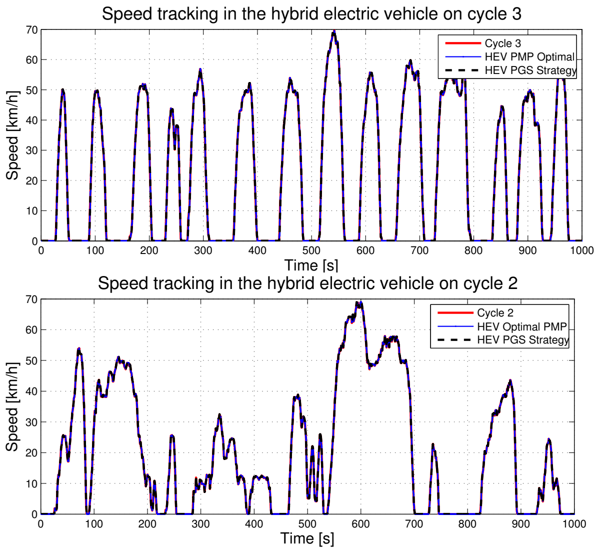


Fig. 9. HEV velocity tracking, high and medium velocity

Fig. 10 and Fig. 11 show simulation results of the state of charge (*soc*) for PMP strategy (blue dashed line) and PGS strategy (red continuous line). Note that the oscillations are bigger for the PMP strategy and that both strategies have the same initial and final state of charge.

If the *soc* initial an final is same for both strategies, the fuel consumption is considered net spending for the comparative index and the vehicle can be repeated the driving cycle as many times as required.

If the battery pack is more small, the oscillations are bigger for the dynamic *soc*.

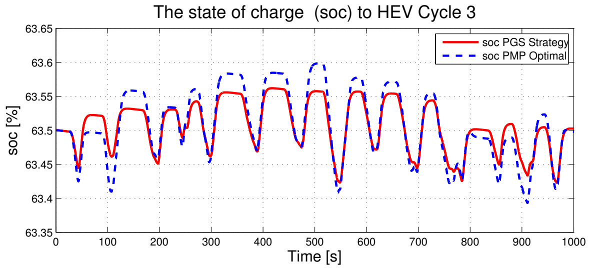


Fig. 10. *soc* for PGS strategy and PMP strategy, cycle c3

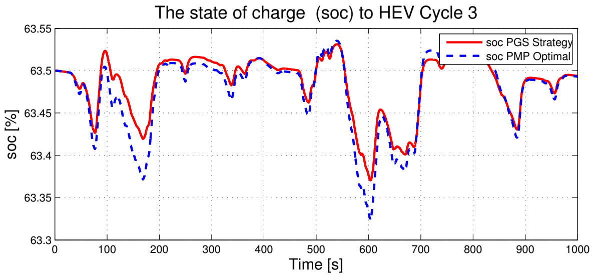


Fig. 11. *soc* for PGS strategy and PMP strategy, cycle c2

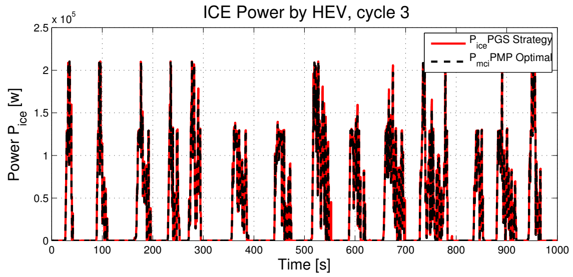


Fig. 12. ICE power compared for PGS and PMP strategies, cycle c3

The resultant ICE power is shown in Fig. 12 and the EM power Fig. 13 for the high velocity driving cycle (MX3), (black dashed line) for PMP strategy and (red continuous line) for the PGS strategy. Power requirements are very similar.

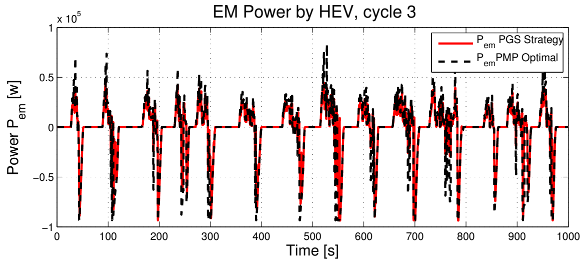


Fig. 13. EM power compared for PGS and PMP strategies, cycle c3

To convey an idea of the fuel consumption reduction provided by the HEV power split strategies, table I compares the total fuel consumption for two driving cycles and for a vehicle equipped only with an ICE. Notice the small difference between the fuel consumption of the two presented strategies.

Table I

COMPARISON OF STRATEGIES

Strategy	Cycle	Consumption (kg)	Consumption (%)
Only ICE	3	17.36	100
PMP	3	10.36	59.68
PGS	3	10.5	60.48
Only ICE	2	10.13	100
PMPl	2	7.976	78.74
PGS	2	8.029	79.26

5. CONCLUSIONS

A pair of power split strategies for HEV were presented and proved by simulations on SIMULINK-MATLAB software. The strategy is designed for a parallel configuration HEV where a planetary gear system is used a power coupling device. Simulations use a detailed model of a HEV that includes a diesel internal combustion engine, an induction electric engine, a planetary gear system, a clutch, batteries and gear transmission.

The first strategy is based on Pontryagin's minimum principle (PMP), while the second is designed around a planetary gear system (PGS). Design procedures are described to obtain power and torque splits. Simulation were performed in such a way that initial and final state of charge of batteries is equal for all driving cycles. Results show that, by appropriate tuning of the weighting matrices in the Hamiltonian of the PMP strategy, both strategies achieve very similar results. However, the information requirements and the computational cost of the PGS strategy are smaller than those of the PMP strategy, making the first strategy more suitable for real time implementation.

Results also indicate excellent tracking of all driving cycles with limited excursions of the battery state of charge during the driving cycle. ICE operates at high efficiency, given the power split determined by both strategies.

An adaptive version and experimental testing of the strategies is ongoing work.

6. ACKNOWLEDGMENT

This research was supported by CONACYT grant 103640, UNAM-PAPIIT grant IN105512 and UNAM-PAPIIT grant IN116012. The first two authors specially acknowledge the support from CONACYT scholarship program.

REFERENCES

Ambarisha, V. K. and R. G. Parcker (2007, January). Nonlinear dynamics of planetary gears using anlytical and finite element models. *Journal of sound and vibration 302*, 577–595.

Delprat, S., J. Lauber, T. M. Guerra, and J. Rimaux (2004, May). Control of a Parallel Hybrid Powertrain: Optimal Control. *IEEE Transactions on Vehicular Technology 53*, 872–881.

Ehsani, M., Y. Gao, and J. M. Miller (2007, April). Hybrid Electric Vehicles: Architecture and Motor Drives. *Proceedins of the IEEE 95*, 719–728.

James, D. and N. Narasimhamurthi (2005, 8-10, June). Design of a optimal controller for commercial trucks. In *American Control Conference*, Portland, Oregon, USA., pp. 1599–1606.

John. M. Miller (2006, May). Hybrid Electric Vehicle Propulsion System Architectures of the e-CVT Type. *IEEE Transactions on Power Electronics 21*, 756–767.

Kessels, J. T. B. A., W. T. Koot, P. P. J. van den Bosch, and D. B. Kok (2008, november). Online Energy Management for Hybrid Electric Vehicles. *IEEE Transactions on Vehicular Technology 57*, 3428–3440.

Kim, N., S. Cha, and H. Peng (2011, September). Optimal control of hybrid electric vehicles based on pontryagin's principle. *IEEE Transactions on Control Systems Technology 19*, 1279–1287.

Koot, M., J. T. B. A. Kessels, B. de Jager, W. P. M. H. Heemels, P. P. J. van den Bosch, and M. Steinbuch (2005, May). Energy Management Strategies for Vehicular Electric Power Systems. *IEEE Transactions on Vehicular Technology 54*, 771–782.

Musardo, C., G. Rizzoni, and B. Sataccia (2005, December). A-ECMS: An Adaptive Algoritm for Hybrid Electric Vehicle Energy Management. In *44th IEEE Conference on Decicsion and Control, and the European Control Conference*, Seville, Spain., pp. 1816–1823.

Outbib, R., X. Dovifaaz, A. Rachid, and M. Ouladsine (2002, May). Speed control of a diesel engine: a nonlinear approach. In *American Control Conference*, Anchorage, Alaska, USA., pp. 3293–3294.

Paganelli, G., T. M. Guerra, S. Delprat, J.-J. Santin, M. Delhom, and E. Combes (2000). Simulation and assessment of power control strategies for a parallel hybrid car. *Journal of automobile engineering 214*, 705–717.

Peresada, S., A. Tilli, and A. Tonielli (2004). Power control of a doubly fed induction machine via output feedback. *Control Engineering Practice 12*, 41–57.

Pisu, P. and G. Rizzoni (2007, may). A Comparative Study Of Supervisory Control Strategies for Hybrid Electric Vehicles. *IEEE Transactions on Control Systems Technology 15*, 506–518.

Szumanowski, A., C. Yuhua, and P. Piórkowski (2005, sept). Analysis of Different Control Strategies and Operating Modes of Compact Hybrid Planetary Transmission Drive. *Vehicle Power and Propulsion 7*, 673–680.

Xiong, W., Y. Zhang, and C. Yin (2009, July). Optimal Energy Management for a Series-Parallel Hybrid Electric Bus. *Energy conversion and management 50*, 1730–1738.

Xiong, W. W. and C. L. Yin (2009, May). Design of Series-parallel Hybrid Electric Propulsion Systems and Application in City Transit Bus. *WSEAS Transaction on Systems 8*, 578–590.

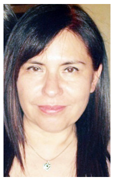
Zou, Y., L. Teng, S. Fengchun, and H. Peng (2013, April). Comparative study of dynamic programming and pontryagin's minimum principle on energy management for a parallel hybrid electric vehicle. *Energies 6*, 2305–2318.

AUTHORS BIOGRAPHY

Guillermo Becerra Was born in Colima, Mexico in 1986. B. S. degree in Mechanical and Electrical Engineering from the Universidad de Colima, Mexico in 2008, the M. S. degree in Electrical Engineering from the Universidad Nacional Autónoma de México UNAM, Mexico in 2010 and he is currently a Ph. D. candidate in (Control) Electrical Engineering at the UNAM, Mexico and since 2011 he joined the Department of Mechatronics Engineering, UNAM, as an Interim Professor. His research interests include nonlinear control theory, mechatronics, optimal control and applications.

Alfonso Pantoja-Vazquez Recived the B. S. degree in instrumentation and process control engineering from the University of Queretaro, Mexico in 2003, the M. S. degree in electrical engineering from the Universidad Nacional Autónoma de México, Mexico in 2006 and he is currently a Ph. D. candidate in Mechatronics Engineering at the Universidad Nacional Autónoma de México, Mexico. He worked for Delphi from 2007 to 2011 as embedded software engineer for automotive devices. He joined the University of Quereraro from 2010 to 2011 as interim professor on the embedded software department and since 2011 he joined the department of mechatronics engineering at the Natinoal University of Mexico, as interim professor. His research interests include mechatronics, embedded software, optimal control and its applications.

Luis Alvarez-Icaza Was born in México City. Obtained his PhD in Mechanical Engineering at University of California in Berkeley. He is Professor at Instituto de Ingeniería of the Universidad Nacional Autónoma de México. He is currently Dean of Graduate Studies of Engineering. His research interests are in the areas of vehicle control, buildings vibration control, wind generators control and traffic control.



Idalia Flores She received her Ph.D. in Operations Research at the Faculty of Engineering of the UNAM. She graduated Master with honors and received the Gabino Barreda Medal for the best average of her generation. She has been a referee and a member of various Academic Committees at CONACYT. She has been a referee for journals such as Journal of Applied Research and Technology, the Center of Applied Sciences and Technological Development, UNAM and the Transactions of the Society for Modeling and Simulation International. Her research interests are in simulation and optimization of production and service systems. She is a full time professor at the Postgraduate Program at UNAM.

MODELING AND SIMULATION OF STRESS FOR CEMENTED CARBIDE CYLINDRICAL END MILL

WU Qiong[(a)], **Li Da Peng**[(b)]

State Key Lab of Virtual Reality Technology and Systems, School of Mechanical Engineering and Automation,
Beihang University, Beijing 100191, China

[(a)]wuqiong@buaa.edu.cn [(b)] lidapeng925@163.com

ABSTRACT

The cutting force and stress of tool is important factor for failure of tool in the cutting process. Predictions of the forces and stress in milling are often needed in order to establish automation or optimization of the machining processes. A theoretical model for milling forces according to a predictive machining theory and the experiment of milling has been presented. The action of a milling cutter is considered as the simultaneous work of a number of single-point cutting tools, and milling forces are predicted from input data of the workpiece material properties, the cutter parameters and tooth geometry, and the cutting condition. Based on cutting force model, the stress of cemented carbide cylindrical end mill is developed. The numerical simulation and the experiment are performed to compare with stress model of mill. Milling tool stress experimental tests and numerical simulation separately were conducted to verify the simulation. The study provides a reliable method for analyzing stress of tool in the milling process.

Keywords: simulation; modeling; finite element analysis; milling

1. INTRODUCTION

Milling is widely used in many areas of manufacturing. During the cutting operation, cutters may break down either at the shank of cutter or at the tip of cutter. Some operation may occur on both conditions. Investigation of failure requires not only knowledge of the state of stress but also failure criteria. The literatures (Kurt, 2009; Jemal, 1992).shows the stresses for metal cutting, it is observed that the analytical method, the finite element method (FEM) and Artificial Neutral Network (ANN), the mathematical modeling method were used in order to analyze the stresses of the orthogonal cutting tools, the cutters for turning and the end mill. Then the experimental stress analysis was used to verify the stress results obtained by FEM, ANN or the analytical method (Kurt, 2009). The orthogonal cutting tool has two types of premature failure: brittle failure, such as edge fracture and edge chipping, and ductile failure, such as edge plastic deformation was discussed by Zhou, Andersson and Stahl (1997).

The estimation of cutting force is an important factor to predict the cutting tool stresses. Some researchers used a series of experimental measurements in order to find the cutting forces (Kumar, Mohan, Rajadurai and Dinakar, 2003; Wu, Zhang and Zhang, 2009). Some researchers presented the analytical cutting force model (Budak, Altintas, and Armarego, 1996). The experimental stress analysis method are the electrical-resistance strain gage and its associated instrumentation, transmission and reflection photo-elasticity, brittle coating, Moiré gratings, Moiré interferometry, x-ray diffraction, holographic and laser speckle interferometry and thermo-elastic stress analysis. From these methods, the electrical-resistance strain gages are widely used in experimental stress analysis (Tsai, 2007). The strain gages have been used to study the deflection and residual stress of the end mill (Kops and Vo, 1990; Rendler and Vigness, 1966).

In this paper, the cutting forces will be analytically predicted using the cutting force formulas. The maximum principal stress, the maximum shear stress and von Mises stress will be investigated by means of FEM commercial software (MSC Patran/Nastran). These results are verified with the results obtained from the electrical-resistance strain gage.

2. MILLING FORCE MODEL

The cutting forces based on the analytical theory(L. Kops and D. T. Vo 1990) are used to investigate the deflection and stresses in the end mill. Tangential ($dF_{t,j}$), radial ($dF_{r,j}$), and axial ($dF_{a,j}$), forces acting on a differential flute element with height dz are expressed in Eq.(1).

$$\begin{aligned} dF_{t,j}(\phi,z) &= \left\lfloor K_{tc}h_j(\phi_j(z)) + K_{te} \right\rfloor dz, \\ dF_{r,j}(\phi,z) &= \left[K_{rc}h_j(\phi_j(z)) + K_{re} \right] dz, \\ dF_{a,j}(\phi,z) &= \left[K_{ac}h_j(\phi_j(z)) + K_{ae} \right] dz. \end{aligned} \tag{1}$$

dF_t, dF_r, dF_a ——Differential tangential, radial and axial forces;

K_{tc}, K_{rc}, K_{ac} ——Shear force coefficients in tangential, radial and axial directions;

K_{te}, K_{re}, K_{ae} ——Cutting force coefficients in the tangential, radial and axial directions;

ϕ —— Rotation angle of the milling cutter;

z —— Height of axial directions

$h(\phi, z)$ ——Chip thickness specified by ϕ , z ;

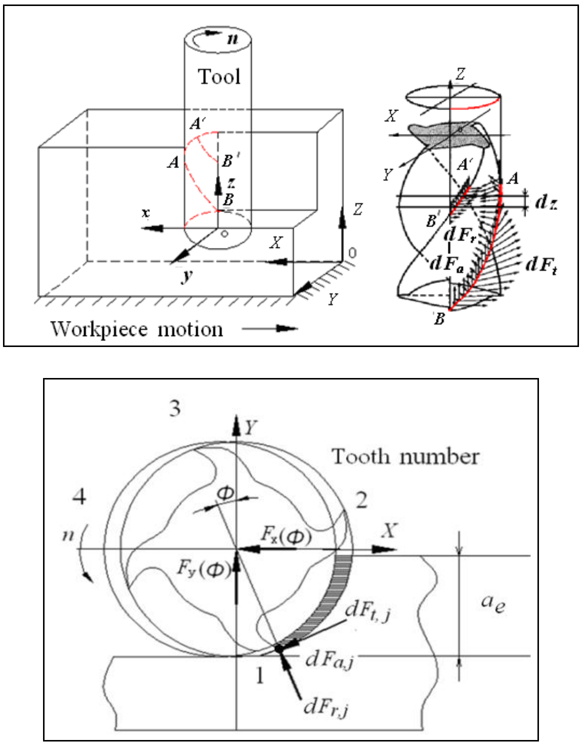


Figure 1: Model of milling force

The elemental forces are resolved into feed (*x*), normal (*y*), axial (*z*) directions using the transformation

$$dF_{x,j} \quad \phi_j(z) \ = -dF_{t,j} \cos\phi_j(z) - dF_{r,j} \sin\phi_j(z),$$
$$dF_{y,j} \quad \phi_j(z) \ = +dF_{t,j} \sin\phi_j(z) - dF_{r,j} \cos\phi_j(z),$$
$$dF_{z,j} \quad \phi_j(z) \ = +dF_{a,j}. \qquad (2)$$

The average cutting forces are

$$\left.\begin{aligned} \bar{F}_x &= \left\{ \frac{Nac}{8\pi} K_{tc} \cos 2\phi - K_{rc}(2\phi - \sin 2\phi) \ + \frac{Na}{2\pi}[-K_{te} \sin\phi + K_{re} \cos\phi] \right\}_{\phi_{st}}^{\phi_{ex}} \\ \bar{F}_y &= \left\{ \frac{Nac}{8\pi} K_{tc}(2\phi - \sin 2\phi) + K_{rc} \cos 2\phi \ - \frac{Na}{2\pi}[K_{te} \cos\phi + K_{re} \sin\phi] \right\}_{\phi_{st}}^{\phi_{ex}} \\ \bar{F}_z &= \frac{Na}{2\pi}[-K_{ac} c \cos\phi + K_{ae}\phi]_{\phi_{st}}^{\phi_{ex}} \end{aligned}\right\} \qquad (3)$$

β ——Helix angle;

a_e —— radial depth of cut

N —— number of teeth

ϕ_{st} —— start angle

ϕ_{ex} —— exit angle

X, Y, Z—— Global stationary coordinate system

3. MILLING FORCE TEST PROCESS

It is assumed that the edge force coefficients (K_{te} , K_{re} and K_{ae}) are constant and have little effect on the resultant forces. By means of Eq. (3), the milling forces are obtained by measuring cutting force device as shown in Figure 2. The parameters of cutting process is shown in Tab.1.

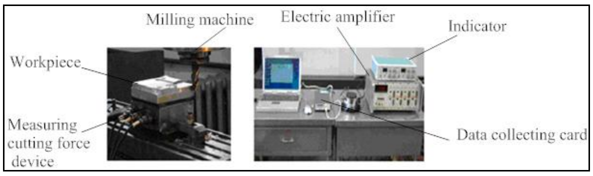


Figure 2 Set Up of Milling Force Test Process

Table 1: Parameters of Cutting Process

Parameters name (unit)	Value	Parameters name (unit)	Value
Rotation speed r/min	3000	Material of tool	Hard alloy
Feed speed mm/min	400	Number of tooth	2
Depth of cutting mm	2	Helical angle	20
Width of cutting mm	6	Length of cantilever mm	120
Material of workpiece	7075	Outer Diameter mm	12
Hardness HRA	90	Rake angle °	20

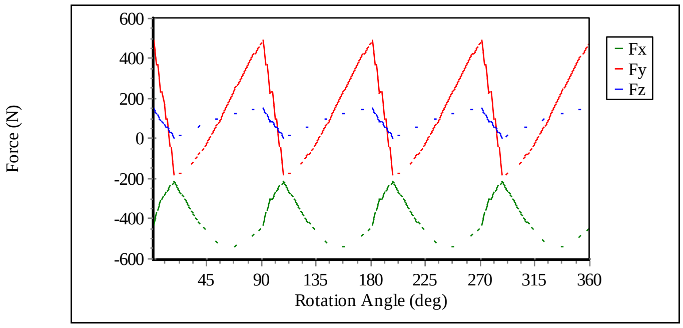


Figure 3: Cutting Force in 360° Rotation Angle

Based on cutting force experimental results as shown in Figure 3, the average cutting forces per tooth can easily be calculated. The milling force coefficients (N/mm^2) of 7075 aluminum alloy were obtained to

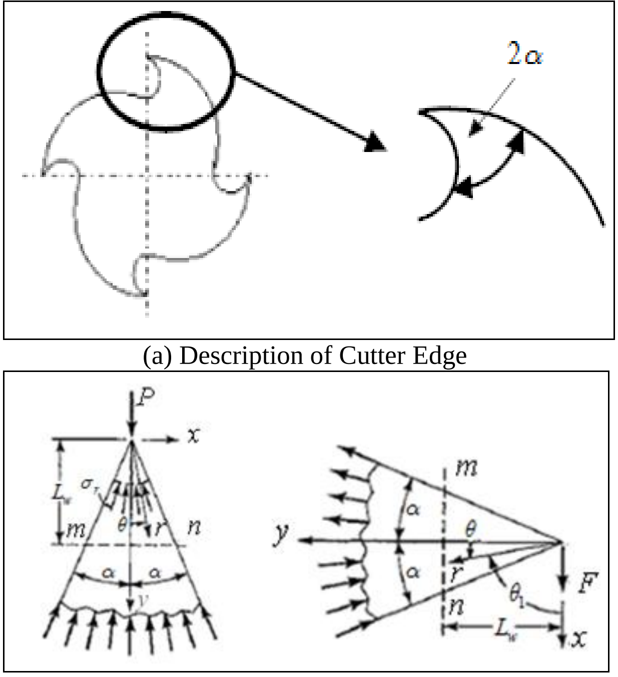
$$K_{rc} = 893.3, \ K_{tc} = 1880.0, \ K_{ac} = 157.1.$$

Apply these coefficients for Eq. (3) to get the milling force model

4. STRESS ANALYSIS OF MILLING MODEL

The shape of the cutter edge as described in Figure 4 (a) is assumed as a semi-infinite wedge .The thickness of the wedge is taken as unity, so *P* or *F* is the load per unit thickness acting at the vertex of a very large or semi-infinite wedge as shown Figure 4 (b).

Assuming that the stress function for *y*-direction is



(a) Description of Cutter Edge

(b) Cutter Edge or Pivot (c) Wedge Cantilever

Figure 4: Description of Cutter Edge and Wedge of Unit Thickness Subjected of Concented Cutting Load

$$\psi = kPr\theta \sin\theta \tag{4}$$

The corresponding stress components are obtained by the following relations.

$$\sigma_r = \frac{1}{r}\frac{\partial\psi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2} \tag{5}$$

$$\sigma_\theta = \frac{\partial^2\psi}{\partial r^2} \tag{6}$$

$$\tau_{r\theta} = -\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\psi}{\partial\theta}\right) \tag{7}$$

By substituting Eq. (4) into Eq.(5) to (7), the radial stresses are：

$$\sigma_r = 2kP\frac{\cos\theta}{r}\ ,\ \sigma_\theta = 0\ ,\ \tau_{r\theta} = 0 \tag{8}$$

The constant *k* can be obtained by using the equilibrium condition at the point O. The force resultant action on a cylindrical surface of small radius, shown by the dotted line in Figure 39 (a), must balance *P*. Since The boundary conditions are expressed by

$$\sigma_\theta = 0\ ,\ \tau_{r\theta} = 0,\ \theta = \pm\alpha \tag{9}$$

$$2\int_0^\alpha (\sigma_r \cos\theta) r d\theta = -P \tag{10}$$

By substituting σ_r in Eq. (5) to Eq. (10)

$$4kP\int_0^\alpha \cos^2\theta d\theta = -P \tag{11}$$

Integration and solving Eq.(11), the constant *k* can be written as

$$k = -\frac{1}{(2\alpha + \sin 2\alpha)} \tag{12}$$

The stress distribution in the edge is therefore

$$\sigma_r = -\frac{P\cos\theta}{r(\alpha + \frac{1}{2}\sin 2\alpha)}\ ,\ \sigma_\theta = 0\ ,\ \tau_{r\theta} = 0 \tag{13}$$

The normal and shearing stresses can be expressed as

$$\sigma_x = \sigma_r \cos^2\theta = -\frac{P\cos^3\theta}{r(\alpha + \frac{1}{2}\sin 2\alpha)} \tag{14}$$

$$\sigma_y = \sigma_r \sin^2\theta = -\frac{P\sin^2\theta\cos\theta}{r(\alpha + \frac{1}{2}\sin 2\alpha)} \tag{15}$$

$$\tau_{xy} = \sigma_r \sin\theta\cos\theta = -\frac{P\sin\theta\cos^2\theta}{r(\alpha + \frac{1}{2}\sin 2\alpha)} \tag{16}$$

By assuming the stress function for *x*-direction

$$\phi = kFr\theta_1 \sin\theta_1 \tag{17}$$

Where θ_1 is measured from the line of action of the force. The equilibrium condition is

$$\int_{\frac{\pi}{2}-\alpha}^{\frac{\pi}{2}+\alpha} (\sigma_r \cos\theta_1) r d\theta_1 = 2kF\int_{\frac{\pi}{2}-\alpha}^{\frac{\pi}{2}+\alpha} \cos^2\theta_1 d\theta_1 = -F \tag{18}$$

After integration

$$k = -\frac{1}{(2\alpha - \sin 2\alpha)} \tag{19}$$

By replacing : $\theta_1 = 90 - \theta$

$$\sigma_r = -\frac{F\cos\theta_1}{r(\alpha - \frac{1}{2}\sin 2\alpha)} = -\frac{F\sin\theta}{r(\alpha - \frac{1}{2}\sin 2\alpha)},$$

$$\sigma_\theta = 0,\ \tau_{r\theta} = 0 \qquad (20)$$

$$\sigma_x = \sigma_r \cos^2\theta = -\frac{F\sin\theta\cos^2\theta}{r(\alpha - \frac{1}{2}\sin 2\alpha)} \qquad (21)$$

$$\sigma_y = \sigma_r \sin^2\theta = -\frac{F\sin^3\theta}{r(\alpha - \frac{1}{2}\sin 2\alpha)} \qquad (22)$$

$$\tau_{xy} = \sigma_r \sin\theta\cos\theta = -\frac{F\sin^2\theta\cos\theta}{r(\alpha - \frac{1}{2}\sin 2\alpha)} \qquad (23)$$

5. STRESS ANALYSIS AND EXPERIMENT

In order to analyze the stresses and deflection of the milling cutter, the cutter model is built by Unigraphics (UG) software. Then the model is imported to MSC Patran/Nastran software and divided into 4-noded tetrahedral elements. And the fixed boundary condition and the concentrated loads are applied to the model as described in Figure 5 The isotropic material properties are input according to cutter material from Tab.4. The solution type is chosen as “Linear statics” from structural analysis. The elemental cutting forces used are at 37.2˚ immersion angle. The distribution of Von Mises stresses in Figure 5 (b). The maximum stress occurs at the cutter edge within the loading region. The location of maximum shank stress can decide from these figures. In addition, the values and locations of maximum and minimum stress for the whole cutter can be obtained.

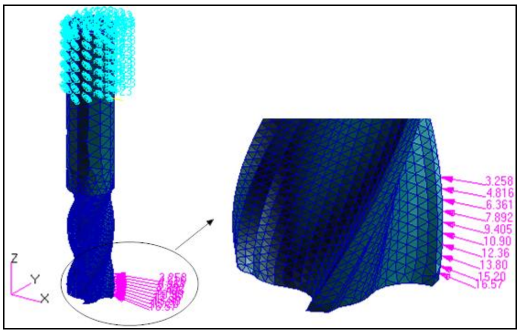


Figure 5(a): Milling Model with FEM Mesh

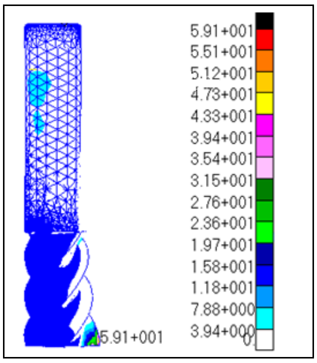


Figure 5(b): Stress Distribution of Cutters

Figure 6 shows equipment used in experiment test to measure strain with electrical-resistance strain gage. The electrical resistance strain gage used in this test is

Figure 6 shows equipment used in experiment test to measure strain with electrical-resistance strain gage. The electrical resistance strain gage used in this test is uniaxial strain gage with 120Ω. The gage factors are 2 and 2.08 for two different size of strain gage. Adhesive is cyanoacrylate. The followings are procedures of strain gage installation.

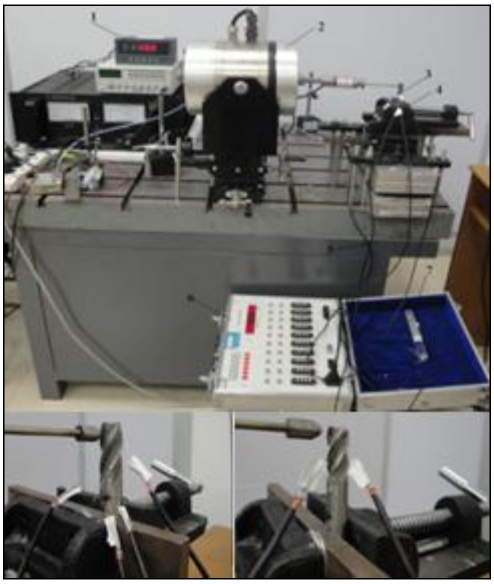


Figure 6: Set up of Experiment

Before the strain gage is bonded a specified site, its resistance should be measured with multitester. If the resistance is not reached 120Ω, it cannot work correctly. Two terminals of strain gage is connected to a piece of wire with proper length by using electronic soldering iron gun tool and tin lead alloy soldering rosin wire reel. The soldered place must be covered with masking tape to prevent possible damage and environment effects. The strain gage bonding site must be cleaned and the strain gage bonding position should be marked. A drop of adhesive is applied to the back of the strain gage and immediately put the strain gage on the bonding site. The strain gage is covered with an appropriate thing and strongly pressed with a thumb for approximately 1 minute. The two terminals of wire are connected to the strain gage indicator. The value of strain in the strain gage is set to zero. Gage factor will be changed to the value of corresponding material of strain gage. The exciter is moved to touch the cutter and up to desired force. The value of force can be read on force indicator. Value of strain on strain gage indicator is changing according to the given force. When the desired force reaches, strain can be read on strain gage indicator. From experiment the results can be obtain as shown in Table.2

Table.2: Parameters of Cutter Strain of Different Directions between Edge and Shank

Force (N)	**Cutter strain on shank**			**Cutter strain at edge**		
	ε_a 180°	ε_b 215°	ε_c 265°	ε_a 60°	ε_b 90°	ε_c 195°
50	-7	4	25	19	2	-137
100	-12	7	46	49	6	-270
150	-19	11	70	78	10	-400
200	-25	14	91	87	14	-550
250	-32	18	115	116	18	-685
300	-37	22	139	145	22	-826
350	-44	26	162	174	26	-960
400	-47	30	183	203	30	-1094

Table.3: Comparison of Cutter Stress in the Three Ways between Edge and Shank

Force	**Cutter stress at edge**				
	Test (MPa)	**FEM (MPa)**	**Error (%)**	**Analytic (MPa)**	**Error (%)**
50	116.16	121.66	4.52	112.72	3.05
100	238.14	243.32	2.13	225.44	5.63
150	357.10	364.98	2.16	338.16	5.60
200	472.51	486.64	2.91	450.88	4.80
250	595.03	608.30	2.18	563.60	5.58
300	721.98	729.96	1.09	676.32	6.75
350	843.82	851.63	0.92	789.04	6.94
400	965.67	973.29	0.78	901.76	7.09
Average error (%)			2.09		5.68
Force	**Cutter stress on shank**				
	Test (MPa)	**FEM (MPa)**	**Error (%)**	**Analytic (MPa)**	**Error (%)**
50	14.36	12.76	12.52	13.81	3.94
100	26.34	25.52	3.21	28.80	8.53
150	40.14	38.28	4.85	43.20	7.08
200	52.25	51.04	2.36	57.60	9.29
250	66.06	63.80	3.53	72.00	8.26
300	79.60	76.57	3.96	86.40	7.87
350	92.85	89.33	3.94	100.80	7.89
400	104.49	102.09	2.35	115.20	9.30
Average error (%)			4.59		7.77

The error between FEM, experiments and analytical are shown in the Table 3. The reason is followings:

Plane stress problem is considered in experiment. Three-dimensional stress is considered in FEM and in analytical method only normal stress in the cross section is considered for shank stress and three normal stresses and two shear stresses are considered for cutter edge stresses. Although the cutter is firmly constrained and the values of stress are zeros in FEM and analytical method, stress and deflection of constrained part in experiment may be small value. The elemental cutting forces of *x*-, *y*- and *z*-directions are used in FEM and analytical method and one-directional point load is used in experiment. In FEM, the mesh density, cutter models and boundary conditions can affect the error. In analytical method, cutters are represented as cantilevered beam and cross sectional diameter of fluted part is 0.9 of shank diameter of cutter. Human error can occur when values from indicators are reading and calculating.

CONCLUSION

This paper shows the model and simulation of stress obtained by FEM and analytical method compared with the experimental results.

1. A theoretical model of cutting force and stress of tool has been established. The correct prediction of stress has been conducted according to the machining parameters.
2. The simulation of stress of tool has been performed by FEM. The stress and deflection has been obtained and compared with the theoretical analysis.
3. The stress of tool was measured with electrical-resistance strain gage. The results are used to verify to theoretical analysis.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the “National Natural Science Fund” (51105025) and “International Science & Technology Cooperation Program of China” (2013DFB70110).

REFERENCES

Kurt A., 2009. Modeling of the Cutting Tool Stresses in Machining of Inconel 718 using Artificial Neural Networks. *Expert Systems with Applications,* 36 (6), 9645–9657.

Tsai C. L., 2007. Analysis and prediction of cutting forces in end milling by means of a geometrical model. *International Journal of Advance Manufacture Technology*, 31, 888–896.

Budak E., Altintas Y., Armarego, E., 2007. Prediction of milling force coefficients from orthogonal cutting data. *Transactions of the ASME*, 118, 216–224.

Jemal G., 1992.*Stress analysis of metal cutting tool,* Thesis (PhD). The University of British Columbia.

Zhou J. M., Andreson M. and St åhl J. E., 1997.Cutting tool fracture prediction and strength evaluation by stress identification, Part I: stress model. *International Journal of Machine Tools & Manufacture.* 12(37), 1691–1714.

Kops L., Vo D.T., 1990. Determination of the equivalent diameter of an end mill based on its compliance. *CIRP Annals Manufacturing Technology*, 39, 93–96.

Kumar M. P., Mohan B., Rajadurai A., Dinakar B. R., 2003. Modeling and analysis of orthogonal cutting of steel using FEM. *Proceedings of International Conference on Mechanical Engineering*. 12

Rendler N. J., Vigness I., 1996. Hole-drilling strain-gage method of measuring residual stresses. *Experimental Mechanics*. 6, 577–586.

Wu Q., Zhang Y. D., Zhang H. W., 2009. Corner-milling of thin walled cavities on aeronautical components. *Chinese Journal of Aeronautics*. 22, 677–684.

AUTHORS BIOGRAPHY

Wu Qiong, now is a lecturer in Beihang university. He has obtained Ph.D. in 2009 and finished Postdoctoral in 2011 at the School of Mechanical Engineering and Automation in Beihang University, China. His main research interest includes analysis and simulation of manufacturing and design as well as FEM.

Li Da Peng, now read for Master Degree at School of Mechanical Engineering and Automation in Beihang University. His main research interest in simulation of manufacturing process.

SIMULATION BASED EVALUATION OF CONCEPTS AND STRATEGIES FOR AUTOMATED STORAGE AND RETRIEVAL SYSTEMS WITH MULTIPLE I/O POINTS

Lantschner Daniel[(a)], Atz Thomas[(b)], Prof. Dr. Günthner Willibald A.[(c)]

[(a) (b) (c)] fml - Institute of Materials Handling, Material Flow, Logistics
Technische Universität München

[(a)]lantschner@fml.mw.tum.de, [(b)]atz@fml.mw.tum.de, [(c)]kontakt@fml.mw.tum.de

ABSTRACT

In recent years, there has been an increasing demand for high speed automated storage and retrieval systems (AS/RS). A possibility to increase the throughput of AS/RS rarely considered so far is the use of more than one I/O point in order to reduce the mean travel distance between I/O and storage locations. In this paper concepts for the alignment of multiple I/O points and strategies for their optimal use are presented. Since no suitable calculation methods for most of those concepts and strategies are available, a simulation model has been developed for their evaluation in terms of cycle times. The simulation results are compared with those of AS/RS with a single I/O point in order to show the benefit of the new concepts.

Keywords: AS/RS, storage, retrieval, travel time, multiple I/O, discrete-event simulation

1. INTRODUCTION

Automated storage and retrieval systems are computer controlled systems for depositing and retrieving loads from defined storage locations (MHI 2013). Warehouses, distribution centres and manufacturing facilities are the most common application areas of AS/RS. The major components of an AS/RS are: the storage rack, the input/output system (I/O system), the storage and retrieval machine (S/R machine) and the warehouse management system.

The performance of an AS/RS can be estimated by the mean cycle time. According to Gudehus (2007) the cycle time depends primarily on the travel time for the partial movement in the three directions in space. To improve the performance, manufacturers of AS/RS continuously develop S/R machines capable of higher speeds and accelerations. The main development directions are weight savings and mostly the use of more powerful drives with a higher power requirement. A solution to increase the throughput without increasing or even reducing the power requirement would be a reduction of the S/R machine's mean travel distance. This can be achieved, for example, by shifting the I/O location from a corner of the rack in a horizontal and/or vertical direction towards the storage rack's centre of area. Such concepts have been analyzed sufficiently in the past (see Gudehus 1972a, Knepper 1980, Eggert et al. 2010, among others) and several calculation models have been developed (see Gudehus 1972b, Bozer and White 1984, among others). In this paper, an approach for a further reduction of the mean travel distance by using more than one I/O location will be examined. Arantes and Kompella (1993) developed analytical expressions for the calculation of the expected cycle time for a specific case of an AS/RS with multiple I/O points which aren't suitable for most of the concepts and strategies proposed by the authors. Therefore, a purpose-built simulation model will be presented.

2. STATE OF THE ART

Simulation is a well-established tool for the analysis of AS/RS. Besides cycle time calculation, simulation is used to optimize the design, analyze different storage assignment and dwell point strategies, to evaluate scheduling rules as well as for several other analyses. Roodbergen and Vis (2009) give a detailed overview of simulation and analytical models and an explanation of the current state of the art in AS/RS design.

3. STORAGE AND RETRIEVAL SYSTEMS WITH MULTIPLE I/O POINTS

With multiple I/O points placed at different locations, the I/O point leading to the shortest travel distance and thus the shortest travel time can be used for each storage or retrieval cycle. Assuming a symmetrical distribution of the I/O points for the racks on either side of an aisle, single racks are shown to illustrate the alignment of the I/O points (see Figure 1).

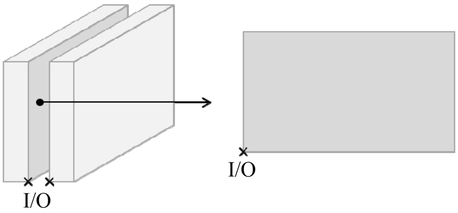


Figure 1: AS/RS aisle and single rack

The shown I/O configuration represents the most frequent I/O alignment in practical applications. A selection of reference concepts with a single I/O point as well as new concepts for the alignment of multiple I/O points in an AS/RS aisle are presented in the following. After that, possible strategies for the selection of an I/O point for storage and retrieval cycles in AS/RS with multiple I/O points will be introduced.

3.1.1. Reference Concepts

As reference for the evaluation of the new concepts, the following alignments of a single I/O point will be selected:

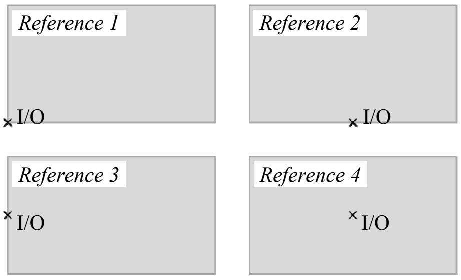


Figure 2. Reference I/O alignments

Reference 1 is the most commonly used alignment. The other alternatives are of less practical importance but will be useful to show the potential of different single I/O locations in comparison to the new concepts with multiple I/O points.

3.1.2. Concepts with multiple I/O points

Multiple I/O points can be aligned in various ways in an aisle. Since each I/O point has to be connected via a conveyor to the pre-storage area on the front side of the rack for the supply and removal of storage goods, only horizontally and vertically aligned I/O points will be considered to allow the use of standardized conveyors or lifters. These I/O points can be places either on the border of the rack face or centred for a further reduction of the mean distance between the storage shelves of the rack and the I/O points. Figure 3 shows those alignments as an example with three I/O points. In the first row concepts with horizontally aligned I/O points are illustrated (*Concept 1* and *Concept 2*), whereas in the second row two concepts with vertically aligned I/O points are shown (*Concept 3* and *Concept 4*):

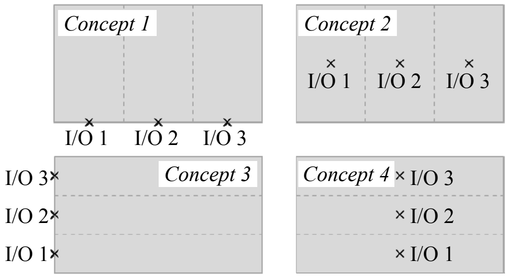


Figure 3: Examples for different I/O configurations

Two or more I/O points are aligned by dividing the rack area into equal subareas and placing the points at mid-length or mid-height of each subarea. It will be assumed that all I/O points of an aisle are equivalent and always available for either picking up or depositing storage goods.

3.2. Strategies

While a conventional AS/RS usually comes with a single I/O point, the presented concepts contain a variety of I/O points. For efficient use of the additional I/O points new strategies are required. Since all of these points are equivalent, each point can be chosen to pick up or put down a storage unit. The strategies define which point will be selected in order to minimize the travel time of the S/R machine. The authors propose three different strategies:

1. Random selection of an I/O point
2. Selection of the nearest I/O point
3. Selection of an I/O point by minimizing the sum of the travel times from a starting shelf to I/O and from I/O to the next target shelf

The difference between the latter two strategies is illustrated in Figure 4. While the second strategy defines a selection of the I/O point for the travel from a storage shelf P1 to I/O independent of the next destination of the S/R machine (P2), the third strategy considers the location of the next target selecting the optimal I/O point.

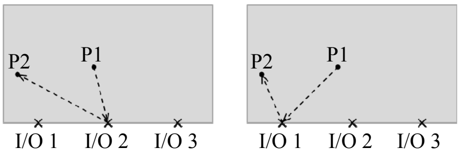


Figure 4: Selection of the nearest I/O point (left) and by minimizing the sum of the two travel times (right)

4. THE SIMULATION MODEL

To calculate the estimated travel times for the new concepts and strategies, a simulation model has been implemented. The model is used to determine the mean cycle time. For the travel time calculation using the simulation model, the following assumptions are made:

- randomized storage, i.e. every shelf of the rack is equally likely to be selected for storage or retrieval
- no empty trips between different I/O points
- constant acceleration and deceleration of the S/R machine
- equal absolute value of acceleration and deceleration

In addition, times such as the travel time between different storage locations in dual command cycles and pick-up and deposit times will be calculated by the simulation model to obtain the mean cycle time.

4.1. Modelling

The model consists of the rack, the S/R machine and conveyors to the I/O points for supply and removal of storage goods. The totality of the storage locations, characterized by its horizontal and vertical coordinates, form the rack. Since the S/R machine travels simultaneously in two directions, the travel time is given by the longer duration of the simultaneous travels in the two axis directions. For the calculation of the travel time in one axis direction, a distinction as to whether or not the maximum travel speed v_{max} will be reached has to be made. The maximum travel speed will be reached if the travel distance s satisfies the following condition (a corresponds to the absolute value of the acceleration and deceleration):

$$s \geq \frac{v_{\max}^2}{a} \qquad (1)$$

Depending on condition (1), the travel time can be calculated as follows:

$$\begin{cases} t = 2 \times \sqrt{\dfrac{s}{a}} & \text{for } s \leq \dfrac{v_{\max}^2}{a} \\ t = \dfrac{s}{v_{\max}} + \dfrac{v_{\max}}{a} & \text{for } s \geq \dfrac{v_{\max}^2}{a} \end{cases} \qquad (2)$$

Using equation (2) the travel times in the horizontal direction t_h and the vertical direction t_v can be calculated using the corresponding values for maximum speed and acceleration. The travel time T between two locations is given by the longer duration of t_h and t_v:

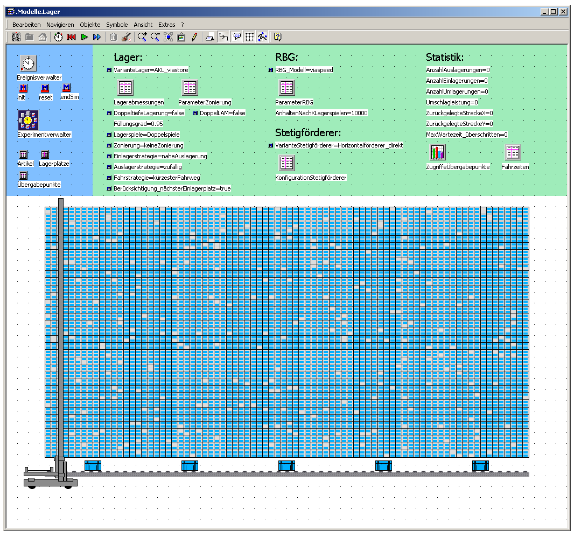
$$T = Max\left(t_h, t_v\right) \qquad (3)$$

The simulation model calculates the mean travel time by complete enumeration or Monte-Carlo-Simulation, depending on the desired operational strategies, using equations (2) and (3). I/O locations are selected according to the proposed strategies. Pick-up and deposit times are independent of the starting and target location and can be included afterwards to calculate the cycle time of an aisle.

The conveyors which connect the I/O points are modelled with constant conveying velocity with identical speeds for supply and removal of storage goods.

4.2. Implementation

The Siemens Plant Simulation software has been chosen for the implementation of the simulation model. The core of the model is a graphic visualization of the complete AS/RS with its storage locations, the S/R machine, the I/O points and the conveyors for the supply (see Figure 5) which has proven to be useful in the verification process and the analysis of the system's behaviour.



VarianteLager=AKL_viastore
DoppeltiefeLagerung=false
DoppelLAM=false
Füllungsgrad=0.95
Lagerspiele=Doppelspiele
Zonierung=keineZonierung
Einlagerstrategie=naheAuslagerung
Auslagerstrategie=zufällig
Fahrstrategie=kürzesterFahrweg
Berücksichtigung_nächsterEinlagerplatz=true

RBG_Modell=viaspeed
AnhaltenNachXLagerspielen=10000

VarianteStetigförderer=Horizontalförderer_direkt

AnzahlAuslagerungen=0
AnzahlEinlagerungen=0
AnzahlUmlagerungen=0
Umschlagleistung=0
ZurückgelegteStreckeX=0
ZurückgelegteStreckeY=0
MaxWartezeit_überschritten=0

Figure 5: Graphic visualization of the simulation model

Result variables and diagrams as well as different user input tables and variables for the model configuration can be found in the upper part of the visualization window. The model and its graphical visualization are generated dynamically based on the user's input. Input parameters are:

- rack and storage shelf dimensions
- accelerations and speeds of the S/R machine
- pick-up and deposit times
- operational strategies for storage assignment and retrieval
- number, position and alignment of the I/O points (vertical, horizontal)
- speed of the supply conveyors

Tables are used for the management of the storage locations and their content as well as for the I/O points and their position.

Actions are triggered by the movement of the S/R machine. Each time the machine reaches a target position, the functions defining the following action are called. If desired, the software also offers the possibility to visualize the machine's movement in real-time.

4.3. Verification and Validation

To verify the implemented model, the travel times between specific locations have been controlled manually. The cycle time calculation has been validated comparing the results for different rack dimensions with a single I/O point calculated using the simulation model to those gained from the analytical expressions proposed by Bozer and White (1984). For this purpose, an S/R machine travelling with constant speed had to be used. As for the later experiments, a horizontal speed of 6 m/s and a vertical speed of 3 m/s have been selected as well as rack dimensions (LxH) of 20x10 m, 20x20 m and 40x10 m with a storage shelf size of 0.5x0.4 m. Table 1 shows the mean travel times between storage shelves and I/O calculated in both ways. The simulation model produces exactly the same results as the analytical calculation.

Table 1: Validation results

Rack size	Simulation	Analytical calculation	Deviation
20x10 m	2.22 s	2.22 s	0.0%
20x20 m	3.61 s	3.61 s	0.0%
40x10 m	3.61 s	3.61 s	0.0%

5. SIMULATION EXPERIMENTS

The aim of the simulation experiments is the evaluation of concepts and strategies for AS/RS with multiple I/O points in terms of cycle time. Target results are:

- cycle times for different AS/RS configurations with multiple I/O points
- influence of the different strategies on the cycle time
- influence of the number of I/O points on the cycle time
- required supply speed for the I/O points

5.1. Setup description

Three rack dimensions with a different width-to-height ratio are used to evaluate I/O configurations. This is necessary since the shown concepts with I/O points distributed on the rack surface are expected to show different gains depending on the ratio of the rack dimensions. Starting with a rack having 20 m length and 10 m height, a second rack with twice the height and a third rack with twice the length will be considered. Table 2 summarizes the rack dimensions:

Table 2: Rack dimensions

Rack name	Length [m]	Height [m]
Rack 1	20	10
Rack 2	20	20
Rack 3	40	10

The distance between the shelves is 0.5 m in the horizontal and 0.4 m in the vertical direction, assuming a mini-load AS/RS. For such systems, S/R machines with horizontal travel speeds up to about 6 m/s are available. Vertical travel speeds are typically lower. Table 3 shows the speeds and mean accelerations of the exemplary device used which represents the current state of technology:

Table 3: Speeds and mean accelerations of the S/R machine

	Horizontal direction	Vertical direction
Max. travel speed	6.0 m/s	3.0 m/s
Mean acceleration	3.0 m/s²	3.0 m/s²

To simplify the matter, it will be assumed that speeds and accelerations are independent of the lift height and the absolute value of the deceleration is equal to the acceleration in the same direction. Pick-up and deposit times are supposed to have an identical duration of 4 s.

Of the concepts with multiple I/O points (see section 3.1.2), configurations with a different number of I/O points will be examined. Preliminary studies showed that the relative gain will decrease for an increasing number of I/O points and become very small for more than two or three I/O points, depending on the AS/RS configuration. To verify this, I/O configurations with up to five I/O points will be considered.

5.2. Results and discussion

A first result of simulations with multiple I/O points was that for most of the examined configurations the speed of conventional conveyors is far from sufficient to supply all of the I/O points from the pre-storage area in time. For this reason, more than one storage unit has to be buffered at the I/O points in order to decouple the supply process from the storage process.

Travel times and cycle times in the following tables are always indicated in seconds.

5.2.1. Comparison of strategies

Table 4 shows the results for the comparison of the strategies introduced in section 3.2 using *Rack 1* (20x10 m) and an I/O alignment according to *Concept 1* with a different number of I/O points. The mean travel time between storage shelves and I/O points is denoted by $t_{shelves - I/O}$; the mean travel time in the opposite direction is denoted by $t_{I/O - shelves}$.

Table 4: Mean travel times for the different strategies

# I/O	Random selection		Nearest I/O		Travel time minimization	
	$t_{shelves - I/O}$	$t_{I/O - shelves}$	$t_{shelves - I/O}$	$t_{I/O - shelves}$	$t_{shelves - I/O}$	$t_{I/O - shelves}$
2	3.26	3.26	2.79	3.26	2.91	2.91
3	3.30	3.30	2.71	3.28	2.84	2.84
4	3.31	3.31	2.68	3.29	2.81	2.81
5	3.31	3.31	2.66	3.30	2.79	2.79

The time for the travel from a storage shelf across an I/O station to the next storage shelf consists of the sum of the two indicated times. As expected, the selection of an I/O point by minimizing the travel time between two targets and I/O will always lead to the shortest total travel time. The selection of the nearest I/O point shows an even shorter travel time between shelves and I/O at the expense of the travel time in the opposite direction. By randomly selecting an I/O point, additional I/O points are not an advantage. The travel time will even increase slightly for an increasing number of I/O points. This strategy is thus not suited for use with multiple I/O points.

Similar results can be expected for other I/O alignments and rack configurations. Therefore, for the following evaluation of different concepts, the travel time minimization strategy will be used.

5.2.2. Comparison of I/O configurations

Different I/O configurations will be compared based on the dual command cycle time. For a dual command cycle starting at I/O the cycle time consists of the following time components:

1. Pick-up time at I/O
2. Travel time from I/O to the storage location
3. Deposit time at the storage location
4. Travel time between storage and retrieval location
5. Pick-up time at the retrieval location
6. Travel time from the retrieval location to I/O
7. Deposit time at I/O (simultaneous to 1.)

With the applied strategy the mean travel times between I/O and storage or retrieval location are identical and thus calculated only once. The mean travel time between storage and retrieval location is independent from the I/O configuration but varies for the examined rack dimensions (see Table 5).

Table 5: Mean travel times between storage and retrieval location

Rack size	Mean travel time
20x10 m	3.06 s
20x20 m	3.79 s
40x10 m	4.19 s

Since both pick-up and deposit times are assumed to have a duration of 4 s and pick-up and deposit at I/O can take place simultaneously in subsequent cycles, a total of 12 s has to be added to the calculated travel times to get the dual command cycle time. Table 6 contains the results of the cycle time calculation for the four reference concepts as well as for the concepts with multiple I/O points in configurations with up to five I/O points for *Rack 1*. This rack is “square-in-time”.

Table 6: Dual command cycle time for different I/O configurations (rack 20x10 m)

I/O alignment	# I/O				
	1	2	3	4	5
Reference 1	22.64 100%	-	-	-	-
Reference 2	21.22 94%	-	-	-	-
Reference 3	22.18 98%	-	-	-	-
Reference 4	20.28 90%	-	-	-	-
Concept 1	-	20.88 92%	20.74 92%	20.68 91%	20.64 91%
Concept 2	-	19.78 87%	19.60 87%	19.52 86%	19.46 86%
Concept 3	-	22.08 98%	22.04 97%	22.04 97%	22.02 97%
Concept 4	-	20.12 89%	20.06 89%	20.04 89%	20.04 89%

The results for *Rack 1* (20x10 m) show that already a differently placed I/O point leads to a significant reduction of the cycle time. By shifting the I/O point towards half the length of the rack the cycle time can be reduced by 6% (*Reference 2*) or by even 10% placing the I/O point in the rack’s centre of area (*Reference 4*), compared to the most frequent I/O alignment in practical applications *Reference 1*. The cycle time is shorter than for some of the concepts with multiple I/O points. Those concepts (*Concept 1* and *3*) have the common ground of I/O points aligned alongside a border of the rack, which is expected to be easier to implement compared to I/O points distributed on the rack surface.

Multiple I/O points on the rack surface might be difficult to implement, but feature shorter cycle times than all of the concepts with a single I/O point (*Concept 2* and *4*). Compared to *Reference 1*, a reduction of the mean cycle time by up to 14% is possible with those concepts.

For all of the examined concepts with multiple I/O points the benefit of using more than two I/O points is very small. The benefit is typically less than 1% in relative terms. Therefore, two I/O points are sufficient for this AS/RS configuration with rack dimensions of 20x10 m. The results for the higher *Rack 2* (20x20 m) are shown in Table 7.

Table 7: Dual command cycle time for different I/O configurations (rack 20x20 m)

I/O alignment	# I/O				
	1	2	3	4	5
Reference 1	25.59 100%	-	-	-	-
Reference 2	24.87 97%	-	-	-	-
Reference 3	23.37 91%	-	-	-	-
Reference 4	21.95 86%	-	-	-	-
Concept 1	-	24.63 96%	24.57 96%	24.53 96%	24.51 96%
Concept 2	-	21.61 84%	21.47 84%	21.41 84%	21.37 84%
Concept 3	-	23.13 90%	23.01 90%	22.97 90%	22.93 90%
Concept 4	-	21.57 84%	21.37 84%	21.29 83%	21.25 83%

Due to the greater rack surface (20x20 m), cycle times are longer than for the first rack (20x10 m). Looking at the results for the reference concepts, it is noticeable that *Reference 3* with an elevated I/O point gives much better results for these rack dimensions than for the first examined rack using the same exemplary S/R machine. The reason for this is the increased rack height. *Reference 2* with an I/O point centred at the bottom of the rack surface has a much smaller advantage in comparison to the first rack. These results show that elevated I/O points are better suited to higher racks, using the same S/R machine. Comparing the results for the concepts with multiple I/O points will confirm this. *Concept 1* with the I/O points located at the bottom of the rack shows the smallest benefit while the other concepts, characterized by differently placed elevated I/O points, allow for cycle time reductions up to 17% compared to *Reference 1*.

Also the results for Rack 2 confirm that it is generally not convenient to use more than two I/O points. Table 8 contains the results for the last of the considered racks, the proportionally long *Rack 3* (40x10 m).

Table 8: Dual command cycle time for different I/O configurations (rack 40x10 m)

I/O alignment	# I/O				
	1	2	3	4	5
Reference 1	26.99 100%	-	-	-	-
Reference 2	23.77 88%	-	-	-	-
Reference 3	26.75 99%	-	-	-	-
Reference 4	23.31 86%	-	-	-	-
Concept 1	-	23.13 86%	22.87 85%	22.75 84%	22.67 84%
Concept 2	-	22.45 83%	22.15 82%	22.01 82%	21.91 81%
Concept 3	-	26.69 99%	26.67 99%	26.67 99%	26.67 99%
Concept 4	-	23.21 86%	23.17 86%	23.17 86%	23.15 86%

As expected, a comparison with the results of *Rack 2* shows that different concepts are suited to a proportionally long rack, always in combination with the given S/R machine. Whereas *Concept 1* offers only a small benefit for *Rack 2*, a noticeable reduction of the cycle time is possible for this rack. Adopting *Concept 2*, a further reduction of the cycle time by up to 19% compared to *Reference 1* will be possible, depending on the number of I/O points. In contrast, alignments with elevated I/O points at the border of the rack (*Reference 3* and *Concept 3*) don't show any significant benefit.

While the optimal strategy could be determined independently from the rack dimensions, the results for different I/O alignments vary for the examined rack dimensions. In particular, concepts with the I/O points aligned along a border of the rack are very sensitive to the rack dimensions. Therefore, a detailed analysis for a given application case is necessary for those concepts and the best suited concept as well as a reasonable number of I/O points have to be determined ad hoc.

6. CONCLUSIONS

In this paper, the idea of reducing the mean travel distance between storage shelves and I/O in an AS/RS by using more than one I/O points has been addressed. The paper proposes different concepts for the alignment of multiple I/O points in an AS/RS aisle. For those concepts, possible strategies for the selection of an I/O point were presented. After choosing an adequate strategy, I/O configurations with a different number of I/O points differently aligned were evaluated in terms of travel times. For this purpose, a simulation model has been developed. The evaluation has been done for three different sizes of rack face. As reference for the evaluation, four systems with differently placed single I/O points were defined.

The simulation results emphasize the benefit of the analyzed concepts. Although a reduction of the mean cycle time compared to an AS/RS with a single I/O point in the corner of the rack can already be achieved by placing the I/O point differently, the implementation of a concept with multiple I/O points yields a further reduction.

In addition, the outcome of the evaluation implies that by using a given S/R machine not every I/O configuration is suitable for a particular examined rack dimension. As a consequence, the obtainable cycle time reduction for a new configuration cannot be generalized but has to be calculated ad hoc using simulation. This is possible with reasonable effort using the developed configurable simulation model.

REFERENCES

Arantes, J. C., Kompella, S., 1993. Travel-time models for AS/RS with multiple I/O stations. *2nd Industrial Engineering Research Conference Proceedings*, IIE, Norcross, GA (USA), 405-409 – ISBN 0-89806-132-6

Bozer, Y. A., White, J. A., 1984. Travel time models for automated storage/retrieval systems. *IIE Transactions*, 16(4), 329-338 – ISSN 0470-817X

Eggert, M., Loschke, C., Schumann, M., 2010. Neuer Ansatz verspricht Effizienzschub. *f+h – Fördern und Heben 7/8*, 264-267 – ISSN 0341-2636

Gudehus, T., 1972a. Wohin mit der Kopfstation? *Materialfluß 2*, Nr. 4, 66-68

Gudehus, T., 1972b. Grundlagen der Spielzeitberechnung für automatisierte Hochregallager, *deutsche hebe- und fördertechnik Sonderheft*, 63-68

Gudehus, T., 2007. *Logistik 2*, Netzwerke, Systeme und Lieferketten. Berlin: Springer Verlag, 640

Knepper, L., 1980. Leistungsverbesserungen in Hochregallagern durch optimale Anordnung der Ein- und Auslager-Bereitstellplätze. *f+h – Fördern und Heben 30*, Nr. 12, 1096-1099

MHI, 2013. MHI glossary. Available from: http://www.mhi.org/glossary [accessed 13 June 2013]

Roodbergen, K.J., Vis, I.F.A., 2009. A survey of literature on automated storage and retrieval systems. *European Journal of Operational Research*, 194(2), 343-362 – ISSN 0377-2217

AUTHORS BIOGRAPHY

Lantschner Daniel, M.Sc., Ph.D. candidate and Research Assistant at the Institute of Materials Handling, Material Flow, Logistics, Technische Universität München. Daniel Lantschner was born in 1983 in Bolzano, Italy. Between 2003 and 2007 he studied Logistics and Production Engineering at the Free University of Bozen - Bolzano and afterwards Mechanical Engineering at the Technische Universität München until 2009.

Atz Thomas, M.Sc., Ph.D. candidate and Research Assistant at the same institute. Thomas Atz was born in 1985 in Bolzano, Italy. Between 2004 and 2007 he studied Logistics and Production Engineering at the Free University of Bozen - Bolzano and afterwards Mechanical Engineering at the Technische Universität München until 2009.

Günthner Willibald A., Prof. Dr., Head of the Institute of Materials Handling, Material Flow, Logistics, Technische Universität München.

MODELING, SIMULATION AND OPTIMIZATION OF THE MAIN PACKAGING LINE OF A BREWING COMPANY

N. Basán [a], L. Ramos [b], M. Coccola [c], C. A. Méndez [d]*

[a] [b] [c] [d] INTEC (UNL - CONICET), Güemes 3450, 3000 Santa Fe, Argentina.

[a]basan.natalia@gmail.com, [b]ramos.lucila@gmail.com, [c]mcoccola@intec.unl.edu.ar, [d]*cmendez@intec.unl.edu.ar

ABSTRACT

Discrete event simulation (DES) techniques cover a broad collection of methods and applications that allow imitating, assessing and predicting the behavior of complex real-world systems. The main purpose of this work is to develop a novel DES model to optimize the design and operation of a complex beer packaging system in order to perform a sensitivity analysis to find one or more alternatives to increase productivity levels. In this way, advanced technologies of modeling, simulation and optimization for system design and operation are applied. The model is developed by using the DES tools provided by the SIMIO simulation software. The proposed tool is able to carry out evaluations of the system using a 3D user-friendly graphical interface that shows the dynamic evolution of the system over time. By using the proposed simulation model, the results of this paper illustrate how the levels of productivity may vary by reducing micro-downtime of machines, when transport rates and other problem features are properly changed with them.

Keywords: simulation, optimization, packaging line, brewing company

1. INTRODUCTION

In modern production processes, quality takes precedence in relation to production volume. A fundamental topic in market requirements is the product presentation. It implies that production lines need an additional number of machines to perform a wider variety of tasks. In this way, all activities must be conducted with the highest possible quality. Therefore, the growing demand and specialization in the presentation of the product make packaging lines more complex. More diversified tasks are performed on them.

This work arises from the need to identify, analyze and reduce the causes affecting the productivity of the main packaging line of an international beer company located in the province of Santa Fe, Argentina. Currently, the efficiency of the company's lines is lower than the level suggested by the managers. This situation has a directly impact on the current production level due to the packaging process is an essential step in the whole production process.

This work aims to optimize the design and operation of the main packaging line of the company in order to improve the global efficiency. In this way, a comprehensive simulation-based model was developed so that different future commitments and changing market conditions can be easily suited in the medium or short term.

2. METHODOLOGY

Having stated the general and specific objectives of the project, all the necessary data from the company under studying was collected by using five different techniques: (i) staff interviews, (ii) in-situ observation, (iii) historical data collection, (iv) reading manuals, and (v) review of equipment and transport.

Once the required information was collected, this one was analyzed, filtered and documented. Such procedure allows us to identify critical points and potential problems to be solved in the current and desired situation of the packaging process.

After understanding the real system that is subject of our simulation study, the conceptual model was developed. If we need to develop a model on a simulator, we need to determine the level of abstraction at which to work. This process of abstracting a model from the real world is known as conceptual modeling. For this particular project, SIMIO simulation software was chosen because this one allows to build animated models in three dimensions (3D), facilitating the verification and validation of the simulation model.

The inherent advantages of the simulation model developed were highlighted by solving three scenarios: (i) theoretical or ideal, (ii) current, and (iii) suggested scenario. A sensitive analysis had to be conducted to determine the more suitable alternatives regarding to the overall performance of the company, and consequently, the expected economical benefits.

2.1. Production Process

The beer production process comprises a series of manufacturing steps depending on the type of beer, varying the amount and type of raw material. Such steps are: malting, malt milling, mashing, cooking, wort cooling and clarification, fermentation, maturation, and the packaging process at the end.

2.1.1. Packaging Process

A packaging line involves a set of machines, equipment units and tools needed to perform the process operations. The success of the line depends on proper coordination of different elements. Similarly, in the case of a packaging line of a brewing company, beer and containers move through a series of processing stages until the final product is obtained. The flowchart is shown in Figure 1.

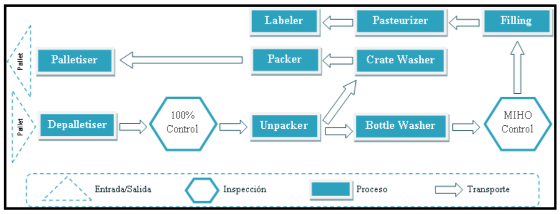


Figure 1: Beer packaging process

The first operation performed in a beer packaging line is depalletizing and disassemblying the pallet that comes from the deposit with empty bottles.

The next step, 100% control, aims to eliminate most of the waste entering with the drawers. This inspection is carried out manually by an operator. He removes items or bottles that could possibly harm the following machines.

Then, the process proceeds to perform the unpacking by a computer that has the function of extracting the empty bottles boxes that feeds the packaging line. The objective of the operation is to separate the drawer containers for subsequent washing operations on the respective machines.

After unpacking, the bottles are guided through the transport system to the washing process. Because all returnable bottles should be sanitized before being filled with beer again, the goal of this stage is to perform the physical and biological cleaning, removing all dirt, labels, adhesive and foil.

While we can assume that all bottles are dirt free in the washing machine, there is a risk this stage has not been able to completely remove all cleaning agents. Therefore, the output of this machine is a containers inspector.

Next, we proceed to perform filling and topped with beer in containers. The aim of the operation is to transfer product from a pot bulk and individual containers with airtight lids seal rolled steel to ensure durability, quality and inviolability of beer. Subsequently, through the process of pasteurization is achieved that beer is kept in ideal state at least until the date of minimum durability, i.e. the primary objective of pasteurization is to avoid possible biological decomposition and lengthen the bottled product.

The next step aims to place the labels presented in the final product. The labeling process begins when filled and capped bottles entering the labeling machine, and ends with a level-cap inspection rejecting bottles that do not meet any of the required characteristics in terms of filling level, internal pressure, and missing state missing labels and cap.

Once the bottle labeling operation is performed encased, which is contrary to the operation of unpacking. The machine is designed for gripping and moving sets of bottles into crates synchronously entering the computer.

Finally, we proceed to make palletizing, i.e. drawers are placed on a wooden stand known as pallet or pallet for easy handling and transportation future. The arrangement of the crates on the pallet is performed in layers according to a set distribution, in order to form a compact load unit capable of supporting stable after storage, transport and distribution.

On the other hand, different units are involved in the packaging process using transport. These are between the various machines in the line, providing a connecting element and synchronism between two of them. The rate is fixed by the same variable speed drives, and startup and shutdown is done by proximity sensor and optical detectors. The elements considered along the packaging process are pallets, crates and bottles.

2.2. Simulation Model

Process simulation is one of the most useful tools of industrial engineering, which is used to represent a complex process by another which makes it simple and understandable.

The proposed simulation model was developed by using the SIMIO programming package, which is one of the most specialized software in the area of process simulation that minimizes the risk and uncertainty in decision making, as well as minimizing the costs by improving the use of resources, reduced time spent and the minimization of the probability of risk.

Likewise, the packaging process involves five classes of modeling elements:

- Bottles that run the line
- Crates of bottles
- Machines that perform the operations necessary to prepare and process the product
- Transports located between different machines that make up the line
- Operators who have assigned different tasks.

2.3. SIMIO Simulation Software

SIMIO is modern flow simulation software for discrete event processes, and procedures based on objects. SIMIO allows conducting a simulation project in a much shorter time than usual. It is the first software that combines simulation modeling speed allowed by the object-oriented technology with the flexibility and power of procedures. This software enables the modeler to build animated models in three dimensions (3D) in one third of the usual time, and thus frees up time to devote to analysis of alternatives and scientifically informed decision-making.

Therefore, to model the packing process, it is necessary to represent the major components of the system, i.e. the products (bottles, boxes, pallets), machines, inspectors, transport and accumulation tables that form.

Different views of the simulation model of the main packaging line, developed in the SIMIO simulation software, are shown in Figures 2-4.

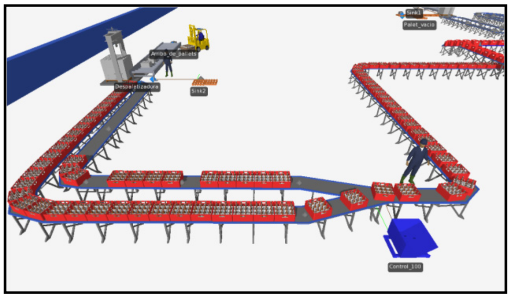


Figure 2: 3D SIMIO model (despalletizer)

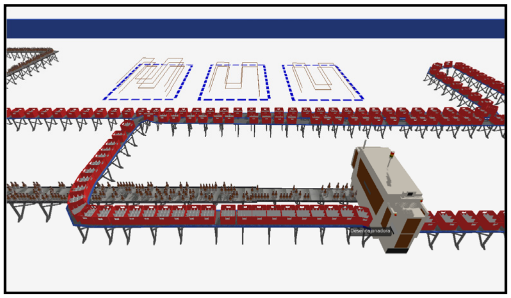


Figure 3: 3D SIMIO model (unpacker)

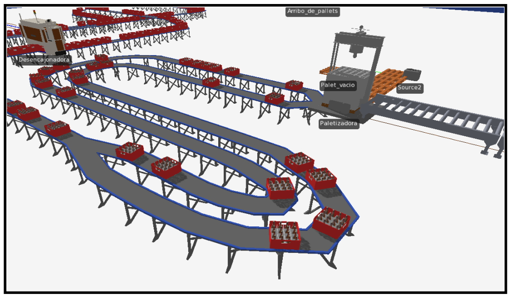


Figure 4: 3D SIMIO model (palletizer)

2.4. Creation of Simulation Model

For the creation of the simulation model, standard elements as source, process and sink were used, connected by a set of paths. The following components were identified in the real packaging line:

2.4.1. Pallets, Boxes and Bottles

As shown Figure 5, the dynamic entities moving through the system are: (i) pallets, (ii) drawers, and (iii) bottles.

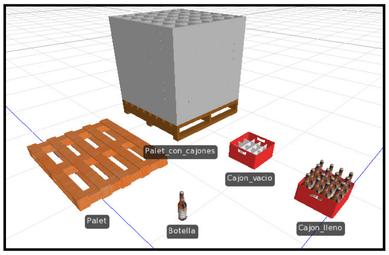
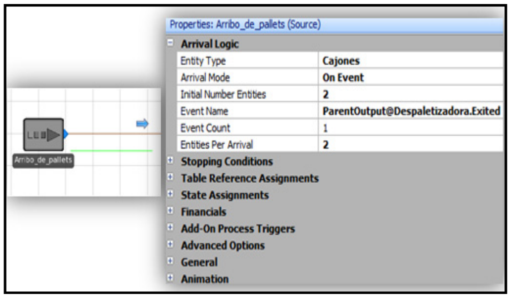


Figure 5: System entities

The "Source" module is used to create entities that arrive to the system. Figure 6 shows as pallet entity arrival is defined.



Properties: Arribo_de_pallets (Source)

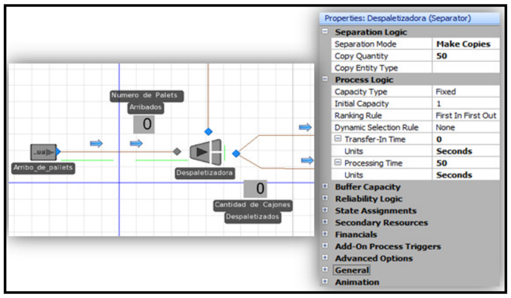
Arrival Logic	
Entity Type	Cajones
Arrival Mode	On Event
Initial Number Entities	2
Event Name	ParentOutput@Despaletizadora.Exited
Event Count	1
Entities Per Arrival	2

Stopping Conditions
Table Reference Assignments
State Assignments
Financials
Add-On Process Triggers
Advanced Options
General
Animation

Figure 6: Pallet entity arrivals.

2.4.2. Depalletizer and Unpacker Machines

The depalletizer is the first equipment unit in the packing process. The module "Separator" provided by SIMIO is used to represent the depalletizer's operation, which is shown in Figure 7. On the right side of this picture, we can see the properties associated with the "Separate" module, i.e. the property "Processing Time" determines that each pallet is processed in a time of 50 seconds. Each pallet that is full of drawers enters to depalletizer to be processed and then 50 new entities representing the drawers are generated by the model.



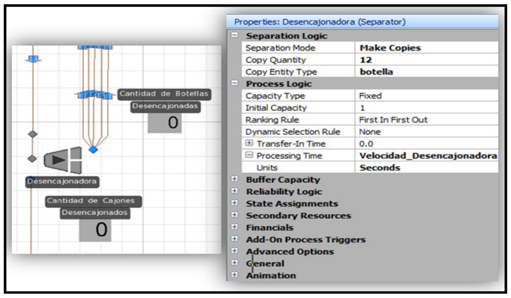
Properties: Despaletizadora (Separator)

Separation Logic	
Separation Mode	Make Copies
Copy Quantity	50
Copy Entity Type	
Process Logic	
Capacity Type	Fixed
Initial Capacity	1
Ranking Rule	First In First Out
Dynamic Selection Rule	None
Transfer-In Time	0
Units	Seconds
Processing Time	50
Units	Seconds

Buffer Capacity
Reliability Logic
State Assignments
Secondary Resources
Financials
Add-On Process Triggers
Advanced Options
General
Animation

Figure 7: 2D SIMIO model (depalletizer machine)

The unpacker machine is modeled in the same way that depalletized equipment. Each drawer that is full of bottles enters to unpacker to be processed and then 12 new entities representing the bottles are generated by the model. Figure 8 shows the "Separator" module associated to the operation of unpacker unit.



Properties: Desencajonadora (Separator)

Separation Logic	
Separation Mode	Make Copies
Copy Quantity	12
Copy Entity Type	botella
Process Logic	
Capacity Type	Fixed
Initial Capacity	1
Ranking Rule	First In First Out
Dynamic Selection Rule	None
Transfer-In Time	0.0
Processing Time	Velocidad_Desencajonadora
Units	Seconds

Buffer Capacity
Reliability Logic
State Assignments
Secondary Resources
Financials
Add-On Process Triggers
Advanced Options
General
Animation

Figure 8: 2D SIMIO model (unpacker machine)

Between two operations describe above, there is an Inspection Process (Control 100%) that controls the drawers entering to the line. The representation in SIMIO of this process is shown in Figure 9. As we can see in the picture, a "Decide Step" that uses a probabilistic distribution is defined so that defective entities can be rejected.

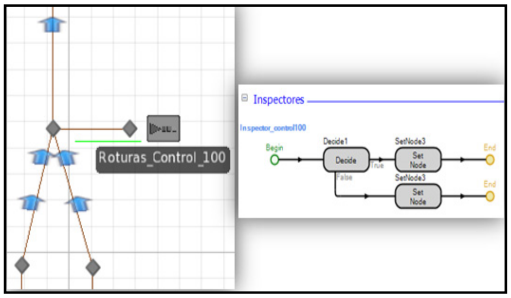


Figure 9: 2D SIMIO model (Control 100%)

2.4.3. Bottles Washer

This operation is built by placing 40 "Conveyor" objects in the SIMIO model. Each conveyor represents a real conveyor belt, which has a capacity of transporting until 710 bottles and a fix speed assuring that the bottles will be in the machine the minimum required time (45 minutes). Figure 10 shows the input/output logic of this stage.

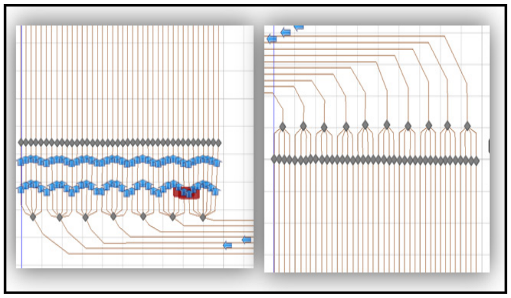


Figure 10: 2D SIMIO model (washing machine)

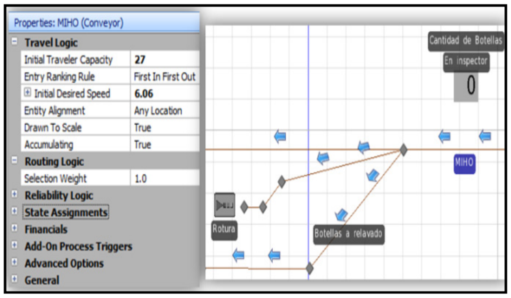
For the creation of the input logic to the washing machine, standard elements as events and timer were used. A group of 40 bottles enters to the washing process every 2 seconds

2.4.4. Empty Bottles Inspector

This stage aims to verify the bottles that previously have been processing in the washing machine. As shown in Figure 11, a basic node is used to represent this operation. Such node has one input path and three outputs path. The first output path receives the bottles that have a physical defect. The bottles that have some dirt are sent by the second output path. Finally, the accepted bottles continue their normal processing by the third output path.

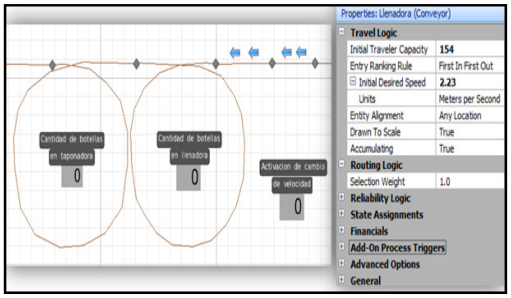
2.4.5. Filling Machine

This processing stage is represented by a conveyor that has a transportation capacity of 154 bottles (equal to the amount of filling valves). The capping machine, which is then, has the same processing capacity too.


Properties: MIHO (Conveyor)

⊟ **Travel Logic**	
Initial Traveler Capacity	**27**
Entry Ranking Rule	First In First Out
⊞ Initial Desired Speed	**6.06**
Entity Alignment	Any Location
Drawn To Scale	True
Accumulating	True
⊟ **Routing Logic**	
Selection Weight	1.0
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	

Figure 11: 2D SIMIO model (bottles inspector)



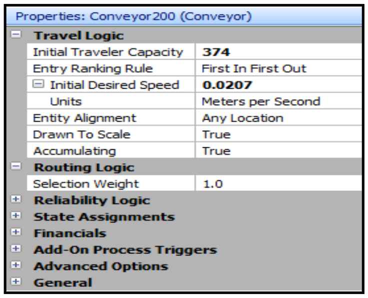
Properties: Llenadora (Conveyor)

⊟ **Travel Logic**	
Initial Traveler Capacity	**154**
Entry Ranking Rule	First In First Out
⊟ Initial Desired Speed	**2.23**
Units	Meters per Second
Entity Alignment	Any Location
Drawn To Scale	True
Accumulating	True
⊟ **Routing Logic**	
Selection Weight	1.0
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	

Figure 12: 2D SIMIO model (filling machine)

2.4.6. Pasteurizer Machine

The pasteurizing machine has two floors which were represented in SIMIO by 60 conveyors working in parallel (processed capacity of the equipment). In this stage, the bottles cross by "rainfall areas" that give water at different temperatures.


Properties: Conveyor200 (Conveyor)

⊟ **Travel Logic**	
Initial Traveler Capacity	**374**
Entry Ranking Rule	First In First Out
⊟ Initial Desired Speed	**0.0207**
Units	Meters per Second
Entity Alignment	Any Location
Drawn To Scale	True
Accumulating	True
⊟ **Routing Logic**	
Selection Weight	1.0
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	

Figure 13: 2D SIMIO model (pasteurizing machine)

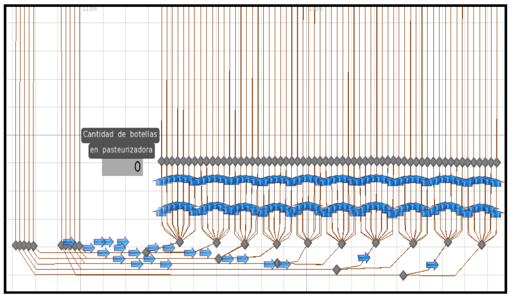
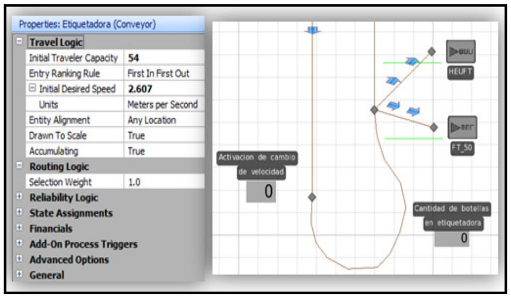


Figure 14: 2D SIMIO model (pasteurizing machine)

2.4.7. Labeler Machine

This equipment unit has an operation similar to the filler so that both processes were modeled in the same way (see Figure 15). Two inspectors look at the bottles to the end of this stage. This control process is defined similarly to Control 100% process described above. Besides, in order to compute the total amount of rejected bottles, two "Sink" component were used in the simulation model (HUEFT and FT_50).



Properties: Etiquetadora (Conveyor)

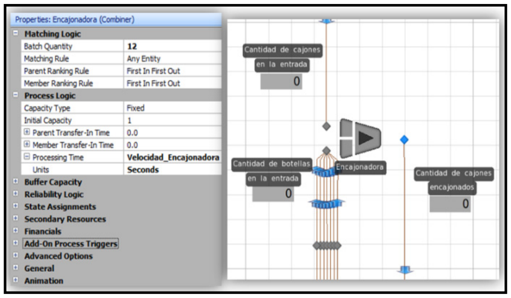
Travel Logic	
Initial Traveler Capacity	**54**
Entry Ranking Rule	First In First Out
⊟ Initial Desired Speed	**2.607**
Units	Meters per Second
Entity Alignment	Any Location
Drawn To Scale	True
Accumulating	True
Routing Logic	
Selection Weight	1.0
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	

HEUFT
FT_50
Activacion de cambio de velocidad
0
Cantidad de botellas en etiquetadora
0

Figure 15: 2D SIMIO model (labeler machine)

2.4.8. Packer and Palletizer Machine

A "Combiner" module was defined in the simulation model to represent the behavior of the packer and palletizer machine (Figure 16 and Figure 17). In the first process, 12 bottles are assembled into a drawer. After that, the palletizer process puts together 50 drawers in a pallet (10 drawers per stack, 5 stacks per pallet). Then, the complete pallets are sent to storage modeled with a "Sink" module.

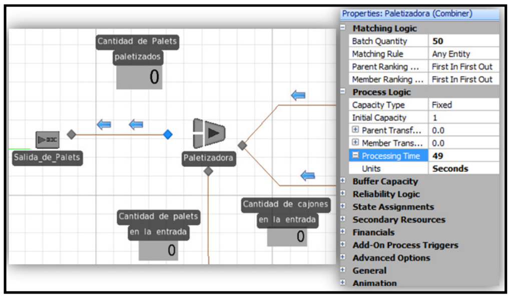


Properties: Encajonadora (Combiner)

Matching Logic	
Batch Quantity	**12**
Matching Rule	Any Entity
Parent Ranking Rule	First In First Out
Member Ranking Rule	First In First Out
Process Logic	
Capacity Type	Fixed
Initial Capacity	1
⊞ Parent Transfer-In Time	0.0
⊞ Member Transfer-In Time	0.0
⊟ Processing Time	**Velocidad_Encajonadora**
Units	**Seconds**
⊞ **Buffer Capacity**	
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Secondary Resources**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	
⊞ **Animation**	

Cantidad de cajones en la entrada
0
Cantidad de botellas en la entrada
0
Encajonadora
Cantidad de cajones encajonados
0

Figure 16: 2D SIMIO model (packer process)



Cantidad de Palets paletizados
0
Salida_de_Palets
Paletizadora
Cantidad de palets en la entrada
0
Cantidad de cajones en la entrada
0

Properties: Paletizadora (Combiner)

Matching Logic	
Batch Quantity	**50**
Matching Rule	Any Entity
Parent Ranking ...	First In First Out
Member Ranking ...	First In First Out
Process Logic	
Capacity Type	Fixed
Initial Capacity	1
⊞ Parent Transf...	0.0
⊞ Member Trans...	0.0
⊟ Processing Time	**49**
Units	**Seconds**
⊞ **Buffer Capacity**	
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Secondary Resources**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	
⊞ **Animation**	

Figure 17: 2D SIMIO model (palletizer process)

2.4.9. Accumulation Tables

The accumulation tables ensure a constant supply of bottles or crates in the equipment that are after them. Since the machines are exposed to internal faults, these tables assured that if an equipment is broken, the rest of machines that are upstream can follow working.

On the packaging line there are three accumulation tables, two for bottles and one for drawers. The first accumulation table is located between the empty bottles inspector and the filling machine (see Figure 18). The second one is located between the pasteurizer equipment and labeling machine. Finally, the drawer accumulator is between the unpacker machine and the packer machine.

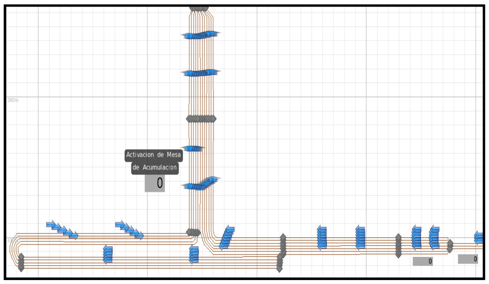
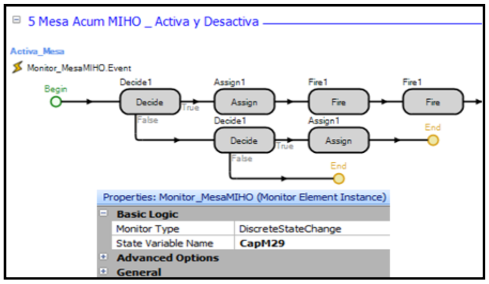


Figure 18: 2D SIMIO model (accumulation table for bottles)

Figure 19 shows as a "Monitor" element can be used to control the capacity of the conveyor that is above the accumulation tables for bottles. If the capacity of conveyor changes, a process called "Activar_Mesa" is trigger by the monitor.



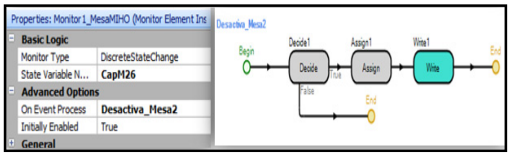
Properties: Monitor_MesaMIHO (Monitor Element Instance)

Basic Logic	
Monitor Type	DiscreteStateChange
State Variable Name	**CapM29**
⊞ **Advanced Options**	
⊞ **General**	

Figure 19: 2D SIMIO model (Monitor element)

In addition, a binary variable named "Activa_Mesa1" determines the current state of table 1. If the table is working, active_mesa1 is equal to 1; otherwise, it is set to zero.

Added to the above, the transport states located before or after of buffer are monitored. In this way, when table tapes are empty, the accumulating table is disabled by stopping transports and assigning a value of 0 to the associated binary variable. Figure 20 shows the monitor of one of the transports mentioned and the process associated with the deactivation of the buffer.



Properties: Monitor1_MesaMIHO (Monitor Element Ins

Basic Logic	
Monitor Type	DiscreteStateChange
State Variable N...	**CapM26**
Advanced Options	
On Event Process	**Desactiva_Mesa2**
Initially Enabled	True
⊞ **General**	

Figure 20: Logic associated with deactivating the first accumulation table

2.4.10. Drawers Combiners

On drawers transport line there are two combiners aiming to join two drawers belts into a single or reversely. The first combiner is located after depalletizer machine, more precisely where Control 100% is performed. Figure 21 shows as in this stage the two drawers lines are combiner into a single belt. On the other hand, the second combiner is situated before palletizer machine and its function is to divide the conveyor belt from the packer machine in two rows.

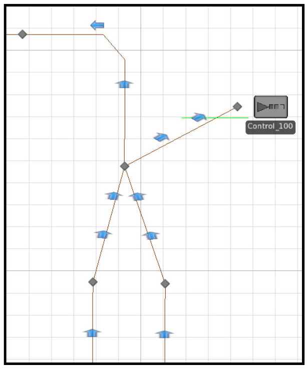


Figure 21: 2D SIMIO model (Drawers combiner)

Each combiner has a predefined logic, which is defined from processes and is associated with the transport involved. The first combiner transport has a longer length and other shorter length in parallel. So it allows passing more drawers with greater capacity in order to achieve a balance in the accumulation of the conveyors involved. It is worth to note that when one of them is moving, the other stops running.

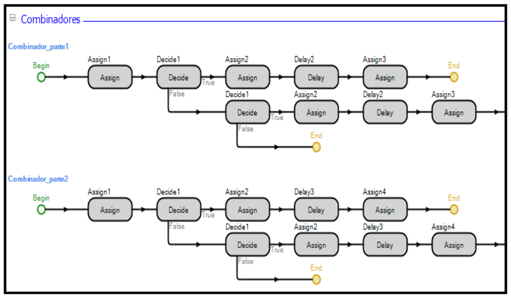
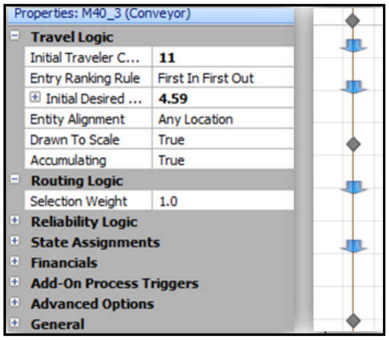


Figure 22: SIMIO processes

2.4.11. Transports

There are two transport lines, one for bottles and other for drawers. "Basic Node" and "Conveyor" elements were used in the simulation model to represent the two transport lines. It is worth to remark that some components of SIMIO have important parameters that must be set by the user. In particular, some "Conveyor" properties are given in Figure 23. From the picture, it follows that these properties might be used to vary things like conveyor speeds, traveler capacity, or the option for accumulating or non-accumulating conveyors


Properties: M40_3 (Conveyor)

⊟ Travel Logic	
Initial Traveler C...	11
Entry Ranking Rule	First In First Out
⊞ Initial Desired ...	4.59
Entity Alignment	Any Location
Drawn To Scale	True
Accumulating	True
⊟ Routing Logic	
Selection Weight	1.0
⊞ Reliability Logic	
⊞ State Assignments	
⊞ Financials	
⊞ Add-On Process Triggers	
⊞ Advanced Options	
⊞ General	

Figure 23: SIMIO simulation software (conveyor properties)

On drawers line there are only single conveyors. Instead, the bottle conveying line has conveyors of different widths. Thus, from one to ten bottles can be transported in parallel. An overview of this variable capacity transport is given in Figure 24.

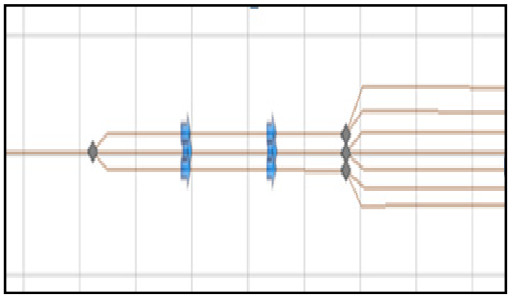
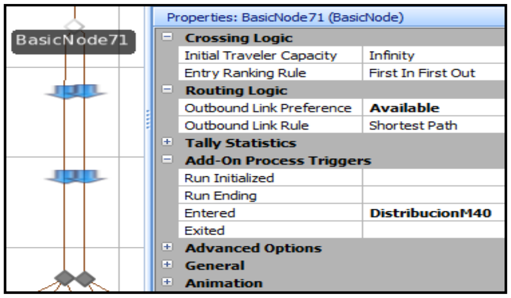


Figure 24: SIMIO simulation software (bottle conveying line)

In order to join transports with different carry capacities, several processes, whose logic is embedded within "Basic Nodes" elements, were defined in the simulation model (see Figure 25). Each process uses a discrete probability so that the bottles can be distributed on conveyors having available capacity. If any of the selected conveyors is on the limit of its capacity, other one in parallel must be chosen.



Properties: BasicNode71 (BasicNode)

⊟ Crossing Logic	
Initial Traveler Capacity	Infinity
Entry Ranking Rule	First In First Out
⊟ Routing Logic	
Outbound Link Preference	Available
Outbound Link Rule	Shortest Path
⊞ Tally Statistics	
⊟ Add-On Process Triggers	
Run Initialized	
Run Ending	
Entered	DistribucionM40
Exited	
⊞ Advanced Options	
⊞ General	
⊞ Animation	

Figure 25: 2D SIMIO model (distribution processes)

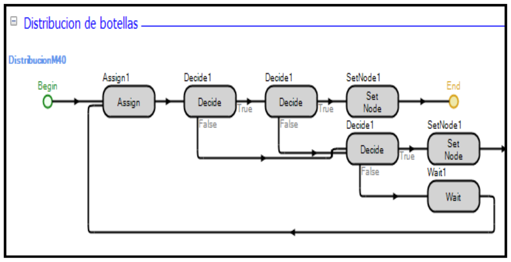
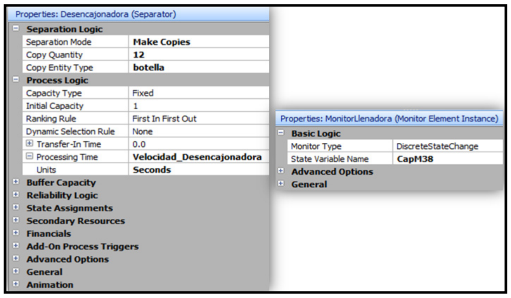


Figure 26: SIMIO processes (Bottles distribution)

2.4.12. Sensors

Several sensors control the number of bottles or drawers that are on the line. Such devices, located on strategic points of conveyor belts, emit signals so that transports or machines can start or stop their activities. These sensors are switches that are activated or deactivated according to whether they are in contact with the object.

To represent the above behavior, “monitor” and “variable” elements were used in the simulation model so that the logic of each machine can be properly defined. For example, the unpacker machine has three possible states: (i) stopped, (ii) low speed or (iii) high speed. A variable was defined to determine the machine state at a given time. The possible values of this variable are: 0 (if the machine is stopped), 1 (if the equipment is operated at low speed) or 2 (if the machine is running at high speed). In addition, three monitors were defined for associated transports. If a capacity change is detected in them, the monitors trigger a process determining the speed at which the equipment should operate. This value is then saved in a predefined variable. On one hand, if there is no accumulation in output transport and there are drawers in input transport, the machine operates at low speed. On the other hand, if there is accumulation in the input conveyor, the machine changes to high speed.



Properties: Desencajonadora (Separator)

⊟ **Separation Logic**	
Separation Mode	**Make Copies**
Copy Quantity	**12**
Copy Entity Type	**botella**
⊟ **Process Logic**	
Capacity Type	Fixed
Initial Capacity	1
Ranking Rule	First In First Out
Dynamic Selection Rule	None
⊞ Transfer-In Time	0.0
⊟ Processing Time	**Velocidad_Desencajonadora**
Units	**Seconds**
⊞ **Buffer Capacity**	
⊞ **Reliability Logic**	
⊞ **State Assignments**	
⊞ **Secondary Resources**	
⊞ **Financials**	
⊞ **Add-On Process Triggers**	
⊞ **Advanced Options**	
⊞ **General**	
⊞ **Animation**	

Properties: MonitorLlenadora (Monitor Element Instance)

⊟ **Basic Logic**	
Monitor Type	DiscreteStateChange
State Variable Name	**CapM38**
⊞ **Advanced Options**	
⊞ **General**	

Figure 27: 2D SIMIO model (unpacker machine properties and accumulation monitor charecteristics)

2.4.13. Model Verification and Validation

In order to perform a verification of the simulation model developed, a detailed analysis of each packaging process operations was accomplished. This assures us that model logic properly represents the sequence of operations of the real process.

In addition, the model validation executes an iterative comparison with the real system, making the necessary adjustments and changes in the model until a satisfactory similarity is achieved.

2.5. Sensitivity Analysis

Therefore, after having identified the major operational problems, a series of changes is proposed to the design and operational model which a priori could increase the production capacity of the company, increasing the level of efficiency and address weaknesses the process. This alternative scenarios raised by changing the values of the factors considered most relevant.

- Scenario 1: system with engine capacities and speeds and actual transport, and percentages of rejections of inspectors,
- Scenario 2: system based on theoretical speeds of the machines, provided the design of the line ("V Line").
- Scenario 3: system considered in scenario 2, where, in addition, use is made of a battery drawer between the packer and unpacker machine,
- Scenario 4: system considered in scenario 2 with the modification logic combiner boxes found after the depalletizing machine,
- Scenario 5: system considered in scenario 2 with the increased transport speed that are in the area of the clean room due to the high rate of accumulation of bottles with those found before the empty bottle inspector .

Throughout the model, we adopt a series of measures for evaluating performance goals achieved by the line and compare its performance against changes that may occur. The proposed changes were modeled using the tools Ape presents the simulator and that were mentioned earlier. Thus, modifications were made on these variables, as they are directly related to the operation of the packaging line and determined the degree of improvement that can be achieved are effected once.

The performance parameters are considered for carrying out the analysis of the system is discussed below.

- Level of limiting Machine Efficiency: determining the average number of bottles processed in the filling machine, which is the limiting resource online, and is made by dividing this number and bottles should be processed in time modeling at the speed that is the filler,
- Effective Efficiency Indicator Global: determining the average number of bottles processed and is performed by dividing this number and bottles that should be processed in time modeling at the speed that is the palletizing machine,
- Load Factor of transport: checks during each work shift the percentage of occupation rate each conveyor involved in the process in order to modify those that have a high occupancy and gain stability in the entire packaging line,

- Number of pallets entered and graduated from the line: the purpose of this parameter is to determine the level of productivity that reaches the line under study for a work shift of eight hours,
- Changes in speed and stability of the machines: they want to reduce machine downtime seeking stability thereof thus explores the reasons why change their speed (or shortage of product accumulation) and proposes improvements.

3. RESULTS

The original design of line speeds under study is based on the concept of the "V" which takes the limiting speed of the machine, i.e. the filler, as a reference for defining machine speeds and hind to the same. The procedure for this calculation is carried out to increase between 10% and 15% speeds as they move away from the filler. Ideal speeds (assuming ideal speed machine bottleneck, 550 bpm) and actual speeds of the machines (considering uptime, downtime and uptime internal) over a month of work were considered. Table 1 and Table 2 show the results obtained from the analysis of speeds detailed above.

Table 1: Analysis of theoretical speeds

	Theoretical Speeds	
Machines	Machines Speeds (bph)	Percentage of capacity limitation machine
Depalletiser	43.260	40
Unpacker	40.170	30
Washer	35.535	15
Filling	30.900	0
Labeler	35.535	15
Packer	40.170	30
Palletizer	43.260	40

Table 2: Analysis of actual speeds

	Real Speeds		
Machines	Machines Speeds (bph)	Percentage of machine capacity preceding	Percentage of capacity limitation machine
Depalletiser	40.440	-9,7	30,9
Unpacker	41.820	8,9	35,3
Washer	38.400	24,3	24,3
Filling	30.900	0,0	0,0
Labeler	36.000	16,5	16,5
Packer	38.160	6,0	23,5
Palletizer	40.740	6,8	31,8

The data in the table are expressed in Figure 28 for a better visualization. It can be seen that the concept of the "V" approaches the ideal nearby machines in the filler, but not at the ends.

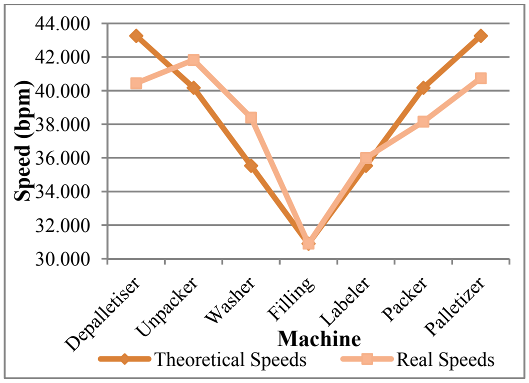


Figure 28: "V" line with ideal and actual speeds

Furthermore, the company determines the productivity of the packaging line from the measurement of the efficiency of their equipment. This is calculated from the values corresponding to the number of bottles produced during an operating period determined in relation to the theoretical amount of bottles that must have occurred during that period (Equation 1). The bottles theoretical amount calculated from the limiting speed of the line which, as mentioned above belongs to the filling machine.

$$Efficiency = \frac{Actual\ number\ of\ bottles\ produced}{Theoretical\ Number\ of\ bottles\ produced} \quad (1)$$

For this calculation, reports were consulted production of 3 consecutive months, of which we obtained the total production time and the volume produced in the same. Thus, Table 3 shows in greater detail the productivity of each month analyzed. From efficiency values shown in the above table is obtained in the same behavior and the present trend in time (Figure 4). Therefore, we can determine that the line has a 66.76% average productivity.

Table 3: Productivity Data

Month	Week	Actually produced bottles	Theoretical produced Bottles	Average efficiency
1	1°	3.187.638	4.752.000	67,58
	2°	3.079.862	4.752.000	
	3°	3.269.809	4.752.000	
	4°	2.985.746	4.752.000	
2	1°	3.082.036	4.752.000	66,55
	2°	3.041.290	4.752.000	
	3°	3.424.400	4.752.000	
	4°	3.054.792	4.752.000	
3	1°	3.155.279	4.752.000	66,17
	2°	3.069.707	4.752.000	
	3°	3.245.196	4.752.000	
	4°	3.108.242	4.752.000	

In this way, the efficiency of the model created can be compared with the one of real system. Actually the company has a line efficiency of 66.77%, which is similar to the obtained by the simulator because it reports a 66.8% value.

In addition, the following productivity indicators were used in order to validate the simulation model developed: (i) number of pallets produced by shift (in both the feed and outlet line) and (ii) production in each machine.

The simulation model developed was executed several times to obtain performance mean values. The information obtained is then compared with historical data of the line. This comparison is showed in Table 4.

Table 4: Performance measures of the real system vs the simulation model

Machine	Real System (pallets per turn)	Simulation Model (Pallets per turn)
Despalletiser	283	282
Empty bottle inspector	280	278
Filling	277	275
Labeler	272	268
Packer	278	274
Palletiser	271	270

Having analyzed all data from many runs, we conclude that the simulation model developed has acceptable apparent validity.

Having analyzed the most relevant scenarios, presented above, key performance measures achieved in each of them are summarized in Table 5 and Table 6. Therefore, it is possible concluding that scenario 5 achieves the highest level of efficiency in terms of the bottleneck resource and also the highest level of overall effective efficiency. This results in a remarkable increase in the production of a rolling line and the use of machines and transports.

Table 5: Summary of results obtained

Scenario	Processed in filling bottles	Processed in bottles depalletiser	Processed in bottles Palletizer
1	165.059	157.200	151.800
2	161.375	160.200	154.200
3	161.512	162.000	155.400
4	177.462	178.800	175.200
5	206.200	211.200	204.000

Table 6: Summary of efficiency indicators

Scenario	Percent Efficiency	Effective Global Efficiency
1	66,8	61,4
2	61,1	58,4
3	61,2	58,9
4	67,2	66,4
5	78,1	77,3

Consequently, a 11.4% increase can be achieved in efficiency by implementing minor changes without incurring large investments while achieving significant improvements in the operation of the line associated with an increase in company profits.

It is noteworthy that the developed simulation model can be easily used for the evaluation of alternative scenarios, i.e. the analysis of proposals for possible future changes in the design and operation of the line.

4. DISCUSSION

The decision variables allowing increasing both the efficiency of the bottleneck machine and the efficiency of the overall line are the speeds of the machines that make up the transport and packaging process as well as the logic conditions programmed in the units.

Furthermore, it is considered that these efficiencies are less sensitive against increasing line capacity drawer, which is associated with an accumulator device for drawers or operator responsible for the same task.

Short stops primarily derived from simple causes can be reduced drastically without complex operations on the machines, although there are also small stalls that can only be removed using sophisticated methods of analysis and operations with high technical content.

The causes that affect the productivity of the packaging line, according to the simulation model carried out, is the modification of the logic of the carriage of the feeding of the packaging process and clean room. Furthermore, the line is sensitive to changes in the speeds of the machines, which are operating at a speed below the nominal speed.

It is noteworthy that for fixed values of speed and transport machines, no investment is needed by the company, because they have the materials and labor necessary for the modification of the same drivers.

Moreover, the study remarks that not always increasing the efficiency ratio on a particular machine line, from the reduction of a kind of loss, produces an increased rate of overall line efficiency. This is because the relationships and interactions in the real system are complex or some degree of uncertainty is present.

It has been essential to have the automatic registration of faults, without which no one could have calculated the time lost in stops. On the other hand, it would be beneficial to all line stoppages would be assigned automatically and easily exported to a spreadsheet. This would avoid much preparation work and subsequent data analysis.

ACKNOWLEDMENTS

Financial support received from ANPCyT under grant PICT-2010-1073, from CONICET under Grant PIP-2221 and from UNL under Grant PI-66-337 is fully appreciated.

REFERENCES

Banks J., Carson J.S., Nelson B.L., Nicol D.M., 2004, *"Discrete-Event System Simulation"*; 4th. Ed.; Prentice-Hall U.S.A.

Christel, M.; Kang, K., 1992, *"Issues in Requirements Elicitation"*, CMU/SEI-92-TR-012, Pittsburgh, EE.UU., Software Engineering Institute, Carnegie Mellon University.

Fishman George S., 1978, *"Conceptos y Métodos en la Simulación de Sistemas Discretos"*. Limusa México.

Gordon Geoffrey, 1981, *"Simulación de sistemas",* 2da. Ed., Prentice Hall, México.

Hepp, P.K.; *"Introducción a la Adquisición de Conocimiento para Sistemas Expertos"*.

Hines W. W., Montogomery D. C., 1990, *"Probability and Statistics in Engineering and Management Science"*, 3rd. Ed., Wiley, New York.

Kelton D., Sadowski, R.P., Sturrok, D.T. 2006, *"Simulation With Arena"* Fourth edition. McGraw-Hill series in Industrial Engineering and Management Science.

Kuo Benjamin; 1975, *"Automatic Control Systems"*; 3rd Ed.; Prentice-Hall, U.S.A.

Law Averil, Kelton W. David; 1991, *"Simulation Modeling and Analysis"*; 2nd. Ed.; McGrawHill, Inc., U.S.A.

Mazzieri G., Obeid Z., 2004 *"Distribución de Mercaderías en Áreas Urbanas desde un Depósito Central"*. Dpto. de Ing. Industrial, Fac. de Ing. Química, Univ. Nac. del Litoral.

Musselman K, 1998, *"Guideliness for Success"* en *"Handbook of Simulation. Principles, Methodology, Advances, Applications and Practice"*. Banks J. Editor. John Wiley & Sons, Inc., USA.

Naylor Thomas, Balintfy Joseph, Burdick Donald, Chu Kong; 1991, *"Técnicas de simulación en computadoras"*; Editorial Limusa, México.

Papoulis Athanasios, 1990, *"Probability and Statistics"*, Prentice Hall, Englewood Cliffs, N.J.

Piattini M., Calvo-Nanzano J., Cervera J., Fernández L., 1996, "*Análisis y Diseño Detallado de Aplicaciones Informáticas de Gestión"*. R. A. M. A. Editorial, Raghavan, S.; Zelesnik, G.; Ford, G.: "Lecture Notes on Requirements Elicitation", CMU.

Renee Thiesing, Christine Watson, Judy Kirby, David Sturrock, 2009, *"SIMIO Reference Guide"*. Versión 2.0.

Ross Sheldon M., 1993, *"Introduction to Probability Models"*, 5th. Ed., Academic Press, N. Y.

Sapag Chain N., Sapag Chain R. 1989, *"Preparación y Evaluación de Proyectos de Inversión"*. Mc. Graw-Hill.

Senn, J.A., 1988*, "Análisis y Diseño de Sistemas de Información",* McGraw-Hill.

Shannon, Robert; 1988, *"Simulación de Sistemas. Diseño, desarrollo e implantación";* Trillas, México.

Walpole Ronald, Myers Raymond, 1989, "*Probability and Statistics for Engineers and Scientists"*, 4th. Ed., Macmillan, N.Y.

AUTHORS BIOGRAPHY

NATALIA BASAN is an Industrial Engineer and PhD student conducting research in hybrid optimization & simulation tools for production planning and scheduling of automated production systems.

LUCILA RAMOS is an Industrial Engineer conducting research in hybrid optimization & simulation tools for production planning and scheduling of automated production systems.

MARIANA COCCOLA is an Information Systems Engineer and a PhD student conducting research in hybrid optimization & simulation tools for production and logistics systems.

Dr. CARLOS A. MENDEZ is a Titular Professor of Industrial Engineering at Universidad Nacional del Litoral (UNL) in Argentina as well as a Researcher of the National Scientific and Technical Research Council (CONICET) in the area of Process Systems Engineering. He has published over 150 refereed journal articles, book chapters, and conference papers. His research and teaching interests include modeling, simulation and optimization tools for production planning and scheduling, vehicle routing and logistics.

FUZZY COGNITIVE MAPS MODELING AND SIMULATION

Evangelia Bourgani[(a)], **Chrysostomos D. Stylios**[(a)], **George Manis**[(b)] **and Voula C. Georgopoulos**[(c)]

[(a)]Dept. of Informatics Engineering, TEI of Epirus, Artas, Greece
[(b)]Dept. of Computer Science & Engineering, University of Ioannina, Ioannina, Greece
[(c)]School of Health and Welfare Professions, TEI of Western Greece, Patras, Greece

[(a)] ebourgani@gmail.com; stylios@teiep.gr, [(b)] manis@cs.uoi.gr , [(c)] voula@teipat.gr

ABSTRACT

Fuzzy Cognitive Maps (FCMs) is a soft computing technique that has been used to model and simulate various and completely different applications from different areas. The FCM modeling approach is symbolic, presenting abstract knowledge and is based on human expert experience and knowledge. This study examines and compares the FCM models as they have been used for various applications.

Keywords: Fuzzy Cognitive maps, modeling

1. INTRODUCTION

Fuzzy Cognitive Maps (FCMs) originated from the combination of Fuzzy Logic and Neural Networks. An FCM models the behavior of a system in terms of interacting concepts; each concept represents an entity, a state, a variable, or a characteristic of the system (Kosko 1986). FCM models are easily understandable as they are similar to the human reasoning procedure, but they require experts' knowledge and contribution during the designing procedure. FCMs constitute a modeling and simulation tool to analyze decision making process for complex systems.

The result of modeling any process depends on the available data, description, information, knowledge and the suggested modeling approach. The fact that FCMs are based on knowledge of experts, which is affected by their experience and background, makes FCMs subjective and potentially vulnerable to possible errors and conflicts. Apart from that, not every possible condition may have been included during the construction of the model, which makes it insufficient. Thus, the results depend on the quality of data that are used to create the model.

FCM modeling creates an abstract representation of a real world system. The modeling and simulation process is simplified, while many assumptions about the system are made, the system's essential relationships are retained and unnecessary detail is omitted. FCMs are able to model and simulate systems in a wide variety of application areas, because of their capability to handle complexity with much and/or even incomplete or conflicting information.

FCMs have been used in many fields, solving a variety of different problems, including social and political sciences, medicine, business and management, engineering, environment and agriculture, information systems and technology, education. Each application has various concepts corresponding to the problem which is under investigation. The large number of concepts makes the system more accurate and completed. However, the less complex a system is the more comprehensible and readable it is. The need for better handling the increased complexity of some applications led to enhance FCMs with learning methods, use of levels and/or separation of the initial complex problem into multiple FCMs and use a supervisor to control the system or use other methods synergistically.

In this work, we describe modeling in various areas, which are completely different from each other, with respect to the concepts and the application requirements. It does not include all the applications, but some of the innovative and useful applications performed by FCMs are described.

2. OVERVIEW OF THE FCM MODEL

FCM is an illustrative causative representation for the description and modeling of any system. It is illustrated as a causal graphical representation consisting of interrelated concepts. FCMs are fuzzy signed directed graphs permitting feedback, where the weighted edge *wij* from causal concept *Ci* to affected concept *Cj* describes the amount by which the first concept influences the latter, as is illustrated in Fig. 1. The values in the graph are fuzzy, so the concepts take values in the range [0,1] and the weights of the arcs are in the interval [–1,1].The value A_i of the concept $C_i$expresses the degree of its corresponding physical value. At each simulation step, the value A_i of a concept C_i is calculated by computing the influence of other concepts C_j's on the specific concept C_i following the calculation rule:

$$A_i^{k+1} = f(A_i^k + \sum_{\substack{j \neq i \\ j=1}}^{N} A_j^k w_{ji}) \quad (1)$$

where A_i^{t+1} is the value of concept Ci at simulation step t+1, A_j^k the value of the interconnected concept Cj at simulation step k, w_{ji} is the weight of the interconnection between concept Cj and Ci, and f is a sigmoid threshold function:

$$f = \frac{1}{1 + e^{-\lambda x}}$$

where λ>0 is a parameter that determines its steepness.

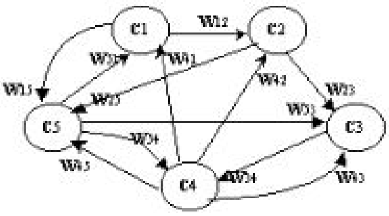


Figure 1: The fuzzy cognitive map model

3. BASIC FCM MODEL

The basic FCM model as described above uses abstract concepts and through the updating equation (1) changes the value of concepts until equilibrium state is reached. Observing the graphical representation (Fig.1), it is clear which concept influences other concepts and it can show the interconnections between concepts.

Concepts can originate from literature, experts and/or non-experts constructing the FCM model for each circumstance that is under investigation. The basic FCM uses the equation (1) for updating their values. Basic FCMs have been enhanced with learning methods or combined with other methods synergistically in order to overcome problems. The results can be linguistic values, which make the use of FCMs more comprehensible and easier to be analyzed and explained.

FCMs use fuzzy logic; hence they can incorporate vagueness and qualitative knowledge and also feedback processes. They can be used to simulate the changes of a system and can also address 'what if' questions. Regarding modeling, FCMs can combine aspects of qualitative methods with the advantages of quantitative methods. FCMs allow dynamically simulating and testing the influence of various scenarios and have been used to reach a decision or to evaluate a procedure or examine management scenarios on system components.
Modeling with FCMs is a simple and transparent way for representing and useful to describe any system in many fields such as engineering, medicine, business and so on. Besides, FCM models can be highly accurate. However, the more complex a system is, the more accurate it is, but complexity decreases the comprehensibility of the system.

The basic FCM model has been used in many applications in various fields and for different purposes. The simulation gave generally satisfactory results. However, each model has its own drawbacks with respect to the field that is used.

For example, in **business and management**,FCMs have been used to model and simulate the information systems strategic planning process (SISP) (Kardaras and Karakostas 1999). For this application, FCMs are used as a means that can combine business and IT perspectives. The concepts of this model are extracted from SISP literature (i.e. case studies and theoretical frameworks) and from relevant practical experiences while the interrelationships are determined by planners using linguistic fuzzy weights. Using the proposed model, planners can develop scenarios and assess alternative ways of applying IT in order to improve organizational performance. It is a dynamic and flexible simulation tool that can handle changes in factors and conflicting assumptions. Planners can simulate different scenarios and when conflict issues are resolved the proposed strategy can be adopted.

In another application FCMs have been used to model third-Part Logistics providers (3PLs) (Huang, et al. 2010), where the concepts were obtained from field visits to selected 3PLs, interviews with experts and literature. This model relies on human reasoning in order to determine the initial concepts and weight values. The output values of the concepts are used to examine scenarios of the company's survivability. In reality, without FCM model's simulation, these features would need many days observation. Thus, this is a useful tool for quick and comprehensive exploration. However, this model uses few concepts and can give an indication of the company's evolution without sudden and unexpected changes. Since concept development is based on three different sources there is a need to include a credibility parameter from each source.

In **education**, basic FCM models have been used to evaluate the teaching-learning process (Laureano-Cruces, Ramírez-Rodríguez and Terán-Gilmore 2004). The proposed model is a reactive environment and it is based on a multi-nodal perspective, a holistic and complete approach (knowledge, abilities, attitudes, values). FCM allow a faster control of the different states of such environments. The cognitive components derive from the expert and the learner. Factors come from the literature, the expert and the learner. This model gives as output different didactic actions according to the event that provoke a condition and the tutor can choose the posed chain of actions or one of them. It is able to include expert's knowledge avoiding the symbolic representation of behavioral reasoning based on rules. Plachebo et.al. used FCM to engineering educational assessment (Pacheco, Carlson and Martins-Pacheco 2004). The need for combining several interrelated aspects makes FCM an efficient method, by means of constant feedback and re-assessment. The proposed tool was used for student assessment. The concepts used for this model were found in students of engineering courses. By activating a concept, positively and negatively, the simulation will give the concepts that are affected, which can be interpreted. The proposed model can be applied to any course, a group of courses, a whole college program an educational department, or to other processes that need to model uncertainty or linguistic imprecision. This is a dynamic model that the user can activate the relative concepts and analyze the outcome. The proposed model is highly based on human intervention.

Another application in education field that FCMs have been used is for modeling educational software adoption, which used in UK secondary schools (Hossain and Brooks 2008). The FCM approach let identifying and modeling both qualitative and quantitative factors and their complex causal relationships in the context of educational software adoption. They used empirically-based FCM, modeling factors in the adoption of educational software in schools. This model can provide information for educational software adoption in schools and can be used as a guide both for educational decision makers and for software developers to which direction they should focus their efforts. This is a model that uses the concept of credibility, assigning different values (weights) in the corresponding parameter, which make it more realistic, while in other applications credibility does not taken under consideration. However, the simulation of FCM by activating a factor may lead to contradictory outcomes, which declares the need for a more dynamic model. Besides, this model is applied using few schools from an area (which means similar characteristics) and as a result the outcome cannot be regarded as general.

For **environmental** issues, FCMs have been used in order to incorporate both experts' and local people's knowledge (Özesmi and Özesmi 2004). FCM's have been used because they are easy to build and can give qualitative results. Experts' knowledge does not require in every field but can be constructed based on simple observations by anybody including indigenous or local people. This is a basic difference to the other applications that are based on experts' opinions and experience. FCMs can analyze the stakeholder's approaches, which are in varying degree of sophistication, requiring in average a low in-depth academic investigation. They do not make quantitative predictions but they can show what will happen under given circumstances, by simulating the system. Hobbs et. al. used FCM to define management objectives for the Lake Erie ecosystem (Hobbs, et al. 2002). Using concepts from many experts and/or organizations, communication among experts with public understanding of ecosystem, the limits and possibilities of management is achieved. However, FCM analysis for defining management objectives gives some guidelines, but the information is insufficient for choosing one single ecosystem objective. This approach should be used complementary with other studies.

In the **technology** domain, FCMs have been used for identifying, classifying and evaluating indicators which related to the success of IT projects. FCM method used to illustrate the applicability and success of a new IT project, the Mobile Payment System(MPS) related to mobile telecommunications (Rodriguez-Repiso, Setchi and Salmeron 2007). Concepts derived from interviews that had the expertise to judge the success (Critical Success Factors, CSF) of an IT project. The initial matrix is converting to another matrix until the final one will obtain the relationships of causality between CSFs. Human factor needs constantly during the process and for analyzing and explaining the final matrix as it may contain misleading data. However, the opinion of an expert may lead to erroneous results, too.

Another application in the technology domain is the use of FCM in telecommunication (Li, et al. 2009). FCMs have been applied for distributed peer-to-peer (P2P) networks, in order to ensure the efficient and successful file sharing. FCMs used for team-centric peer selection and analyzed for improving the network performance. The approach used was compared with the traditional process (min number of hops). FCMs were constructed for candidate and intermediate peers. The concepts have been collected from the literature for the intermediate peer, while some of them were selected by the parameter collecting module for constructing the candidate FCM. With the FCM several important parameters can be considered so the best candidate peer is selected. However, the success of this method is highly dependable on the candidate peer selection. The results showed that FCM approach gives better results compared to the tradition min-hop selection. The output is real positive numbers corresponding to transfer rate and transfer time.

Basic FCM model has been used to simulate circumstances in almost every field. Because of its capability to combine and take into consideration various concepts, without direct relation, it has become a tool in every field. However, the basic FCM model face a number problems which some of them overcame using other methods. These are examples of the basic FCM model on which others are developed. Some applications, such as medical, modeled on more complex FCM structures discussed later on. For the other areas basic FCM models are highly applicable.

4. ENHANCING FCM MODELS USING LEARNING METHODS

The construction of FCMs is based on experts, which make FCM modeling subjective. This is one of the basic weaknesses; another one is the potential convergence to undesired steady states. Learning procedures constitute means to increase the efficiency and robustness of FCMs by updating the weight matrix so as to avoid convergence to undesired steady states, while other are used to make FCM models more automatic for decreasing human factor to the overall process.

The initial basic models of FCMs have been enhanced with learning methods. Each learning method influences the weight matrix with aim to optimize the result. Each leaning method exploits different sources such as historical data. Some learning methods require much human intervention some other less, determined by the under investigation field.

The desired steady state is characterized by values of an FCM's output concepts that are accepted by the experts. Thus, in order to overcome the weaknesses, FCMs have been combined with various learning methods, such as Non Linear Hebbian (NHL), Active Hebbian Learning (AHL), Genetic Algorithms (GAs), Particle Swarm Intelligence (PSO), Differential Evolution (DE). Models based on FCMs, using unsupervised learning methods, that is NHL and AHL, are semi-automated meaning that they require initial human intervention. In order to have more automated models, genetic algorithms (GAs) have been used. More recently, algorithms from the fields of Swarm Intelligence and Evolutionary Computations have been used to train FCM based on historical data, reducing even more the human intervention. They take into consideration all the initial knowledge suggested by human experts and not only the initial elements of the weight matrix. Each of the learning methods contributes and enhances the FCM model attributing additional characteristics such as making them more transparent and readable or less expert dependent.

The advantage of NHL is that it updates only those weights that experts determined, that is, the non-zero weights. The weight values of FCM are updated synchronously. The AHL algorithm adapts all the weights of the FCM model using an acyclic fragment approach for concepts (asynchronous activation and interaction among concepts based on the initial experts' knowledge). The AHL algorithm increases the FCMs' effectiveness, flexibility and robustness, and creates advanced FCMs with dynamic behavior and great modeling abilities where new features can easily be introduced, added or deleted allowing a model to continuously evolve. However, these learning methods require human intervention before the learning process starts. The Hebbian algorithm provides a small change to the weights in the direction of reducing prediction errors. On the contrary, genetic algorithm is a repetitive weight trial and error method that iterates the process of trying a new weight matrix until the prediction error is minimized. GAs are fully automatic in contrast to NHL and AHL methods and do not require input from a domain expert, thus leading to more objective models.

Many applications enhanced their results using learning methods to model their tool. In **medicine** FCM-NHL has been used to model the process to make a decision for the final dose of radiation (Papageorgiou, Stylios and Groumpos 2003). The principle of this model is that all the concepts in FCM model trigger synchronously at each iteration step and only those weights that experts determined are updated. Initial concepts and weight values are determined by experts. The final value should fulfill the requirements: deliver the highest volume of beam to tumor and keep the dose level at the minimum for health tissues and critical organs. This is a complex procedure that many factors should be taken under consideration. The simulation of this tool gave satisfactory results. This model was also used in another application for tumor grading with high accuracy (Papageorgiou, Spyridonos, et al. 2003). It is also a versatile modeling and grading tool, offering a degree of transparency, so the experts have some insight to the system behavior. AHL has been used for classification problems in medical applications, giving better results compared to FCM-NHL, under the same conditions. For tumor grading, better results are owed to the fact that the AHL algorithm can determine new FCM causal links between all the concepts in order to increase classification capabilities of the FCM. In this way the AHL algorithm increases the FCMs' effectiveness, flexibility and robustness, and creates advanced FCMs with dynamic behavior and great modeling abilities. Both the AHL and NHL algorithms are problem-dependent and they use the initial weight matrix. However, both processes are independent of the initial values of concepts and the system's output concepts manage to converge to the desired equilibrium points with appropriate learning parameters.

An extension of GAs is the Real Coded Genetic Algorithm used for prediction in medical cases. The RCGA performs linear transformation for each variable of the solution to decode it to the desired interval. Its main advantages are ability to be used with highly dimensional and continuous domains, and richer spectrum of evolution operators that can be applied during the search process. It has been used for long-term prediction of the patient state after a period of time following a suggested therapy plan for the individual patient. Specifically, FCM-RCGA has been used for prediction of prostate cancer (Froelich, et al. 2012). The simulation had real number output, which is the estimated prediction error, giving promising results. In another application, PSO & DE have been used to optimize the weight matrix of an FCM. The combination of these two algorithms used in radiation therapy to give a more reliable decision for the final dose. This fully automatic model (uses historical data) has been used in the hierarchical structure to optimize the weight matrix of the supervisor-FCM model.

In **business**, the improvement of FCM combined with GA has been used to evaluate forward-backward analysis of Radio Frequency Identification (RFID) supply chain (Kim, et al. 2008). This application tries to mine bidirectional cause-effect knowledge from the state of data. The input of the FCM is obtained by a linguistic method and uses GA to adjust the weight matrix, while the output estimates the fuzzification error between the real and predicted cause-effect. The simulation gave large errors, which is justified by the randomly selected initialization.

In **agricultural**, FCMs (enhancing with NHL) has been used for crop yield prediction (Papageorgiou, Markinos and Gemptos 2009) making a decision for the yield (if it is low or high). Experts determined the concepts and the threshold value to evaluate the procedure. Besides, it has been simulated using NHL algorithm and better results succeeded, showing that FCM-NHL gives better outcomes compared to the basic FCM model.

In **education**, the weight matrix of the FCM has been used in combination with Interactive Evolutionary Computing (E-FCM). This model uses as basis the FCM, but the E-FCM allows a different update of the concepts' values.This model incorporates characteristics from Evolutionary Computing (mutation, crossover) but it demands expert to choose candidate E-FCMs in order the optimization process be completed. It can model not only fuzzy causal relationship but also probabilistic causal relationship among variables. E-FCMs have been applied to a serious game for science learning, which let children learn about different diseases by exploring in a virtual world (Cai, et al. 2010). The E-FCM shows the evolution of the state in real-time. Concepts of this model represent the variables of interest in a real-time system. Generally, E-FCMs use attributes from both FCMs and Bayesian network (use and the conditional probability for representing the causality). The output is converted into linguistic when steady state is reached.

In **engineering** FCMs learning with Hebbian algorithm have been used to model structural damage detection (Beena and Ganguli 2011).For this application some concepts include the difference between measurement frequencies that declare damage and those for undamaged, while others contain the possible damage location. It is simulated with the basic FCM approach and FCM using Hebbian learning with the second one have better results. The output gives with high success rate the damage location. However, this approach depends on the input and structure selection and as the number of input and outputs are increased the system is becoming more complex.

For every application area, experts or non-experts contribute differently to the result. As not all experts/non-experts have the same experience and background, a 'credibility' parameter should be inserted in order to become a more reliable model. Additionally and merely for medical and business-management applications, time may be a major factor that can change the output. Therefore, the lack of the concept of time, regarding the order and the reaction time of a change, may provoke important changes in patient state or influence the output in a strategic/economic problem.

5. CREATING SYNERGISTICAL MODELS BASED ON FCMS

For improving modeling and simulation results the basic form of FCM has been supplemented with other approaches, such as Case Base Reasoning (CBR), evolutionary algorithms, Decision Trees (DT). Synergistical models can handle the data in a more efficient way as they can combine and take advantage of the characteristics of two or more methods in order to optimize a process and obtain more reliable results. The models that have been proposed succeeded in making FCMs less human independent, overcoming one of the main FCMs' weaknesses. The use of CBR and evolutionary algorithms wield the historical data, while Decision Trees can enhance FCMs by letting processing both qualitative and quantitave data.

Modeling using Competitive FCMs (Georgopoulos, Malandraki and Stylios 2003) has been used for decision making, ensuring that there will be only one result. CFCMs are capable on their own to perform a comparison and lead to a decision based on expert knowledge and experience. They are based on 'competitiveness' which will give a 'winner' concept. They consist of two types of concepts: factor and decision. These concepts are determined by experts. The output of modeling is linguistic. CFCMs have been enhanced with various methods in order to infer more reliable results, overcome weaknesses that have the initial model, cope with the increasing number of concepts and become less human dependent. FCMs supplemented with CBR can use information from previous cases and in that way they can face problems such as no activation of nodes because of human underestimation. CBR bases on the fact that similar problems have similar solution. This model has been enhanced with additive methods in order to reduce the simulation time. Another enhancement is the combination of FCMs with Evolutionary Algorithms which are population-based algorithms, and are efficient when they are applied to solve optimization problem. Decision Trees have been also used in order to face the amount of data that may be qualitative and quantitative.

Synergistical models have been used for **medical** circumstances. CFCMs have been used for describing differential diagnosis among Specific Language Impairment, dyslexia and autism (Georgopoulos, Malandraki and Stylios 2003). The simulation reaches a decision, however, this may not be a clear one since in certain situations the output concepts have very close values. CFCMs supplemented with CBR giving the Augmented Competitive FCM (Georgopoulos and Stylios 2005). The main idea of CBR is the assumption that similar problems usually have similar solutions. Thus, this model takes in consideration historical data as it assumes that similar problems usually have similar solutions, reducing the human intervention and making a more automated model. CBR is not called every time, thus it can decrease simulation time. CFCM-CBR has been applied for diagnosis in speech and Language pathology and specifically developed as a differential diagnostic System for Specific Language Impairment, Autism and Dyslexia. It is applied in comparison with the results without CBR and the results showed the advantages of the new system. The difficulty in inferring a distinct result in some cases led to the Complementary CFCM-CBR (CBR enhanced), which uses lateral inhibition and gives an even clearer decision. This model uses an additive method in order to enhance the differences between different decisions/diagnoses and emphasize boundaries. It has been used applied successfully to model and test two decision support systems, one a differential diagnosis problem from the speech pathology area for the diagnosis of language impairments and the other for decision making choices in external beam radiation therapy (Georgopoulos and Stylios 2008). As regard the speech pathology, the simulation of the Complementary CFCM-CBR with the same concepts as the previous CFCMs resulted to better diagnosis of the most probable disorder. The External Beam Radiotherapy

Decision Support is another example that CBR-enhanced CFCMs contribute to the final choice for the optimum distribution of beam so as the healthy cells are not be affected. The use of CFCMs with GA decreases the human intervention. GAs are designed to evaluate existing potential solutions as well to generate new (improved) solutions to a problem for evaluation. Based on this method the Genetic Algorithm Factor Interaction Competitive FCM (GAFI-CFCM) a diagnostic support model has developed (Georgopoulos and Stylios 2009) giving even better results. GAFI-CFCM is also applied for labor modeling (Stylios and Georgopoulos 2010).In this case the result was promising.

FCMs have been also combined with Decision Trees(DT) FCM (Papageorgiou, Stylios and Groumpos 2006). This technique includes a DT (created by any decision tree algorithm) with the FCM. It can handle different types of initial data and can be used differently depending on the type of input data, that is, if data include only quantitative or only qualitative or both. Depending on the type of input data, it is activated with the use of DT or the construction of FCM according to experts' opinions or both; the outputs can be combined in order to reach decision. For a large number of input data, qualitative and quantitave data are used to construct a FCM separately. Their combination constructs the enhanced FCM model. The FCM enriched with IF-THEN rules to assign weights direction and values. The NHL unsupervised training algorithm is used to reach to the proper decision. This model offers better handling of large data, while the FCM's flexibility is being enhanced by the introduction of the decision tree rules that specify weight assignment through new cause-effect relationships. This approach reduces the human intervention however it remains highly dependable on it. It has been used for medical purposes, too (grading bladder tumor).Concepts derived from experts, while quantitative data from the DT. The inductive knowledge from the DT has been used for deriving a set of association rules. This tool has been applied under the same conditions with the FCM tool and the results showed that the FCM-DT approach succeeded more accurate and clear results. The output is linguistic.

These models that have been developed for medical purposes lack the temporal concept. This may be important for certain medical application.

For **political and strategic** issues, FCMs have been applied for crisis management and political decision. The flexibility of FCM has been improved by allowing a variety of Activation Levels (AL) of each concept thus creating Certainty Neuron Fuzzy Cognitive Maps (CNFCM) (Andreou, Mateou and Zombanakis 2005). This model has been also enhanced with GA and applied for the Cyprus issues trying different scenarios. It offers the ability not only to design multi-objective scenarios, but also to predict the dynamics of a future realization. The advantage of this model is that GAs are used and aim at solving the problem of the invariability of the weights and the inability of the method to model a certain political situation following the change of a certain weight or group of weights. Concepts derived from questionnaires and interviews and experts determined the final concepts for modeling. The decision-maker is able to consider hypothetical scenarios by defining the target activation level of a concept in focus and to study the resulting weight values and AL once the model has reached equilibrium. However, it is possible to get into limit cycle or chaotic behavior. This model can be used as an assistant tool for the political analysts and decision makers.

These models are based on human (expert or non-expert). However there is no credibility parameter, as with the most basic models, discussed above.

6. HIERARCHICAL STRUCTURES

For better handling cases of large complex systems, an approach is the decomposition into sub-systems. This technique has been used extensively on conventional modeling and simulation approaches. When subsystems have many elements in common decomposition is not easy and that prohibit the simplified approach of summing up the individual components behavior.

Hierarchical models consist of a number of levels. In theory, the levels can be infinite; in practice, however, they have a limited number. The output of a level is usually the input to another one, which can activate another part of the system. FCMs have been used to construct subsystems of a larger system. The generic purpose of FCMs as subsystems is to exchange information among all the subsystems in order to accomplish a task, make decisions and to plan strategically. Hierarchical models are ideal for processes that involve a large number of factors with complex interrelations. FCM modeling and simulation offers a tool to cope with complexity of large systems.

The advantage of a multi-level system is that factors that correspond to a process can be organized into groups and each group constitutes a (sub)FCM, reducing the complexity. The sub-FCM should be handled both as an autonomous FCM system and as part of a general interactive system. Human intervention is needed. Hierarchical models have been in various application areas, in both medicine and business. The output can be linguistic or real numbers and can be used either to reach a decision or to compare FCMs' values to those estimated by experts.

In **medicine**, the hierarchical architecture has been used in obstetrics, for making a decision whether a natural delivery or caesarian section is to be applied. This model also has been used for decision making in radiation therapy (Papageorgiou, Stylios and Groumpos 2003). The hierarchical models that have been developed for medical purposes use up to two levels: the lower and the upper, which is referred to as a Supervisor-FCM. Thus, the upper level –the supervisor- can perform some of the tasks that a human operator successfully performs in supervising systems. It can handle and express qualitative information and have knowledge about the process structure and determine about the acceptance of a result. Especially, in a two-level hierarchical structure the FCM at the upper level can receive inputs from the lower, interact with the whole FCM and send back to the lower level new set of values until steady state is reached.

Another medical application is in radiotherapy. The hierarchical architecture is used to make decisions about the final dose of beam for radiation therapy. Concepts are determined by radiotherapists and physicists taking under consideration the basic beam data from experimental measurements and patient information. The concepts of the FCM model for radiotherapy treatment are divided into three categories: Factor-concepts, Selector-concepts, and Output-concepts. Input (factors and selectors) concepts; represent treatment variables with given or measured or desired values and taking their values from the real system and/or sensors and their measurements, with transformation to fuzzy interval. On the other hand, there are the output concepts that their value is influenced and determined by the value of the input concepts with the corresponding causal weight and the decision making process is the determination of their value. The values of the output concepts lead to the final decisions. Values of concepts are described using linguistic variables and they are transformed in numerical values using a defuzzification method.

In the **business** and management domains a hierarchical model has been used to propose an effective business modeling support tool that can also drive process change activities (Xirogiannis and Glykas 2004). In contrast to the proposed two layer hierarchical model for medical applications, the proposed model for Business Process Reengineering (BPR) consists of four layers. Each layer can contain one or more sub-FCMs that represent the map categories. Business models and strategic BPR plans are used as input and during the process an expert - redesigner is called to determine business scenarios thus modifying the fuzzy weights and then reasoning about the business performance. Interconnections are determined using linguistic notion by experts and a defuzzification method is used to produce weights. Each sub-FCM has an output that triggers another sub-FCM. This tool aid supplements the strategic planning and business analysis phases of typical BPR and simulates the operational efficiency of complex process models with imprecise relationships and quantifies the impact of the reengineering activities to the business model. The outputs of the proposed model were real numbers that were examined and compared with the values of some concepts, between the FCM approach and the experts' decision. The results gave very close results. However, more real life experiments are needed. Another application is the modeling of e-business maturity (Xirogiannis and Glykas 2007). This model was developed according to the previous hierarchical model for BPR. It consists of four levels and each level has one or more sub-FCMs that group the relative characteristics of each category. Each level is abstract, while each sub(FCM) can be dynamically reconfigurated. The modeling and simulation for this case aim at simulating complex strategic models with imprecise relationships while quantifying the impact of strategic changes to the overall e-business efficiency. Concepts and weight values are obtained by business models and/or financial planning. A team of experts provide linguistic variables for the causal weights, the concept values and the coefficients values to let the FCM algorithm reason about the impact of potential change initiatives. The experts also provide their independent expert estimates (using similar linguistic variables) of the impact of the strategic change choices to specific maturity metrics. The output of the model is expressed in real numbers as described previously.

Hierarchical models in both application areas demand human intervention during the process. In medicine it is needed from the beginning to construct the concepts, while in business it is needed during the important phase of redesigning.

In **education,** hierarchical structure has been used to model an adjustable tool for Learning Style Recognition (Georgiou and Botsios 2007). It is a three layer FCM Schema that allows educators to interfere, tune up and adjust the system parameters in order to order to contribute on the accuracy of the recognitions. The inner layer contains the learning styles, the middle one the learning activity factors and the outer the 48 statements of the learning style inventory. The factors come from the literature. The teacher can tune up system's weights using his/her own diagnosis on a learner's Learning Activity Factors. An algorithm has been used in order to eliminate fault implication, enhancing the outcome. The possible output concepts comes from Kolb's learning theory (Kolb 1984). The result is linguistic.

In **engineering** field, (Stylios and Groumpos 2004) used FCMs to model a heat exchanger system. This FCM has been used as the supervisor of a hierarchical model. Concepts were defined by experts as well as the connections between the concepts. The supervisor-FCM model can be expanded to include advanced features or planning and decision making characteristics, improving the overall performance of system's performance. A two-level hierarchical structure was also proposed to handle modeling of complex systems, where the supervisor is modeled as a FCM. The simulation gives results, if they are not acceptable experts have to redesign the model. The proposed supervisor-FCM contains few concepts however its simulation gave satisfactory results. This model demands experts' opinion for the overall construction of the system.

Hierarchical models can control the lower levels that they may consist of individual FCMs, grouped according to their similar characteristics. This approach gives the opportunity to better handle more complex and as result more realistic models with better observation and control.

7. CONCLUSION

This paper compares FCM modeling and simulation approaches for various non-overlapping application areas, which have different requirements in data handling and concept outputs. FCMs provided the opportunity to model and simulate many problems that require decision making or classification or predicition/checking of scenarios. The fact that FCMs can handle vague, missing or not available information renders them a versatile tool. The table in appendix gathers the described applications according to the used FCM approach for modeling and simulation.

For each field there are different problems as each one has different requirements. Time, for example, with respect to the concepts' content and simulation time, in medical and business applications is one significant factor that contributes and may change the final result. For the other applications, however, time is not such an important parameter. In contrast, a credibility parameter, proportional to experts' experience and knowledge, is a common requirement for all the fields. Generally, FCMs offer a tool that should be used as assistance to decision makers as this methodology may not be a sufficient model of the system because the human factor is not always reliable.

ACKNOWLEDGEMENT

This work was supported by the E.U. FP7–PEOPLE–IAPP–2009, Grant Agreement No. 251589, Acronym: SAIL.

APPENDIX

Models based on FCMs	Applications area	Application	Solving problem/type of output
Basic FCM	Business	3PLs	Predict company's survival; linguistic result
		SISP	Decision making; applying various scenarios; linguistic output
	Education	Evaluation teaching-learning process	Didactic tactics; linguistic output
		Engineering education assessment	Observes tendencies by provoking a situation; analyze the outcome linguistically
		Modeling educational software adoption	Visual medium for investigating factors that affect educational software; Linguistic output
	Technology	Modeling IT project success for Mobile Payment System(MPS)	Decision making; Critical Success Factor (indicators); Linguistic output
		Team-centric peer selection scheme for distributed wireless P2P Networks	Decision making; real positive output
	Environment	Define management objectives for the Lake Erie ecosystem	Decision making; linguistic output; complementary method
FCMs models using learning methods : - Hebbian learning	Engineering	Structural damage detection	Decision making; linguistic result
- FCM-NHL	Medicine	Radiotherapy	Decision making; Linguistic output
		Tumor grading	Classification; linguistic output
	Agricultural	Yield prediction	Make decision; linguistic output
- FCM-AHL	Medicine	Tumor grading	Classification; linguistic output
		Predict autistic disorders	Prediction; linguistic output
- FCM-GA	Medicine	Prostate cancer (RCGA)	Prediction; Real number output
	Business	Forward-backward of RFID supply chain	Evaluation; Fuzzification error

- FCM-PSO&GE	Medicine	Radiation therapy	Optimize Supervisor’s weight matrix
- Evolutionary-FCM (with Interactive Evolutionary Computing)	Education	Serious game for science learning	Explore virtual world; linguistic output
Synergistic models : - CFCM - Augmented CFCM with CBR	Medicine	Dyslexia and autism	Differential diagnosis; linguistic output
- Complementary CFCM-CBR		Language impairment	
		External beam radiation	Decision making; Linguistic output
- GAFI-CFCM		Labor modeling	
- DT-FCM		Tumor grading	Classification; handling both qualitative and quantitative input data; linguistic output
- Genetically Evolved Certainty Neuron FCM	Political and Strategic issues	Cyprus issues	Decision making; linguistic output; applicable to many strategic scenarios
Hierarchical structure	Medicine	Labor Radiotherapy	Decision making ; linguistic output
	Business	Business modeling support tool (Business Process Reengineering-BPR)	Compare bibliographic and expert values; Real numbers output
	Education	Learning Style Recognition	Decision making;Linguistic output
	Engineering	Heat exchanger system	Decision making;Linguistic output

REFERENCES

Andreou, A. S., Mateou, N. H. and Zombanakis, G. A., 2005. Soft computing for crisis management and political decision making:the use of genetically evolved fuzzy cognitive maps. *Soft Comput.,9,*194–210.

Beena, P. and Ganguli, R., 2011. Structural damage detection using fuzzy cognitive maps and Hebbian learnin. *Applied Soft Computing 11.* 2011. 1014–1020.

Cai, Y., Miao, C., Tan, A.-H., Shen, Z. and Li, B., 2010. Creating an Immersive Game World with Evolutionary Fuzzy Cognitive Maps. *IEEE Journal of Computer Graphics and Applications*, 30(2), 58-70.

Froelich, W., Papageorgiou, E., Samarinas, M. and Skriapas, K., 2012. Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer. *Soft Computing*, 3810–3817.

Georgiou, D. A. and Botsios, S.D., 2007. Lerning Syle Recognition A Three Layer Fuzzy Cognitive Mao Schema. *IEE International Confererence on Fuzzy Systems*, 2202-2207.

Georgopoulos, V.C. and Stylios, C.D., 2008. Complementary case-based reasoning and competitive fuzzy cognitive maps for advanced medical decisions. *Soft Comput.12(2),*: 191-199.

Georgopoulos, V. C., Malandraki, G. A. and Stylios, C. D., 2003. A Fuzzy Cognitive Map Approach to Differential Diagnosis of Specific Language Impairment. *Journal of Artificial Intelligence in Medicine* v.29, 221-278.

Georgopoulos, V.C. and Stylios, C.D., 2005. Augmented fuzzy cognitive maps supplemented with case based reasoning for advanced medical decision support. In *Soft Computing for Information Processing*, by M. Nikravesh, L. A Zadeh and J. Kacprzyk, 391-405. Springer.

Georgopoulos, V.C. and Stylios, C.D., 2009. Diagnosis Support using Fuzzy Cognitive Maps combined with Genetic Algorithms. *31st Annual International Conference of the IEEE EMBS.*

Hobbs, B. F., Ludsin, S. A., Knight, R. L., Ryan, P. A., Biberhofer, J. and Ciborowski, J.J.H., 2002. Fuzzy cognitive mapping as a tool to define management objectives for complex ecosystems. *Ecolog.Appl. 12.* 1548-1565.

Hossain, S. and Brooks, L., 2008. Fuzzy cognitive map modelling educational software adoption. *Computers & Education* 51, no. 4. 1569-1588.

Huang, Y.K., Feng, C.M., Yeh, W.C. and Lin, L.Y., 2010. A fuzzy cognitive map modeling to explore the operation dynamics of third-party logistics providers. *Logistics Systems and Intelligent Management.* 1266 - 1270.

Kardaras, D. and Karakostas, B., 1999. The use of fuzzy cognitive maps to simulate the information systems strategic planning process. *Information and Software Technology,* 197-210.

Kim, M.C., Kim, C.O., Hong S. R. and Kwon, I. H., 2008. Forward-backword analysis of RFID-enabled supply chain using fuzzy cognitive and genetic algorithm. *Expert Systems with Applications,* 1166-1176.

Kolb, D., 1984. Experiential learning: Experience as the source of learning and development. *Englewood Cliffs, NJ: Prentice-Hall.*

Kosko, B., 1986. Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies*, 65-75.

Laureano-Cruces, A. L., Ramírez-Rodríguez J. and Terán-Gilmore, A., 2004. Evaluation of the Teaching-Learning Process with Fuzzy Cognitive Maps. *Lecture Notes in Computer Science* 3315, 922-931.

Li, X., Ji, H., Zheng, R., Li, Y. and Yu, F. R., 2009. A Novel Team-Centric Peer Selection Scheme for Distributed Wireless P2P Networks.*Wireless Communications and Networking Conference,WCNC.*

Özesmi, U. and Özesmi, S. L., 2004. Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, 43-64.

Pacheco, R. L., Carlson, R. and Martins-Pacheco, L. C., 2004. Engineering Education Assessment System Using Fuzzy, 4867-4881.

Papageorgiou, E. I., Markinos A. and Gemptos, T., 2009. Application of fuzzy cognitive maps for cotton yield management in precision farming. *Expert Systems with Applications*, no. 36,12399–12413.

Papageorgiou, E. I., Stylios, C. D. and Groumpos, P. P., 2003. An Integrated Two-Level Hierarchical System forDecision Making in Radiation Therapy Based on Fuzzy Cognitive Maps. *IEEE Transactions on Biomedical Engineering*.

Papageorgiou, E. I., Spyridonos, D. D., Stylios, C. D., Nikiforidis, G.C. and Groumpos, P. P., 2003. Grading Urinary Bladder Tumors Using Unsupervised Hebbian Algorithm for Fuzzy Cognitive Maps. *Biomedical Soft Computing and Human Sciences Vol.9, No.2*, 33-39.

Papageorgiou, E., Stylios, C. and Groumpos, P. 2003. Fuzzy Cognitive map learning based on nonlinear Hebbian Rule. *Gedeon, T.; Fung, L.C.C.;.* Heidelberg: Springer, 256-268.

Papageorgiou, E., Stylios,C. and Groumpos P., 2006. A Combined Fuzzy Cognitive Map and Decision Trees Model for Medical Decision Making. *IEEE EMBS Annual International Conference.* New York,6117-6120.

Rodriguez-Repiso, L., Setchi, R. and Salmeron J. L., 2007. Modelling IT projects success with Fuzzy Cognitive Maps. *Expert Systems with Applications 32*, 543–559.

Stylios, C. D. and Groumpos, P.P., 2004. Modeling Complex Systems Using Fuzzy Cognitive Maps. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans.*155-162.

Stylios, C.D. and Georgopoulos, V.C., 2010. Fuzzy Cognitive Maps for Medical Decision Support – A Paradigm from Obstetrics. *32nd Annual International Conference of the IEEE EMBS.* Buenos Aires, Argentina.

Xirogiannis, G. and Glykas, M., 2007. Fuzzy Cognitive Maps in Business Analysis and Performance-Driven Change, *IEEE Transactions on Engineering Management*, 334-351.

Xirogiannis, G. and Glykas, M., 2007. Intelligent Modelling of e-Business Maturity. *Experts Systems with Applications*, 687-702.

AUTHORS BIOGRAPHY

Evangelia Bourgani is collaborator at Knowledge and Intelligent Computing Lab, Dept. of Informatics Engineering, TEI of Epirus and she is a Ph.D student in Depart. of Computer Science & Engineering, University of Ioannina. She received her diploma in Electrical & Computer Engineering and her M.Sc. degree in Information Processing from Depart. of Computer and Informatics Engineering and Physics, University of Patras. Her research interests are Soft Computing and Decision Support Systems.

Chrysostomos D. Stylios is an Associate Professor at Dept. of Informatics Engineering, TEI of Epirus; he is a senior researcher at Telematics Center Department of Computer Technology Institute & Press. He received his Ph.D from the Dept. of Electrical & Computer Engineering University of Patras (1999) and diploma in Electrical & Computer Engineering from the Aristotle University of Thessaloniki (1992). He has published over 100 journal and conference papers and book chapters. His main scientific interests include: Fuzzy Cognitive Maps, Soft Computing, Computational Intelligence Techniques, Neural Networks, Knowledge Hierarchical Systems and Decision Support Systems

George Manis is an Assistant Professor at Dept. of Computer Science & Engineering, University of Ioannina. He received his Ph.D from the Dept. of Electrical & Computer Engineering, National Technical University of Athens (N.T.U.A). He received his M.Sc. in Advanced Methods in Computer Science from QMW College, University of London and diploma in Electrical & Computer Engineering from the N.T.U.A. His main scientific interests include: Biomedical Engineering and Computing Systems.

Professor Voula Georgopoulos holds a Diploma Engineer, from the University of Thrace, Greece, an M.S., from MIT, USA, and a Ph.D., from Tufts University, USA, all in the field of Electrical Engineering. She is Professor of Informatics in the School of Health and Welfare Professions of TEI of Western Greece. Her current research interests are in Medical Applications of Artificial Intelligence Systems.

NEW FORMALISM FOR PRODUCTION SYSTEMS MODELING

Guido Guizzi[(a)], Daniela Miele[(b)], Liberatina Carmela Santillo[(c)], Elpidio Romano[(d)]

[(a), (b), (c), (d)] University of Naples "Federico II", Naples – Italy

[(a)] g.guizzi@unina.it, [(b)] daniela.miele@unina.it, [(c)] santillo@unina.it, [(d)] elromano@unina.it

ABSTRACT

This paper aims to highlight the usefulness of the simulation, analyzing in particular, two simulative techniques: the Discrete Event Simulation and the System Dynamics. The main objective is to propose a simulation methodology to use to model, analyze and control any type of system. This approach is supported by three studies, belonging to different sectors, which demonstrate the utility of adopting a simple and common scheme of analysis.

Keywords: Simulation, Decision making, System Dynamics

1. INTRODUCTION: THE SIMULATION

The simulation is a methodology for experimental analysis of dynamical systems and in particular of complex dynamic systems.

The term "system" refers to a set of entities that, individually distinct, interact through interdependent relationships or reciprocal connection (Forrester 1961).

The systems in question are dynamic: they are characterized by the evolution over time. The emphasis is placed not only on the analysis of the system as a state of equilibrium, but focuses on the process through which change over time. Technically, the dynamics of a system is defined as the succession of its states over time, where the system state is a set of measurable quantities. The further attribute is the complexity. A complex system has its own characteristics, which do not correspond to the sum of the parts that constitute it. In other words, the network of relationships between entities produces non-linear effects that can't be explained by studying each component separately (Bertalanffy 1969).

Therefore, the presence of non-linearity, dominant feature in complex systems, leads to adopt the simulation as interesting alternative to analytical models in the study of complex dynamic systems.

The simulation is a methodology that is part of the so-called experimental mathematics. It is a representation of the system, realized through a computer language, which allows to use the computer to calculate numerically the behavior. This methodology provides a valid alternative analysis: "The simulation models thus represent a significant response to the demands of flexibility and adaptability descriptive, on the one hand, and of the possibility of computation, on the other. A computer code has formal, adaptability and flexibility and computability requirements."

The adaptability of the simulation is referring to the fact that the programming languages allow to define the properties of the system in great detail, determining the behavior dynamically, based on its current state. Through the conditional constructs, typical of programming languages (if ... then ... else), it is possible easily introduce such behaviors conditionals in simulation models. There are also disadvantages: the simulation implies difficulty in generalizing the results. For example, the task to extract general rules from a simulation model is more difficult compared to the case in which these rules must be extracted from an analytical model. In fact, in the analytic case, the solution of the equation system allows you to have full information on the system represented. While the simulation is only able to provide information relating to particular demands of the possible future path of the model, often determined by the initial parameters. This methodology does not promise to deliver the same quality and information content of an analytical solution, but certainly allows to analyze and formalize complex systems, otherwise intractable.

2. THE SIMULATION TECHNIQUES MOST WIDESPREAD

Below, there are the two most common techniques of numerical simulation based on the computer:

- The Discrete Event Simulation (DES);
- The System Dynamics (SD).

Each of them is characterized by a specific formalism for the representation of the entities, relationships and time

2.1. The Discrete Event Simulation (DES)

The Discrete Event Simulation is based on a dynamic ordering of events in time. The system evolves through a succession of leaps in time, at which an event occurs and changes the status of the system. The discrete event simulation is based on a dynamic ordering of events in time (Caputo, Gallo and Guizzi 2009; Guerra, Murino and Romano 2009).

Certain events are scheduled at the beginning of the simulation, others are generated during execution. The simulated experiment consists in the reproduction sequence of status changes. This simulation methodology is very useful for analyzing the utilization rate of resources (production units) and to highlight the eventual critical points (bottlenecks) in the process (Gallo, Montella, Santillo and Silenzi 2012).

This methodology adopts a graphic symbols. The processes are sequences of activities described by graphic symbols, linked by sequential relationships (the lines connecting them). The entities, said token, flow within the chain described by the process. The token is a placeholder that moves in the process and occupies the possible queues. The token is a placeholder that moves in the process and occupies the possible queues. In addition to the token, also the information can flow. So, through a different symbology, it is possible to distinguish the routes taken by the information and those made by the token. For each token may be associated state variables that are normally handled by the units. Thanks to the graphical representation and the rich library of symbols (building block), available in a discrete event simulator, it is possible to construct models with a reduced use of programming. The adoption of programmable blocks allows to realize sophisticated models with relative simplicity and clarity of expression. Such logic design is similar to that used to draw the electronic circuits. Employing the integrated circuit, capable of performing complex functions, it is possible to construct a complicated circuit by adopting a scheme very simple.

2.2. The System Dynamics (SD)

Among the techniques of simulation, this is the one that is closest to the mathematical formalism, in fact, is based on differential equations. The system dynamics is based on a useful perspective to represent the relations of cause and effect in the dynamic phenomena (Revetria, Catania, Cassettari, Guizzi, Romano, Murino, Improta and Fujita 2012). Compared to other simulation techniques, it enables a reduced use of programming languages, allowing extreme rapidity in the design of the models. In the formalism of system dynamics, there are three types of variables:

- the level variables (also called stocks);
- the flow variables;
- the auxiliary variables.

The Level variables relate to stock or endowment of a good at a given time t, acting as containers that are filled and emptied during the evolution of the system. The flow variables represent the rate with which a variable level changes over time. The net rate of change of a stock is the sum of all inflows minus the sum of all the outflows. Mathematically, the stock integrate their net flows, while the net flow is the derivative of the stock. Obviously the rate represented by the flow variables can be expressed as a constant value, a function stochastic or can depend on other variables of the model.

Another fundamental element for the System Dynamics is the delay. The delays are divided into two categories:

- The material delay. It postpones the flow of goods in output from a variable level, ensuring that the total of what enters the stock is equal to the total of what will come out;
- The delay of information. It does not guarantee that the sum of the information in input is equal to that in output. In the cognitive process, in fact, the most recent information may overlap with those previously perceived.

The System Dynamics uses two types of diagrams, useful in describing the system in analysis:

- The Causal Loop Diagram, which allows to represent in a direct way the system from the mere point of view of the relations of cause and effect;
- The Stock and Flow Diagram, which allows to represent the system as a function of the variables of stock, the flow variables and auxiliary areas.

2.3. SD vs DES

The SD and DES are two basic simulation techniques, both used as a tool for decision support and therefore, both adopted to analyze the evolution of the system over time and its behavior according to the variation of some parameters.

In fact, there are substantial differences in terms of modeling approaches: the SD traces the problem, on the basis of its general structure, emphasizing the causal links between the variables, while the DES attempts to trace the path followed not by the system, but by a single element forming part of it.

The table I presents a clear overview on the aspects that characterize and differentiate the two approaches.

The SD models are adopted to study complex systems and offer the possibility to aggregate a large number of individual objects in the flows. The SD allows the evaluation of the behavior of the system for long periods of time, responding to the needs especially strategic (Converso, De Carlini, Guerra and Naviglio 2012; Gallo, Aveta, Converso and Santillo 2012). While the DES, is usually adopted to model business processes, which require specific performance measures, such as the levels of production output or levels of customers served (Gallo, Guerra, Guizzi and 2009).

Stahal (1995), in his studies, shows that, due to the high level of aggregation, the models in SD, tend to be relatively small in terms of number of elements considered, on the contrary, the models in DES tend to be rather complex, as each process is modeled in detail, until the single working units. The level of detail, then, in DES, is a critical factor: a very detailed model takes a long time to realize it and may be less reliable. The first step to define a pattern in DES is the mapping of the process, through which to define logical relationships among the elements. The mapping process, realized according to the logics of DES, may be sufficient to understand the system, without necessarily proceed with the simulation. The DES, as previously mentioned, it is appropriate to conduct a detailed analysis of a specific system, that is well defined and linear, so as to provide estimates of performance measures statistically valid, such as the number of entities or pieces in the queue (Greasley, 2009). The previous statements confirm the choice of System Dynamics, as a simulation tool for the study of the behavior of a flow shop, whose production follows the logic of the Make to Order, according to the priority rule FIFO. Furthermore, the deficiency, in the literature, of similar works, compared to DES in studies on production systems, has led again to make use of the System Dynamics to conduct the work mentioned.

Table 1: Criteria for selection of modeling approach

Factor	SD	DES
Target	Investigate the behavior model of a system.	Investigate the operating performances of some processes.
Determination of the behavior	The behavior of the systems is determined by structures of accumulation and feedback.	The behavior of the systems is determined by the stochastic nature and the interdependence of the processes.
Uniqueness of the problem	The problem is related to a recurring behavior in all the system.	The problem is unique.
Level of the implementation	The level is usually managerial and strategic.	The level is usually operational and tactical.
Time scale of analysis	From days to months / years.	From minutes to days.
Presentation of the results	Statistics and graphs showing the behavior of the system	Statistics showing the performances and the paths of the single elements
Level of aggregation	The single elements are grouped in layers.	Every single item can be modeled.
Dimensions of the model	Small.	From medium to large.
Conceptual model	Influence diagrams	Process map

3. METHODOLOGICAL SCHEME FOR THE SIMULATION IN SD

The core of this paper is to describe a methodology to schematize, analyze, manage and control a process of every kind, whether belonging to the manufacturing world and the service sector, in an optimal manner (Guizzi, Chiocca and Romano 2012; Guizzi, Murino and Romano 2012). After a long and careful study path it was possible to observe that the reality can be easily schematized through the aid of three elements. This means that in any type of system it is possible to find three basic tools, which allow the dynamic developments control of the same. The items under discussion are as follows:

- A time system, the Hourglass;
- A system of evolution, the Chain of Events;
- A system of routing, the Route.

The Hourglass is the time constraint that the system under analysis must respect. An hourglass has the function to mark the time to perform a given operation. When the time runs out, the operation is considered ended and only then, eventually, if all other constraints have been met, the entity can move to the next stage. The schema of the hourglass is characterized by a variable level "Time object", that increments and decrements itself thanks to flows of input and output, respectively "Load time" and "Unload Time", indicating, for example, the rate at which carries out the operations of loading and unloading. Obviously, the level variable influence some auxiliary variables, such as the "Time Remaining".

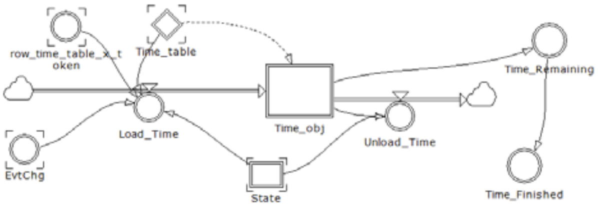


Figure 1: Hourglass

The Chain of events is the core of the simulation, through the construction of this it is possible to trigger events that allow the advancement of the entities in the simulation model. The structure is characterized by a variable level "State" that increments and decrements itself thanks to the flows "Shift_in" and "Shift_out", indicating the rate of entry and exit from the particular state.

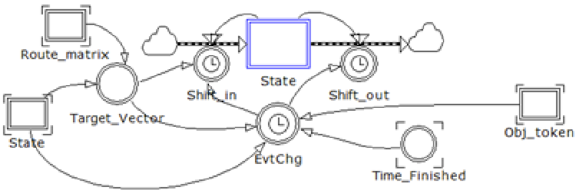


Figure 2: Chain of events

The Route: the structure is characterized by a variable level, "Route_matrix", through which it is possible to identify the possible routes that entities can undertake.

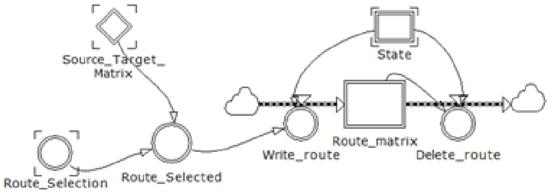


Figure 3: Route

In order to make clearer the proposed methodology to schematize a system in SD, below it is possible to analyze three different studies conducted with this approach. Specifically, the works just mentioned, regarding the port sector, the airport sector and the productive sector. In these three cases, the system represents the case in which the route is unique and is identified by the chain of events, then all three constitute the case more "trivial."

3.1. Port Model (Guizzi, Santillo and Romano 2013)

A port terminal is a node in the freight network both container and dry bulk. A port terminal is a complex system to manage, in fact inside, often encounter criticality difficult to decipher and to overcome. In this context, there are two types of problems: structural and logistical For example the first type belongs the size of the access channel. The access channel is not the same for each terminal port, otherwise, each channel has its length, but especially its width, depends on the morphology of the place. Obviously an access channel to the terminal with smaller width, presents major complications compared to a channel with a greater width. From the width of the channel depends on the possibility to pass two or more vessels together. The case in which the transit of the channel is constrained to a single ship implies, for example, congestion problems: thus a structural constraint becomes a logistical constraint. Another critical, which has a high complexity, is related to the safety distance, that vessels must maintain between them while they run through the channel. These problems in addition to other difficulties were analyzed by the method mentioned above. In the image below it is possible to observe the chain of events of a port system.

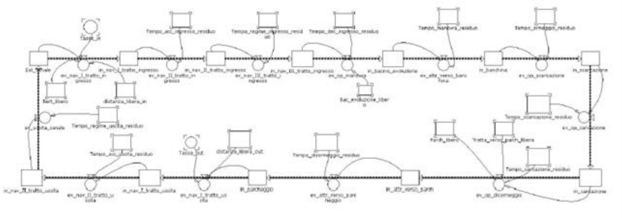


Figure 4: Port Model: Event Chain

The chain of events is the core of the simulation, through the construction of this chain it is possible to trigger events that allow the advancement of the vessels in the circulation model. In this case, the route that the vessels must follow is that indicated by the chain of events and is, therefore, fixed. This means that ships can’t carry out an alternative route. This obstacle is overcome by the introduction of the matrix of routes. This chain of events has been built using the logic of Petri nets and allows to track the movement identified in the context of analysis. In this chain of events, the levels are operations undertaken by the ships in the harbor, from the moment of entry into the channel at the time of exit from the same channel. The flows of the chain of events represent the different events that must be activated to switch from an operation to the next. The constraints are graphically represented by arrows and must be satisfied so that the events are triggered in the chain and chain operations can proceed. The logic used to trigger the events is of the type "if-then": if the constraints are satisfied, then the event is active, otherwise the event remains inactive until the combination of the constraints is not satisfied. The main constraints related to events, are dimensional constraints and temporal, ie the dimensional constraints are related to the ability of a certain area of the port, such as the quay, to be able to accommodate only a limited number of vessels at the same time because of the limited size of that area, while the timing constraints are represented by the time necessary to make and terminate an operation that precedes a subsequent activity. The timing constraints are represented by means of “hourglasses”, used in such a way as to exhaust the remaining time of a certain task. In this way, only when an hourglass runs out then the system advances to the next step. For example in the case of the operation of maneuver is possible to consider a timing pattern of this type:

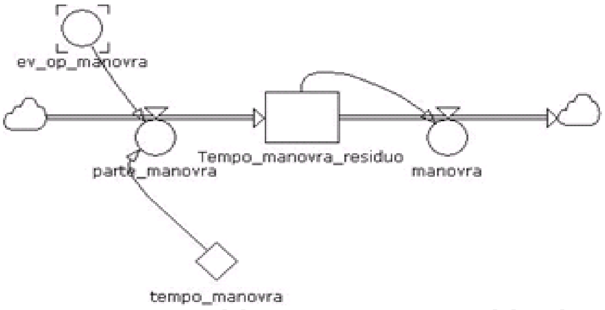


Figure 5: Port Model: Time consuming Model in the case of the operation of maneuver

3.2. Production model

The model schematizes a production system "Flow Shop" and the sequencing of its activities, under the rule of dispatching, F.I.F.O. type. The model is composed of two submodels: one for schematize the productive system and one for schematize the sequencing of activities. In this context, the second submodel, just mentioned, is shown. From the figure below it is possible to see the chain of events.

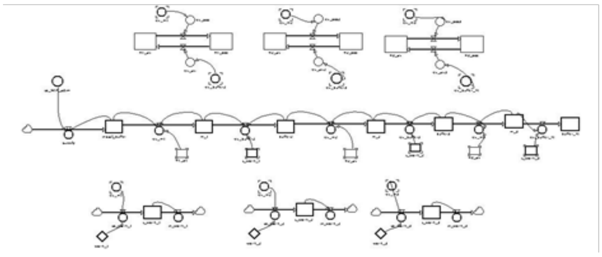


Figure 6: Production model: Event chain

The scheme involves a production system "flow shop", consisting of a single production line, where the operations necessary for the realization of products must be carried out on the same set of machines and in the same order of precedence: the flow of elements along the line is unidirectional and there are precedence constraints between operations. There are four types of product and each of these must be tried on 3 different machines: M1, M2, M3. Each operation will have a different duration by virtue of the product being processed on the specific resource. Furthermore, the machines are "dedicated", then two successive operations must be performed on two different machines. The module respects the following constraints:

- operations must be specified in the order defined;
- each machine must perform at most one operation at a time;
- each operation must be carried out, at most, by a machine at a time, that is, an operation on a machine may commence only after completion of processing on the previous machine.

Upstream of each machine, the module provides a buffer: this means that the buffer downstream of the machine M1 contains the piece, that has been processed by the machine M1, from which it is taken to undergo secondary processing and so on until the end of the process, where there is a buffer of finished products. The levels of the model belong to two categories: some are indicative of the operations that are performed on different machines, others are indicative of the buffer upstream and downstream of a certain resource. The flows, represent the events: to transit from an operation to the next, or from one buffer to another, the constraints must be satisfied, if these are not verified, the event is not activated and the flow does not allow the unit in question to transit from the previous level to the next. In addition, the FIFO rule is implemented: the first element to enter the layer upstream of the chain, will be the first to be worked, and so on all subsequent.

The following constraints were considered:

- dimensionless, they are deprived of measurement units and related to resources;
- temporal, they are representative of processing times for each item, on each machine and determine the beginning and the end of the individual machining operations. An item, can pass to a subsequent processing, for example on the machine M2, only if the first machining operation on the resource M1 has been completed: the completion of this operation is defined by a specific level that represents the processing time remaining.

The time required to perform each operation is controlled via time constraints: the hourglasses. An hourglass is designed to scan the time to devote to an operation: When the time has run out, the transaction is considered completed and if all other constraints have been met, the product being processed can move on to the next resource to undergo a further processing. The structures of hourglasses, are similar for each type of operation, so for brevity, hereinafter, the structure of the hourglass relative to M1.

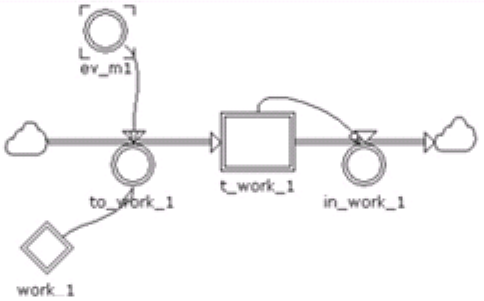


Figure 7: Production model: Hourglass time

3.3. Airport Model

The Airport is an interchange intermodal and can be considered an integrated system of infrastructure, devices and equipment. The operational functionality of airports must be guaranteed by the capacity of its components, which must be dimensioned and must operate at least according to the standard level of work of the airport. In the context of airport operations and structures are usually divided into two areas: “Landside” and “Airside”. The factor that distinguishes these two areas is the capacity: the capacity of the landside is measured in number of passengers served per unit of time, while that of the airside is measured by the number of operations (takeoffs or landings) per unit of time. Between the two subsystems, the airside is the system most likely can generate bottlenecks. This means that this study is focused on the capacity of the airside. In detail, the three sections: the runway, taxiways and parking areas, are in series with each other, so the capacity of the entire subsystem will be equal to the lesser of the three values, and in this regard the critical resource can be runway. The study and implementation of the appropriate measures to increase the capacity of the slopes, however, must not be separated from the consideration of the capacity of the other parts of the system, in order to avoid that these, entering saturation, undermine efforts to increase the capacity of entire airport system. The capacity of a track depends essentially on the distancing between aircraft and the runway occupancy time (landings, takeoffs, or both). In addition, the occupation time of the track employee of the following elements: configuration of the slopes (single, parallel, crossed, etc..), coefficient of utilization of the airport, interference of the slopes between them, the number and location of fast exits from the runway. The airport infrastructure, analyzed, has a single runway, dedicated exclusively to landing operations, in addition, this track is used mainly to the use of aircraft weight class Medium. This infrastructure is also equipped with a Holding Stack consists of a number of circuits equal to 4, arranged at different heights, and such that each circuit can be engaged by only one aircraft at a time. The management of the stack is FIFO type. Thanks to the model created it is possible to identify levels of capacity and delay beyond which the loss of efficiency of the system will reflect itself negatively on the infrastructure creating congestion. For this purpose, the implementation of an appropriate scheduling for managing the flow of aircraft can make a significant contribution to the increase of efficiency and safety.

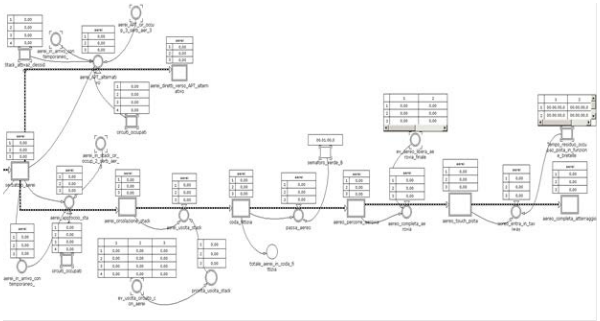


Figure 8: Airport Model: Event Chain

The physical flow of the simulation model precisely describes the path of the aircraft until landing on runway. This flow allows to follow the individual aircraft in all phases that precede the landing: from the descent phase to the landing phase, and then liberation of the track. The FIFO stack implies that the second aircraft, of the waiting sequence in the "stack", can go down to the minimum level only when the first aircraft made free this level and sufficient time to dissipate the turbulence generated is spent. Similarly the subsequent aircraft may descend into the "stack" of waiting to lower levels when these levels were made free from the aircraft above. The aircraft went out from the stack, will have to cross the "final approach path" before starting the landing.

4. CONCLUSIONS

The simulation appears to be a good system of analysis, monitoring and evaluation of real systems, since it offers the possibility to create "experiments" at low cost.

This advantage must be accompanied by a good modeling capabilities, otherwise the simulation approach can be an obstacle to the activities of synthesis and analysis. For this purpose a long process of investigation and study has led to the need to identify a pattern methodological simple and easy to apply for anyone who intends to use the System Dynamics.

The advantage of this approach lies in the possibility to outline and analyze systems of different nature, with the help of three instruments that represent the dynamism of all reality.

REFERENCES

Bertalanffy, V., 1969. *General System Theory*. New York: George Braziller, pp. 139-1540.

Caputo, G., Gallo, M., Guizzi, G., 2009, Optimization of production plan through simulation techniques, *WSEAS Transactions on Information Science and Applications*, 6 (3), pp. 352-362.

Converso, G., De Carlini, R., Guerra, L., Naviglio, G., 2012, Market strategy planning for banking sector: an operational model, *Advances in Computer Science: 6th WSEAS European Computing Conference (ECC '12)*, pp. 430-435, September 24-26, 2012, Prague, Czech Republic.

Forrester, J., W., 1961, *Industrial Dynamics*, Pegasus Communications.

Gallo, M., Aveta, P., Converso, G., Santillo, L.C., 2012, Planning of supply chain risks in a make-to-stock context through a system dynamics approach, *Frontiers in Artificial Intelligence and Applications*, 246, pp. 475-496, IOS PRESS.

Gallo, M., Guerra, L., Guizzi, G., 2009, Hybrid remanufacturing/manufacturing systems: Secondary markets issues and opportunities, *WSEAS Transactions on Business and Economics*, 6 (1), pp. 31-41.

Gallo, M., Montella, D.R., Santillo, L.C., Silenzi, E., 2012, Optimization of a condition based maintenance based on costs and safety in a production line, *Frontiers in Artificial Intelligence and Applications*, 246, pp. 457-474, IOS PRESS.

Greasley, 2009, *A Comparison of System Dynamics and Discrete Event Simulation*, Aston Business School, Aston University, Birmingham, United Kingdom.

Guerra, L., Murino, T., Romano, E., 2009, Reverse logistics for electrical and electronic equipment: A modular simulation model, Proceedings oh the 8th WSEAS International Conference on System Science and Simulation Engineering, ICOSSSE '09, pp. 307-312, October 17-19, 2009,Genoa, Italy.

Guizzi, G., Chiocca, D., Romano, E., 2012, System dynamics approach to model a hybrid manufacturing system, *Frontiers in Artificial Intelligence and Applications*, 246, pp. 499-517, IOS PRESS.

Guizzi, G., Murino, T., Romano, E., 2012, An innovative approach to environmental issues: The growth of a green market modeled by system dynamics, *Frontiers in Artificial Intelligence and Applications*, 246, pp. 538-557, IOS PRESS.

Guizzi, G., Santillo, L.C., Romano, E., 2013, A new model to manage vessels flow in a port Terminal, *7th International Conference on Applied Mathematics, Simulation, Modelling (ASM '13)*, January 30-February 01, 2013.

Revetria, R., Catania, A., Cassettari, L., Guizzi, G., Romano, E., Murino, T., Improta, G., Fujita, H., 2012, Improving healthcare using cognitive computing based software: An application in emergency situation, *Lecture Notes in Computer Science*, 7345 LNAI, pp. 477-490

Stahl, J., E., 1995, *New Product Development: When Discrete simulation is Preferable to System Dynamics*, Elsevier Science.

ASSIGNING CLASSES TO TEACHERS IN UNIVERSITIES VIA MATHEMATICAL MODELLING: USING BEAM SEARCH METHOD AND SIMULATION IN JAVA

Anibal Tavares de Azevedo[(a)], Andressa Fernanda S. M. Ohata[(b)], Joni A. Amorim[(c)], Per M. Gustavsson[(d)]

[(a)]University of Campinas – UNICAMP, Brazil
[(b)]State University of São Paulo – UNESP, Brazil
[(c)] Högskolan i Skövde – HiS, Sweden
[(d)]Saab Group, Sweden

[(a)]anibal.azevedo@fca.unicamp.br / atanibal@gmail.com, [(b)]andressa.matsubara@hotmail.com, [(c)]joni.amorim@his.se / joni.amorim@gmail.com, [(d)]per.m.gustavsson@saabgroup.com

ABSTRACT

In universities, before the beginning of each school year, it is held the distribution of classes among the available teachers. For such task, different constraints must be fulfilled like preventing a teacher to teach in two different places at the same time and avoid solutions in which some teachers have more class hours than others. This process, if performed manually, is time consuming and does not allow viewing other combinations of assignment of classes to teachers. In addition, it is subject to error. This study aims to develop a decision support tool for the problem of assigning teachers to classes in universities. The project includes the development of a computer program using the concepts of object orientation as a way to implement a search algorithm called Beam Search which explores the combinatorial nature of the problem. The programming language used is Java and the program has a graphical interface for insertion and manipulation of the relevant data.

Keywords: beam search, combinatorial optimization, teaching, timetable.

1. INTRODUCTION

Timetable is an event table that specifies who will participate, who will be held where and when such an event occurs. Thus a timetable should satisfy all constraints that are simultaneously involved and there should be no conflict in the schedule.

The Educational Timetabling problems can be classified in two categories: exam and course timetabling (Al-Yakoob, Sherali, and Al-Jazzaf, 2010; Carter and Laporte, 1998).

When constructing the course timetabling of a university, there is a great difficulty to relate the different variables such as students, teachers and classrooms. In special, it is necessary to consider prerequisites established by the university, individual preferences of teachers/students for certain disciplines to be taught/routed and, most often, the downtime between classes should be avoided. Added to this, there are risks of errors in the definition of the grids and these may be detected only when the classes have already begun (Al-Yakoob, Sherali, and Al-Jazzaf, 2010).

According to (Carter and Laporte, 1998) the course timetabling problem can be divided into five subproblems: teacher assignment, class-teacher timetabling, course scheduling, student scheduling and classroom assignment. The teacher assignment problem only allocates teachers to courses without using the information about the courses allocation to time periods. Course scheduling problem often uses a given allocation of teachers (Gunawan, Ng, and Poh, 2013).

This work will address the Problem of Assignment of Classes to Teachers (PACT) that combines teacher assignment and scheduling problem simultaneously within a university. A special feature had been considered in the model in order to consider teacher´s preference for classes with the same subject. That is, as general purpose, teachers must teach the classes with the least possible effort and different from the one considered in recent literature (Al-Yakoob and Sherali, 2013).

PACT is part of the set of combinatorial optimization problems (Schaerf, 1999; Willenmen, 2002) which justify the development of heuristics and meta-heuristics. Another contribution of this work is to develop and apply a Beam Search method for the PACT in a manner that the optimal solution or at least a very close solution to the optimal one is produced and constraints are all satisfied.

This paper is structured as follows. Section 2 presents the mathematical model of PACT, while section 3 presents the proposed solution method. Section 4 presents computational results and section 5 addresses conclusions and future work.

2. MATHEMATICAL MODEL

Some initial considerations are necessary for the development of the mathematical model for PACT:

- classes does not exceed their limit in terms of maximum number of students;

- the physical distance between classes is negligible, i.e., the time required for the teacher to go from one classroom to another can be neglected.

To better illustrate the developed approach, this work will employ a little number of teachers and classrooms, without loss in reliability of the results. The following data should be available:

- all classes should have the identification number of the group to which it belongs, the course initials and the name applied to the discipline, the workload and the time set in the grid;
- all teachers should have some kind of registration number, a full name and defined maximum workload.

The PACT can be formulated as follows. Let *m* classes to be assigned to *n* teachers. The cost of *k* teachers preferences for certain classes of subjects *i (i = 1, ..., m, k = 1, ..., n)* is given by P_{ik}. The variable cost of similarity between the subjects of the classes, i.e., classes *i* and classes *j (i, j = 1, ..., m),* is given by the variable S_{ij}. The demand of hours per week for each available class *i (i = 1, ..., m)* is CT_i and the workload of each teacher per week *k (k = 1, ..., n)* is CP_k.

Let:

- x_{ik} be such that the defined variable assumes the value 1 if *i* is assigned to the class teacher *k* and 0 otherwise.
- y_{ik} be the auxiliary variable defined such that it assumes the value 1 if the classes *i* and *j* are assigned to the same teacher *k* and 0 otherwise.

Thus, x_{ik} and y_{ik} are related by equation (1).

$$y_{ij} = \sum_{k=1}^{n} x_{ik} x_{jk} \text{ , } i, j = 1, ..., m \quad (1)$$

From these variables it is possible to derive the constraints of the problem as follows:

(a) Each class should be assigned to a single teacher.

$$\sum_{k=1}^{n} x_{ik} = 1, \; i = 1, ..., m \quad (2)$$

(b) There is a maximum of hours for each teacher (workload) should be respected, i.e., the sum of the hours of classes assigned to the teacher must be less or equal to the weekly workload.

$$\sum_{i=1}^{m} CT_i x_{ik} \le CP_k \text{ , } k = 1, ..., n \quad (3)$$

Besides these constraints related to teachers, it is necessary to check the compatibility of the allocation of classes to a given teacher *k* in terms of the time they occupy in its timetable. Thus, a timetable is divided in various *slot(r,c)´s* which corresponds to the interval *r* in the day *c*. Then a variable $h_i(r,c)$ is used represent if the *slot(r, c)* is allocated (it assumes the value 1) or not (value 0) to the class *i*. This new variable must obey the constraints given by equations (4) and (5).

(c) Respect the total number of hours in a week for each class *i*, as given by Equation (4).

$$\sum_{r=1}^{R} \sum_{c=1}^{C} h_i(r,c) = CT_i \text{ , } i = 1, ...,m \quad (4)$$

(d) Avoid the conflict of time between classes allocated to the same teacher. This means that a *slot(r, c)* occupied by a class *i* cannot be shared by another class assigned to the same teacher *k*. Otherwise, there will be a conflict of time between the classes. This constraint is represented by equation (5).

$$\sum_{r=1}^{R} \sum_{c=1}^{C} \sum_{i=1}^{I} h_i(r,c) x_{ik} \le 1, \; k = 1, ..., n \quad (5)$$

The objective of this problem is to assign each class to a teacher in order to minimize the total cost of the assignments according to: individual discipline preferences and avoidance of allocation of many different disciplines to the same teacher. Then:

(a) The total cost according to individual discipline preference:

$$\sum_{i=1}^{m} \sum_{k=1}^{n} P_{ik} x_{ik}$$

(b) The total cost of similarity between different disciplines will be:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} S_{ij} y_{ij}$$

The total cost is the sum **(a)** and **(b)**, and is given by Equation (6).

$$\sum_{i=1}^{m} \sum_{k=1}^{n} P_{ik} x_{ik} + \sum_{i=1}^{m} \sum_{j=1}^{m} S_{ij} y_{ij} \quad (6)$$

Thus, the problem to be solved is to minimize the objective function given by Equation (6) subject to the constraints corresponding to Equations (1)-(5). It is possible to modify the formulation of the problem in order to eliminate the variable y_{ij} from the objective function by replacing Equation (1) in Equation (6). Thus, we obtain a formulation with only variables x_{ik} and $h_i(r, c)$ as given by the mathematical model (7).

Min $$\sum_{i=1}^{m}\sum_{k=1}^{n} P_{ik}x_{ik} + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{n} S_{ij}x_{ik}x_{jk}$$

S.a.:
$$\sum_{k=1}^{n} x_{ik} = 1,\ i = 1, \dots, m$$
$$\sum_{i=1}^{m} CT_i x_{ik} \le CP_k,\ k = 1, \dots, n \qquad (7)$$
$$\sum_{r=1}^{R}\sum_{c=1}^{C} h_i(r,c) = CT_i,\ i = 1, \dots, m$$
$$\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I} h_i(r,c)x_{ik} \le 1,\ k = 1, \dots, n$$
$$x_{ik = 0 \text{ or } 1},\ i = 1, \dots, m, ,\ k = 1, \dots, n$$
$$h_i(r,c) = 0 \text{ or } 1,\ i = 1, \dots, m$$

3. BEAM SEARCH METHOD

The problem of allocating *m* classes for *n* teachers at a university is a problem that can generate many different combinations as a result. There is no exact method able to find an optimal solution to the problem in reasonable time. The only way to guarantee an optimal solution is through an exhaustive search. In this case, it is necessary to examine the entire space of possible solutions, which is not feasible due to the large amount of solutions. For example, for a university with 50 classes and 10 teachers, the number of possible solutions is (10^{50}). If a computer can be used to examine a solution to every 1 nanosecond, it would take 3.17 $x10^{33}$ years to examine all the results for the values mentioned. Therefore, for this problem it is convenient to use heuristics to find one feasible and good solution.

To solve the PACT, the Beam Search will be used. The Beam Search approach is a heuristic based on complete enumeration (Azevedo et al., 2012; Ribeiro and Azevedo, 2009; Sabuncuoglu and Bayiz, 1999; Ow and Morton, 1988; Valente and Alves, 2005).

Before presenting the proposed algorithm, it will be developed a simplified example of the problem in order to make it easier to see the decision tree (assignments), considering that all possible solutions. Thus, one can identify the best solution and compare it with the solution found by the developed Beam Search.

3.1. Numerical example

Let $M = 4$ be the number of classes that should be assigned to $n = 2$ teachers. Each class T_i requires a workload CT_i. Thus, the number of hours the class T_1 demand is CT_1 and CT_2 is the number of hours demanded by the class T_2, and so on. These hours (periods) will be occupied by a teacher P_k, if the lessons of the class T_i are taught by teacher P_k. Each teacher has a workload CP_k available to take classes. So, CP_1 is the workload of the teacher available P_1, CP_2 is the workload of the teacher available P_2, and so on. Each class can only be taught by a single teacher and there can be two or more classes in the same slot allocated the timetable of a teacher. Figure 1 shows the problem as a bipartite graph.

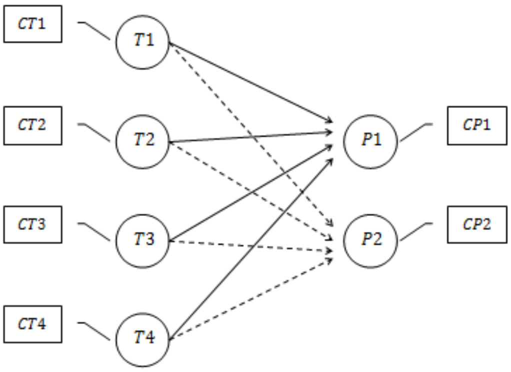


Figure 1: Representation of PACT as a graph.

As a way for the program to reach an optimal solution, it is necessary to seek information values of constraint satisfaction (maximum number of hours for each teacher and total number of hour for a class) and cost of each assignment (preferably by cost discipline and cost similarity between disciplines). Tables 1, 2, 3 and 4 show these values, respectively.

Table 1: Class demand of hours (T for "Total").

Class	Total in Week (hours)
T1	4
T2	2
T3	3
T4	3

Table 2: Teacher capacity in hours (P for "professor").

Teacher	Maximum (hours)
P1	8
P2	7

Table 3: Cost of preference for discipline.

Class\Teacher	P1	P2
T1	1	5
T2	2	3
T3	6	1
T4	6	1

Table 4: Cost of similarity between disciplines.

Class\Class	T1	T2	T3	T4
T1	0	5	12	12
T2	5	0	3	3
T3	12	3	0	0
T4	12	3	0	0

3.2. Tree solutions for the numerical example

The solution tree enumerates all possible solutions to the problem. To perform this enumeration it is necessary to establish some definitions:

- the solution tree has levels and the level *h* is assigned to the *h*-th class to a teacher;
- a complete solution is obtained only by setting all attribute classes to teachers to level *m*;
- each new attempted assignment is necessary to check if the workload of the teacher allows the assignment of the workload of the *h*-th class, considering all assignments *h-1* previous assignments;
- in each level *h*, when performing the assignment of *h*-th class to a teacher, it is essential that the costs are accounted preferably by discipline and by similarities between disciplines, considering all previous assignments *h-1*.

Figure 2 shows the tree with all possible solutions to the problem. In this figure, the numbers of teachers are represented within the nodes and the numbers related with classes are represented alongside the letter *T*, for example, T_1 represents the class 1. Each node of the tree represents an assignment, for example, the first level node N_1 symbolizes the assignment to the teacher of the class T_1 P_1. After the assignment of the root nodes in the case, P_1 and P_2, other duties are performed for the following nodes, and each node should be given the cost value considering the previous assignments.

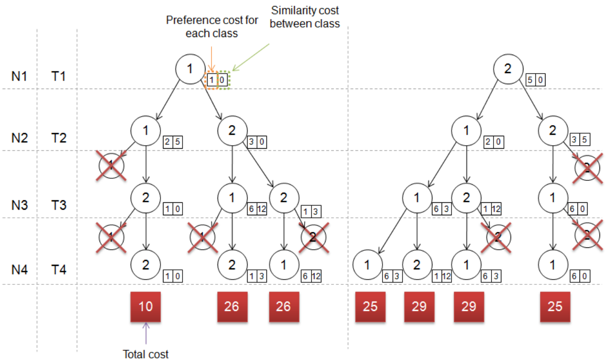


Figure 2: Possible solutions to the roots P_1 (left tree) and P_2 (right tree).

In Figure 2, at each node, the values of the cost of individual preference and similarity, in according to the allocation made to the node, are represented. Nodes marked with a cross (X) provide a solution infeasible and therefore are eliminated from the process. This is because after a few assignments, some teachers reach their maximum workload. Thus, it is understood that although the total number of solutions to be equal to $n^m = 2^4 = 16$, only 7 of these solutions are feasible.

It is also important to understand that assigning a class to a teacher held each level provides only a cost of partial similarity between disciplines. The total cost of similarity will only be achieved when all classes are assigned to a teacher. For example, at level 1 the assignment of class T_1 is done to each teacher (roots), but still cannot account for the cost of similarity between disciplines. The reason is that they do not know the assignment of the remaining classes.

At level 2, the allocation of T_2 class to each teacher is made, and it may already be counting the cost of similarity between subjects according to assignments made for each teacher. Thus, in the case of both classes T_1 and T_2 are assigned to the same teacher, and if the subjects are different, the cost will be some non-zero value. This represents the cost of preparing disciplines with different subject at the time. So, if the subject taught is the same or classes T_1 and T_2 are assigned to different teachers, the additional cost of similarity will be zero.

Similarly, at level 3, the cost of similarity will be given between classes T_1 and T_3 and between T_2 and T_3 groups. If the three groups of subjects are different, the cost of similarity between the classes should be added.

Thus, a solution will only be complete when you reach the last level of the tree (the last node) with all assignments completed and accounting costs computed.

3.3. Beam Search Method

The algorithm does not generate all the nodes of the tree as shown in the previous example. Therefore, the enumeration of all feasible solutions is not made while avoiding the exponential growth of the tree. Accordingly, we have created a few rules for the generation of nodes. These rules are also intended to prevent the creation of infeasible solutions.

First, it should be borne in mind that each node should be set up store the following information:

- identification of the teacher;
- identification of the class;
- sum of the costs of similarity between subjects from the root node;
- partial cost of the node, or adding cost, preferably in the discipline and the cost of similarity between subjects given assignments made to that node.

In addition, to create a node, it should be checked to load the remaining teacher. A node will only be created at level *h* if the workload CT_i class T_i is less than or equal to the workload $CP_k^{(h)}$ remaining teacher P_k on level *h*. So if $CP_k^{(h)} < CT_i$, all branches that originate from this node will not exist, because the class T_i cannot be attributed to the teacher P_k.

To select the nodes created, only the β nodes with smaller solutions, as pointed by a greedy algorithm, will remain as a part of the tree. From β we will create the bundles, and each development level of the tree, only still part of each beam node to generate the lowest cost solution and therefore will remain β nodes per level after the overall evaluation process. Thus, at the end of the process only β solutions will remain in the tree.

Briefly the algorithm should follow the steps.

i. Initially the algorithm is at level zero solution as no assignment has been made so far, then level = 0.

ii. Then it is level = level + 1 and nodes are created at this level, one for each teacher, taking into account their workload. For each node are calculated the cost preference for matter and similarity between subjects, and partial cost given by the sum of the other two most part calculated the cost to the previous level.

Observation: If the number of nodes created in step ii is less than or equal to the search width β back to step ii, and all nodes are expanded at the next level.

iii. For each node created in step ii, it is:

- Store the class and teacher identification;
- Store the partial node cost;
- Store the remaining capacity for each teacher, i.e., once a class is assign to a teacher his/her corresponding capacity is reduced in a number equal to total number of class hours.

iv. If the level set in step ii is less than the number of classes, applies the greedy algorithm, considering the costs and workloads partial remaining until then. The details of the greedy algorithm is as following:

(a) For the actual nodes level (parent nodes) makes it a tree structure, i.e., it creates child nodes whose tasks are feasible considering the remaining workloads of each teacher and calculate the costs of preference and similarity;

(b) For each parent node, order up the child nodes as the sum of costs and it selects the lowest cost node;

(c) Store up for the resulting nodes in (b), the partial cost and workload of the remaining teachers;

(d) If the level in question is less than the number of classes, it returns to the step (a), if not, whether the β-select solutions that generated the lowest cost. If the development of the tree is in bundles, you must select a node beam;

(e) Among the nodes created in step ii, pick up the β we generated the greedy lower-cost solutions obtained in step (c), and the others are discarded.

v. If the level set in step ii is equal to the number of classes the algorithm terminates, since the tree has reached the last level. If not, return to step ii.

In the proposed algorithm, the value of the total cost of a branch of the tree obtained from the greedy solution, acts as an upper bound (cutoff value) to generate or not the other nodes of the tree in step ii. This upper bound must be updated as a branch is found with value less cost. Thus, two criteria have been cut or not to generate a node of the tree, the remaining workload of teachers and the lowest total cost solution obtained by the greedy.

3.4. Applying the Beam Search Method to the numerical example

A detailed resolution to the example of Section 3.1 was developed, for a better understanding of the algorithm shown in Section 3.3

Recalling that m = 4 is the number of classes that should be assigned to n = 2 teachers. Furthermore, one must consider β = 2, and Tables 1, 2, 3 and 4 for query workload of classes, workload of teachers, teachers' costs preferred by subjects and costs similarity between disciplines, respectively.

In figure 4.3 we have set up the first level as the workloads of tables 1 and 2. Thus, for example, the level 1 node N_1 are: load the class and T_1 = 4 teacher load P_1 = 8 as $CP_1 > CT_1$ is possible to perform the assignment. Preferably costs (value of square left) and similarity (value of the square to the right) were obtained according query to nodes in Tables 3 and 4: the cost of the teacher's preference for the subject class P_1 is 1 and the value T_1 similarity is zero because it is the first assignment. As the number of nodes obtained is equal to the search width β = 2, it is not necessary to apply the greedy algorithm. Develop, then the nodes N_2 and feasible level are calculated costs. Up to this point were carried out only steps ii and iii.

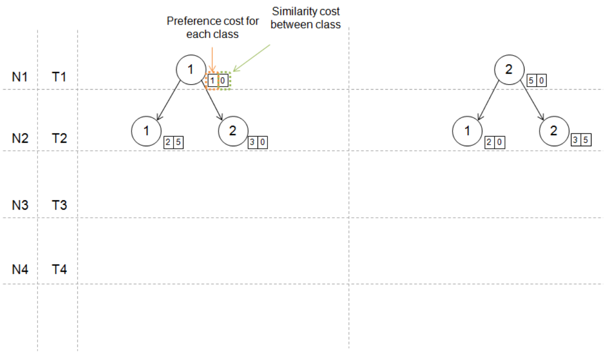


Figure 3: Development levels 1 and 2 corresponding to steps ii and iii of the Beam Search.

Figure 4 is made from the tree structure created in the standard parent nodes N_2. As can be seen in the figure, two nodes were eliminated due to the workload of the class that exceeded the remaining workload of teachers.

These nodes are marked with an X and drawn slightly above the rest. The cost preferred and similarity was calculated, obtaining thus the partial cost of each node.

As the greedy algorithm allows only one child node for each parent node, only one child node was maintained for each parent node and the others were eliminated.

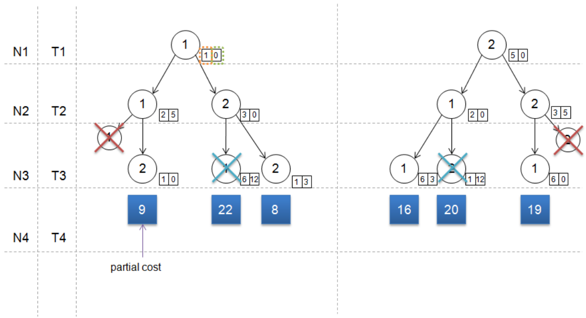


Figure 4: Tree structure of nodes of level 2 corresponding to the greedy algorithm of step iv.

Likewise, in Figure 5 was made from the tree structure of nodes that are left in the previous layer, calculate the costs and partial nodes are eliminated unnecessary.

As reached the top level of the tree, that is, level = the number of divisions = 4, β = 2 only the lowest total cost solutions remain part of the tree.

The two solutions are marked with an arrow below the total cost.

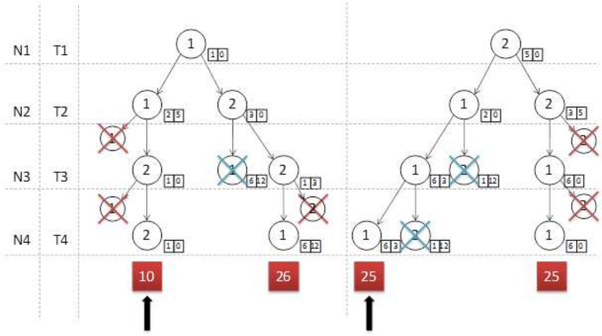


Figure 5: Tree structure of nodes of level 2 corresponding to the greedy algorithm of step iv to the sub step d.

Figures 4 and 5 correspond to step iv by the greedy algorithm of sub step d.

In Figure 6 are shown the level of the selected nodes N_2 for β = 2 greedy best solutions obtained in Figure 5, corresponding to step subsection and iv. Furthermore, we have been developed in the level N_3 and the costs calculated (steps ii and iii). In Figure 7, the greedy algorithm is applied again now to level nodes N_3 and selected the best solutions β = 2 (step iv to sub step d).

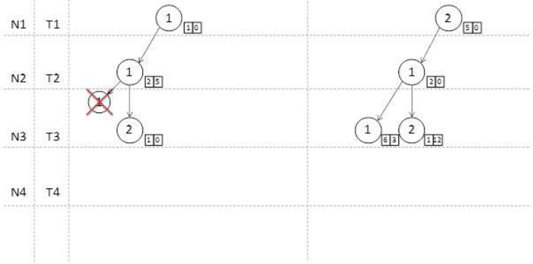


Figure 6: Selecting the nodes of level 2 (step iv sub step e) and development of nodes of level 3 (steps ii and iii).

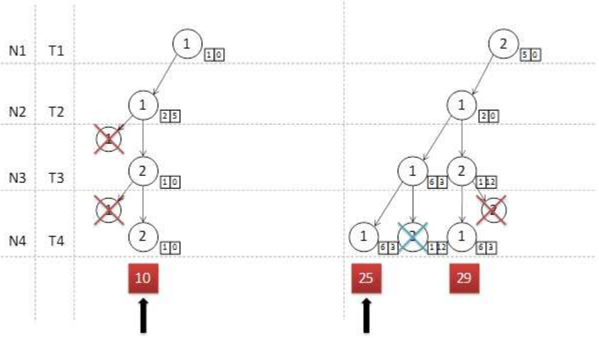


Figure 7: Arborescence level 3 (step iv to sub step d).

Figure 8 shows the selected nodes to continue the level N_3, chosen by the two best solutions greedy (step iv subsection e). Develop, too, the nodes N_4, and its cost level (steps ii and iii). Note that it is not needed to use the greedy algorithm, because the tree has reached its final level. Therefore, the calculated costs are the total costs of solutions, and among them will be selected only the best solutions β = 2, one for each beam.

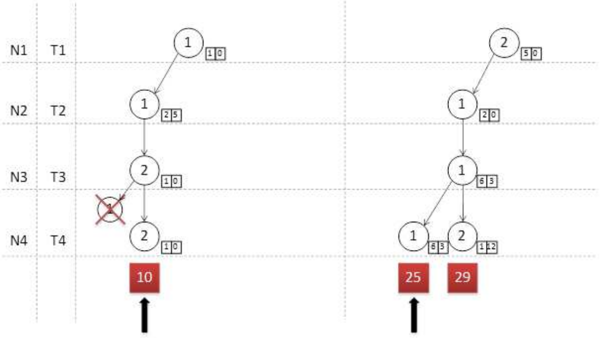


Figure 8: Selecting the nodes of level 3 (step iv to sub step e) and development of the nodes of level 4 (steps ii and iii).

Finally, Figure 9 shows the solution to the allocation problem completely, since the algorithm has reached the end N_4 which is the last level of the tree. Thus, among the best solutions β = 2, only one will be considered the final solution, which has a lower total cost. So, the final solution will be: classes T_1 and T_2 assigned to the teacher P_1 and classes T_3 and T_4 assigned to the teacher P_2.

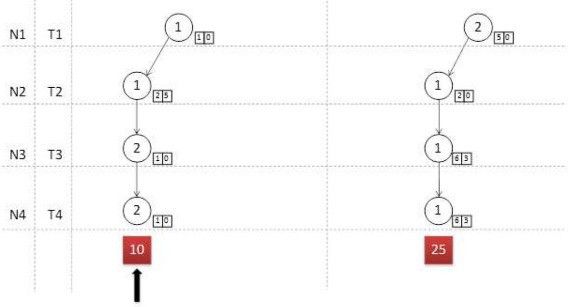


Figure 9: Final solution of the problem.

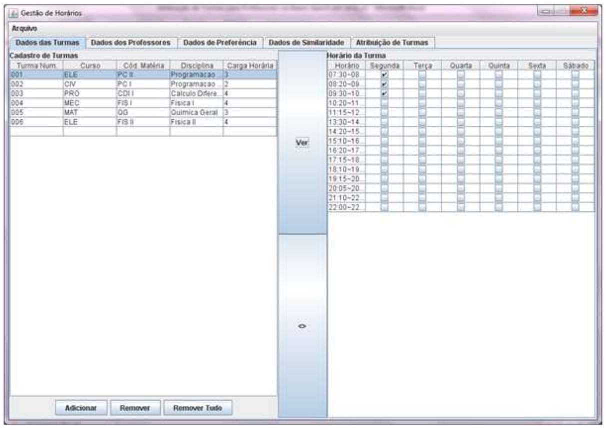
4. COMPUTATIONAL RESULTS

From the algorithm developed we present a computational tool for the PACT, using a set of classes implemented in Java. The Java language was chosen due to the ease in developing GUIs (JAVA, 2012). And for the development of the program it was used the IDE JCreator (http://www.jcreator.com/).

To archive data from teachers and class, the costs of preference for similarity between subjects and disciplines, and for the solution of the problem, it was used a set of ".txt" files. A machine-readable version of a MySQL database also began to be developed, but the time was spent to make changes to the database for use in small tests was too long. Moreover, as the PACT is only a part of a large project, how should be configured the database and its tables depend on previous steps of the project. Thus, it was decided to work with the files in ".txt", eliminating the configuration of a database and user authentication, and making quick and convenient transportation of the program and its data.

The computer used to develop the program and its testing was a Notebook, processed with Intel Core2 Duo T6500 2.10GHz, with 3GB of RAM. The operating system used was Windows 7 32-bit.

Some screens of the graphical interface developed for the developed program are provided in Figures 9 and 10.



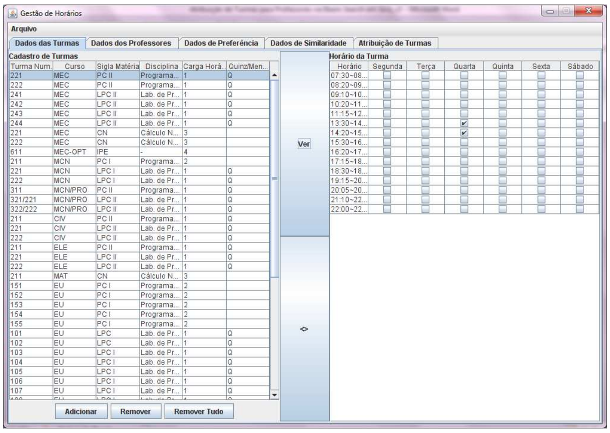
Cadastro de Turmas

Turma Num.	Curso	Cód. Matéria	Disciplina	Carga Horária
001	ELE	PC II	Programacao ...	3
002	CIV	PC I	Programacao ...	2
003	PRO	CDI I	Calculo Difere...	4
004	MEC	FIS I	Fisica I	4
005	MAT	QG	Quimica Geral	3
006	ELE	FIS II	Fisica II	4

Horário da Turma

Horário	Segunda	Terça	Quarta	Quinta	Sexta	Sábado
07:30~08	☑	☐	☐	☐	☐	☐
08:20~09	☑	☐	☐	☐	☐	☐
09:30~10	☑	☐	☐	☐	☐	☐
10:20~11	☐	☐	☐	☐	☐	☐
11:15~12	☐	☐	☐	☐	☐	☐
13:30~14	☐	☐	☐	☐	☐	☐
14:20~15	☐	☐	☐	☐	☐	☐
15:10~16	☐	☐	☐	☐	☐	☐
16:20~17	☐	☐	☐	☐	☐	☐
17:15~18	☐	☐	☐	☐	☐	☐
18:10~19	☐	☐	☐	☐	☐	☐
19:15~20	☐	☐	☐	☐	☐	☐
20:05~20	☐	☐	☐	☐	☐	☐
21:10~22	☐	☐	☐	☐	☐	☐
22:00~22	☐	☐	☐	☐	☐	☐

Figure 9: Graphical interface for data entry.



Cadastro de Turmas

Turma Num.	Curso	Sigla Matéria	Disciplina	Carga Horá...	Quinz/Men...
221	MEC	PC II	Programa...	1	Q
222	MEC	PC II	Programa...	1	Q
241	MEC	LPC II	Lab. de Pr...	1	Q
242	MEC	LPC II	Lab. de Pr...	1	Q
243	MEC	LPC II	Lab. de Pr...	1	Q
244	MEC	LPC II	Lab. de Pr...	1	Q
221	MEC	CN	Cálculo N...	3	
222	MEC	CN	Cálculo N...	3	
611	MEC-OPT	IPE	-	4	
211	MCN	PC I	Programa...	2	
221	MCN	LPC I	Lab. de Pr...	1	Q
222	MCN	LPC I	Lab. de Pr...	1	Q
311	MCN/PRO	PC II	Programa...	1	Q
321/221	MCN/PRO	LPC II	Lab. de Pr...	1	Q
322/222	MCN/PRO	LPC II	Lab. de Pr...	1	Q
211	CIV	PC II	Programa...	1	Q
221	CIV	LPC II	Lab. de Pr...	1	Q
222	CIV	LPC II	Lab. de Pr...	1	Q
211	ELE	PC II	Programa...	1	Q
221	ELE	LPC II	Lab. de Pr...	1	Q
222	ELE	LPC II	Lab. de Pr...	1	Q
211	MAT	CN	Cálculo N...	3	
151	EU	PC I	Programa...	2	
152	EU	PC I	Programa...	2	
153	EU	PC I	Programa...	2	
154	EU	PC I	Programa...	2	
155	EU	PC I	Programa...	2	
101	EU	LPC	Lab. de Pr...	1	Q
102	EU	LPC	Lab. de Pr...	1	Q
103	EU	LPC I	Lab. de Pr...	1	Q
104	EU	LPC I	Lab. de Pr...	1	Q
105	EU	LPC I	Lab. de Pr...	1	Q
106	EU	LPC I	Lab. de Pr...	1	Q
107	EU	LPC I	Lab. de Pr...	1	Q

Horário da Turma

Horário	Segunda	Terça	Quarta	Quinta	Sexta	Sábado
07:30~08...	☐	☐	☐	☐	☐	☐
08:20~09...	☐	☐	☐	☐	☐	☐
09:10~10...	☐	☐	☐	☐	☐	☐
10:20~11...	☐	☐	☐	☐	☐	☐
11:15~12...	☐	☐	☐	☐	☐	☐
13:30~14...	☐	☐	☑	☐	☐	☐
14:20~15...	☐	☐	☑	☐	☐	☐
15:30~16...	☐	☐	☐	☐	☐	☐
16:20~17...	☐	☐	☐	☐	☐	☐
17:15~18...	☐	☐	☐	☐	☐	☐
18:30~18...	☐	☐	☐	☐	☐	☐
19:15~20...	☐	☐	☐	☐	☐	☐
20:05~20...	☐	☐	☐	☐	☐	☐
21:10~22...	☐	☐	☐	☐	☐	☐
22:00~22...	☐	☐	☐	☐	☐	☐

Figure 10: Results obtained for real data.

The program showed satisfactory performance for the test with real data where it is necessary to perform the allocation of 63 classes to 11 teachers. The detailed results of the program and its comparison with the manual allocation are provided in Appendix.

5. CONCLUSIONS AND FUTURE WORKS

The PACT is a problem of combinatorial nature that is part of a more complex problem of Course Timetabling is made where the efficient management of educational resources. PACT aims to make the distribution of teachers between classes at a university in order to respect the weekly class, restrictions workload of teachers and the preferences of the same subjects to be taught, assign subjects to a similar same teacher, do not assign a class to more than one teacher and not allow a teacher to be allocated to different classes in the same time. A special feature had been considered in the model in order to consider teacher´s preference for classes with the same subject in a manner that is not considered in recent literature.

To solve the problem we developed a totally new computational tool that is a heuristic for the automatic termination of the combinatorial problem. The tool contributes mainly by the speed and efficiency in decision making for the allocation of teachers, and prevents some teachers being overloaded. The developed algorithm is a heuristic based on complete enumeration technique through the search tree and the greedy algorithm.

A graphical interface was developed to facilitate data capture to problem resolution. It is possible, through the interface, create, open, save and edit files containing data of teachers, classes of data, preference values for subjects and similarity values between disciplines. Another facility that provides graphical interface is the possibility to simulate various assignments with various widths search soon. This facility brings a very big advantage is that the visualization and comparison of different configurations of the timetable, including performing rapid changes in the values of preference and similarity to determine the effect on assignments. At the end of the assignment, the program also lets you make manual adjustments in the result, if a teacher wants to make a simple change that does not result in shocks or extrapolation of the time course load.

This GUI can be expanded in the future to integrate data capture and resolution of the remaining issues in managing educational resources related to PACT. Furthermore, one can add other features such as printing files in formats ideal for viewing and distribution among teachers, and also improve the logical assignment to get results closer to the manual distribution. The development of the interface in additional languages is also a relevant future work to be accomplished.

ACKNOWLEDGMENTS

The authors would like to thank the following organizations for their support: Saab Group (www.saabgroup.com), University of Skövde (www.his.se), University of São Paulo (www.usp.br), University of Campinas (www.unicamp.br) and National Counsel of Technological and Scientific Development (www.cnpq.br).

APPENDIX

The following are the results obtained by manual and program.

Table 5: Manually obtained results.

Results without the use of computer simulations	
Discipline	**CHD**
Teacher 1	**15/9h**
PC II - MEC(222)	2*
LPC II - MEC(243)	2*
LPC II - MEC(244)	2*
LPC II - MCN(321)/PRO(221)	2*
LPC II - MCN(322)/PRO(222)	2*
LPC II - CIV(222)	2*
PC II - MCN(311)/PRO(222)	2*
PC II -CIV (211)	2*
LPC II - CIV (221)	2*
Teacher 2	**0/8h**
PC I - EU(155)	2
PC I - ESP(131)	2
LPC I - EU(109)	2*
LPC I - EU(110)	2*
LPC I - EU(161)	2*
LPC I - EU(162)	2*
Teacher 3	**14/10h**
CDI	2
CDI	2
PC I - MCN (211)	2
CDI	2
LPC I - MCN(221)	2*
LPC I -MCN(222)	2*
Teacher 4	**3/3h**
PC II - ELE(211)	2*
LPC II - ELE(221)	2*
LPC II - ELE(222)	2*
Teacher 5	**13/9h**
MAC - ESP(111)	2
PC I - ESP(132)	2
PC II - MEC(221)	2*
LPC II - MEC(241)	2*
LPC II - MEC(242)	2*
LPC I - ESP(163)	2*
LPC I - ESP(164)	2*
Teacher 6	**13/13h**
CN - MEC(222)	3
CN - MAT(211)	3
IPE - MEC-OPT(611)	4
CN - MEC (221)	3
Teacher 7	**0/8h**
PC I - EU(151)	2
PC I - EU(152)	2
LPC - EU(101)	2*
LPC I - EU(104)	2*
LPC - EU(102)	2*
LPC I - EU(103)	2*
Teacher 8	**12/12h**
PC I - EU(154)	2
LPC I - EU(108)	2*
CCN - ESP(211)	2
LPC I - EU(107)	2*
CCN - ESP(211)	2
PC I - ESP(133)	2
LPC I - ESP(165)	2*
LPC I - ESP(166)	2*
Teacher 9	**7,5/4h**
CN - LMN(211)	2
LCN - LMN(211)	2
Teacher 10	**3/3h**
PC - LMN(111)	2
LPC - LMN(111)	2*
Teacher 11	**6/4h**
PC I - EU(153)	2
LPC I - EU(105)	2*
LPC I - EU(106)	2*

Table 6: Results provided by the software in Java.

Computer based simulations	
Discipline	**CHD**
Teacher 1	**15/15h**
PC II - MEC(222)	2*
LPC II - MEC(241)	2*
LPC II - MEC(242)	2*
LPC II - MEC(244)	2*
PC II - MCN(311)/PRO(222)	2*
LPC II - MCN(321)/PRO(221)	2*
LPC II - MCN(322)/PRO(222)	2*
PC II -CIV (211)	2*
LPC II - CIV(222)	2*
PC II - ELE(211)	2*
LPC II - ELE(222)	2*
PC I - EU(151)	2
PC I - EU(152)	2
Teacher 2	**0/0h**
Teacher 3	**14/13h**
CN - MEC (221)	3

PC I - MCN (211)	2
PC - LMN(111)	2
CDI	2
CDI	2
CDI	2
Teacher 4	**3/3h**
LPC II - MEC(243)	2*
LPC II - CIV (221)	2*
LPC II - ELE(221)	2*
Teacher 5	**13/13h**
LPC I - MCN(221)	2*
LPC I -MCN(222)	2*
PC I - EU(153)	2
PC I - EU(154)	2
LPC I - EU(103)	2*
PC I - ESP(131)	2
PC I - ESP(133)	2
LPC I - ESP(163)	2*
LPC I - ESP(164)	2*
Teacher 6	**13/12h**
PC II - MEC(221)	2*
CN - MEC(222)	3
IPE - MEC-OPT(611)	4
CN - LMN(211)	2
LCN - LMN(211)	2
Teacher 7	**0/0h**
Teacher 8	**12/11h**
LPC I - EU(104)	2*
LPC I - EU(105)	2*
LPC I - EU(106)	2*
LPC I - EU(107)	2*
LPC I - EU(108)	2*
LPC I - EU(109)	2*
LPC I - EU(110)	2*
LPC I - EU(161)	2*
LPC I - EU(162)	2*
LPC I - ESP(165)	2*
LPC I - ESP(166)	2*
Teacher 9	**7,5/7h**
CN - MAT(211)	3
CCN - ESP(211)	2
CCN - ESP(211)	2
Teacher 10	**3/3h**
LPC - EU(101)	2*
LPC - EU(102)	2*
LPC - LMN(111)	2*
Teacher 11	**6/6h**
PC I - EU(155)	2
MAC - ESP(111)	2
PC I - ESP(132)	2

REFERENCES

Al-Yakoob, S.M., Sherali, H.D., A column generation mathematical programming approach for a class-faculty assignment problem with preferences, *to appear in Computational Management Science*, pp. 1-22, 2013.

Al-Yakoob, S. M. , Sherali, H.D., Al-Jazzaf, M., A mixed-integer mathematical modeling approach to exam timetabling, *Computational Management Science*, Vol. 7(1), pp. 19-46, 2010.

Azevedo, A.T., Ribeiro, C.M., Sena, G.J.D., Chaves, A.A., Neto, L.L.S., Moretti, A.C.: Solving the 3D Container Ship Loading Planning Problem by Representation by Rules and Beam Search. ;In ICORES, pp.132-141, 2012.

Deitel, H. M.; Deitel, P. J. Java: How to Programm. 6. ed. Bookman, 2006.

Della Croce, F.; T'kindt, V., A Recovering Beam Search Algorithm for the One-Machine Dynamic Total Completion Time Scheduling Problem, *Journal of the Operational Research Society*, vol 54, 2002, pp. 1275-1280.

Gunawan, A., Ng, K.M., Poh, K.L., Solving the Teacher Assignment-Course Scheduling Problem by a Hybrid Algorithm, CiteSeer. Available in: < http://130.203.133.150/viewdoc/download?doi=10.1.1.193.3646&rep=rep1&type=pdf>. Access: 21 jun. 2013.

Java™ , Sun Microsystems, Platform, Standard Edition 6 API Specification. Available in: <http://java.sun.com/javase/6/docs/api/>. Access: 20 jan. 2012.

Michael, W.C., Laporte, G., Recent Developments in Practical Course Timetabling, *Selected papers from the Second International Conference on Practice and Theory of Automated Timetabling II*, Springer-Verlag, London, UK, pp. 3-19, 1998.

Sabuncuoglu, I.; Bayiz, M., Job Shop Scheduling with Beam Search, *European Journal of Operational Research*, vol. 118, 1999, pp. 390-412.

Ow, P.S; Morton T.E., Filtered Beam Search in Scheduling, *International Journal of Production Research*, vol. 26, 1988, pp. 35-62.

Ribeiro, C.M., Azevedo, A.T., Teixeira, R.F.,Problem of assignment cells to switches in a cellular mobile network via Beam Search Method, WSEAS Transactions on Communications, Vol. 9(1): pp.11-21, 2009.

Schaerf, A. A Survey of Automated Timetabling. Dipartimento di Informatica e Sistemistica, Università di Roma “La Sapienza”, 1999. Available in: <http://www. diegm.uniud.it/satt/papers/Scha99.pdf>. Access: 16 jun. 2011

Valente, J. M. S; Alves, R. A. F. S., Filtered and Recovering Beam Search Algorithm for the Early/Tardy Scheduling Problem with No Idle Time, *Computers & Industrial Engineering*, vol. 48, 2005, pp. 363-375.

Willenmen, R. J. School timetable construction:

algorithms and complexity. Technische Universiteit Eindhoven, 2002. Available in: < http://alexandria.tue.nl/extra2/200211248.pdf >. Access: 10 jul.2011.

AUTHORS BIOGRAPHY

Anibal Tavares de Azevedo. PhD in Engineering. Dr. Azevedo teaches at Universidade Estadual de Campinas – UNICAMP (http://www.unicamp.br/), in Brazil. Previously, he worked as a researcher and as a teacher at Universidade Estadual Paulista – UNESP (http://www.unesp.br/), Brazil. Dr. Azevedo graduated in Applied and Computational Mathematics from UNICAMP (1999), holds a Master's degree in Electrical Engineering from UNICAMP (2002) and received his Ph.D. degree in Electrical Engineering from UNICAMP (2006). He has experience in software development and in mathematical modeling for Production Engineering and Planning, for Scheduling of Power System Operation and for Education. His research has an emphasis on Linear Programming, Nonlinear Programming and Mixed Dynamics in the following topics: interior point methods, planning and production control of manufacturing, flows in networks, linear programming, graph generalized combinatorial optimization, allocation of cells to cellular centrals, loading and unloading of containers on ships 2D and 3D, genetic algorithms, beam search and simulated annealing. < http://lattes.cnpq.br/9760457138748737 >.

Andressa Fernanda Saemi Matsubara Ohata. Engineer. Previously worked as a researcher at Universidade Estadual Paulista – UNESP (http://www.unesp.br/), Brazil.

Joni A. Amorim. PhD in Engineering. Postdoctoral Fellow at the University of Skövde, or Högskolan i Skövde – HiS (http://www.his.se/english/), in Sweden, in collaboration with SAAB Training and Simulation (http://www.saabgroup.com/en/training-and-simulation/). The University of Skövde offers first-class programs and competitive research, which attracts research scientists and students from all over the world. The University of Skövde is one of the most specialized universities in Sweden and its research is focused on the development and use of advanced information technology systems and models. Dr. Amorim previously worked as a researcher and as a teacher at Universidade Estadual de Campinas – UNICAMP (http://www.unicamp.br/), in Brazil. Dr. Amorim collaborates with researchers at UNICAMP, a university with more than 3,600 original scientific publications published in 2009, 78% of which in journals indexed in the ISI/Web of Science. Dr. Amorim collaborates with researchers at Universidade de São Paulo – USP (http://www.usp.br/), the major institution of higher learning and research in Brazil. His research has an emphasis on multimedia production management, project portfolio management, distance education and training based on serious games and simulations. < http://lattes.cnpq.br/3278489088705449 >.

Per M. Gustavsson. PhD in Computer Science. Dr. Gustavsson works as a Research Scientist at Saab Group (http://www.saabgroup.com/). Saab serves the global market with products, services and solutions ranging from military defence to civil security. Dr. Gustavsson also works at the Swedish National Defence College – SNDC (http://www.fhs.se/en/), in Sweden. At SNDC, research is carried out in diverse, but inter-related subject areas and subsequently disseminated to other interested sectors of society both nationally and internationally; the College trains and educates military and civilian personnel in leading positions, both nationally and internationally as part of the contribution to the management of crisis situations and security issues. < http://se.linkedin.com/in/pergu >.

MODELING AND SIMULATION OF AN ELECTRIC CAR SHARING SYSTEM

Monica Clemente[a], Maria Pia Fanti[b], Giorgio Iacobellis[c] , Walter Ukovich[d]

[a]Department of Engineering and Architecture, University of Trieste, Trieste, Italy
[b]Department of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy
[c] Department of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy
[d] Department of Engineering and Architecture, University of Trieste, Trieste, Italy

[a]monica.clemente@phd.units.it, [b],[c]{fanti,iacobellis}@deemail.poliba.it, [d]ukovich@units.it

ABSTRACT

Electric Vehicles (EVs) represent an effective response to the pressing environmental and economic problems linked to the mobility in the urban areas. However, limited drive range and high purchase costs limit strongly their popularity. The deployment of EVs in the car sharing (CS) fleets is an effective strategy to overcome these initial drawbacks. In order to make it competitive with the traditional private mobility forms an accurate planning is required. With the aim of obtaining a tool able to highlight the impact of the EVs adoption on the performances of a CS service and evaluate different operative conditions, this paper develops a discrete event simulation model of a generic electric-CS system. In order to assess the validity of the approach, a real case study is analyzed and simulated.

Keywords: Electric Vehicles, Car Sharing, Discrete-event Simulation.

1. INTRODUCTION

In the last decades, the pressing need of reducing the energy dependency of the mobility on fossil fuels and the growing environmental problems highlighted the importance of identifying concrete alternatives to the traditional Internal Combustion Engine (ICE) vehicles. In particular, Electric Vehicles (EVs) are reaching increasing interest and the Governments of many Countries all over the world have been investing huge amounts of money in the research and the development of this technology. In particular, the drawbacks that mainly affect the popularity of the EVs are high prices and limited driving range (which causes the so-called *range anxiety*): concrete answers to these concerns have to be given. In this context, an opportunity is represented by the *shared-use vehicle systems*, i.e., systems in which a company makes available to registered users a common fleet of vehicles and customers pay only for the actual utilization of the rented vehicle (Barth and Shaheen 2002). Thanks to this new form of mobility, therefore, users can share the fixed costs usually associated to the ownership of a private means of transport and, at the same time, the overall mobility behavior becomes more rational, since customers are more aware of the effective costs associated to the conformation of their trips. Moreover, among the other forms of shared-used vehicle systems, the *car sharing* (CS) seems to be the most suitable for the adoption of the EVs. In this kind of organization users can rent a car also for limited time periods and, usually, for trips within a specified urban area. Hence, the distances typically travelled in this context are compatible with the EVs driving ranges.

However, the deployment of EVs in CS fleets introduces some management complexities that have to be faced carefully in order to guarantee service efficiency and flexibility and, therefore, to overtake the undeniable competitive advantages of the traditional private mobility forms. In this context, a simulation approach can be useful to identify the best operative strategies and to highlight the critical issues of the considered service.

This paper is part of this framework and deals with the development of a simulation model useful to analyze the behavior of a generic CS service and to assess the impact of the EVs adoption on the overall system performances.

Moreover, the used approach can be outlined as follows. First, a detailed analysis of the peculiarities characterizing the management of a generic CS service and the deployment of the EVs is conducted. Second, the results of this phase are formalized through the Unified Modeling Language (UML) and, in particular, class and activity diagrams are developed. Finally, a simulation model is realized in the Arena environment, a discrete event simulation software. In order to assess the validity of the developed simulator and the effectiveness of different operative conditions, a real case study involving the electric-CS service of Pordenone, a city of the North of Italy, is taken into account, and different simulations are carried out.

The remainder of the paper is structured as follows. In section 2 the electric-CS problem is analyzed and formalized through UML class and activity diagrams. Section 3 describes a specific case study, while in Section 4 the Arena model is developed and the system behavior under different operative conditions is simulated. Finally, in Section 5 the conclusions are summarized.

2. THE ELECTRIC-CAR SHARING PROBLEM

2.1. UML Class Diagram

In order to make a generic CS service efficient and competitive, several parameters must be analyzed and calibrated properly. More in detail, the main key issues to consider are:

1. the *optimal fleet size* (George and Xia 2011, Nakayama, Yamamoto and Kitamura 2002);
2. the *location of the parking areas* (Correia and Antunes 2012);
3. the *pricing policies*;
4. the *service accessibility* (Barth, Todd, and Xue 2004);
5. the *rental rules*, intended as the possibility or not for the users to pick up the vehicle in a parking area and to return it in a different one (Wang, Cheu and Lee 2010, Ken, Chew, Meng and Fung 2009, Uesugi, Mukai and Watanabe 2007, Mukai, Watanabe 2005).

Moreover, when EVs are deployed in the service fleet, additional concerns must be taken into account (Nakayama, Yamamoto and Kitamura 2002, Xu, Miao, Zhang and Shi 2013, Chen, Kockelman and Khan 2013):

6. the *number of the charging stations*;
7. the *location of the charging stations*;
8. the *recharging policies*, that is, when and how the EVs must be recharged (every time the users return them after the rental, or when the battery SoC is below a certain threshold, and so on).

In order to formalize these aspects in a synthetic manner, the UML formalism is adopted. In particular, to represent the structure of a generic CS service whose fleet consists also of EVs, a class diagram is developed. This kind of diagram is useful to highlight the different types of objects that the considered system have and the relationships between them. More in detail, each class represents a set of objects characterized by the same attributes, operations and relationships. Every class is depicted by a rectangular box divided into compartments: the first compartment holds the class name, which must be unique and distinguishes the classes from each other; the second one holds the class attributes, which are the qualities that describe its characteristics; the last compartment holds the operations, that is, features that specify the behavior of the class. Between classes various types of relationships are possible and they are represented with different graphic connections: association (solid line), aggregation (solid line with a clear diamond at one end), composition (solid line with a filled diamond at one end), generalization (solid line with a clear triangle at one end), realization (dashed line with an arrow at one end) and dependency (dashed line with an arrow at one end). In addition, labels on the lines express the multiplicity of the considered relationship, that is, how many instances of a particular class are involved in a relationship.

In Fig. 1 the class diagram describing the CS problem is depicted. As can be seen, in a generic CS system the following structural components can be pointed out:

- *Management System*: it deals with the monitoring and the coordination of the system and with the system state information collection;
- *Reservation Management System*: system dealing with the management of the vehicles reservation operations;
- *Payment Management System*: centralized system that, once collected travel data of a specific rented vehicle, calculates how much the customer has to pay and, then, manages the payment phase;
- *Relocation Activities Management System*: if the user can pick-up and return the vehicle in different parking areas, this system monitors the distribution of vehicles among the parks and, if necessary, starts the vehicle relocation activities;
- *Pricing Policies Determination System*: system that determines which pricing policy must be chosen and if a mechanism of economic incentives to the users has to be started;
- *Emergencies Management System*: module that handles any emergency that users communicate to the system through ICT tools installed on board each vehicle;
- *Registrations Management System*: centralized system which manages the registration phase of new customers;
- *Maintenances Management System*: module that deals with the maintenance and repair of fleet vehicles;
- *Car Sharing Company*: CS company on the whole;
- *Operator*: employee of the CS organization;
- *User*: once registered into the system, he can rent a vehicle;
- *Vehicle*;

- *Traditional Vehicle*: child class of "Vehicle", represents any car with an Internal Combustion Engine (ICE);
- *Electric Vehicle*: child class of "Vehicle", represents any kind of Electric Vehicle (EV);
- *Parking Area*: station at which it is possible to rent/return a vehicle;
- *Traditional Parking Area*: child class of "Parking Area", identifies simply the place where the vehicles are available to be rented or can be returned;
- *Charging Station*: child class of "Parking Area", identifies a parking equipped with an EV charging infrastructure and, then, a parking area characterized by a greater management complexity.

Moreover, the following association classes can be stressed:

- *Rental* (between the classes *User, Parking Area* and *Vehicle*);
- *Relocation* (between the classes *Parking Area, Vehicle* and *Operator*);
- *Maintenance* (between the classes *Vehicle* and *Operator*);
- *Emergency* (between the classes *Vehicle* and *Operator*);
- *Purchase/Substitution* (between the classes *Vehicle* and *Operator*);
- *Share a Vehicle* (between different instances of class *User*).

2.2. UML Activity Diagrams

The values of the attributes of the classes described in the previous subsection represent the state of the system: the attributes update rules and the system behavior is modeled by the activity diagrams. Indeed, UML activity diagram is useful to highlight the set of events that, coming in succession, determine any process occurring in a given system.

More in detail, each activity diagram is divided in columns, the so-called *swim lanes*, in order to clearly stress which actor is responsible for which action. The main elements characterizing this kind of diagram are: the *initial activity* (depicted with a solid circle), the *final activity* (represented with a bull's eye symbol) and *generic activities*, depicted with rectangles with rounded edges. *Flows* are represented through arcs connecting activities, while *alternative flows*, and, therefore, *decisions*, are denoted with diamonds. Finally, concurrent activities respectively beginning and ending at the same time (*forks* and *joins*) are depicted with a tick horizontal line.

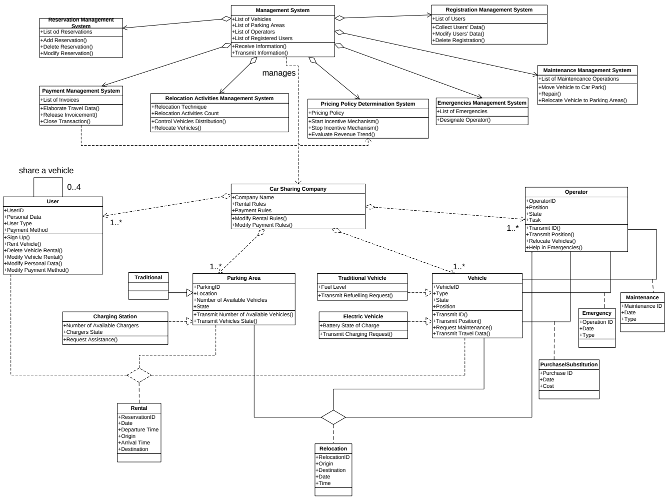


Figure 1: Electric-CS service class diagram

In this context, among the other process characterizing the evolution of an electric-CS service, we consider the car renting process and its rules, since it is the most affected by the introduction of the EVs. When EVs are deployed in the service fleet the management complexity increases, since the wait for an available charger and the charging time are additional delays that can reduce drastically the number of served users and, so, the company gain.

In Fig. 2 the activity diagram of the car renting process is reported. Such a process involves three actors: the *user*, the *vehicle*, and the *management system*. Seven main phases characterize it:

1. the *vehicle request* phase, representing the request and the user waiting for an available vehicle;
2. the *checking vehicle availability* phase, during which the management system checks if there are available vehicles at the parking area;
3. the *rental and use of the vehicle* phase: note that the *"Travel time determination*" and the "*Destination parking determination*" activities are simple schematization of the users' decision process and not required declarations to the system.
4. the *vehicle restitution* phase;
5. the *charging* phase: each vehicle, once given back in a parking area, waits for an available charger and starts its recharging process. In this case it is assumed that the charging takes place after each rental period, as it is reasonable to suppose that users will be more confident if they start their travel with a fully-charged vehicle.
6. the *maintenance* phase, which involves only the vehicles that need maintenance operations after the rental period;
7. the *payment* phase, after which the user leaves the system.

3. A CASE STUDY

In order to analyze more comprehensively the effects of the introduction of EVs in a CS fleet, a particular case study is considered. From data derived from the experience of the electric-CS service of Pordenone, a town of the North of Italy, the following system is studied.

A number of 10 electric vehicles is made available to the users for rental in two parking areas, named respectively P1 and P2. Furthermore, we assume that the parking area P1 is characterized by a greater attractiveness (and, so, a greater mobility demand) than the other one. The service fleet consists of small EVs, with a driving range of about 40 km and a maximum recharge time of 1.5 hours.

The following rules, relevant for the purpose of this work, manage the service:

- it is not allowed to leave the municipality territory;
- rented vehicles must be given back by 8:15 pm;
- rented vehicles can be returned in any parking area of the system.

Moreover, the following additional assumptions are considered:

- *maximum waiting time*: we assume that users are not willing to wait more than 10 minutes to rent a vehicle and so, after this time interval, they leave the system without being served;
- *maintenance*: we consider the possibility that a vehicle, after the rental period, needs a repair service. Moreover, two different types of maintenance are taken into account: the first one is of about 8 hours, the second one takes 1 hour;
- *initial distribution of the vehicles*: we consider that initially vehicles are equally distributed between the two parking areas;
- *vehicle recharging operations*: we assume that vehicles are recharged after each rental and that the users start their travels with a fully charged vehicle;
- *number of available chargers*: initially we assume that both P1 and P2 are equipped with 2 chargers.

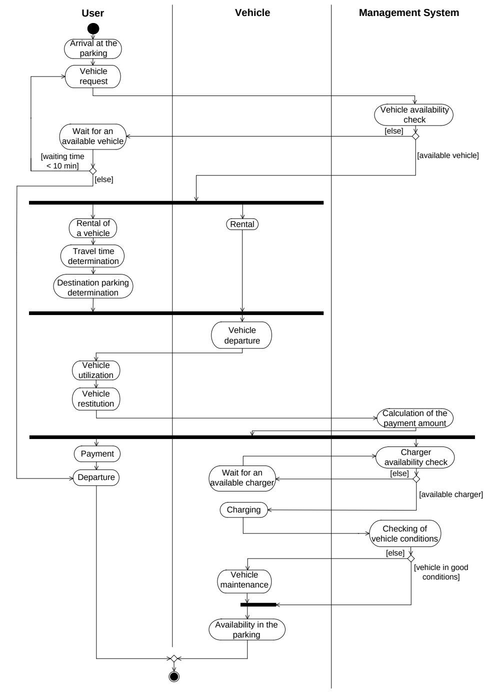

User
Vehicle
Management System
Arrival at the parking
Vehicle request
Vehicle availability check
[else]
Wait for an available vehicle
[available vehicle]
[waiting time < 10 min]
[else]
Rental of a vehicle
Rental
Travel time determination
Destination parking determination
Vehicle departure
Vehicle utilization
Vehicle restitution
Calculation of the payment amount
Payment
Charger availability check
Wait for an available charger
[else]
Departure
[available charger]
Charging
Checking of vehicle conditions
[else]
[vehicle in good conditions]
Vehicle maintenance
Availability in the parking

Figure 2: UML activity diagram of the car renting process

4. THE SIMULATION MODEL AND RESULTS

4.1. The Simulation Specification

In order to analyze the behavior of the considered system, a simulation model is implemented in the Arena environment, a discrete-event simulation software. More precisely, each activity of the UML diagram of Fig. 2 is modeled by a discrete event of the simulation. Users and vehicles are entities requiring and seizing resources and the management system is the actor that determines the system evolution rules.

Users' inter-arrival times (named A1 and A2 for parking area P1 and P2, respectively) are modeled by an exponential distribution of mean λ time units (t.u.), where the minute is considered as t.u. In order to express the operation time, the rental, maintenances and charging operations have triangular distribution.

Two possible travel times, denoted by T1 and T2 respectively, are considered and different modeling approaches are used: T1 is modeled by a triangular distribution, T2 is modeled by an exponential distribution.

The parameters of all the mentioned distributions are reported in Table 1: the third column reports the average values of the exponentially distributed times (EXPO) and the modal (δ), minimum (d_δ) and maximum (D_δ) values of the triangular distributions.

Table 1: Used Probability Distributions and Parameters

Delay	Distribution	Parameter [min]
A1	EXPO	λ=24
A2	EXPO	λ=30
Rental	TRIA	$(d_\delta,\delta,D_\delta)$=(2,3,4)
Return and payment	TRIA	$(d_\delta,\delta,D_\delta)$=(3,4,5)
T1	TRIA	$(d_\delta,\delta,D_\delta)$=(13,20,30)
T2	EXPO	λ=60
Move to car park	EXPO	λ=60
Maintenance (short)	TRIA	$(d_\delta,\delta,D_\delta)$=(48,60,72)
Maintenance (long)	TRIA	$(d_\delta,\delta,D_\delta)$= (384,480,576)
Charging	TRIA	$(d_\delta,\delta,D_\delta)$=(30,60,90)

The performance index defined to evaluate the system behavior is the *Level of Service* (*LOS*), defined as the fraction of served users as follows:

$$LOS = \frac{number\ of\ served\ users}{total\ number\ of\ users\ in\ the\ service}. \qquad (1)$$

The metric *LOS* are evaluated by a long simulation run of 21600 t.u., with a warm-up period of 30 t.u.. The estimates are deduced by 50 independent replications with a 95 % confidence interval, whose half width is about 2.2 % in the worst case.

Note that the average CPU time for a simulation run is about 15 seconds on a PC with a 1.40 GHz processor and 6 GB RAM: the presented simulator can be therefore applied to larger and more complex systems.

4.2. The Simulation Results

In order to analyze the system behavior as the level of request changes, a 5 minutes-step variation of users' average inter-arrival times is considered. More specifically, the mean values of the exponential distributions of the users' arrival times vary in the following intervals: $\lambda\in[5, 59]$ in P1 and $\lambda\in[10, 65]$ in P2.

In Fig. 3 the comparison between system LOS of a traditional CS service (i.e., a CS with ICE vehicles) and of an electric-CS service is depicted. It is apparent that there is a general worsening of the system performances when EVs are deployed in the fleet, since vehicles are not available for rental for longer time periods (on average, 21% less users are served).

Moreover, with the aim of assessing how the number of available chargers influences the *LOS*, we consider one more available charger in P1 (which is the most attractive destination). System performances under this new operative condition are reported in Fig. 4. The results highlight that the increase of the number of available chargers leads to a growth of the *LOS*. Moreover, we point out that when the level of congestion of the system is low, this solution is not incisive and still 15% of users are not served. At the same time, even when the system is really congested, the availability of this new resource leads to a moderate increase (of about 5%) of the fraction of served users. Therefore, such a kind of action is not sufficient to overcome the system inefficiency.

Finally, we finally use the simulation to determine the number of EVs that has to be initially available in the service in order to obtain the same *LOS* of the case with ICE vehicles. In particular, for this analysis we consider in P1 and in P2 the average inter-arrival times reported in Table 1. As can be seen in Fig. 5, a value of *LOS* equal to 0.92 (which is the value of *LOS* in the corresponding scenario when ICE vehicles are considered) is reached when there are 23 available EVs, which means when the size of the service fleet has been doubled.

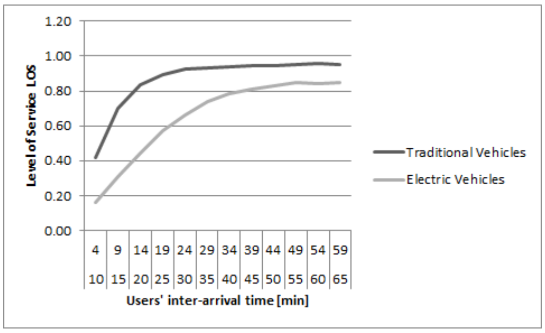


Figure 3: Level of Service comparison: ICE vs. electric vehicles.

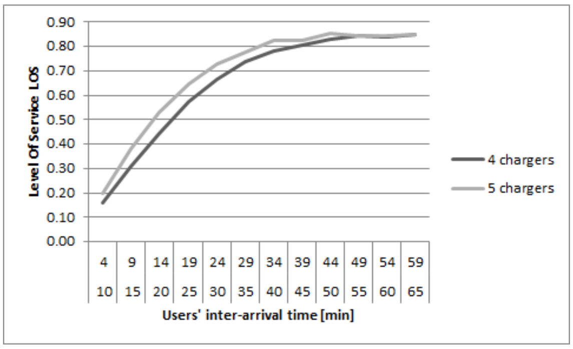


Figure 4: Level of Service comparison: 4 vs. 5 chargers.

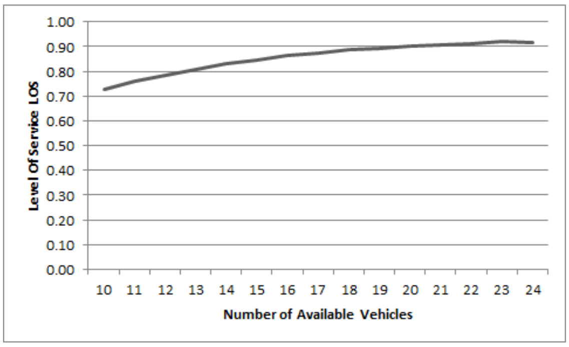


Figure 5: Level of Service in function of the number of available vehicles.

5. CONCLUSIONS

The aim of this paper is to formalize the issues characterizing the adoption of EVs in CS fleets and to develop a tool useful to assess system performances and evaluate different possible operative scenarios. In particular, a discrete event simulation in Arena environment is realized and a specific electric-CS service based on a real CS experience is analyzed.

Simulations results enhance the management complexities introduced by the EVs deployment. There is a general deterioration of the system performances and different parameters have to be re-calibrated in order to obtain the same level of service guaranteed in the traditional case. However, environmental benefits and the incentive that an electric-CS service can give to the diffusion of the EVs, encourage finding solutions able to ensure the competitiveness and the efficiency of such a kind of mobility form. The realized simulation is useful to perform this type of analysis. Moreover, considering its modular structure and the average CPU time for a simulation run, the simulation model can be easily extendable to greater and more complex systems.

Future researches will investigate about the possibility of improving system performances by influencing users' mobility behavior and applying suitable pricing strategies.

ACKNOWLEDGMENTS

This paper is supported by the project "Un electric Car Club per il FVG", financed by the Friuli Venezia Giulia region, in Italy, within the framework "contributi a sostegno della ricerca, dello sviluppo, dell'innovazione e del trasferimento tecnologico per lo sviluppo di sistemi per la mobilità individuale finalizzati alla riduzione di consumi e di emissioni, ai sensi dell'articolo 16 della Legge Regionale 11 agosto 2010 n.14".

REFERENCES

Correia, G. H., Antunes, A. P.,2012. Optimization Approach to Depot Location and Trip Selection in One-Way Carsharing Systems. *Transportation Research Part E,* No. 48, pp. 233-274.

Barth, M., Shaheen, S. A., 2002. Shared-Use Vehicle Systems: Framework for Classifying Car sharing, Station Cars and Combined Approach. *Transportation Research Record* 1791, Paper No. 02-3854, pp. 105-112.

Bart, M., Todd, M., Xue, L., 2004. User-Based Vehicle Relocation Techniques for Multiple-Station Shared-Use Vehicle Systems. *Proceedings of the Transportation Research Board Annual Meeting*. January 2004, Washington D. C. (USA).

Chen, T. D., Kockelman, K. M., Khan, M., 2013. The Electric Vehicle Charging Station Location Problem: A Parking-Based Assignment Method for Seattle. *Proceedings of the 92° Annual Meeting of the Transportation Research Board*, January 2013, Washington D. C. (USA).

George, D. K., Xia, C. H., 2011. Fleet-sizing and Service Availability for a Vehicle Rental System via Closed Queuing Networks. *European Journal of Operational Research*, Vol. 211, Issue 1, pp. 198-207.

Kek, A. G. H., Cheu, R. L., Meng, Q., Fung, C. H., 2009. A Decision Support System for Vehicle Relocation Operations in Carsharing Systems. *Transportation Research Part E,* No. 45, pp. 149-158.

Mukai, N., Watanabe, T., 2005. Dynamic Location Management for On-Demand Car Sharing System. *Proceedings of the 9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES 2005*. September 14-16 2005, Melbourne (Australia).

Nakayama, S., Yamamoto, T., Kitamura, R., 2002. Simulation Analysis for the Management of an Electric Vehicle-Sharing System – Case of the Kyoto Public-Car System. *Transportation Research Record* 1791, Paper No. 02-2653, pp. 99-104

Uesugi, K., Mukai, N., Watanabe, T., 2007. Optimization of Vehicle Assignment for Car Sharing System. *Proceedings of the 11th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES 2007, XVII Italian Workshop on Neural Networks*, September 12-14 2007, Vietri sul Mare (Italy).

Wang, H., Cheu, R., Lee, D. H., 2010. Dynamic Relocating Vehicle Resources Using a Microscopic Traffic Simulation Model for Carsharing Services. *Proceedings of the 2010 Third International Joint Conference on Computational Science and Optimization*. May 28-31, Huangshan (China).

Xu, H., Miao, S., Zhang, C., Shi, D., 2013. Optimal Placement of Charging Infrastructures for Large Scale Integration of Pure Electric Vehicles Into Grid. *Electrical Power and Energy Systems* No. 53, pp. 159-165.

AN ANALYTICAL HIERARCHY PROCESS METHODOLOGY TO EVALUATE IT SOLUTIONS FOR ORGANIZATIONS

Spiros Vasilakos[a], Chrysostomos D. Stylios[a],[b], John Garofalakis[c]

[a]Dept. of Telematics Center, Computer Technology Institute & Press "Diophantus"
[b]Dept. of Informatics Engineering, Technological Educational Institute of Epirus, Arta, Greece

[a]vasilakos@westgate.gr, stylios@teiep.gr [b]stylios@westgate.gr, [c]garofala@cti.gr

ABSTRACT

Nowadays organizations continuously update, adopt or introduce new Information Technology Systems within their structure. This requires great resources in capital, staff and time. Selection and adoption of the proper information system and/or technologies must be performed in a way which will ensure that the system will meet the requirements and will fit in the organization's procedures and structure. This paper proposes a methodology for selecting the appropriate solution for an organization among the available options. It describes the requirement analysis phase where the selection criteria are defined and the available solutions are evaluated. For the evaluation of the different solutions, the Bipolar Analytic Hierarchical Process (BAHP) is proposed. The best solution is selected so as to fulfill the functional requirements and the requirements of the organization regarding the maintenance needs and the cost. The proposed approach is tested for the case of the Port Authority of Igoumenitsa, Greece.

Keywords: Information Systems, Modeling, Bipolar Analytical Hierarchical Process

1. INTRODUCTION

The ICT infrastructure of any organization has great importance. It supports the handling of any information and determines the effectiveness and performance of the organization. The adoption or update of a new IT System is an essential issue that may lead to higher efficiency both in customer services and in the organization's internal processes. But this selection is a complex process with great importance and any delay or failure to successfully adopt the new IT system within the organization's structure may lead to loss of critical amounts of resources. On the other hand, any failure to adapt to new technologies almost certainly means loss of competitiveness.

Thus, selection of new IT systems for any organization has to be done in a structured way. All the parameters that affect the choice of a new system must be taken into account. These parameters refer to the funds that the organization will invest, to the capabilities that the system must offer, to the technologies used as well as to how well the system can adapt to internal procedures and the philosophy of the organization and vice versa how some procedures of the organization can adapt to the capabilities of the system.

In this paper, we present a methodology to select the suitable IT technology that it is tested for the case of introducing an IT system at the port of Igoumenitsa, Greece. The first stage of the proposed methodology is the detailed domain analysis and requirements specification. Then, all the possible alternatives and the criteria that affect the decision are determined, regarding both the system requirements as well as criteria related to the organization's specificities. After the phase of alternatives and criteria recognition, there is the stage of eliminating alternatives that not fulfill some of the requirements and criteria required for the final choice. At the final stage, an updated version of the Analytical Hierarchy Process (AHP) (Saaty 1990) called Bipolar Analytical Hierarchy Process (BAHP) (Millet and Schoner 2005) is applied to prioritize the criteria and evaluate the different alternative solutions.

The rest of the paper is as follows: section 2 describes the problem of IT selection and the use of AHP in the specific area; section 3 describes the case study of the port of Igoumenitsa and its specificities. Then sections 4 and 5 describe the implementation of the proposed methodology and the results obtained, and finally section 6 concludes the paper.

2. IT SELECTION PROBLEM AND AHP

2.1. IT System Selection

The constant growth of demand for better services and/or products as well as cost reduction and better customer satisfaction, combined with the growth of complexity at all the operations of any organization makes the adoption of new technologies an absolute necessity. The adoption of a new IT system by an organization requires a high amount of resources both in capital investment as well as time for selection, installation and user training. The phase of adoption demands careful planning so that the influences to the business processes are as smooth as possible.

IT system selection is a Multi Criteria Decision Problem with many factors competing each other. First of all, one main factor is the capital investment. The cost of acquisition of any IT system is not the only factor that must be taken into account. Costs like maintenance, support and license must be considered as well. The capabilities of the product must be also examined in detail in order to verify if they fulfill both functional and non-functional requirements of the organization. Non-functional requirements demand special attention as in many cases they are qualitative and cannot be easily, or are impossible to, be quantified. Finally, the problem of IT system selection becomes even more complicated when the option to develop the system exists along with the option to select an ‘of the shelf’ commercially available product.

Many methodologies have been proposed for the selection of the appropriate IT system. Jadhav and Sonar (2009) reviewed a large number of methodologies for software selection, evaluation techniques and criteria; among others AHP, Feature Analysis, Weighted Average Sum (WAS) and Fuzzy-based approaches have been proposed. A general selection approach consists of six steps, beginning from the domain analysis and proceeding to gradually decomposing the criteria until quantifiable measures are used (Franch and Carvallo 2002). A set of actions that include the determination of alternatives and steps for their qualification was proposed by Jadhav and Sonar (2011). Stamelos and Tsoukias (2003) proposed the categorization of the software selection problem in seven categories.

2.2. The Analytic Hierarchy Process (AHP) and the Bipolar AHP (BAHP)

AHP introduced in the ‘70s (Saaty 1980, Saaty 1990) and since then it has found a wide adoption and use (Saaty 2008, Saaty and Vargas 2012). Especially at the domain of IT system selection AHP has been widely applied (Cebeci 2009; Lai, Wong, and Cheung 2002; Wei, Chien, and Wang 2005). AHP is a Multi Criteria Decision Methodology (MCDM) that uses a hierarchy to formulate the problem. At the top of the hierarchy the goal of the decision is placed. The second level includes the criteria that are used for comparison. Each criterion may have sub-criteria that are placed at the consequent levels. At the final level, all the alternative choices are placed.

A 1-9 scale is used to determine the relative importance between the criteria. Their meaning is shown in Table 1.

Table 1: The Range from 1 to 9 used in AHP to Determine Relative Importance among Criteria

Relative Importance	Value
equal	1
moderate	3
strong	5
very strong	7
extreme	9
intermediate values	2,4,6,8

For *n* criteria a comparison matrix $\mathbf{A}_{(n\,\mathrm{x}n)}$ is formed. In the a_{ij} position of the matrix the relative importance of the i_{th} criterion compared to the j_{th} criterion is placed and consequently in the a_{ji} position the $1/a_{ij}$ value is placed. So, a reciprocal square matrix is formed where value 1 is placed in the diagonal a_{ii}

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ 1/a_{12} & 1 & a_{23} & \dots & \\ 1/a_{13} & 1/a_{23} & 1 & \dots & \\ \dots & & & \dots & a_{(n-1)n} \\ 1/a_{1n} & & & 1/a_{(n-1)n} & 1 \end{pmatrix}$$

The relative weights of the criteria are the normalized eigenvector v_n of comparison matrix $\mathbf{A}$. The same procedure is followed with the alternatives. So, for *m* alternatives that are compared to *n* criteria a $\mathbf{W}_{(mxn)}$ matrix is formed where $w_{n,m}$ is the ranking of alternative *m* in relation to criterion *n*. The final ranking of each alternative $r_i^{(i=1:m)}$ is calculated with equation (1):

$$r_i = \sum_{j=1}^{n} w_{i,j} v_j \qquad (1)$$

When numerical values are available, then they are used instead of the 1-9 scale. AHP have the ability to combine both qualitative and quantitative criteria. Due to the nature of IT system selection, both kinds of criteria have to be used, so AHP is used in many cases (Jadhav and Sonar 2009).

In many multicriteria problems, along with factors that contribute positively to the decision; there may exist factors that have negative impact. Common factors of this category can be cost, time, and required effort. The strictly positive additive values of equation (1) in the final ranking of AHP make handling of such negative factors problematic. The standard procedure is to use inversion of these values. But inversion of a positive number also leads to a positive number, in this way factors with great negative contribution are treated as factors with very small, but still positive, contribution, which may lead to incorrect ranking. For these reasons, it is preferable the use of BAHP (Millet and Schoner 2005). BAHP is an extension of AHP that allows the incorporation of negative factors into the AHP calculations. Factors that have negative contribution to the decision are treated as negatives in the final ranking calculations and are not inverted.

3. METHODOLOGY DESCRIPTION

For any IT system selection there are two main options either to choose among available commercial systems or to develop an IT system from scratch when there are not available commercial systems that fulfill all requirements. Thus, here we include both of these options.

The proposed methodology is analytical and precise, but also flexible enough so that it is able to adapt to a variety of scenarios (Figure 1).

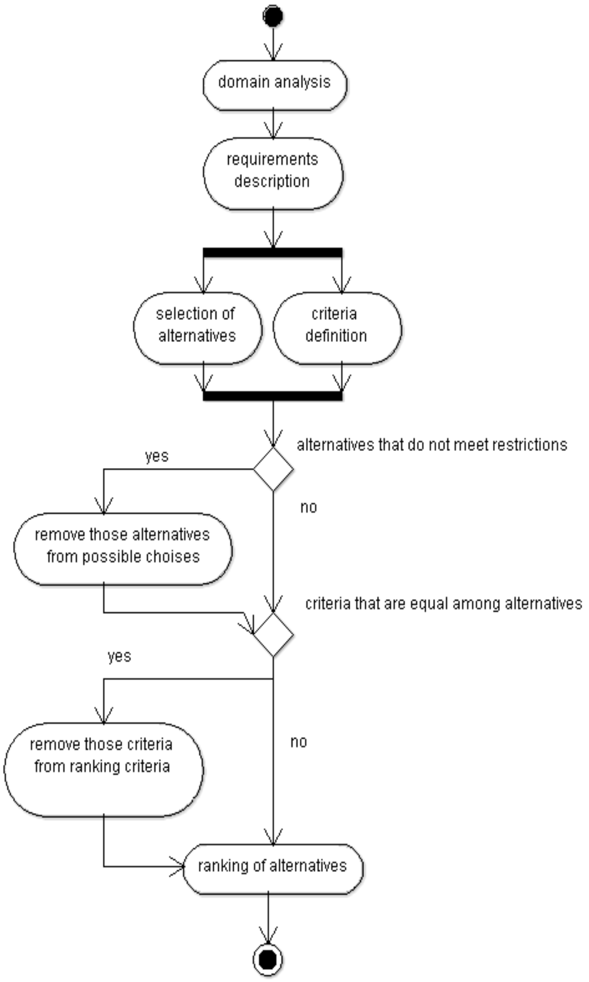

domain analysis
requirements description
selection of alternatives
criteria definition
alternatives that do not meet restrictions
yes
no
remove those alternatives from possible choises
criteria that are equal among alternatives
yes
no
remove those criteria from ranking criteria
ranking of alternatives

Figure 1: UML Activity Diagram describing the Sequence for IT system Selection Methodology

The first step for every IT system selection is to perform a detailed domain analysis. Sometimes, this is an overlooked step, but it is of high importance. The domain of interest must be analyzed so that to describe in detail all the procedures that the system has to support. The analysis must be done in many perspectives, covering all the spectrum of user categories. Overlooking this step and considering the procedures description and analysis as trivial can lead to unpredicted situations and to the selection of an IT system that does not fulfill all the requirements and procedures within the organization.

Next step is the description of the requirements, both functional and non-functional. The term “functional requirements” describes specific tasks that the IT system has to perform. There are includes the intended behavior of the system that is expressed as tasks and/or behaviors. Non-functional requirements are used to describe criteria and goals of the system rather than certain behavior and can be qualitative attributes.

The next two steps can be performed in parallel. The first one is the selection of alternatives and the second one is the determination of comparison criteria. There are two main alternatives either to select existing commercial IT system or to develop of a new IT system. In the second case, there are some more subcategorized alternatives regarding the technologies to be used. .

In addition to this, the different criteria for the comparison of the alternatives have to include all aspects of the system’s implementation, usage and functionality. These criteria can be divided into four categories: i) managerial, ii) user-related, iii) technology-related and iv) vendor-related.

Managerial criteria are mostly related to cost and to required implementation time of the project. User-related criteria refer to the capabilities of the system (whether it satisfies the functional and nonfunctional requirements) as well as the ease of use and learning cycle of the IT system. Technology-related criteria refer to technologies used by the system, both software- and hardware-related. Finally, vendor-related criteria refer to the vendor’s reputation, expertise and stability.

After the selection of alternatives and criteria, the phase of elimination follows. Firstly, the existing alternatives are compared to an initial set of criteria. These criteria do not regard the functionality of the system, since all the alternatives must fulfill the system requirements. The initial set of criteria is usually related to management, and we name them “hard criteria”, they are mostly constraints that are used to eliminate alternatives from the set of choices and not to compare them. Such criteria can be cost or time of deliverance and any other criteria that impose certain restrictions to a particular project. For example, all alternatives up to a certain cost are among the candidates for selection and the cost will be calculated as a criterion in the final decision, but alternatives exceeding a certain cost are unacceptable and are eliminated from the selection procedure.

The criteria elimination procedure is the following: criteria that have the same value among different alternatives are eliminated. The elimination of alternatives and criteria is important as it reduces the amount of required calculations and furthermore it eliminates the potential to choose an unacceptable alternative.

The final step is the ranking of alternatives using a Multi Criteria Decision Methodology (MCDM). As aforementioned, we use the Bipolar Analytic Hierarchical Process (BAHP). BAHP has the main advantage to use negative values in the ranking calculation for criteria that have negative impact on the final decision. The BAHP that has described above forms a hierarchy, with the four criteria and their subcriteria at lower levels of the hierarchy and it is shown at Figure 2.

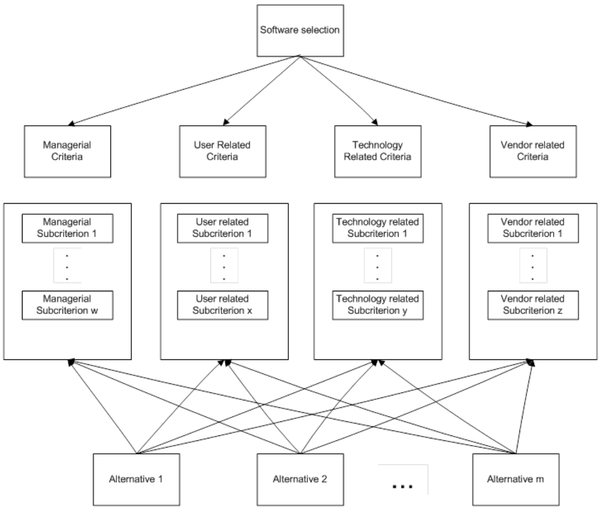

Software selection
Managerial Criteria
User Related Criteria
Technology Related Criteria
Vendor related Criteria
Managerial Subcriterion 1
Managerial Subcriterion w
User related Subcriterion 1
User related Subcriterion x
Technology related Subcriterion 1
Technology related Subcriterion y
Vendor related Subcriterion 1
Vendor related Subcriterion z
Alternative 1
Alternative 2
...
Alternative m

Figure 2: The Hierarchy formed for IT system Selection

4. CASE STUDY

4.1. Problem Description

The port of Igoumenitsa is a very important transport hub of Northwestern Greece. It connects Greece to Italy and mainly provides ship docking, passenger and vehicle traffic services. It focuses on passenger traffic through ferry connections to domestic and foreign destinations, while goods are transported mainly by trucks. It serves a high amount of vehicle and passengers in relation to its size. In 2012, it served a total of 1.436.239 passengers, 338.355 private vehicles and 79.814 trucks for domestic destination and 79.814 passengers, 120.409 private vehicles and 68.702 trucks for abroad destinations (OLIG S.A. 2013).

Our team collaborates with Igoumenitsa Port Authority (OLIG) within the Generalized Automatic Exchange of Port Information Area project (GAIA). We provide OLIG with consulting services by describing the functionality and organization for an Integrated Information System which will supply the Port Authority with the desired functionality for passenger, vehicle and authorized personnel trafficking in the port areas and also the itineraries to and from the port. Every passenger, in order to enter the port facilities, is supplied with a boarding card. These cards will be edited by the shipping agents to the customers and they will contain all the necessary information allowing the passengers to pass any security checks. Similar procedure will be followed for vehicles. The authorized personnel and their vehicles will make use of security cards in order to access the port area. In addition, the different IT systems of the shipping agents, responsible for the tickets/boarding passes editing, will communicate with the overall port's system in order to provide the required data.

The Integrated Information System will be able to process and store data regarding all procedures of the port operations related to the itineraries and passenger and vehicle traffic. Also a Database Management System (DBMS) will be included in the system that will provide the required scalability for the storage and processing of large amount of data and the tools to monitor and tune the database.

4.2. Approach and Criteria Selected

4.2.1. Methodology Used

For the selection of the appropriate technology, a detailed domain analysis was initially performed. This phase involved interviews with the interested parties/Stakeholders that are somehow influencing the system, such as staff of Port Authority, of Shipping Companies, of Customs Office and of the Coast Guard.

After this domain analysis, the user main categories were identified. Regarding, Port Authority, two categories exist: one from a managerial perspective and one from users perspective. The first category is interested more in factors regarding the cost of the system and its value as a long-term investment. The second category was mainly concerned about the user-friendliness of the system and its capabilities regarding its everyday usage and administration. The Shipping companies, in order to provide the necessary data, will connect to the system through their operational IT systems. Their main concern was about the compatibility of the system and the ease of the interconnection implementation. Passengers are the final beneficiaries as they will use the system to enter the port area and during their boarding procedure. Their main concerns were about the speed and the user friendliness of the system.

At the next stage, the functional and non-functional requirements were described. In addition to this, the architectural description of the system was described, including the system's functionality, the actors of the system, the components and their interactions. After this phase, the main criteria regarding all the aspects of the system's implementation, usage and functionality and the available alternatives were determined.

Then, there is the first phase of the selection, where we followed the elimination procedure, i.e. any choice that did not cover certain criteria was eliminated. Also criteria whose values were the same among the alternatives were eliminated. In the second phase of the selection procedure the alternatives were ranked according to the remaining criteria based to the BAHP.

4.2.2. Evaluation Criteria

The evaluation criteria are divided into four main categories: i) Managerial, ii) User-related, iii) technology-related, iv) vendor-related.

The managerial criteria include mostly cost-related criteria and the implementation time of the project. The selected criteria are: cost of the project which includes development cost, maintenance cost, support cost, hardware cost, usage of already existing equipment and delivery time of the system.

User-related criteria mainly regard the capabilities of the system (whether it would satisfy the functional and nonfunctional requirements) as well as ease of use and learning curve of the system. The actors related to these criteria are the back end users that monitor the traffic in the port, retrieve data and reports and perform administrative tasks. The criteria for this category are: fulfillment of requirements, user-friendliness for operators, user-friendliness for passengers, learning curve for operators, and accessibility from different devices.

Technology-related criteria refer to the technologies that will be used to implement the system or that the commercially available system uses, both software- and hardware-related. The main actors concerned with these criteria are the company that will develop the system, the shipping companies whose IT systems will interact with the system, and both the management and users of the Port Authorities regarding software/hardware support and warranties. The criteria for this category are: reliability, speed, database capabilities, security, ease of upgrade and maintenance, hardware compatibility, ease of integration with other systems and support and warranty.

Finally, vendor-related criteria concern the choice of the candidate vendors and include vendor expertise, experience and stability. The criteria for every category are shown in table 1.

Table 1: Selection Criteria for each category

Category	Criteria	Sub-criteria
Managerial	Cost	development cost; maintenance cost; support cost; hardware cost; software (DBMS) cost; use of existing equipment
	Delivery time	
User-related	Fulfillment of requirements	
	User-friendliness	user-friendliness for operators; user-friendliness for passengers; learning curve for operators
	Accessibility from different devises	
Technology-related	Capabilities	Reliability; Speed;, database capabilities; security; ease of upgrade and maintenance; hardware compatibility; ease of integration with other systems
	Support and Warranty	
Vendor-related	Expertise	
	Experience	
	Stability	

4.2.3. Alternatives

Four main alternatives were considered: i) acquisition of an existing commercial system ii) updating the existing IT system that does not meet the functional and non-functional requirements iii) design and develop the system from scratch using proprietary technologies and software and iv) develop an new system using free/open source technologies and software.

4.2.4. Elimination of Alternatives and Criteria

In this phase, the different alternatives were examined. The acquisition of an existing system was excluded, as none of the available systems fully covered the requirements and the acquisition cost was significantly higher than the cost of the other three alternatives. Also difficulties would arise in using existing hardware equipment consisted of barcode/RFID readers, license plate recognition cameras and the acquisition of new would further increase the cost. This choice would be delivered in a quite short period of time, but this advantage by itself is not so important to consider this alternative as a choice.

Regarding the criteria, the fulfillment of requirements was eliminated as it is a prerequisite, since the requirements of the system were described and the final developed system will cover all the requirements. Also user-friendliness for the passengers was removed from the list since in all cases, the passengers will interact with the system in the same way regardless the used technologies. Both the programming language and the DBMS in the three cases are compatible with the majority of hardware, so the criterion of compatibility was removed. Finally, in our case-study, the vendor-related criteria were also removed, as only one vendor was final candidate. Table 2 shows the criteria after the elimination phase.

Table 2: Criteria used after the Elimination of the Criteria that would not affect the selection

Category	Criteria	Sub-criteria
Managerial	Cost	development cost; maintenance cost; support cost; hardware cost; software (DBMS) cost; use of existing equipment
	Delivery time	
User-related	User-friendliness	user-friendliness for operators; learning curve for operators
Technology-related	Capabilities	Reliability; Speed;, database capabilities; security; ease of upgrade and maintenance; ease of integration with other systems
	Support and Warranty	

Figure 3 illustrates a diagram with the remaining, after the elimination, criteria and their hierarchy.

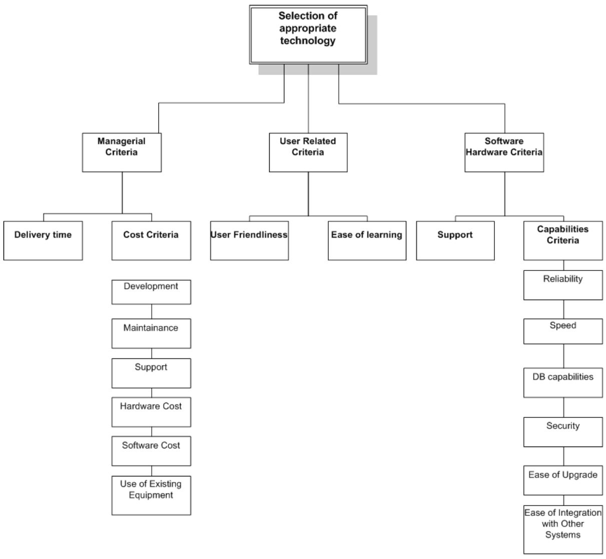


Figure 3: The hierarchy of the criteria after the elimination phase

5. CALCULATIONS AND RESULTS

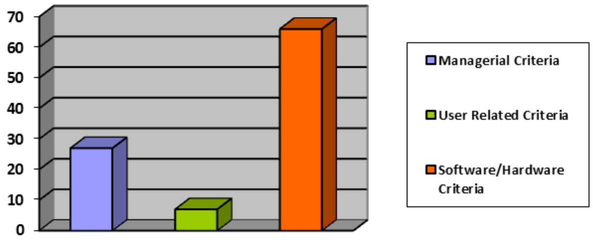
5.1. Calculations

The values and priorities regarding the criteria were acquired by the Port Authority staff that acted as the domain experts. They explained the importance of all criteria and, after a brief explanation of the BAHP methodology, they suggested their estimations about the relative weights for each factor. In this case, mostly qualitative criteria were used. In table 3 the comparison matrix between the main criteria is shown and figure 4 illustrates a chart with the calculated importance of each criterion.

Table 3: Values among Main Criteria

	Managerial	User-related	Technology-related
Managerial	1	5	0,333
User-related	0,2	1	0,125
Technology-related	3,000	8	1

Figure 4: The percentage weight of each of the three main factors



Due to space restrictions, we cannot present all the tables that have been formed and the calculations performed for each of the sub-criteria. The final weights of all sub-criteria are presented in Table 4, as well as whether the contribution to the decision has positive or negative affect.

Table 4: The Weights of the Criteria and Sub-criteria in Percentage Values and Positive or Negative Contribution of the Criteria

Managerial Criteria	Contribution	%
delivery time	negative	9
development	negative	6,65
maintenance	negative	1,80
support	negative	1,44
hardware cost	negative	3,78
software cost	negative	2,16
use of existing equipment	positive	2,16
Total Cost percentage		17,98
Total percentage of managerial criteria		27

User-Related Criteria	Contribution	%
User-friendliness for operators	positive	4,66
learning curve for operators	positive	2,33
Total percentage of user-related criteria		7

Technology-related Criteria	Contribution	%
support	positive	21,98
reliability	Positive	16,70
speed	Positive	2,07
DB capabilities	Positive	7,91
security	Positive	7,91
ease of upgrade and maintenance	Positive	3,52
ease of integration with other systems	Positive	5,85
Total capabilities criteria		43,96
Total percentage of technology related criteria		66

Moreover, there were asked experts from vendors to estimate the time of development and experts from Computer Technology Institute & Press estimated the software and hardware capabilities of the specific technologies. In addition, the costs and actual prices of DBMS licenses and support, as well as the maintenance cost were estimated.

After the estimation of the corresponding criteria, the calculation of the alternatives ranking was performed. Table 5 presents the scores of the three alternatives regarding the available support for each solution.

Table 5: The scores of the alternatives regarding the support criterion

Update of existing system	**27%**
Development with proprietary technologies	**60%**
Development with open-source technologies	**13%**

The final calculations were performed using the BAHP and Table 6 presents the score for each alternative.

Table 6: Scores of each of the alternatives

Alternative	Score
Update of existing system	**0,05141**
Development with proprietary technologies	**0,32606**
Development with open source technologies	**0,15386**

5.2. Result Analysis

The alternative of improving the existing system received the lowest ranking. The main advantage of this solution was the utilization of the existing equipment. But it would require updating the existing programming code produced by a non-up-to-date programming language. Moreover, the existing system did not met the requirements so it would require more development time that introduce further risk for the final system not to meet the functional requirements.

Poor project planning is recognized as one of the main reasons for software projects failure (Han and Huang 2007, Whittaker 1999). Incorrect or incomplete system requirements are another one of the top five reasons for software projects failure (Baccarini, Salm, and Love 2004, Han and Huang 2007). Since the existing system did not fulfill the requirements, attempt to use code that has not been properly developed and was poorly designed in the first time is likely to require more development time or even to project failure. Also, since the used programing language is not-up-to-date, it lacked from the two other solutions in quality characteristics.

The development with proprietary technologies scored highest and was recommended as the appropriate solution. Development with open source solution had lower cost, mainly related to the lack of license costs. Also regarding some quality characteristics, it is equal and better than the proprietary solution. However, the development with proprietary technologies has the advantage of support and guaranties that come along with the use of proprietary solutions. The lack of standard support in the case of using open source technologies could potentially lead to greater cost when an update of the system would be required or in case of system malfunction. This is very critical for the case under study, because OLIG has not qualified and experienced IT team to support and maintain the system.

Our results come in accordance with other researches results; actually, the success of open source technologies adoption from an organization greatly depends on the organization's IT capabilities and its experience in using open source software (Lin 2008; Goode 2005; Spinellis and Giannikas 2012). Also, the lack of reliable technical support is recognised as one of the reasons that Open Source Software is rejected (Goode 2005).

6. CONCLUSION

This paper describes a procedure for the selection of appropriate technologies for the development of an IT system regarding the passenger, vehicle and authorized personnel in a Greek port. After the determination of the functionality requirements, the possible solutions and the corresponding criteria were identified. A two-stage procedure was followed. At first stage, solutions that did not meet certain requirements and criteria that would not affect the selection (having the same characteristics among the remaining solutions) were eliminated.

Commercial available system was removed from the candidate solutions, because its price exceeded the available budget of the project and in addition to this not all the requirements were met. For the selection of the appropriate solution, the Bipolar Analytic Hierarchical Process (BAHP) was implemneted so that criteria with negative impact would be incorporated. From the examined solutions, the further development of an earlier non-functional system received the lowest score. The solutions with the higher scores were the development of the system from the beginning with proprietary or with open source technologies. Although the capabilities and characteristics of the technologies had a high impact on the final results, characteristics related to the software guaranties and support played a fundamental role in the final selection of the proprietary solution. Our results are similar to other studies and point out that lack of experienced IT teams is one of the factors for which organizations prefer proprietary solutions, even if open source solutions with similar capabilities are available.

ACKNOWLEDGMENTS

This work was supported by the project "Generalized Automatic Exchange of Port Information Area - GAIA", Code I1.22.01, implemented within the framework of the European Territorial Cooperation Programme Greece-Italy 2007-2013, co-funded by the European Union (ERDF) and by National Funds of Greece and Italy.

REFERENCES

Baccarini D., Salm G., Love P.E.D., 2004. Management of risks in information technology projects. *Industrial Management & Data Systems* 104 (4):286-295.

Cebeci U., 2009. Fuzzy AHP-based decision support system for selecting ERP systems in textile industry by using balanced scorecard. *Expert Systems with Applications* 36 (5):8900-8909.

Franch, X., Carvallo, J.P., 2002. A quality-model-based approach for describing and evaluating software packages. *Proceedings of IEEE Joint International Conference on Requirements Engineering*, 104-111. 9–13 September 2002, Essen, Germany

Goode S., 2005. Something for nothing: management rejection of open source software in Australia's top firms. *Information & Management* 42 (5):669-681.

Han W.M., Huang S.J., 2007. An empirical analysis of risk components and performance on software projects. *Journal of Systems and Software* 80 (1):42-50.

Jadhav A.S., Sonar R.M, 2011. Framework for evaluation and selection of the software packages: A hybrid knowledge based system approach. *Journal of Systems and Software* 84 (8):1394-1407.

Jadhav A.S., Sonar R.M., 2009. Evaluating and selecting software packages: A review. *Information and Software Technology* 51 (3):555-563.

Lai V.S., Wong B.K., Cheung W., 2002. Group decision making in a multiple criteria environment: A case using the AHP in software selection. *European Journal of Operational Research* 137(1):134-144.

Lin L., 2008. Impact of user skills and network effects on the competition between open source and proprietary software. *Electronic Commerce Research and Applications* 7 (1):68-81.

Millet I., Schoner B., 2005. Incorporating negative values into the Analytic Hierarchy Process. *Computers & Operations Research* 32 (12):3163-3173.

OLIG S.A., 2013. *Passenger Port Statistics*. Igoumenitsa Port Authority S.A. Available from: http://www.olig.gr/?q=en/node/443 [Accessed 3 May 2013]

Saaty T.L., Vargas L.G., 2012. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. 2nd ed. New York: Springer.

Saaty, T.L., 1980. *The Analytic Hierarchy Process*. New York: McGraw Hill

Saaty, T.L., 1990. How to make a decision: The analytic hierarchy process. *European Journal of Operational Research* 48 (1):9-26.

Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences* 1 (1):83-98

Spinellis D., Giannikas V., 2012. Organizational adoption of open source software. *Journal of Systems and Software* 85 (3):666-682.

Stamelos I., Tsoukias A., 2003. Software evaluation problem situations. *European Journal of Operational Research* 145 (2):273-286.

Wei C.C., Chien C.F., Wang M.J.J., 2005. An AHP-based approach to ERP system selection. *International Journal of Production Economics* 96 (1):47-62.

Whittaker B., 1999. What went wrong? Unsuccessful information technology projects. *Information Management & Computer Security* 7 (1):23-30.

AUTHORS BIOGRAPHY

Spiros Vasilakos is research collaborator at Computer Technology Institute & Press “Diophantus”. He received his diploma from the Department of Computer Engineering & Informatics, University of Patras. His scientific interests include: System Modeling and Analysis, Decision Support Systems and Multicriteria Decision Methodologies.

Chrysostomos D. Stylios is an Associate Professor at Dept. of Informatics Engineering, TEI of Epirus; he is a senior researcher at Telematics Center Department of Computer Technology Institute & Press. He received his Ph.D from the Dept. of Electrical & Computer Engineering University of Patras (1999) and diploma in Electrical & Computer Engineering from the Aristotle University of Thessaloniki (1992). He has published over 100 journal and conference papers and book chapters. His main scientific interests include: Fuzzy Cognitive Maps, Soft Computing, Computational Intelligence Techniques, Neural Networks, Knowledge Hierarchical Systems and Decision Support Systems

John Garofalakis, born in 1959, obtained his Ph.D. from the Department of Computer Engineering and Informatics (CEID), University of Patras, Greece, in 1990, and his Diploma on Electrical Engineering from the National Technical University of Athens, Greece, in 1983. He is currently Professor in CEID and manager of the Telematics Center Department at the Research and Academic Computer Technology Institute of Greece. His research interests include performance evaluation, distributed systems and algorithms, Internet technologies and applications. He has published over 100 papers in various journals and refereed conferences and is author of several books and lecture notes in the Greek language.

A MODEL FOR IRREGULAR PHENOMENA IN URBAN TRAFFIC

Luigi Rarità

Department of Information Engineering, Electric Engineering and Applied Mathematics,
University of Salerno, Via Giovanni Paolo II, 132, 84084, Fisciano (SA), Italy

lrarita@unisa.it

ABSTRACT

In this paper, we consider an analysis of car dynamics and its optimization on urban networks of City type, namely rectangular networks with roads of unequal length. In particular, we study the traffic variations due to changes of permeability parameters, that describe the amount of flow allowed to enter a junction from incoming roads. On each road, we distinguish a free and a congested regime, characterized by an arrival and a departure flow, respectively. Dynamics at nodes of the network is solved maximizing the through flux. The evolution on the whole network gives rise to very complicated equations, as car traffic at a single node may involve time – delayed terms from all other nodes. Hence, the network solution is found by an alternative hybrid approach, via the introduction of additional logic variables. Finally, simulations on a portion of the Salerno network, in Italy, allows to test the obtained results.

Keywords: traffic dynamics, control theory, simulation, optimization.

1. INTRODUCTION

Urban areas are often characterized by strange phenomena for car traffic: high car densities, leading to various congestion types; reductions of velocities for transport vehicles; pollutions problems, mainly due to fuel consumption. From a more specific point of view, traffic flows, especially in cities, are basic examples of material flows, mostly organized in networks.

Traffic flows have been modeling for years via several approaches (Bretti et al. 2006; Daganzo, 1995b; Helbing et al. 2005; Herty and Klar, 2003; Herty and Klar, 2004; Herty et al. 2006; Hilliges et al. 1995; Lebacque and Khoshyaran, 2005); some of them are based on conservation laws (Coclite et al. 2005; Garavello et al. 2006). The reason for such a choice is simple: the solutions of these equations have nonlinear characteristics, very useful to describe almost all the dynamic effects of car traffic, especially for vehicle queues. Although it is proved that conservation laws are a possible valid alternative for urban traffic models, road network modeling always represents a hard task, considering that the adoption of conservation laws does not always guarantee that: phenomena of daily lives are well described, such as traffic jams in some road sections (see Daganzo, 1995a; Helbing, 2001; Kerner, 2004; Schönhof et al. 2006); it is not always possible to define a total solution for the overall urban networks and, as a consequence, a global optimization procedure for traffic flows. This last problem is highly non trivial. In fact, although it often happens that traffic congestions have to be reduced in some little portions of networks (see Cascone et al. 2007; Cascone et al. 2008; D'Apice et al. 2011), the necessity of a total redesign of network roads and junctions is often required, with necessity of finding solutions of global type and extendible to various network topology. For this reason, we need a model that, beside all advantages due to conservation laws formulation, is able to focus on the overall network dynamics.

In order to achieve this aim, in this work we use a two – phase model for flows on roads (see Helbing et al. 2007; Rarità et al. 2010). In particular, the road is decomposed into road sections (links) of homogeneous capacity and nodes for their connections. Traffic dynamics along the road sections are assumed to follow the Lighthill – Whitham – Richards (briefly, LWR) model (see Lighthill et al. 1955; Richards, 1956; Whitham, 1974), but with a simplified representation, reducing the Partial Differential Equation (PDE) approach to a delayed Ordinary Differential Equation (ODE) one. For each road, two regimes are considered: free and congested. The lengths of the corresponding areas determine the exact dynamics of cars. This two – phase model, beside the obvious mathematical simplification, allows either the representation of all phenomena described via conservation laws, or the analysis of some real effects in urban traffic, such as transitions from free to congestion traffic flows due to lack of capacity, the propagation speeds of vehicles in congested traffic, spill – over effects and traffic jams, the last ones expressed by a suitable equation.

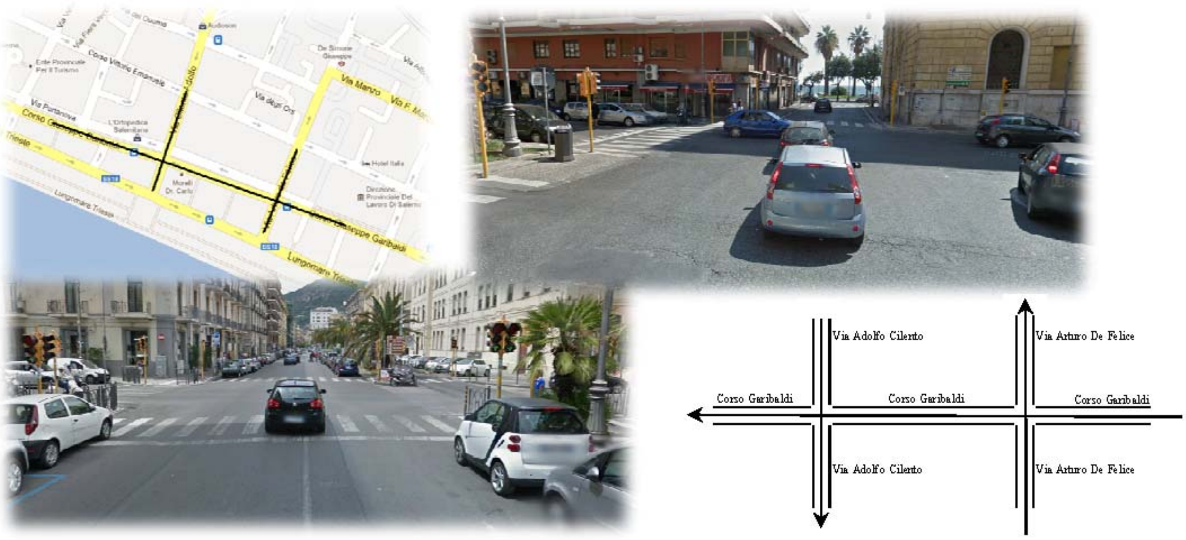


Figure 1: a portion of the real network of Salerno (up, left), consisting of two road junctions (up, right; bottom, left) and its graph (bottom, right)

Flows at nodes are regulated by permeability parameters γ, that indicate the amount of cars allowed to enter the junction from incoming roads. In Helbing et al. 2007, such parameters are assumed either zero or one, modeling the possibility of traffic lights at intersections. Here, following the approach in Rarità et al. 2010, permeabilities can also vary between zero and one. This adds the following interesting further interpretation: $0 < \gamma < 1$ indicates the situation in which the traffic is free to circulate, but only a part of it is allowed to enter the junction. This is quite normal in the usual traffic conditions, due to queues on roads that imply velocity reductions and delays in crossing the road junctions. Notice that $0 < \gamma < 1$ represents not only the possibility of modeling traffic lights, but also the normal traffic at not light – controlled road junctions.

Dynamics at nodes is, according to Coclite et al. 2005, described by the following two rules:

(A) The incoming traffic distributes to outgoing roads according to fixed (statistical) distribution coefficients;

(B) Drivers behave in order to maximize the through flux.

Here, we consider permeability parameters as controls in order to optimize the dynamics of complex networks, of “City” type, namely rectangular networks with roads of unequal length. Precisely, the variations of permeability parameters allows to establish some optimization criteria. In particular, we focus on the minimization of a cost functional, which represents the sum of queue lengths, i. e. number of delayed vehicles or lengths of congested areas. It is described that queues on roads can influence the dynamics on the whole network, leading to a “nested” equation, which cannot be solved in an analytical way.

Hence, the total solution of the network, also in terms of optimization procedures, is found using additional logic variables that represent the emptying of queues or filling the road segments. Such variables are influenced by delayed and non delayed continuous variables (queue lengths, arrival and departure flows). Indeed, they themselves influence the continuous quantities, leading to a particular system of hybrid type.

Considering the Pontryagin Maximum Principle (Bressan and Piccoli, 2007), we consider needle variations of permeability parameters, and the hybrid modeling allows either a rich description of all phenomena connected to the car traffic or the definition of a procedure to state, for the overall network, the optimal solution of minimizing car queues on roads.

The obtained results for the hybrid dynamics are tested by simulation using a modified Runge – Kutta numerical algorithm that considers delayed terms for incoming and outgoing flows into road sections. Numerical results are analyzed for a real case: a portion of Salerno urban network, which is one of the most suitable examples of rectangular network in the south of Italy. The topology of the network, represented in Figure 1, consists of three principal roads: Corso Garibaldi, Via Adolfo Cilento, Via Arturo De Felice. Road junctions are in this case of 2×2 type, namely they are characterized by two incoming roads and two outgoing roads. For this last case, it is proved (details are found in Rarità et al. 2010) that a simple needle variation of a permeability parameter provokes a wealth of variations in the other quantities at nearby nodes. This situations indicates that the hybrid approach permits, on one side, the description of the network evolution with nodes dynamics having separate equations; and, on the other hand, it keeps all characteristics of the original system.

The paper is organized as follows. Section 2 shows the model for roads, while Section 3 concerns city networks and descriptions of dynamics at nodes. In Section 4, the optimal control problem and the nested equations are analyzed; logic variables are introduced in order to define a hybrid dynamics; the variational equations, useful to define how permeability parameters influence the overall network flows, are described. Simulations for the case study are presented in Section 5. Conclusions and future perspectives are reported in Section 6.

2. FLOWS MODELLING ON ROADS

This section focuses on the car traffic behaviour on each single road of a traffic network, while the dynamics at nodes is analyzed for the special case of City type networks in next Section. The following assumptions are made: (A1) The road network consists of road sections of homogeneous capacity (links) and nodes describing their connections; (A2) A first order approach (such as LWR) gives a good description of car traffic on roads; (A3) On each road section, the fundamental diagram (density-flux graph) is well approximated by a triangular shape, with an increasing slope V_i^0 (i.e. maximum speed of vehicles, corresponding to the freed speed or speed limit on road section *i*) at low densities and a decreasing slope $c = \left(\rho_{\max} T\right)^{-1}$ in the congested regime, where: $\rho_{\max}$ denotes the maximum vehicle density in car queues; *T* is the safe time headway, which is constant along the road section; (A4) Who enters a road section first exits first (FIFO principle); (A5) Each road section has a first subsection in free phase and a second subsection in congested phase.

A road is characterized by (see Helbing et al. 2007; Rarità et al. 2010): the *arrival flow* $A_j(t)$, which indicates the inflow of vehicles into the upstream end of road section *j*; the *departure flow* $O_j(t)$, which is the flow of vehicles leaving road section *j* at its downstream end; the maximum in – or outflow of road sections *j*, $\hat{Q}_j = \left[T + \left(V_j^0 \rho_{max}\right)^{-1}\right]^{-1} = cV_j^0 \rho_{max}\left(c + V_j^0\right)^{-1}$. All the above quantities refer to flows *per lane*, indicating by I_j the number of lanes and by L_j the length of road section *j*. Moreover, the length $l_j(t) \le L_j$ represents the length of the congested area on link *j* (measured from the downstream end), and ΔN_j is the number of stopped or delayed vehicles. An ideal representation of road section *j* is in Figure 2.

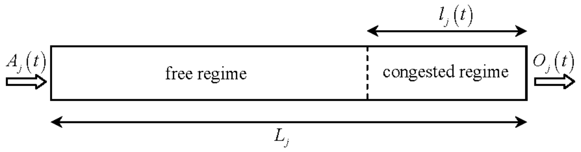


Figure 2: road section *j*

Functions $A_j(t)$ and $O_j(t)$ are also assumed upper limited by $\hat{A}_j(t)$ and $\hat{O}_j(t)$, respectively. In order to define these bounds, we refer to the following:

Definition. *For road section j, the function* $\gamma_j(t) \in [0,1]$, $t \ge 0$, *is said "permeability parameter". It defines the amount of cars that goes out from road section j.*

Remark. *Following the approach used in Rarità et al. 2010, the permeability parameter for road section j has the following interpretations:* $\gamma_j = 0$ *implies a red or amber light;* $\gamma_j = 1$ *corresponds to a green light, and all cars can flow out from road section j;* $0 < \gamma_j < 1$ *represents the green light for a situation in which not all cars can go out immediately from road section j, thus indicating that unvanished queues are still present nearby the road junction, with consequent non perfect migration of cars. Notice that* $0 < \gamma_j \le 1$ *is also useful to indicate situations in which no traffic lights are present at road intersections, but cars are free to circulate according to some yielding rules.*

We assume that $A_j(t)$ is bounded by the maximum inflow $\hat{Q}_j$, if road section *j* is not fully congested, namely $l_j(t) < L_j$; otherwise, if road section *j* is full ($l_j(t) = L_j$), $A_j(t)$ is limited by $O_j\left(t - L_j/c\right)$ a time period L_j/c before. Hence, we have $0 \le A_j(t) \le \hat{A}_j(t)$, with:

$$\hat{A}_j(t) := \begin{cases} \hat{Q}_j, & if\ \ l_j(t) < L_j, \\ O_j\left(t - L_j/c\right), & if\ \ l_j(t) = L_j. \end{cases} \qquad (1)$$

Moreover, the potential departure flow $\hat{O}_j(t)$ of road section *j* is given by its permeability $\gamma_j(t)$ times the maximum outflow $\hat{Q}_j$ from this road section, if there is a queue of vehicles, namely $\Delta N_j(t) > 0$; otherwise, if road section *j* is empty ($\Delta N_j(t) = 0$), the outflow is limited by the permeability times the arrival flow $A_j\left(t - L_i/V_i^0\right)$ a time period L_i/V_i^0 before. We get that $0 \le O_j(t) \le \hat{O}_j(t)$, with:

$$\hat{O}_j(t) := \gamma_j(t) \begin{cases} A_j\left(t - L_i/V_i^0\right), & if\ \ \Delta N_j(t) = 0, \\ \hat{Q}_j, & if\ \ \Delta N_j(t) > 0. \end{cases} \qquad (2)$$

2.1. An equation for traffic jams

As for traffic jams, we consider one of the suggested approaches in Helbing et al. 2007. Setting $AO^t_{j,L_j/V_j^0} := A_j\left(t - L_j/V_j^0\right) - O_j(t)$, the number of delayed vehicles for road section *j*, ΔN_j, is given by the following equation:

$$\Delta\dot{N}_j(t) = \begin{cases} AO^t_{j,L_j/V_j^0}, \ if \ t < \bar{t} \ or \\ t \geq \bar{t} \ and \ AO^t_{j,L_j/V_j^0} < 0, \\ L_j \rho_{\max}, \ if \ t \geq \bar{t} \ and \ AO^t_{j,L_j/V_j^0} \geq 0, \end{cases} \quad (3)$$

where $\bar{t}$ is the time instant for which $\Delta N_j(\bar{t}) = L_j \rho_{\max}$. It is proved in Rarità et al. 2010 that (3) represents an alternative exhaustive formulation of LWR model for roads modelling.

Remark. *Notice that dynamics of traffic queues is also considered in (2), replacing $l_j(t) < L_j$ by $\Delta N_j(t) < N_j^{\max} := L_j \rho_{\max}$ and $l_j(t) = L_j$ by $\Delta N_j(t) = N_j^{\max}$. This corresponds to a situation in which the vehicles would not queue up along the road section, but at its downstream end.*

3. CITY NETWORKS

A City network is given by a rectangular network, seen as a matrix with $\mathcal{N}$ rows and $\mathcal{M}$ columns. In particular, the network is described by the couple $(\mathcal{I}, \mathcal{J})$, where $\mathcal{I}$ and $\mathcal{J}$ indicate, respectively, the set of roads and junctions. Moreover $\mathcal{I} = \mathcal{I}_C \cup \mathcal{I}_R$, where $\mathcal{I}_C$ and $\mathcal{I}_R$ represent, respectively, the set of vertical and horizontal roads (columns and rows of the network graph). Each node is identified by a couple $(i, j) \in \mathcal{J}$, with $i \in \mathcal{N}$ and $j \in \mathcal{M}$, and has two incoming and two outgoing roads (junction of 2×2 type). At node (i, j), vertical roads are labelled as C_{ij} (entering) and C_{i+1j} (exiting), while horizontal ones are indicated by R_{ij} (entering) and R_{ij+1} (exiting), as in Figure 3.

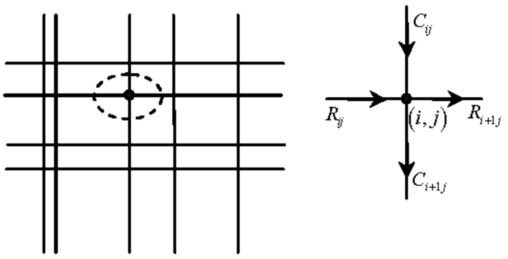


Figure 3: City network (left) and zoom on a portion (right)

To simplify the notation, we make the following assumption: **(CN)** All roads $C_{ij} \in \mathcal{I}_C$, $R_{ij} \in \mathcal{I}_R$, $i \in \mathcal{N}$, $j \in \mathcal{M}$, have the same maximum in – and outflow, i.e. $\hat{Q}_k = \hat{Q} \quad \forall \, k \in \mathcal{I}$ and free speed: $V_k^0 = V_0 \quad \forall \, k \in \mathcal{I}$. Notice, however, that we consider possibly different lengths of roads $L_{C_{ij}}$ and $L_{R_{ij}}$.

Permeability parameters of roads C_{ij} and R_{ij} are indicated by $\gamma_{C_{ij}}(t)$ and $\gamma_{R_{ij}}(t)$. As the two roads belong to the same node (i, j), we assume that $0 \leq \gamma_{C_{ij}}(t) + \gamma_{R_{ij}}(t) \leq 1$.

3.1. Traffic at nodes

The dynamics at nodes is defined by solutions to Riemann problems, i.e. Cauchy problems with initial constant data on each road. The map, which associates to every initial data the corresponding fluxes at the node, is called Riemann Solver and indicated by *RS*. The solution depends on initial fluxes and on the number of delayed vehicles (resp. length of congested zone) of all roads meeting at the node. Now, we consider two rules (see Coclite et al. 2005; Garavello and Piccoli, 2006) to define uniquely the solution to an *RS*: (A) At each node $(i, j) \in \mathcal{J}$ drivers distribute according to fixed coefficients, given by a matrix X; (B) Respecting (A), drivers behave so as to maximize the flux through node (i, j).

Remark. *Considering road junctions of 2×2 type, we assume that:*

$$X = \begin{pmatrix} \alpha_{ij} & \beta_{ij} \\ 1 - \alpha_{ij} & 1 - \beta_{ij} \end{pmatrix}, \quad (4)$$

where $0 \leq \alpha_{ij}, \beta_{ij} \leq 1$ and α_{ij} (resp. β_{ij}) represents the percentage of traffic that, from road C_{ij} (resp. R_{ij}), goes to road R_{ij+1} (resp. C_{ij+1}).

Remark. *Using both rules (A) and (B) under the assumption $\alpha_{ij} \neq \beta_{ij}$, we get a rich set of possible solutions for the dynamics at road junctions, depending on the state of roads, namely if they are empty, almost congested or totally congested. Details are in Rarità et al. 2010.*

From formulas (1) and (2), we also get that the 4 – tuple $\left(A_{C_{i+1j}}, A_{R_{ij+1}}, O_{C_{ij}}, O_{R_{ij}}\right)$, defined by the *RS* for the node $(i, j) \in \mathcal{J}$, is essentialy determined by: α_{ij}; β_{ij}; $\gamma_{R_{ij}}$; $\gamma_{C_{ij}}$; ΔN of roads connected to (i, j); delayed (A, O) for other nodes.

4. AN OPTIMAL CONTROL PROBLEM AND A HYBRID DYNAMIC

Now, we consider an optimal control problem for City Networks. The dynamics over the network is represented as a control system of the form:

$$\dot{x} = f(x, \gamma, \gamma_\delta), \quad (5)$$

where x is the state (the number of delayed vehicles ΔN), γ is the control (the permeability parameters) and γ_δ are delayed controls. Introduce the variable y_k such that $\dot{y}_k = \Delta N_k\left(x,\gamma,\gamma_\delta\right)$, $y_k\left(0\right)=0$, $i \in \mathcal{I}$. For a class Γ of admissible controls, we state the following optimal control problem:

$$\min_{\gamma\in\Gamma} \sum_{k\in\mathcal{I}} y_k\left(t\right) \tag{6}$$

for a fixed initial condition $\bar{x}$. Notice that (6) represents the minimization of delayed vehicles over the whole network in terms of permeability parameters. Now, we consider the dynamics in detail.

4.1. Queues on roads

Fix a generic node $\left(i,j\right)\in\mathcal{J}$. The dynamics for the whole network is described by the system (5), where $x=\left(y_{C_{ij}},y_{R_{ij}},\Delta N_{C_{ij}},\Delta N_{R_{ij}}\right)$.

First, assume that, for roads C_{ij} and R_{ij}, $\Delta N_{C_{ij}}>0$ and $\Delta N_{R_{ij}}>0$. Omitting, for simplicity, the dependence on traffic distribution coefficients, which are not dependent on time, we get the following equations, where the evolution of queues is function, through *RS*, of delayed and non delayed controls at node $\left(i,j\right)$:

$$\begin{aligned}
&\dot{y}_{C_{ij}} = \Delta N_{C_{ij}},\ \ \dot{y}_{R_{ij}} = \Delta N_{R_{ij}},\\
&\Delta\dot{N}_{C_{ij}} = RS\left(\left(\gamma_{C_{ij}},\gamma_{R_{ij}}\right)\left(t\right),\left(\gamma_{C_{i-1\,j}},\gamma_{R_{i-1\,j}}\right)\left(t-L_{C_{ij}}/V_0\right)\right), \quad (7)\\
&\Delta\dot{N}_{R_{ij}} = RS\left(\left(\gamma_{C_{ij}},\gamma_{R_{ij}}\right)\left(t\right),\left(\gamma_{C_{ij-1}},\gamma_{R_{ij-1}}\right)\left(t-L_{R_{ij}}/V_0\right)\right).
\end{aligned}$$

Consider now that roads C_{ij} and R_{ij} are empty, namely $\Delta N_{C_{ij}}=\Delta N_{R_{ij}}=0$. Dropping as usual the dependence on traffic distribution coefficients, for road C_{ij} we have:

$$\begin{aligned}
&\dot{y}_{C_{ij}} = \Delta N_{C_{ij}},\\
&\Delta\dot{N}_{C_{ij}} = g_1\left(A_{C_{ij}}\left(t-L_{C_{ij}}/V_0\right),O_{C_{ij}}\left(t\right)\right),
\end{aligned} \tag{8}$$

where g_1 is some function. Considering for simplicity the only presence of nodes $\left(i-1,j\right)$ and $\left(i-2,j\right)$ inside the network, $A_{C_{ij}}\left(t-L_{C_{ij}}/V_0\right)$ is written as:

$$\begin{aligned}
&A_{C_{ij}}\left(t-\frac{L_{C_{ij}}}{V_0}\right)=RS\left(g_2\left(RS\right),(O_{R_{i-1\,j}},\gamma_{C_{i-1\,j}},\gamma_{R_{i-1\,j}})\left(t-\frac{L_{C_{ij}}}{V_0}\right)\right),\\
&g_2\left(RS\right)=g_2\left(RS\left(\left(O_{C_{i-2\,j}},O_{R_{i-2\,j}},\gamma_{C_{i-2\,j}},\gamma_{R_{i-2\,j}}\right)\left(t-\frac{L_{C_{i-1\,j}}+L_{C_{ij}}}{V_0}\right)\right)\right),
\end{aligned} \tag{9}$$

where g_2 is a function different from g_1. Notice that (9) represents a "nested equation", as phenomena at $\left(i,j\right)$ are dependent on other nodes, namely the evolution of $y_{C_{ij}}$ and $\Delta N_{C_{ij}}$, expressed by (8), is written in terms of all nodes of the network. For road R_{ij}, we have similar equations.

4.2. A hybrid dynamic and needle variations

Here, we consider a hybrid dynamic to avoid the nested equations in case of empty queues. Continuous equations involving the whole network can be replaced introduced some extra logic variables. The latter, in turn, are affected and affect the continuous variables evolution.

Define the logic variables $\varepsilon_{C_{ij}}$ as:

$$\varepsilon_{C_{ij}} := \begin{cases} -1, \text{ if } \Delta N_{C_{ij}}=0,\\ 0, \quad \text{if } 0<\Delta N_{C_{ij}}<\Delta N_{C_{ij}}^{\max},\\ +1, \text{ if } \Delta N_{C_{ij}}=\Delta N_{C_{ij}}^{\max}. \end{cases} \tag{10}$$

We set the following: $\tilde{\gamma}:=\left(\gamma_{C_{i-1\,j}},\gamma_{R_{i-1\,j}}\right)$, $\underset{\sim}{\gamma}:=\left(\gamma_{C_{ij}},\gamma_{R_{ij}}\right)$, $\tilde{O}:=\left(O_{C_{ij}},O_{R_{i-1\,j+1}}\right)$, $\underset{\sim}{O}:=\left(O_{C_{i+1\,j}},O_{R_{ij+1}}\right)$, $\tilde{A}:=\left(A_{C_{i-1\,j}},A_{R_{i-1\,j}}\right)$, $\underset{\sim}{A}:=\left(A_{C_{ij}},A_{R_{ij}}\right)$, $\tilde{\varepsilon}:=\left(\varepsilon_{C_{i-1\,j}},\varepsilon_{R_{i-1\,j}},\varepsilon_{C_{ij}},\varepsilon_{R_{i-1\,j+1}}\right)$, and $\underset{\sim}{\varepsilon}:=\left(\varepsilon_{C_{ij}},\varepsilon_{R_{ij}},\varepsilon_{C_{i+1\,j}},\varepsilon_{R_{ij+1}}\right)$. A complete hybrid dynamics for node $\left(i,j\right)$ is given by the following equations (for simplicity, the dependence of distribution coefficients on time is omitted, while the exponent δ indicates a delayed dependence on time):

$$\begin{aligned}
&\dot{y}_{C_{ij}} = \Delta N_{C_{ij}},\ \ \Delta\dot{N}_{C_{ij}} = A_{C_{ij}}^{\delta}-O_{C_{ij}},\\
&A_{C_{ij}} = RS\left(\tilde{\gamma},\tilde{A}^{\delta},\tilde{O}^{\delta},\tilde{\varepsilon}\right),\ O_{C_{ij}} = RS\left(\underset{\sim}{\gamma},\underset{\sim}{A}^{\delta},\underset{\sim}{O}^{\delta},\underset{\sim}{\varepsilon}\right).
\end{aligned} \tag{11}$$

For $\varepsilon_{R_{ij}}$, the definition is similar, substituting *C* with *R*. Moreover, also for road R_{ij} we have equations similar to (11). Suitable differences are already explained in Rarità et al. 2010.

The dynamic of control parameters γ (and distribution coefficients α or β) influence the evolution of the couple $\left(A,O\right)$ through *RS*. In turn, the values of $\left(A,O\right)$ influence themselves through *RS* and determine the continuous dynamics of ΔN. The dynamics of ΔN defines that of y and discrete changes, through ε, of the couple $\left(A,O\right)$. In Figure 4, a summarizing scheme is reported, where *c* and *d* indicate, respectively, if the dynamics is continuous or delayed.

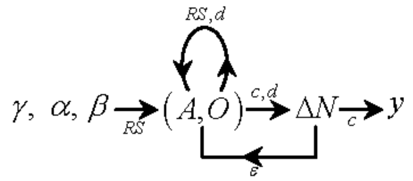


Figure 4: scheme of the hybrid dynamics

Now, we consider the sensitivity of the control system (5) with respect to control variations, adopting the point of view of Pontryagin Maximum Principle (PMP), see Bressan at al. 2007. In particular, we consider special variation of controls, called "needle variations", and variational equations along trajectories to determine the relative effects on the dynamics.

Consider the control system (5), where $\gamma_\delta(t) = \gamma(t-\delta)$ and $\delta > 0$. Fix a candidate optimal control $\Gamma \ni \gamma^* : [0,T] \mapsto U = [0,1]$ and let x^* be the corresponding trajectory, starting from a given point $\bar{x}$. A needle variation is defined as follows:

Definition *(Needle Variation). Consider the map* $\varphi : t \mapsto f\left(x^*(t), \gamma^*(t), \gamma^*_\delta(t)\right)$ *and let* τ *be a Lebesgue point for* φ. *Given* $\omega \in U$, *define a family of controls* $\eta_\gamma(t,\tau,\zeta,\omega)$, $\zeta \in [0,\tau[$ *in the following way:*

$$\eta_\gamma(t,\tau,\zeta,\omega) := \begin{cases} \gamma^*(t), & \text{if } t \in [0,\tau-\zeta[, \\ \omega, & \text{if } t \in [\tau-\zeta,\tau[, \\ \gamma^*(t), & \text{if } t \in [\tau,T]. \end{cases} \tag{12}$$

and let $\eta_x(t,\tau,\zeta,\omega)$ *be the trajectories corresponding to* η_γ *with* $\eta_x(0,\tau,\zeta,\omega) = \bar{x}$. *We call the couple* $(\eta_\gamma,\eta_x) = (\eta_\gamma,\eta_x)(\tau,\omega)$ *a needle variation of* $\left(x^*,\gamma^*,\gamma^*_\delta\right)$. *If the trajectories are uniquely determined by controls we use the simplified notation* $\eta_\gamma(\tau,\omega)$.

Given a needle variation, for every time $t \geq \tau$ it is defined a curve of points $\eta_x(t,\tau,\zeta,\omega)$ that are reached at time *t* by admissible controls $\eta_\gamma \in \Gamma$. In particular, at the final time, the points $\eta_x(T,\tau,\zeta,\omega)$ are reached. If the cost is given as in (6), for γ^* to be optimal we need that: $\nabla\left(\sum_{k\in\mathcal{I}} y^*_k(T)\right) \cdot v(T) \geq 0$, where $v(t)$ is the tangent vector to the curve $\eta_x(T,\tau,\zeta,\omega)$ at $\zeta = 0$, equal to $v(t) = \left.\dfrac{d\eta_x(t,\tau,\zeta,\omega)}{d\zeta}\right|_{\zeta=0}$.

The vector *v*, for $t > \tau$, satisfies the variational equation $\dot{v} = D_x f\left(x^*,\gamma^*,\gamma^*_\delta\right) \cdot v$, with initial condition:

$$v(\tau) = f\left(x^*(\tau),\omega,\gamma^*_\delta(\tau)\right) - f\left(x^*(\tau),\gamma^*(\tau),\gamma^*_\delta(\tau)\right), \tag{13}$$

that presents a jump at time $\tau + \delta$, see Rarità et al. 2010. For a City network, if a variation of $\gamma_{C_{ij},R_{ij}}$ occurs at node (i,j), we have to consider the tangent vectors $v^{y}_{C_{ij},R_{ij}}$ and $v^{\Delta N}_{C_{ij},R_{ij}}$ for the variables $y_{C_{ij},R_{ij}}$ and $\Delta N_{C_{ij},R_{ij}}$, respectively. Hence, the variational equations are described by $\dot{v}^{y}_{C_{ij}} = v^{\Delta N}_{C_{ij}}$, $\dot{v}^{y}_{R_{ij}} = v^{\Delta N}_{R_{ij}}$ and $\dot{v}^{\Delta N}_{C_{ij}} = \dot{v}^{\Delta N}_{R_{ij}} = 0$.

Needle variations of permeability parameters (controls) generate other needle variations for the arrival and departure flows, which in turn provokes jumps in the variational vectors for delayed vehicles. In Table 1, we summarize an exhaustive scheme of jumps due to needle variations. Notice that column 1 shows which is the parameter (γ, *A* or *O*) for which a needle variation occurs; columns 2 indicates what are the quantities on which the needle variation provokes jumps.

Table 1: scheme of needle variations and jumps

1	**2**
$\gamma_{C_{ij}}$	$A_{C_{i+1\,j}}$, $A_{R_{ij+1}}$, $O_{C_{ij}}$
$\gamma_{R_{ij}}$	$A_{C_{i+1\,j}}$, $A_{R_{ij+1}}$, $O_{R_{ij}}$
$O_{C_{ij}}$	$A_{C_{ij}}$, $A_{R_{i-1\,j+1}}$, $O_{C_{i-1\,j}}$, $O_{R_{i-1\,j}}$ if $\Delta N_{C_{ij}} = \Delta N^{\max}_{C_{ij}}$
$O_{R_{ij}}$	$A_{C_{i+1\,j-1}}$, $A_{R_{ij}}$, $O_{C_{ij-1}}$, $O_{R_{ij-1}}$ if $\Delta N_{R_{ij}} = \Delta N^{\max}_{R_{ij}}$
$A_{C_{ij}}$	$A_{R_{i-1\,j+1}}$, $O_{C_{i-1\,j}}$, $O_{R_{i-1\,j}}$ if $\Delta N_{C_{ij}} = 0$
$A_{R_{ij}}$	$A_{C_{i+1\,j-1}}$, $O_{C_{ij-1}}$, $O_{R_{ij-1}}$ if $\Delta N_{R_{ij}} = 0$

The interpretations of Table 1 is the following: it is sufficient the variation of just one permeability parameter to provoke jumps in incoming and outgoing flows. Notice that some jumps occur only if roads are empty (case of incoming flows) or full (case of outgoing flows).

Remark. *For sake of space, we omit dynamics of jumps for logic variables, which is in Rarità et al. 2010.*

5. SIMULATIONS

We aim to illustrate the effect of a needle variation on a single permeability parameter for a given node in a network. For this reason we present some simulations of a real road network, proving that a unique little variation can provoke some cascade effects on incoming flows, and car queues.

We run some simulation for a City type network, which is a portion of the real network of Salerno (see Figure 1). In particular, according to the notations of Section 3, we label by (i,j) the junction between Corso Garibaldi and Via Adolfo Cilento; hence, $(i,j+1)$ indicates the intersection between Corso Garibaldi and Via Arturo De Felice. In particular, Corso

Garibaldi is identified by the three road segments R_{ij+1}, R_{ij}, and R_{ij-1}; Via Arturo De Felice by the road segments C_{ij+1}, and C_{i-1j+1}; Via Adolfo Cilento by C_{ij}, and C_{i+1j}. Figure 5 shows the topology of the considered network.

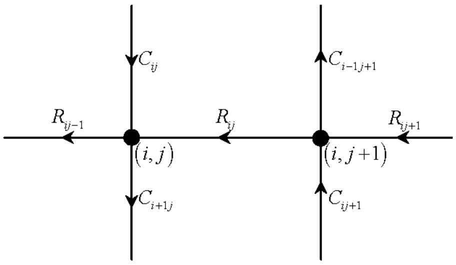


Figure 5: topology of the network

A fourth Runge – Kutta scheme is used, with temporal step $h=0.01$ and a total simulation time $T=25\text{ min}$. We assume that: $L_{R_{ij+1}}=6$; $L_{C_{ij+1}}=5$; $L_{C_{i-1j+1}}=L_{R_{ij}}=L_{C_{i+1j}}=4$; $L_{C_{ij}}=L_{R_{ij-1}}=3$; for all roads, $V_0=c=2$, $\rho_{\max}=1$, hence $\hat{Q}=1$; incoming fluxes:

$$A_{R_{ij+1}}(t)=A_{C_{ij+1}}(t)=A_{C_{ij}}(t)=\begin{cases}0.5, & \text{if } t\geq 0,\\ 0, & \text{otherwise;}\end{cases} \tag{14}$$

distribution matrices $X^{(i,j)}$ and $X^{(i,j+1)}$ at nodes (i,j) and $(i,j+1)$ equal to:

$$X^{(i,j)}=\begin{pmatrix}0.3 & 0.3\\ 0.7 & 0.7\end{pmatrix};\ X^{(i,j+1)}=\begin{pmatrix}0.2 & 0.2\\ 0.8 & 0.8\end{pmatrix}; \tag{15}$$

initial conditions for queues are: $\Delta N_{R_{ij+1}}(0)=3$; $\Delta N_{C_{ij+1}}(0)=\Delta N_{C_{i-1j+1}}(0)=\Delta N_{R_{ij}}(0)=\Delta N_{C_{i+1j}}(0)=2$, and $\Delta N_{C_{ij}}(0)=\Delta N_{R_{ij-1}}(0)=1$; constant permeability parameters, with the exception of $\gamma_{R_{ij}}(t)$, for which a needle variation occurs, namely we have that: $\gamma_{R_{ij+1}}=\gamma_{C_{ij+1}}=\gamma_{C_{ij}}=0.5$; $\gamma_{C_{i-1j+1}}=\gamma_{C_{i+1j}}=0.3$, and $\gamma_{R_{ij-1}}=0.7$; finally:

$$\gamma_{R_{ij}}(t)=\begin{cases}\gamma^*_{R_{ij}}, \text{ if } t\in[0,t_1]\cup]t_2,T],\\ \omega_{R_{ij}}, \text{ if } t\in]t_1,t_2],\end{cases} \tag{16}$$

with $\gamma^*_{R_{ij}}=0.5$ and $\omega_{R_{ij}}=0.2$, $t_1=11\text{ min}$ and $t_2=13\text{ min}$.

Remark. *Notice that lengths of roads, velocities in free and congested regimes, initial conditions for queues and the maximal densities are normalized with respect to the length* $L\simeq 116$ *meters, measured on the real network that we are considering.*

In Figure 6, we present the evolution of $O_{R_{ij}}(t)$, while $A_{R_{ij}}(t)$ and $\Delta N_{R_{ij}}(t)$ are represented in Figures 7 and 8, respectively.

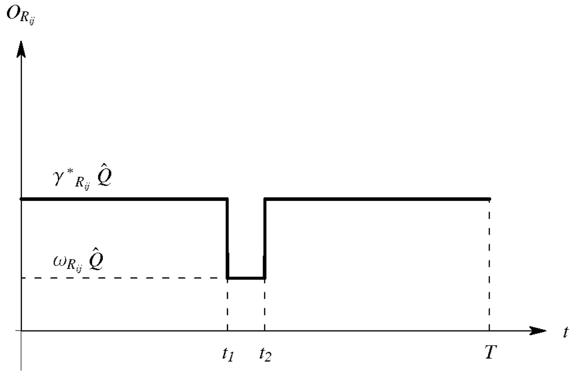


Figure 6: $O_{R_{ij}}(t)$ due to a needle variation for $\gamma_{R_{ij}}(t)$

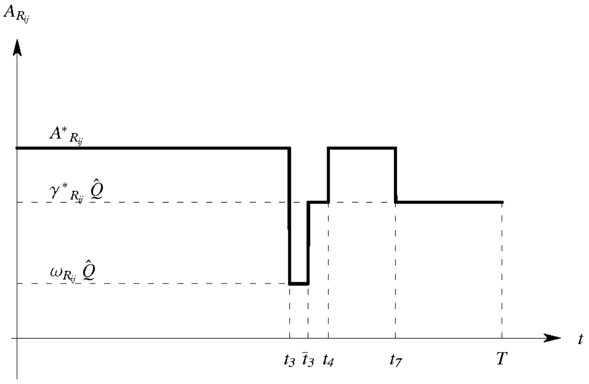


Figure 7: $A_{R_{ij}}(t)$ due to a needle variation for $\gamma_{R_{ij}}(t)$

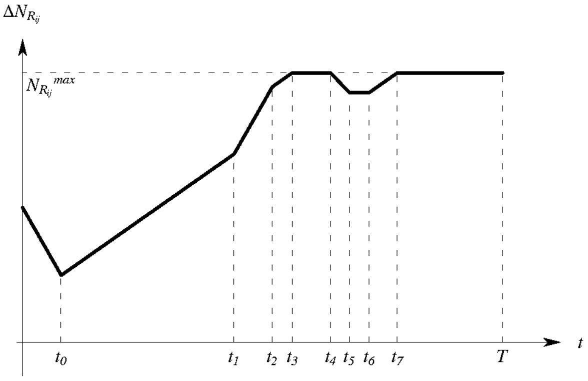


Figure 8: $\Delta N_{R_{ij}}(t)$ due to a needle variation for $\gamma_{R_{ij}}(t)$

Notice that: for $t\leq t_0=L_{R_{ij}}/V_0=2$, $\Delta N_{R_{ij}}(t)$ decreases, as the solutions of *RS* at nodes $(i,j+1)$ and (i,j) imply, respectively, $A_{R_{ij}}(t-t_0)=A^*_{R_{ij}}=0.7$ and

$O_{R_{ij}}(t)=\gamma_{R_{ij}}(t)\hat{Q}=0.5$, with $A_{R_{ij}}(t-t_0)-O_{R_{ij}}(t)>0$. At $t=t_1$, the needle variation for $\gamma_{R_{ij+1}}(t)$ generates a needle variation for $O_{R_{ij}}(t)$, that provokes an immediate change of slope for $\Delta N_{R_{ij}}(t)$. At $t=t_2$, the needle variation for $\gamma_{R_{ij+1}}(t)$ vanishes, hence we get an immediate further change of slope for $\Delta N_{R_{ij}}(t)$ while $O_{R_{ij}}(t)$ comes back to the nominal value imposed by *RS* at node (i,j). At $t_3\simeq 14\text{ min}$, $\Delta N_{R_{ij}}(t)=\Delta N_{R_{ij}}^{\max}$ and $A_{R_{ij}}(t)$ follows the delayed $O_{R_{ij}}(t)$, namely:

$$A_{R_{ij}}(t)=O_{R_{ij}}\left(t-\frac{L_{R_{ij}}}{c}\right)=\begin{cases}\omega_{R_{ij}}\hat{Q}=0.2, \text{ if } t\in\left[t_3,\bar{t}_3\right],\\ \gamma_{R_{ij}}^{*}\hat{Q}=0.5, \text{ if } t\in\left]\bar{t}_3,t_4\right],\end{cases}\qquad(17)$$

where $\bar{t}_3=t_2+t_0$, $t_4=t_3+t_0$. At t_4, $\Delta N_{R_{ij}}(t)$ starts to decrease as $A_{R_{ij}}(t-t_0)-O_{R_{ij}}(t)<0$, and $A_{R_{ij}}(t)=A_{R_{ij}}^{*}=0.7$, the value imposed by *RS* at node $(i,j+1)$; $\Delta N_{R_{ij}}(t)$ becomes constant for $t\in\left[t_5,t_6\right[$, with $t_5=\bar{t}_3+t_0$, $t_6=t_4+t_0$, as $A_{R_{ij}}(t-t_0)-O_{R_{ij}}(t)=0$. At t_6, $\Delta N_{R_{ij}}(t)$ starts to increase, and it grows until $t_7\simeq 19.5\text{ min}$, for which $\Delta N_{R_{ij}}(t)=\Delta N_{R_{ij}}^{\max}$ and, as a consequence, $A_{R_{ij}}(t)=O_{R_{ij}}\left(t-\frac{L_{R_{ij}}}{c}\right)=\gamma_{R_{ij}}^{*}\hat{Q}=0.5$. Moreover, $\Delta N_{R_{ij}}(t)$ remains at its maximal value, as $A_{R_{ij}}(t-t_0)-O_{R_{ij}}(t)=0$ for $t\geq t_7+t_0$.

A further analysis can also be made. For the network of Figure 5, we want to solve the optimization control problem (6). From a theoretical point of view, it is necessary to find a set of permeability parameters such as to minimize the sum of queues for all roads, namely $Y(t)=\sum_{k\in\mathcal{I}}y_k(t)$, $k\in\mathcal{I}$, where $\mathcal{I}$ is the set of roads, $\mathcal{I}\in\left\{R_{ij+1},R_{ij},R_{ij-1},C_{ij+1},C_{i-1j+1},C_{i+1j},C_{ij}\right\}$. As we introduced logic variables and defined an hybrid framework to avoid nested equations, the minimization of $Y(t)$ is simply found analyzing needle variations of the only permeability parameters, as their only variation provokes cascade effects on the whole network in terms of incoming flows, outgoing flows and queues on roads. For the portion of the real network of Salerno, traffic at nodes (i,j) and $(i,j+1)$ is regulated through $\gamma_{R_{ij+1}}(t)$, $\gamma_{C_{ij+1}}(t)$, $\gamma_{R_{ij}}(t)$, and $\gamma_{C_{i+1j}}(t)$. A suitable choice of such parameters allows to optimize the performances on the whole network in terms of delayed vehicles.

Assume, for simplicity, that $\gamma_{R_{ij+1}}(t)+\gamma_{C_{ij+1}}(t)=1$ and $\gamma_{R_{ij}}(t)+\gamma_{C_{i+1j}}(t)=1$. Then, the choice for the optimization clearly depends only on $\gamma_{R_{ij+1}}(t)$ and $\gamma_{R_{ij}}(t)$. Using the numerical software Mathematica, it is possible to use a steepest descent method in order to find the couple $\left(\gamma_{R_{ij+1}}^{*},\gamma_{R_{ij}}^{*}\right)$ that solves problem (6). In our case, with the same simulation parameters we have considered before, we get that $\left(\gamma_{R_{ij+1}}^{*},\gamma_{R_{ij}}^{*}\right)\simeq(0.415,0.318)$ in eight iterations, starting from $\left(\gamma_{R_{ij+1}}^{0},\gamma_{R_{ij}}^{0}\right)=(0.65,0.25)$. The cost functional $Y(t)$ decreases from 13 to 3.6. In Figures 9, 10 and 11, we report how $\gamma_{R_{ij+1}}(t)$ and $\gamma_{R_{ij}}(t)$ vary according to the different steps of the numerical minimization method and, finally, the cost functional, that decreases until the steady state minimum value.

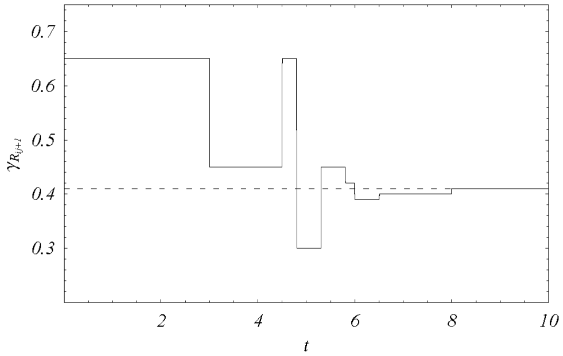


Figure 9: variations of $\gamma_{R_{ij+1}}(t)$ in different steps of the numerical minimization algorithm

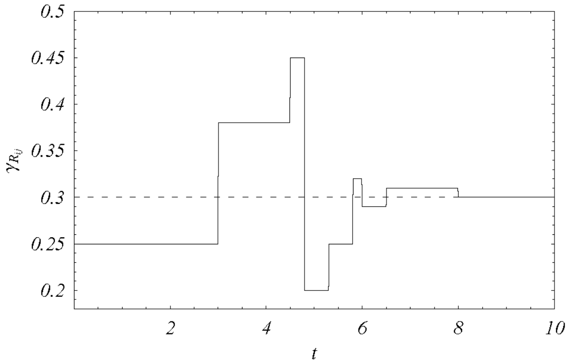


Figure 10: variations of $\gamma_{R_{ij}}(t)$ in different steps of the numerical minimization algorithm

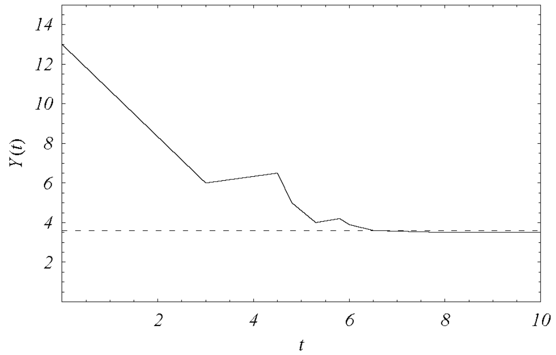


Figure 11: variations of $Y(t)$

Although it is evident that the minimization of queues is achieved, it is not possible to erase all queues on roads. A such phenomenon is not surprising, as dynamics at nodes does not always allow a complete emptying of queues, and this is the classical situation, that arises in normal traffic in cities.

6. CONCLUSIONS

We considered a delayed – ODE approach to describe car traffic in road networks of City type. The minimization of the number of vehicles was studied in terms of permeability parameters, which regulate the inflows at nodes. As the overall dynamics gives rise to nested equations, logic variables were introduced and a hybrid framework was thus obtained. A sensitivity analysis, based on needle variations, was developed for permeability parameters. The total effects of variations, also in terms of optimization of traffic performances, were described and then verified by simulations of a portion of the real network of Salerno, Italy.

Further research should be developed to achieve more information on optimal controls, e. g. using necessary conditions for hybrid control systems. This problem was not completely solved yet from a theoretical point of view.

From a numerical point of view, large scale simulations, extended to the overall network of big cities, are nowadays giving meaningful results for the optimization of traffic performances.

REFERENCES

Bressan, A., Piccoli B., 2007. *Introduction of Mathematical Theory of Control*. Appl. Math. Ser., 2, American Institute of Mathematical Sciences.

Bretti, G., Natalini, R., Piccoli, B., 2006. Numerical approximations of a traffic flow model on networks. *Netw. Heterog. Media*, 1 (1), 57 – 84.

Coclite G., Garavello M., Piccoli, B., 2005. Traffic flow on road networks. *SIAM J. Math. Anal.*, 36, 1862 – 1886.

Cascone, A., D'Apice, C., Piccoli, B., Rarità, L., 2007. Optimization of traffic on road networks. *Math. Models Methods Appl. Sci.*, 17 (10), 1587 – 1617.

Cascone, A., D'Apice, C., Piccoli, B., Rarità, L., 2008. Circulation of car traffic in congested urban areas. *Comm. Math. Sci.*, 6 (3), 765 – 784.

D'Apice, C., Manzo, R., Rarità, L., 2011. Splitting of traffic flows to control congestion in special events. *Int. Journ. Math. and Math. Sci.*, Article ID 563171, 18 pages.

Daganzo, C. F., 1995. Requiem for second – order fluid dynamic approximations of traffic flow. *Transport. Res. B*, 29, 277 – 286.

Daganzo, C. F., 1995. The cell transmission model, Part II: Network Traffic. *Transport. Res. B*, 29, 79 – 93.

Garavello, M., Piccoli, B., 2006. *Traffic Flow on Networks*. American Institute of Mathematical Sciences, Springfield.

Helbing, D., 2001. Traffic and related self – driven many – particle system. *Rev. Modern Phys.*, 73, 1067 – 1141.

Helbing, D., Lämmer, S., Lebacque J. P., 2005. Self – organized control of irregular or perturbed network traffic, *C. Deissenberg, R. F. Hartl (Eds.), Optimal Control and Dynamic Games, Springer, Dordrecht*, 239 – 274.

Helbing, D., Siegmeier, J., Lämmer, S., 2007. Self – organized network flows. *Netw. Heterog. Media,* 2 (2), 193 – 210.

Herty, M., Klar, A., 2003. Modeling, simulation and optimization of traffic flow networks. *SIAM J. Appl. Math.*, 64 (2), 565 – 582.

Herty, M., Klar, A., 2004. Simplified dynamics and optimization of large scale traffic flow networks. *Math. Models Methods Appl. Sci.*, 14 (4), 579 – 601.

Herty, M., Moutari, S., Rascle, M., 2006. Optimization criteria for modelling intersections of vehicular traffic flow. *Netw. Heterog. Media*, 1 (2), 275 – 294.

Hilliges, M., Weidlich, W., 1995. A phenomenological model for dynamic traffic flow in networks. *Transport. Res. B.*, 29, 407 – 431.

Kerner, B., 2004. *The Physics of Traffic*. Springer, Berlin.

Lebacque, J. P., Khoshyaran, M. M., 2005. First – order macroscopic traffic flow models: Intersection modeling, network modeling. *Proceedings of 16th International Symposium on Transportation and Traffic Theory*, pp. 365 – 386, H. S. Mahmasani (Ed.), Elsevier.

Lighthill, M. J., Whitham, G. B., On kinematic waves: II. A theory of traffic on long crowded roads. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 229, 317 – 345.

Rarità, L., D'Apice, C., Piccoli, B., Helbing, D., 2010. Sensitivity analysis of permeability parameters for flows on Barcelona networks, *Journ. Diff. Equat.*, 249 (12), 3110 – 3131.

Richards, P. I., 1956. Shock waves on the highway. *Oper. Res.*, 4, 42 – 51.

Schönhof, M., Helbing, D., 2006. Empirical features of congestion traffic states and their implications for traffic modelling. *Transport. Sci.*, 41 (2), 135 – 166.

Whitham, G. B., 1974. *Linear and Nonlinear Waves*, Wiley, New York.

AUTHORS BIOGRAPHY

LUIGI RARITÀ was born in Salerno, Italy, in 1981. He graduated cum laude in Electronic Engineering in 2004, with a thesis on mathematical models for telecommunication networks, in particular tandem queueing networks with negative customers and blocking. He obtained PhD in Information Engineering in 2008 at the University of Salerno, discussing a thesis about control problems for flows on networks. He is actually a research assistant at the University of Salerno. His scientific interests are about numerical schemes and optimization techniques for fluid – dynamic models, queueing networks, and Knowledge models for the Cultural Heritage area.
His e-mail address is lrarita@unisa.it.

ADVANCED TOOL FOR PREDICTIVE DIAGNOSIS AND MAINTENANCE USING CASE-BASED REASONING AND FUZZY LOGIC

Nikolinka Christova[a], Atanas Atanassov[b]

[a]Dept. of Automation of Industry, University of Chemical Technology and Metallurgy,
1756 Sofia, Bulgaria
[b]Dept. of Programming and Computer System Application, University of Chemical Technology and Metallurgy,
1756 Sofia, Bulgaria

[a]nchrist@uctm.edu, [b]naso@uctm.edu

ABSTRACT

The application of computational intelligence in condition-based maintenance and diagnosis plays a leading role in the technology development of intelligent manufacturing systems. Case-Based Reasoning (CBR) is mostly used in designing the real time application having the decision support capability. In this study implementation of fuzzy logic in the CBR systems that deriving effective knowledge representation schemes has been described. The benefits of the approach have been presented. The applications of the developed advanced tool based on fuzzy logic and CBR for solving the real problems of predictive diagnosis and maintenance in industrial systems have been discussed.

Keywords: diagnosis, maintenance, fuzzy logic, Case Based Reasoning (CBR)

1. INTRODUCTION

Failure prognostic is emerging as the next logical step towards improved system condition based maintenance, beside classic fault detection and diagnostics techniques. These methods form system health management platforms which contribute to longer and reliable operation of systems enable them forecasted maintenance intervals, remaining useful life of system components, system reconfiguration, optimization, etc. (Tenchev and Kondev 2006).

The past three decades have witnessed an explosion of renewed interest in the areas of Computational Intelligence (CI) (Karray and De Silva 2004, Konar 2005) – a technology that involves advanced information processing methodologies and techniques for analyzing, designing and developing intelligent systems.

The combination of (two or more) different problem solving and knowledge representation methods is a very active research area in artificial intelligence (Karray and De Silva 2004, Konar 2005). The aim is to create combined formalisms that benefit from each of their components. If the methods (ontologies, agents, rule-based reasoning, and case-based reasoning) and the techniques (fuzzy logic, neural networks, genetic algorithms, and swarm optimization) are presented at two levels, horizontal and/or vertical integration of them could be implemented. It is generally believed that complex problems are easier to solve with hybrid or integrated approaches. The effectiveness of various hybrid or integrated approaches has been demonstrated in a number of application areas (Aha 2006; Boshnakov, Boishina, and Hadjiiski 2011; Chan 2005; Hadjiski and Boishina 2010; Karray and De Silva 2004; Konar 2005; Prentzas and Hatzilygeroudis 2009).

The methodology of Case-Based Reasoning (CBR) involves solving new problems by identifying and adapting solutions to similar problems stored in a library of past experiences. This approach utilizes the experience gained from solving past problems (Aamodt and Plaza 1994).

Fuzzy set theory (Zadeh 1983) provides an approximate but effective and flexible way of representing, manipulating, and utilizing vaguely defined data and information. It can also describe the behaviors of systems that are too complex or too ill-defined (Karray and De Silva 2004; Konar 2005).

In this paper combination of CBR and fuzzy logic-based techniques into a generic tool capable of handling problems in which an existing case base would be used to build solutions to new cases. The developed advanced tool is based on the investigation in (Atanassov and Antonov 2012) where the main purpose of the carried out analysis is to determine the rate of applications of the software frameworks for development of CBR-software platforms for the tasks of predictive diagnosis and maintenance.

2. CASE-BASED REASONING (CBR)

Case-Based Reasoning (CBR) is a method that compares the present problem with previous ones and applies the problem solving of the past to the present problem (Aamodt and Plaza 1994). CBR techniques have been widely applied to various real applications. A successful case-based reasoning system requires a high-quality case base, which provides rich and efficient solutions for solving real problems (Avramenko and

Kraslawski 2006; Mitra and Basak 2005; Yang, Farley, and Orchar 2008).

Case-based representations store a large set of previous cases with their solutions in the *case base* (or case library) and use them whenever a similar new case has to be dealt with (Aamodt and Plaza 1994).

The stages of reasoning in CBR systems, based on cases, are known as classical R^4 cycle. Cases are the main object in CBR systems. They can be represented as free text, in conversational type when each case is represented as a list of question and answers, or in structural type when the cases are represented as a data base (case base).

All structural cases are described as the pair problem-solution (Aamodt and Plaza 1994). The problem $\boldsymbol{p_i = (a_i, v_i)}$ is organized as a structure of attributes and values, described by the attribute vector $\boldsymbol{a_i=(a_{i1},a_{i2},\dots,a_{ir})}$ and the value vector $\boldsymbol{v_i = (v_{i1},v_{i2},\dots,v_{ir})}$.

The solution $\boldsymbol{s_i}$ is represented as vectors, defined by the specific tasks. In multidimensional supervised control tasks, the decision includes two vectors $\boldsymbol{s_i = (sp_i, pr_i)}$, where the first vector $\boldsymbol{sp_i = (sp_{i1}, sp_{i2},\dots,sp_{iq})}$ consists of controllers sets on first hierarchical level, and the second $\boldsymbol{pr_i = (pr_{i1},pr_{i2},\dots,pr_{im})}$ – values of the target parameters, corresponding to the sets.

For solving an actual problem, the following 4 main tasks of CBR R^4 cycle are iteratively performed (Fig. 1) (Aamodt and Plaza 1994):

- *Retrieve* similar previously experienced cases, whose problem has similar definition
- *Reuse* the cases by integrating the solutions from retrieved cases
- *Revise* or adapt the retrieved solution(s) in order to solve the new problem
- *Retain* the new solution in the case base for future usage.

Retrieve – process of extraction of one (nearest neighbor) or a group of cases (*k*-nearest neighbors) having closest definition to the current problem. The global similarity between the problems of these cases (the new $\boldsymbol{p_{new}}$ and the one in the case base $\boldsymbol{p_j}$) is presented by following expression:

$$sim(p_{new}, p_j) = \sum_{i=1}^{n} w_i sim_i(p_{newi}, p_{ji}), \sum_{i=1}^{n} w_i = 1, \qquad (1)$$

where $\boldsymbol{w_i}$ is the weight of $\boldsymbol{i}$-th attribute $0 \le \boldsymbol{w_i} \le 1$ and $sim(p_{newi}, p_{ji})$ is the local similarity between $\boldsymbol{i}$-th attributes of the cases.

For global similarity measure the following metrics are most used: weighted Euclidian distance, Manhattan's metric, Humming's metric, Tversky's metric, Tchebishev's metric, minimum or maximum metrics, etc. (Aamodt and Plaza 1994; Avramenko and Kraslawski 2006).

In the *reuse phase*, a solution for the new case is created based on the retrieved most relevant case(s).

The *revise phase* validates the correctness of the proposed solution. This verification is mostly done by an expert or it is made based on simulation researches if there is a mathematical model available.

Finally, the *retain phase* decides whether the knowledge learned from the solution of the new case is important enough to be incorporated into the system. Quite often the solution contained in the retrieved case(s) is adapted to meet the requirements of the new case.

Usual adaptation methods are substitution, transformation and derivational replay (Aamodt and Plaza 1994; Mitra and Basak 2005; Yang, Farley, and Orchar 2008). For the adaptation task, domain knowledge, usually in the form of rules, is used. Incorporation of knowledge during the operation of a case-based system enhances its reasoning capabilities. This is a major advantage, since the knowledge base of intelligent systems employing other representations remains rather static during operation.

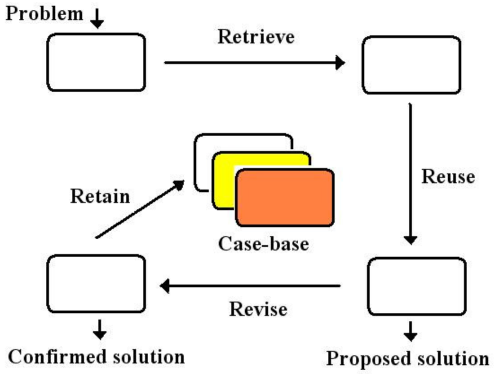


Figure 1: The Classical R^4 Cycle of CBR

The case base size is closely associated with two competing efficiency parameters: mean retrieval time and mean adaptation time. As the case base size increases, retrieval time becomes progressively greater and savings in adaptation time progressively less. There is a saturation point in the case base size after which the increases in the retrieval time are not offset by savings in adaptation time (Aamodt and Plaza 1994; Mitra and Basak 2005). To deal with this problem there can be three ways: restricted insertion of new cases to the case base, carefully devised indexing schemes to guide search and proper case base maintenance policies.

In the literature the question "Is CBR a technology, such as linear programming, neural networks, genetic algorithms, fuzzy logic, and probabilistic reasoning, or just a methodology for problem solving similar to structured systems analysis and design methodology?" has been under discussion (Pal, Dillon, and Yeung 2001). Janet Kolodner (Kolodner 1993) raised this question. She proposed the idea that CBR is both a cognitive model and a method of building intelligent systems. Then Ian Watson published an article explicitly arguing that CBR is a methodology, not a technology (Watson 1999). In examining four very different CBR applications he showed that CBR describes a methodology for problem solving but does not prescribe specific technology. He pointed out that different techniques could be used and applied in various phases of the CBR problem-solving life cycle. For example, nearest-neighbor techniques, induction algorithms such as ID3 and C4.5, fuzzy logic, and database techniques can all be applied to the retrieval phase of a CBR system.

Inductions and many clustering algorithms, such as c-means clustering, Kohonen's self-organized network, and Fuzzy similarity matrix, could be used to partition a case library for similarity assessment (Jeng and Liang 1995, Watson 1997). These techniques generally use three indexes as a measure of the clustering performance: intercluster similarity, intracluster similarity, and the total number of clusters.

CBR can be effectively combined with other intelligent methods (ontologies, agents, rule-based reasoning) (Boshnakov, Boishina, and Hadjiiski 2011; Chan 2005; Hadjiski and Boishina 2010; Karray and De Silva 2004; Konar 2005; Pal, Dillon, and Yeung 2001; Prentzas and Hatzilygeroudis 2009). Two main trends for CBR combinations can be discerned. The first trend involves embedded approaches in which the primary intelligent method (usually CBR) embeds one or more other intelligent methods to assist its internal online and offline tasks. The second combination trend involves approaches in which the problem solving process can be decomposed into tasks for which different representation formalisms are required or available. In such situations, a CBR system as a whole (with its possible internal modules) is integrated "externally" with other intelligent systems to create an improved overall system (Aamodt and Plaza 1994; Chan 2005; Mitra and Basak 2005).

3. COMBINING FUZZY LOGIC TECHNIQUES AND CASE-BASED REASONING

Unlike conventional sets, fuzzy sets include all elements of a universal set but with different membership values in the interval [0, 1] (Karray and De Silva 2004; Konar 2005; Zadeh 1983). Fuzzy set theory has been applied successfully to computing with words or the matching of linguistic terms for reasoning. In the context of CBR, using quantitative features to create indexes involves conversion of numerical features into qualitative terms for indexing and retrieval. Moreover, one of the major issues in fuzzy set theory is measuring similarities in order to design robust systems. Another application of Fuzzy Logic (FL) to CBR is the use of fuzzy production rules to guide case adaptations. Fuzzy production rules may be discovered by examining a case library and associating the similarity between problem and solution features of cases (Prentzas and Hatzilygeroudis 2009).

FL is enabled through:

- *Case Representation*: Approximate or incomplete knowledge of case attributes can be represented by fuzzy intervals or sets, which in turn can be associated with linguistic terms stored as text.
- *Case Retrieval*: A concept of "neighborhood" or partial match has been implemented for numeric attributes. Non-numeric attributes (such as fuzzy linguistic terms) can either be handled by adjusting the distance calculation or by extending the current components.
- *Case Similarity*: Distance calculation is highly customizable. A fuzzy similarity based on the Generalized Bell function exists. Alternative fuzzy similarity measures can also be coded and used.

A fuzzy set ***A*** is a collection of objects drawn from the universal set ***U***, with a continuum of grades of membership where each object ***x*** ($\boldsymbol{x} \in \boldsymbol{U}$) is assigned a membership value that represents the degree to which ***x*** fits the imprecise concept represented by the set ***A*** (Karray and De Silva 2004; Konar 2005; Zadeh 1983). Formally, it is written as follows:

$$\boldsymbol{A} = \{\boldsymbol{\mu_A(x)/x}, \boldsymbol{x} \in \boldsymbol{U}\}, \tag{2}$$

where the membership function $\boldsymbol{\mu_A}(\boldsymbol{x})$ is defined as

$$\boldsymbol{\mu_A}: \boldsymbol{U} \rightarrow [0, 1]. \tag{3}$$

The number of linguistic terms for each attribute in a case can be assumed to be five, usually referred to as negative big, negative small, zero, positive small, and positive big, or NB, NS, ZE, PS, and PB. Their membership functions can be expressed in many forms, such as in trapezoidal, Gaussian, and generalized bell shapes (Karray and De Silva 2004; Konar 2005; Zadeh 1983). The most commonly used membership functions are triangular in shape, as shown in Figure 2.

Fuzzy linguistic representation of patterns: Let a pattern (object) $\hat{e}$ be represented by ***n*** numeric features (attributes) (i.e., $\hat{e} = [F_1, F_2, ..., F_n]$). Each feature is described in terms of its fuzzy membership values, corresponding to three linguistic fuzzy sets: low (L), medium (M), and high (H) (Figure 3). Thus, an ***n***-dimensional pattern vector is represented as a ***3n***-dimensional vector (Karray and De Silva 2004; Konar 2005; Zadeh 1983).

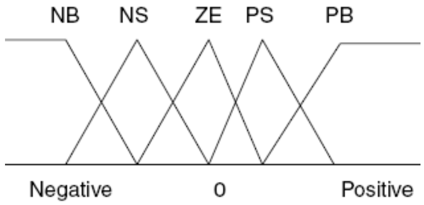


Figure 2: Fuzzy Membership Functions

A vector of triplets is used to represent a case. The elements of this vector describe the property, its importance (weight) within this case, and its value:

$$e = \{e_1, e_2, ..., e_k\} \quad e_i = (a_i, w_i, v_i) \tag{4}$$

Concept of fuzzy sets in measuring similarity: one of the features of cases in a CBR system may be described by such linguistic terms as low, medium, and high (Karray and De Silva 2004; Konar 2005; Zadeh 1983). Then for implementing the process of case matching and retrieval, one needs to define an appropriate metric of similarity. The traditional definition of similarity is obviously not valid and at least not effective to deal with this difficulty. Here the concept of fuzzy set provides a good tool to handle the problem in a natural way.

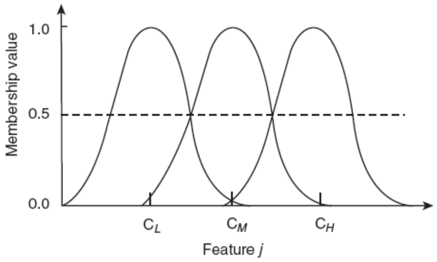


Figure 3: Bell Fuzzy Membership Functions for Linguistic Property Sets

In fuzzy set theory, the linguistic term as a fuzzy number, which is a type of fuzzy set, may be considered (Jeng and Liang 1995; Pal, Dillon, and Yeung 2001; Prentzas and Hatzilygeroudis 2009; Watson 1997). Then a membership function is determined with respect to the given linguistic term. When a real value of the feature of a given problem is input, the corresponding values of membership to different linguistic terms are obtained through the membership functions.

That is, after approximate matching, the real-valued features are transformed to linguistic features. Then, depending on the problem, to select the best-matching case or the best set of cases, one needs to define some similarity measures and algorithms for computing fuzzy similarity. Before we define them, we provide a mathematical framework that signifies the relevance of fuzzy similarity in case matching.

Inference based on a fuzzy case rule can be divided into two stages (Jeng and Liang 1995; Pal, Dillon, and Yeung 2001; Prentzas and Hatzilygeroudis 2009). In the first stage, an inference is based on how well the facts of a new case correspond to the elements associated with a (precedent) case rule. This is judged using a criterion yes or no, which is evaluated according to the degree of fuzzy membership between the facts and elements. In the second stage, the inference from the precedent case to the new case is drawn, and this is directed by the similarity between the cases.

The conclusions obtained from both these stages are compared with that of the precedent case. If they are identical with the conclusion of the precedent case, the new case has the same result as the precedent. If they are not identical with that conclusion, a decision concerning the new case cannot be supported by the precedent. When a judgment on the correspondence between the facts of the new case and the elements of a (precedent) case rule (that is represented by the fuzzy membership function) is made, a yes or no judgment is unnecessary for inference by case rule. Accordingly, the center of gravity of the fuzzy membership function of these cases can be defined as

$$\mathrm{CG}(A_i) = \frac{\int_{c_1}^{c_2} x\mu_{A_i}(x)\,dx}{\int_{c_1}^{c_2} \mu_{A_i}(x)\,dx} \tag{5}$$

where $\boldsymbol{U} = [\boldsymbol{c_1}, \boldsymbol{c_2}]$, $\boldsymbol{A_i}$ is the fuzzy set that describes the judgment on the correspondence between the elements of a case rule ($\boldsymbol{i}$) and the facts of a new case. $\boldsymbol{\mu_{Ai}}$ is the membership function of $\boldsymbol{Ai}$. $\mathbf{CG}(\boldsymbol{A_i})$ lies in [0, 1]. Considering 0.5 as the threshold, if the value of the center of gravity is greater (or less) than 0.5, the judgment is yes (or no).

The *distance* between two centers of gravity, $|\mathbf{CG}(\boldsymbol{A}) - \mathbf{CG}(\boldsymbol{B})|$, is used to describe the degree of similarity. To satisfy the conditions of similarity relations, the degree of similarity $\mathbf{SM}(\boldsymbol{A}, \boldsymbol{B})$ is calculated using

$$\mathbf{SM}(\boldsymbol{A}, \boldsymbol{B}) = 1 - |\boldsymbol{CG}(\boldsymbol{A}) - \boldsymbol{CG}(\boldsymbol{B})| \tag{6}$$

The conceptual similarity of an elemental item within the cases is assessed as

$$\Delta\mathrm{SM} = e^{-\beta\Delta d^2} \tag{7}$$

where $\boldsymbol{\beta}$ ($\boldsymbol{\beta} > 0$) denotes amendment accuracy, which should be fixed beforehand. The formulation of the provision acceptance depends on the elemental item that belongs to this issue $\boldsymbol{j}$. The value $\boldsymbol{\Delta d}$ is the distance between the relevant items from the two cases ($\boldsymbol{e_p}$, $\boldsymbol{e_q}$), and it can be computed as

$$\Delta d = \left|\mathrm{CG}(e_p) - \mathrm{CG}(e_q)\right| \tag{8}$$

The similarity of the issue $\boldsymbol{j}$ is assessed using the similarity of the associated elemental items as

$$\begin{aligned}\mathrm{SM_j} &= \min\{\Delta\mathrm{SM}_1, \Delta\mathrm{SM}_2, \ldots, \Delta\mathrm{SM}_i, \ldots, \Delta\mathrm{SM}_n\},\\ &\Delta\mathrm{SM}_i \in [0,1], n \in N\end{aligned} \tag{9}$$

where $\boldsymbol{n}$ is the number of elemental items that belong to the issue $\boldsymbol{j}$. As a general rule, more than one issue can be compared between two cases. The algorithm applied when there is more than one relevant issue should also be considered.

In this situation, a weight $\boldsymbol{w_i}$ is introduced into the case-based retrieval. The average similarity is then weighted. It is calculated as

$$\overline{\mathrm{SM}} = \frac{\sum (w_i \mathrm{SM}_i)}{\sum w_i} \tag{10}$$

Let each frame of a precedent case and a new case be represented as follows:

$$\text{Precedent}: \quad A = \{A_i\}_{i=1}^{n}$$
$$\text{New case}: \quad B = \{B_i\}_{i=1}^{n}$$

where ***A*** is the frame that represents the precedent, ***B*** the frame that represents the new case, $\boldsymbol{A_i}$ the fuzzy set that describes the judgment concerning the elements of the precedent case rule, and ***n*** the quantity of slots in a frame.

The similarity assessment is performed as follows: Let the membership functions of $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$ be $\boldsymbol{\mu_{Ai}}$ and $\boldsymbol{\mu_{Bi}}$, respectively. The center of gravity of $\boldsymbol{A_i}$ i and $\boldsymbol{B_i}$ can be computed using equation (5). Let $\mathbf{SM}(\boldsymbol{A_i}, \boldsymbol{B_i})$ be the degree of similarity between $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$. Then the degree of similarity between ***A*** and ***B*** can be obtained from

$$\mathrm{SM}(A, B) = \min(\mathrm{SM}(A_1, B_1), \ldots, \mathrm{SM}(A_n, B_n)) \quad (11)$$

If the degree of similarity is greater than the threshold (which was determined in advance), the conclusion is that frame ***B*** is the same as frame ***A.*** For example, if there is a conclusion that ''the proposal is sufficiently definite'' in a precedent, the conclusion of new case is also ''the proposal is sufficiently definite.'' If the degree of similarity is less than the given threshold, the conclusion is that frame ***B*** cannot arrive at the same conclusion as that of ***A***. This does not necessarily mean that the new case has an opposite conclusion to the precedent. Perhaps it is possible to reach the same conclusion using another precedent.

There are several methods for computing the similarity between cases (Jeng and Liang 1995; Pal, Dillon, and Yeung 2001; Prentzas and Hatzilygeroudis 2009):

- Numeric combination of feature vectors (properties, attributes), representing the known cases, using different combination rules.
- Similarity of structured representations, in which each case is represented as a structure, such as a directed graph, and thus the similarity measure takes into account the structure of the different attributes of the case and not only the attribute value.
- Goal-driven similarity assessment, in which the attributes of the cases that are to be compared with those of a new case depend on the goal sought. This means that some attributes of a case are not important in the light of a certain goal and thus should not be taken into account in the similarity calculation.
- Rule-based similarity assessment, in which the cases in the case base (CB) are used to create a set of rules on the feature vector of the cases. This rule set is then used to compare the cases in the CB and to solve the new case.
- Aggregation of the foregoing methods according to application-specific hierarchies.

The similarity measures are used for case matching and retrieval through classification or clustering of cases under supervised and unsupervised modes, respectively. In general, in the process of case matching and retrieval, the searching space is the entire case base, which not only makes the task costly and inefficient, but also sometimes leads to poor performance.

To address such a problem, many classification and clustering algorithms are applied before selection of the most similar case or cases. After the cases are partitioned into several sub-clusters, the task of case matching and retrieval then boils down to matching the new case with one of the several sub-clusters, and finally, the desired number of similar cases can be obtained. Thus, various classification/clustering algorithms, such as fuzzy ID3 and fuzzy c-means, play an important role in this process (Jeng and Liang 1995; Pal, Dillon, and Yeung 2001; Prentzas and Hatzilygeroudis 2009).

4. IMPLEMENTATION OF FUZZY LOGIC TECHNIQUES IN CBR TOOL

On the base of previous comparative analysis in (Atanassov and Antonov 2012) the above described fuzzy logic techniques are implemented in software platform ***myCBR***. It is one of the most popular CBR software platforms with certain capabilities and limitations. The platform has open source code written on ***Java*** and can be easily modified by the users depending on the purpose (Atanassov and Antonov 2012). The usage of ***myCBR*** could minimize the efforts to create specific customer CBR applications. For its normal use, without modifying the source code, no programming skills are required, but expertise in a specific CBR-developed applications. The framework ***myCBR*** supports description of cases with various attributes: numeric, character and string, logical, class type, etc. The templates of the cases are generated as classes or subclasses with a number of attributes, called slots.

The CBR cases are objects of the class described by its attributes. Each attribute can participate in the class with its value and a weight that determines the significance of the attribute in relation to others. An attribute weight of zero (0) is not considered when searching the case-base DB.

In ***myCBR*** the opportunity to edit the similarity functions (SFs) on class level (global SFs) and on an attribute level (local SFs) are given. At the class level the SFs are: weighted sum, Euclidean difference, maximum or minimum. On attribute level the SFs can be modified through the GUI and they can be symmetrical, asymmetrical, step-type or smooth step-type, linear or polynomial.

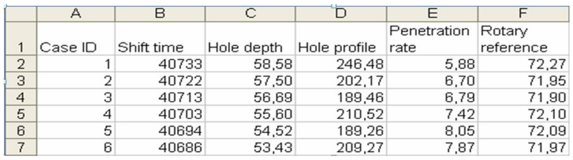
With regard to maintenance the CBR R^4 cycle phases ***myCBR*** supports only *Retrieve* and *Retain*. During the Retrieve phase all precedents are extracted. They are presented sorted by degree of similarity based on the chosen global SFs. The Query to the case-base DB could be done on the basis of all or part of the attributes, describing the case. Fuzzy similarity measures based on usage of membership functions of the defined linguistic variables are implemented.

On Retain phase ***myCBR*** allows to save the Query as a new case, also to use an old case as a basis for new Query. ***myCBR*** is entirely based on GUI, providing a ready-windows templates and forms for defining classes, attributes, SFs, queries to the case-base DB, visualization of found results and more.

myCBR does not work with external DB. It stores the cases in text file or in XML file. Because ***myCBR*** can not support the case indexation and clusterization an additional module based on fuzzy logic has been developed and included in the platform to solve the tasks of diagnosis problems.

To validate the capabilities of the developed CBR tool it is applied for solving diagnostics problem of drill machine in mine industry (Atanassov and Antonov 2012). The description of case base dataset is given in Table 1. Columns A, B, C and D in Table 1 are the problem attributes of the cases and columns E and F – the decision attributes. All data is processed using the methodology described in (Tenchev and Kondev 2006).

Table 1: Case-Base Data Set



	A	B	C	D	E	F
1	Case ID	Shift time	Hole depth	Hole profile	Penetration rate	Rotary reference
2	1	40733	58,58	246,48	5,88	72,27
3	2	40722	57,50	202,17	6,70	71,95
4	3	40713	56,69	189,46	6,79	71,90
5	4	40703	55,60	210,52	7,42	72,10
6	5	40694	54,52	189,26	8,05	72,09
7	6	40686	53,43	209,27	7,87	71,97

Figure 4 shows how Predictive Diagnosis class (case) and its attributes are defined, as well the definition of the type and range of these attributes.

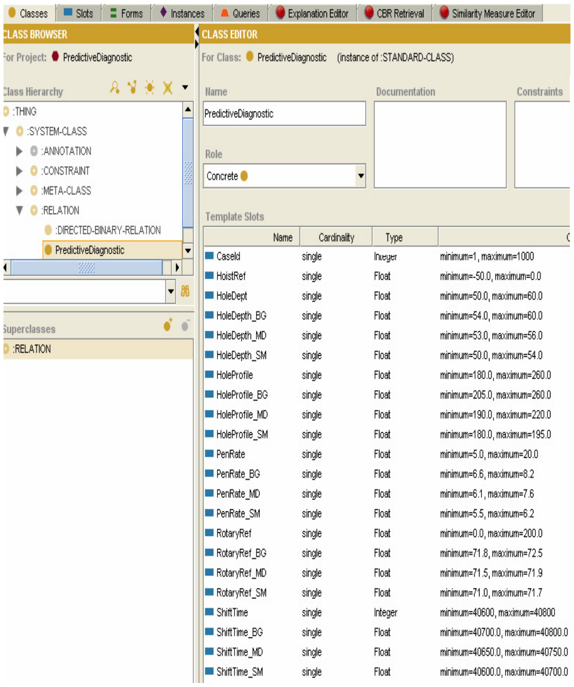
In order to support fuzzy logic three extra attributes related to the defined linguistic variables (Small – SM, Medium – MD and Big – BG) and the corresponding membership functions are presented in *PredistiveDiagnostic* class.

The defined fuzzy membership functions of a selected attribute (*Hole Depth*) are illustrated at Figure 5.

The results of the query to the case-base DB are given in Figure 6. All cases are sorted in ascending way on the base of their proximity to the queried case. In the estimations of the proximity the local and global SF are taken into account.

Figure 7 presents the form used to insert data for each instance of the class in Case Base. It suppresses inserting of values that are out of range, defined for each slot (attribute).

As can be seen from the example the CBR tool has more options for weights definition of the attributes and for selection or modification of similarity functions on attribute and on class levels. This is of great importance for query adjustment and refining to the case base.



For Project: PredictiveDiagnostic

For Class: PredictiveDiagnostic (instance of :STANDARD-CLASS)

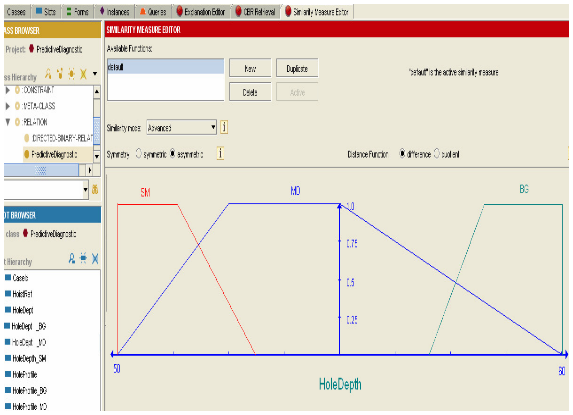
Name: PredictiveDiagnostic

Role: Concrete

Template Slots

Name	Cardinality	Type	
CaseId	single	Integer	minimum=1 , maximum=1000
HoistRef	single	Float	minimum=-50.0, maximum=0.0
HoleDept	single	Float	minimum=50.0, maximum=60.0
HoleDepth_BG	single	Float	minimum=54.0, maximum=60.0
HoleDepth_MD	single	Float	minimum=53.0, maximum=56.0
HoleDepth_SM	single	Float	minimum=50.0, maximum=54.0
HoleProfile	single	Float	minimum=180.0, maximum=260.0
HoleProfile_BG	single	Float	minimum=205.0, maximum=260.0
HoleProfile_MD	single	Float	minimum=190.0, maximum=220.0
HoleProfile_SM	single	Float	minimum=180.0, maximum=195.0
PenRate	single	Float	minimum=5.0, maximum=20.0
PenRate_BG	single	Float	minimum=6.6, maximum=8.2
PenRate_MD	single	Float	minimum=6.1, maximum=7.6
PenRate_SM	single	Float	minimum=5.5, maximum=6.2
RotaryRef	single	Float	minimum=0.0, maximum=200.0
RotaryRef_BG	single	Float	minimum=71.8, maximum=72.5
RotaryRef_MD	single	Float	minimum=71.5, maximum=71.9
RotaryRef_SM	single	Float	minimum=71.0, maximum=71.7
ShiftTime	single	Integer	minimum=40600, maximum=40800
ShiftTime_BG	single	Float	minimum=40700.0, maximum=40800.0
ShiftTime_MD	single	Float	minimum=40650.0, maximum=40750.0
ShiftTime_SM	single	Float	minimum=40600.0, maximum=40700.0

Figure 4: Predictive Diagnostic Class with its Attributes



Project: PredictiveDiagnostic

class PredictiveDiagnostic

Available Functions: default

"default" is the active similarity measure

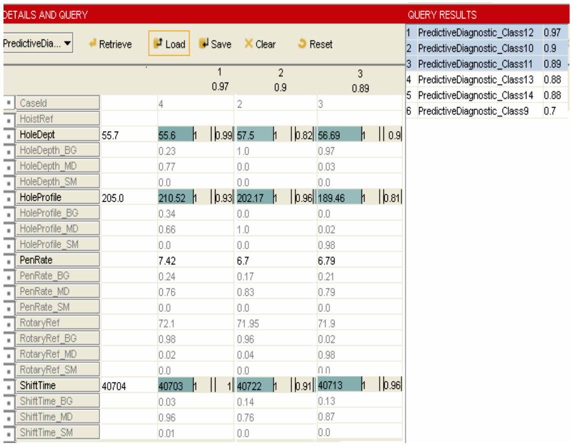
Similarity mode: Advanced

Symmetry: ○ symmetric ● asymmetric

Distance Function: ● difference ○ quotient

Figure 5: Fuzzy Membership Functions of Attribute *Hole Depth*

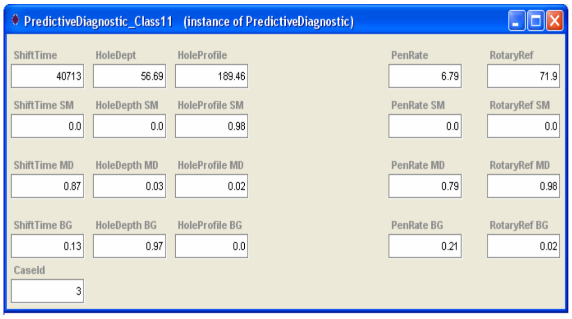
jCOLIBRI can be used as a basis for complex CBR applications development with full CBR R^4 cycle, using various data bases. Development of this kind of applications however requires excellent programmer knowledge, time for requirements definition, development of software architecture, complicated graphical user interface, data base configuration and time for implementation, test, adjustment and verification (Atanassov and Antonov 2012). Based on the examples, given above, it is obvious, that ***myCBR*** interface overmatches ***jCOLIBRI***`s and gives more options for weights and SF type modification of attributes and cases. This is of great importance for query adjustment and refining to the case base.



		1 0.97	2 0.9	3 0.89
CaseId		4	2	3
HoistRef				
HoleDept	55.7	55.6 1 0.99	57.5 1 0.82	56.69 1 0.9
HoleDepth_BG		0.23	1.0	0.97
HoleDepth_MD		0.77	0.0	0.03
HoleDepth_SM		0.0	0.0	0.0
HoleProfile	205.0	210.52 1 0.93	202.17 1 0.96	189.46 1 0.81
HoleProfile_BG		0.34	0.0	0.0
HoleProfile_MD		0.66	1.0	0.02
HoleProfile_SM		0.0	0.0	0.98
PenRate		7.42	6.7	6.79
PenRate_BG		0.24	0.17	0.21
PenRate_MD		0.76	0.83	0.79
PenRate_SM		0.0	0.0	0.0
RotaryRef		72.1	71.95	71.9
RotaryRef_BG		0.98	0.96	0.02
RotaryRef_MD		0.02	0.04	0.98
RotaryRef_SM		0.0	0.0	0.0
ShiftTime	40704	40703 1 1	40722 1 0.91	40713 1 0.96
ShiftTime_BG		0.03	0.14	0.13
ShiftTime_MD		0.96	0.76	0.87
ShiftTime_SM		0.01	0.0	0.0

1	PredictiveDiagnostic_Class12	0.97
2	PredictiveDiagnostic_Class10	0.9
3	PredictiveDiagnostic_Class11	0.89
4	PredictiveDiagnostic_Class13	0.88
5	PredictiveDiagnostic_Class14	0.88
6	PredictiveDiagnostic_Class9	0.7

Figure 6: Retrieval Results Sorted by Their Local (Left Side) and Global (Right Side) Similarity


PredictiveDiagnostic_Class11 (instance of PredictiveDiagnostic)

ShiftTime	HoleDept	HoleProfile	PenRate	RotaryRef
40713	56.69	189.46	6.79	71.9
ShiftTime SM	HoleDepth SM	HoleProfile SM	PenRate SM	RotaryRef SM
0.0	0.0	0.98	0.0	0.0
ShiftTime MD	HoleDepth MD	HoleProfile MD	PenRate MD	RotaryRef MD
0.87	0.03	0.02	0.79	0.98
ShiftTime BG	HoleDepth BG	HoleProfile BG	PenRate BG	RotaryRef BG
0.13	0.97	0.0	0.21	0.02
CaseId				
3				

Figure 7: Form for Attributes Values Input in a Case Base

For development of our CBR tool some suggestions have been taken into account:

- *Suggestions for **myCBR** usage* – the Java code of ***myCBR*** to be expanded with additional module to work with external data bases as proposed above. It can be intended to read external data base and to convert all cases in the format used in ***myCBR***, as well to ensure back-way conversion.
- *Suggestion for development of new own CBR software application* – which can support groups of data bases – one for the cases and the solutions, and other one – for on-line data of the diagnosis object or system. Also the development of specific software intended for input/output, for case retrieval from case base DB, for filtration, adaptation, etc. is recommended. The advantages of data bases are that they can keep complex cases in tables with relations to other tables with graphical and/or picture information or relations to tables with lectures, that contain decisions and recommendations for solving specific problems.

5. CONCLUSIONS AND FUTURE WORK

In this paper an advanced tool based on CBR and fuzzy logic techniques has been proposed. It can be successfully applied to solve the complex tasks of predictive diagnosis and maintenance.

The tool was developed as an extension of available ***myCBR*** software platform. In the work an example of CBR tool application has been presented. This study is in the beginning stage and further research will be in progress in order to carry out the diagnostics problems in real industrial systems.

The further investigations will be carried out on a pellet production plant (Hadjiski, Christova, and Valova 2013). The focus will be at the development of a new method for estimation and prediction a degradation level of most loaded elements in extruding part. The main indexes of the pellets quality are hardness, durability and calorific value, which determine the pellets price. The existing own operating experience and available literature data will allow to create a case base and corresponding rules for selecting the matrices in a specific combination of parameters of feed extrusion dried biomass. Under consideration will be an aggregation of the various partial optimizing potentials. The usefulness of the condition based maintenance of pellet mill with implementation of CBR and fuzzy logic based procedures for current state inference of the rollers and Remaining Useful Life (RUL) prediction of the pair die/rollers will be discussed.

As future step the movement from our own CBR tool to available business intelligence platform ***MicroStrategy*** is planned (Tomova, Atanassov, and Boshnakov 2012). This way the possibility to work with many databases and definition of own CBR similarity and fuzzy membership functions can be realized that will improve the capability of more precise data analysis and prognostics maintenance.

ACKNOWLEDGMENTS

This work has been supported by the National Science Fund of Bulgaria, Ministry of Education, Youth and Science under the project "Diagnostic and Risk Assessment Predictive Asset Maintenance" No. DVU-10-0267/10.

REFERENCES

Aamodt, A., Plaza, E., 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, *Artificial Intelligence Communications*, 7 (1), 39–59.

Aha, D. W., 2006. Advances in conversational case-based reasoning, *The Knowledge Engineering Review*, 20 (3), 247–254.

Atanassov, A., Antonov, L., 2012. Comparative Analysis of the Capacities of Case Based Reasoning Software Frameworks jCOLIBRI and myCBR, *JCTM*, 47 (1), 83–90.

Avramenko, Y., Kraslawski, A., 2006. Similarity Concept for Case-Based Design in Process

Engineering, *Computers and Chemical Engineering*, 30, 548–557.

Boshnakov, K., Boishina, V., Hadjiiski, M., 2011. Multiagent fault-tolerant supervising control of wastewater treatment plants for wastewater, *Int. Conf. "Automatics & Informatics'11"*, October 3-7 2011, Sofia, Bulgaria.

Chan, F. T. S. 2005. Application of a Hybrid Case-Based Reasoning Approach in Electroplating Industry, *Expert Systems with Applications*, 29 (1), 121–130.

Hadjiski, M., Boishina, V., 2010. Enhancing Functionality of Complex Plant Hybrid Control System Using Case-Based Reasoning, *5th IEEE International Conference on Intelligent Systems (IS)*, 7-9 July 2010, London, 25–30.

Hadjiski, M., Christova, N., Valova, M., 2013. Incremental Re-design of Control System of Small-Scale Pellet Production Plant, *IFAC SWIIS 2013*, Prishtina, Kosovo.

Jeng, B. C., Liang, T. P., 1995. Fuzzy indexing and retrieval in case-based systems, *Expert Systems with Applications*, 8(1), 135–142.

Karray, F. O., De Silva, C., 2004. *Soft Computing and Intelligent Systems Design: Theory, Tools and Applications*, Addison-Wesley.

Kolodner, J. L., 1993. *Case-Based Reasoning*, Morgan Kaufmann, San Francisco.

Konar, A., 2005. *Computational Intelligence: Principles, Techniques and Applications*, Springer, New York.

Mitra, R., Basak, J., 2005. Methods of case adaptation: A survey, *International Journal of Intelligent Systems*, 20 (6), 627–645.

Pal, S. K., Dillon, T. S., Yeung, D. S. (eds.), 2001. *Soft Computing in Case-Based Reasoning*, Springer-Verlag, London.

Prentzas, J., Hatzilygeroudis, I., 2009. Combination of case-based reasoning with other intelligent methods, *International Journal of Hybrid Intelligent Systems*, 6 (4), 189–209.

Tenchev, D., Kondev, G., 2006. *Total Maintenance of Equipment*, MJ Publishing Technical University of Sofia, Sofia.

Tomova, F., Atanassov, A., Boshnakov, K., 2012. The opportunities of software platform MicroStrategy for intelligent data processing, *Proc. of the International Conference Automatics & Informatics'12*, Sofia.

Watson, I., 1997. *Applying Case-Based Reasoning: Techniques or Enterprise Systems*, Morgan Kaufmann, San Francisco.

Watson, I., 1999. Case-based reasoning is a methodology, not a technology, *Knowledge-Based Systems*, 12, 303–308.

Yang, C., Farley, B., Orchar, B., 2008. Automated case creation and management for diagnostic CBR systems, *Applied Intelligence*, 17–28.

Zadeh, L. A., 1983. The role of fuzzy logic in the management of uncertainty in an expert system, *Fuzzy Sets and Systems*, 11, 199–227.

AUTHORS BIOGRAPHY

NIKOLINKA G. CHRISTOVA was born in Pazardjik, Bulgaria. She received MS degree in Industrial Automation and Ph.D. on Methods and Algorithms for Data Reconciliation and Diagnosis of Measurement Errors in Technological Systems from the University of Chemical Technology and Metallurgy (UCTM) – Sofia, in 1982 and 1999 respectively. She obtained European Master Degree in Environmental Protection and Sustainable Development at the University of Chemical Technology and Metallurgy – Sofia in collaboration with Universities from UK and Belgium in 2011. She received Course Certificates on "The Effective Manager", "Managing Customer & Client Relations", "Accounting for Managers", and Professional Certificate in Management from the Open University, Business School, Sofia, in 1996. Now Dr. N. Christova has a position of Associate Professor at the Department of Automation of Industry, UCTM – Sofia and gives lectures on Intelligent Control Systems, Industrial Management, Quality Control, and Integrated Control Systems. Her main research interests are in the field of Computerized Integrated Industrial Control and Environmental Management, Fuzzy Logic and Neural Network Applications to Simulation, Control and Fault Diagnosis in Industrial Systems, Decision Support Systems for Business Management, Energy Efficiency and Renewable Energy Sources.

ATANAS V. ATANASSOV was born in Bourgas, Bulgaria. He graduated MSc. Degree Automation and Telecommunications from Technical University – Sofia in 1985 and becomes Ph.D. on Parallel Control of Real-Time Processes in 2009. From 1985 till now he is working at Computer Science (CS) Department at University of Chemical Technology and Metallurgy (UCTM) – Sofia. Currently he is Assoc. Prof. and head of CS at UCTM and gives lectures in Informatics and Microprocessor Systems. His scientific interests work are oriented to Programming Languages, Robot Control, Parallel Control Systems, Real-Time Operating Systems, Postal Automation Systems, Automatic Number Plate Recognition Systems, Case-Based Reasoning Systems intended to Predictive Diagnostics and Maintenance of Technological Systems, Learning and Test Systems. He lead or took part in lots of projects with industry and Hi-Tech companies as Siemens Logistics (Germany), FedEx-Ground (USA Minnesota), Die Post (Swiss), Knowledge Support Systems (UK), Logosol (USA California) and with many Bulgarian ministries and firms.

A MATHEMATICAL MODEL FOR DETERMINING TIMETABLES THAT MINIMIZES THE NUMBER OF STUDENTS WITH CONFLICTING SCHEDULES

Anibal Tavares de Azevedo[a], Alexander Kameyama[b], Joni A. Amorim[c] , Per M. Gustavsson[d]

[a]Universidade Estadual de Campinas – UNICAMP, Brazil
[b]Universidade Estadual Paulista – UNESP, Brazil
[c]Högskolan i Skövde – HiS, Sweden
[d]Saab Group, Sweden

[a]anibal.azevedo@fca.unicamp.br / atanibal@gmail.com, [b]alexkameyama@hotmail.com, [c]joni.amorim@his.se / joni.amorim@gmail.com, [d]per.m.gustavsson@saabgroup.com

ABSTRACT

With the increasing complexity of educational initiatives, several challenges arise as to appropriately allocate human and material resources or how to select alternative investments within a portfolio. Due to the complexity, solutions based on intuition are risky, which leads to a search for less intuitive and more reliable ways of solving educational problems. An engineering approach to the problem may lead to operations research mathematical modeling as a way to help on finding the solution for timetabling. In this case, a timetable is a schedule or, more precisely, a list or table of events arranged according to the time when they take place. This text features new and useful software to optimize the use of human resources. The software is based on a mathematical model for determining the timetable while minimizing the number of students with conflicting schedules.

Keywords: mathematical modeling, optimization, scheduling, timetable.

1. INTRODUCTION

Timetable is an event table used to specify who will participate, who will be held where and when such an event occurs. Generally, build an adequate timetable is not an easy task since some kind of constraints should be fulfilled in a manner there is no conflict in the schedule. That is it; Timetable is so hard that find a feasible solution is even difficult.

One category of Timetable problem is the Educational Timetabling. This problem can be classified in two sub-categories: exam and course timetabling (Al-Yakoob, Sherali, and Al-Jazzaf, 2010; Carter and Laporte, 1998).

To propose a course timetabling of a university is necessary to relate the different variables and interests related with students, teachers and classrooms. Some of examples of them are prerequisites established by the university for each discipline, individual preferences of teachers/students for certain disciplines to be taught/routed and, most often, the downtime between classes should be avoided. Added to this, there are risks of errors in the definition of the grids and these may be detected only when the classes have already begun (Al-Yakoob, Sherali, and Al-Jazzaf, 2010).

According to (Carter and Laporte, 1998) the course timetabling problem can be divided into five subproblems: teacher assignment, class-teacher timetabling, course scheduling, student scheduling and classroom assignment. The student scheduling problem only allocates students to courses without using the information about the courses allocation to time periods. Student scheduling problem often uses a given allocation of teachers and courses (Gunawan, Ng, and Poh, 2013).

This work will address the Problem of Assignment of Classes to Students (PACS) that combines student scheduling and courses allocation problem simultaneously within a university. A special feature had been considered in the model in order to consider student incompatibility to attend for more than one class. That is, as general purpose, students must attend the classes with the least possible incompatibilities and different from the one considered in recent literature (Al-Yakoob and Sherali, 2013).

PACS is part of the set of combinatorial optimization problems (Schaerf, 1999; Willenmen, 2002) which justify the development of heuristics and meta-heuristics. Another contribution of this work is to develop and apply an approach that enables solving of real large-scale problems.

This paper is structured as follows. Section 2 presents the mathematical model of PACS, while section 3 presents the proposed solution method and the developed software. Section 5 addresses conclusions and future work.

In this research report, in the section devoted to materials and methods, we present a mathematical model for determining the timetable while minimizing the number of students with conflicting schedules. This model is useful for determining the number of prospective students to be met for the subject matters of a university course that employs the system of credits.

Then, in the section with results, it is both presented and discussed software with features that aim to simplify the use of the method, especially when using an interface that allows a user without basis in advanced mathematics to enter the relevant information and perform the optimization of the use of resources. Finally, a section presents the main conclusions and suggests future work.

2. MATHEMATICAL MODEL

To determine the possible number of students to be served by the subjects of a college course that used the credit system, the mathematical model given below was developed.

Table 1: Materials in a course that employs the system of credits.

Number	Subject matter	Prerequisites
1	Calculus I	-
2	Analytic Geometry	-
3	Calculus II	Calculus I and Analytic Geometry
4	Calculus III	Calculus II

A constraint that must be met is to check whether a given student has the prerequisites required to perform a given subject matter. In other words, there are subject matters that can only be routed after verifying that the student has been approved in another one. For example, assume a course whose contents are described in Table 1 and whose prerequisites are given in Figure 1. Subject matter 3 can only be followed if the student has attended and successfully passed the subject matters 1 and 2.

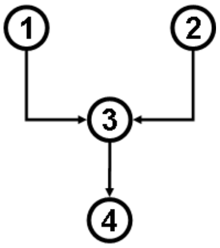


Figure 1: Indication of the relationship between prerequisites between subject matters of Table I.

From Figure 1 it is possible to determine whether a given student may or may not attend a subject matter based on the approval history. For example, if a student has been approved only in subject matter 1, then this one cannot attend the subject matters 3 or 4. The same statement can be made if the student has been approved only in the subject matter 2. It is important to note that Figure 1 is a graph where the nodes represent the materials and the relationship arcs of prerequisites between them. Figure 1 also can be used to show the evolution of the student during the course in each time period as given in Figure 2.

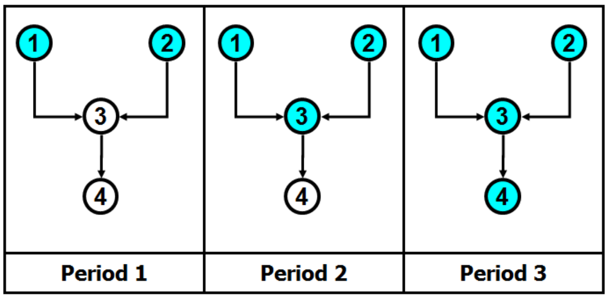


Figure 2: Illustration of a student's progress throughout the course considering data shown in Table 1.

Figure 2 illustrates the evolution of a student who has been approved in the subject matters 1 and 2, which allowed the same in Period 2 to attend the subject matter 3 and finally with the approval of the latter to attend subject matter 4 in period 3.

In order to facilitate the storage of information relating to the approval history of a variable number of students, y_{jkt}, a strategy that can be employed follows. The variable y_{jkt} is equal to 1 if the student *k* is approved in the subject matter *j* at the end of period *t*; y_{jkt} = 0, otherwise. The example of Figure 2 for one student corresponds to the values given in Figure 3.

	Subject matter 1	Subject matter 2	Subject matter 3	Subject matter 4
Period 1	1	1	0	0
	y_{111}	y_{211}	y_{311}	y_{411}
Period 2	1	1	1	0
	y_{112}	y_{212}	y_{312}	y_{412}
Period 3	1	1	1	1
	y_{113}	y_{213}	y_{313}	y_{413}

Figure 3: Representation of the evolution of a student, given in Figure 2, in the variable y_{jkt}.

From the variable y_{jkt} it is possible to express in mathematical terms whether a student may attend a subject matter based on his/her history of approvals and prerequisites between the subject matters. To do so, it must be also defined a new variable x_{ikt} indicating whether the student *k* in period *t* can attend the subject matter *i*. A matrix M_{ij} may be defined such that if M_{ij} = 1, then, the subject matter *j* is prerequisite for the subject matter *i*.

It is noticeable that the array of prerequisites is not only invariant with respect to the period, but that it is also supposedly unique. If there are students from different courses, then you need it is necessary to consider different matrices according to the courses of each student. The matrix M_{ij} of prerequisites associated with the example of Figure 2 is given in Figure 4.

	Subject matter 1	Subject matter 2	Subject matter 3	Subject matter 4
Subject matter 1	0	0	0	0
	M_{11}	**M_{12}**	**M_{13}**	**M_{14}**
Subject matter 2	0	0	0	0
	M_{21}	**M_{22}**	**M_{23}**	**M_{24}**
Subject matter 3	1	1	0	0
	M_{31}	**M_{32}**	**M_{33}**	**M_{34}**
Subject matter 4	0	0	1	0
	M_{41}	**M_{42}**	**M_{43}**	**M_{44}**

Figure 4: Matrix prerequisites associated with the example in Figure 2.

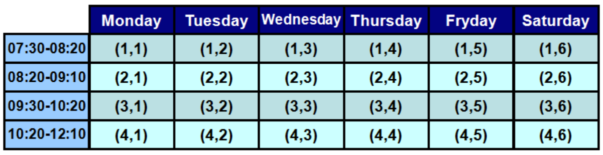
Notice that the third row of Figure 4 shows that, for the subject matter 3 to happen, it is necessary to have an approval in the subject matters 1 (M_{31} = 1) and 2 (M_{32} = 1). The value x_{ikt} can be obtained from y_{jkt} and M_{ij} by means of (1).

$$x_{ikt} = \prod_{j=1}^{J} 1 - \left| M_{ij} - y_{jkt} \right| \qquad (1)$$

An additional constraint on the variable y_{jkt} is that this should be such that after a student *k* be approved in a matter *j* in period *t* this approval shall be considered in subsequent periods. This constraint can be represented by (2).

$$y_{jkt} \geq y_{jk(t+1)} \qquad (2)$$

In addition to (1) and (2) it is necessary to consider, too, that every subject matter should be offered according to a given workload. Therefore, it is necessary to set a timetable such that the spaces in the grid, called slots, indicate whether in a given day and time a subject matter will be given. In order to facilitate the appropriate reference to these spaces in the grid hours, the slots are associated with a number pair *(r, c)*, where *r* indicates the time interval and *c* indicates the day as given in Figure 5.



	Monday	Tuesday	Wednesday	Thursday	Fryday	Saturday
07:30-08:20	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
08:20-09:10	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
09:30-10:20	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
10:20-12:10	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)

Figure 5: Numerical correspondence between the pair *(r, c)* and gaps (slots) of the timetable.

To represent that a particular slot *(r,c)* of the timetable is occupied by a subject matter *i* followed by student *k* in a period *t*, the variable h_{ikt}(r,c) with a value of 1 indicates that it is occupied while the value of 0 indicates that the slot is empty. This new variable, however, must meet two constraints: (1) meet the workload of a subject matter, and (2) avoid the conflict between subject matters. The representation of these two constraints, in mathematical terms, is given below.

For constraint 1, which addresses the workload of a subject matter, we have that a particular subject matter *i* should answer a weekly workload CH_i. Mathematically, this is given by (3).

$$\sum_{r=1}^{R}\sum_{c=1}^{C} h_{ikt}(r,c) = CH_i \qquad (3)$$

For the second constraint on the conflict zone between subject matters, we have a slot *(r, c)* occupied by a subject matter *i* followed by a student *k* in a period *t* that cannot be shared by other subject matter. Otherwise, there will be a conflict of time between subjects and the student must choose to take only one of them. In mathematical terms, this restriction is represented by (4).

$$\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I} h_{ikt}(r,c) x_{ikt} \leq 1 \qquad (4)$$

Finally, the objective function is such that it should minimize the number of students that despite the prerequisite to study a subject matter *i* cannot do so because the slots that such subject matter *i* occupies in the timetable are conflicting with one or more subject matters. A possible objective function is one such that it only counts the number of slots in which mismatch occurs between the *I* subject matters for all *K* students in all *T* periods as given by (5).

$$Min \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=i+1}^{J} \begin{pmatrix} h_{ikt}(r,c) x_{ikt} \\ \bullet\, h_{jkt}(r,c) x_{jkt} \bullet M_j \end{pmatrix} \qquad (5)$$

The complete mathematical model is given by (6). Note that the model given by (6) is a nonlinear, integer and stochastic problem. The stochasticity arises from the fact that it is not possible to know in advance the approval history of a student, ie, the value of y_{jkt}. To this end, there are two possible solutions: (a) assuming that the value y_{jkt} is known only to the period *t* and that the optimization process is performed only for the period immediately follows, or period *(t+1)*, and (b) assuming that the best estimate provided by a predictor or a process of random generation based on the history of previous students is employed.

$$Min \quad \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=i+1}^{J}\begin{pmatrix} h_{ikt}(r,c)x_{ikt} \\ \bullet h_{jkt}(r,c)x_{jkt} \\ \bullet M_{j} \end{pmatrix}$$

$$S.a.: \quad x_{ikt} = \prod_{j=1}^{J} 1 - \left|M_{ij} - y_{jkt}\right|$$

$$y_{jkt} \geq y_{jk(t+1)} \qquad (6)$$

$$\sum_{r=1}^{R}\sum_{c=1}^{C} h_{ikt}(r,c) = CH_{i}$$

$$\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I} h_{ikt}(r,c)x_{ikt} \leq 1$$

To solve this problem, in this work, we adopted the second hypothesis. Having the variable's value y_{jkt}, it is possible to determine the variable x_{ikt} by (1) and the problem can be reformulated as given by (7).

$$Min \quad \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=i+1}^{J}\begin{pmatrix} h_{ikt}(r,c)x_{ikt} \\ \bullet h_{jkt}(r,c)x_{jkt} \\ \bullet M_{j} \end{pmatrix}$$

$$S.a: \quad \sum_{r=1}^{R}\sum_{c=1}^{C} h_{ikt}(r,c) = CH_{i} \qquad (7)$$

$$\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I} h_{ikt}(r,c)x_{ikt} \leq 1$$

Notice that the model (7) is a combinatorial problem whose complexity is to allocate schedules of materials in the timetable and requires exponential computational effort to find the optimal solution. Referring to the example of Figure 5, there are 8 time slots to be placed in one of 24 positions, namely 24^{8}, which is approximately 10^{11}, or almost one trillion possible solutions. The enumeration of all possible solutions with subsequent evaluation of the degree of incompatibility of each is not a suitable alternative to real problems in that the number of subjects and the timetable is greater.

To solve the model (7), it is proposed as an approach to apply heuristic methods in which there is no guarantee of obtaining the optimal solution. Despite this, good quality solutions can still be found in a suitable computational time.

However, even with the use of heuristics there is the problem of the amount of information needed to encode a given solution. Note that, for the case of Figure 7, it is necessary to employ H x C x T x I binary variables, or 4 x 6 x 2 x 1 = 48.

There is, however, an alternative in which the number of variables for each solution depends only on the number of periods T and, for Figure 5, it would result in employing only a single integer variable. This alternative is the representation of the solution by rules, which will not be detailed in this section. However, it employs two key concepts: (a) the representation of occupation timetable through a matrix and (b) this matrix modification may suffer depending on the application of a rule to fill the timetable. The most important aspect of the approach is that the set of subject matters provided for each period *i* is represented by a matrix B and that this array is filled to ensure obtaining a feasible timetable provided that all prerequisites are met. It is of interest to note that the proposed algorithm facilitates the application of heuristics in solving the model given by (7).

3. COMPUTATIONAL SYSTEM

Instead of direct solving the binary mathematical model given by Eq. (7), one option is to use a simulation procedure combined with rules of filling the slots of timetable in a unified framework. The complete procedure is as follows:

(1)Apply filling rules that can be based on some type of criteria like teacher assignment, or in some pedagogic feature.

(2)Evaluate how many students could not attend to a course, although he or she already the prerequisites. This may happen because two courses that a student could attend is sharing one or more slots in Timetable.

(3) Return to (1) and apply different rules.

This combination of simulation and rules framework is necessary to avoid increasing of computational burden for real large-scale problems and was successfully applied for other types of problems with more complex binary model for problems with matrix structure like Stowage Planning (Azevedo et al., 2012).

The developed program used this features that aim to simplify the use of solving the problem. This becomes possible by using an interface that allows a user without advanced mathematics knowledge to enter the relevant information and perform the optimization. Figure 6, below, shows the main screen of the software.

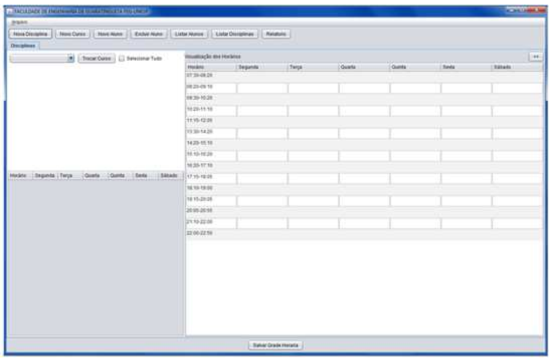
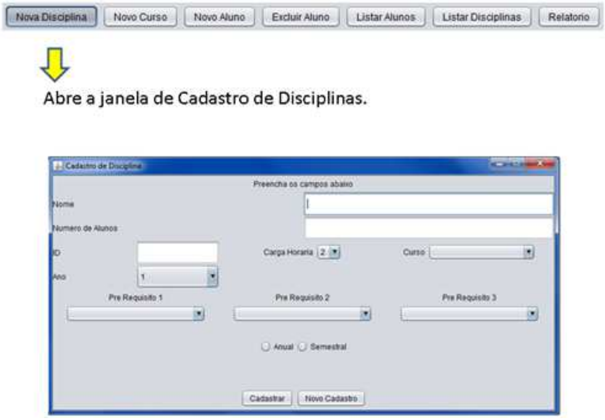


Figure 6: Software interface for determining the timetable.

The software was developed in Java (Deitel and Deitel, 2006). In the upper left corner of the window shown in Figure 6, on "File", we have the following: (a) "new", which allows you to open the window for registration of disciplines, opens the window for registration of students and allows you to open the window registration of each course; (b) "save", which lets you save the timetable of the selected course; and (c) "leave", which closes the program.

In Figure 7, the bar is highlighted with seven buttons perceived in Figure 6. An example is shown in Figure 7, with the window associated with a new discipline where the corresponding fields must be filled out with information such as the name of discipline, number of students, course load, course prerequisites.



Abre a janela de Cadastro de Disciplinas.

Figure 7: Bar with buttons and window associated with the registration of a new discipline.

The system takes care of managing if a student has the prerequisites to attend a particular matter and presents a list of possible subjects that the student can attend next semester. This list is automatically generated for all registered students so that it is possible to evaluate each suggestion that, despite having the prerequisite to attend a course, cannot do it, because it is in conflict at that schedule with another.

4. CONCLUSIONS AND FUTURE WORKS

It can be concluded that the OR has great potential for application to problems of administrative nature, including those related to initiatives in education. However, mathematical modeling often inhibits the use of tools and optimization techniques given the need for deeper knowledge of the procedures and algorithms.

In this perspective, this research report presents software with various features that allows, through a direct understanding interface, perform the optimization of resource use, in this case having been focused on determining timetable minimizing the number of students with scheduling conflict. The determination timetable is a significant problem for different types of institution, but especially for those responsible for larger initiatives involving a larger number of students.

Future work will involve the survey and analysis of data on the use of the software here presented in real situations, with a view to improving procedures and algorithms in use. Since the algorithm developed to determine the timetable facilitates the application of heuristics through rules and simulation instead of direct solving the binary model explained by equations. Further investigations may also involve the exploitation of the potential of Beam Search (Della Croce and T'kindt, 2002; Ribeiro and Azevedo, 2009; Valente and Alves, 2005) for search the better combination of rules application, among other types of meta-heuristics combined with rules application with simulation.

ACKNOWLEDGMENTS

The authors would like to thank the following organizations for their support: Saab Group (www.saabgroup.com), University of Skövde (www.his.se), University of São Paulo (www.usp.br), University of Campinas (www.unicamp.br) and National Counsel of Technological and Scientific Development (www.cnpq.br).

REFERENCES

Al-Yakoob, S.M., Sherali, H.D., A column generation mathematical programming approach for a class-faculty assignment problem with preferences, *to appear in Computational Management Science*, pp. 1-22, 2013.

Al-Yakoob, S. M. , Sherali, H.D., Al-Jazzaf, M., A mixed-integer mathematical modeling approach to exam timetabling, *Computational Management Science*, Vol. 7(1), pp. 19-46, 2010.

Azevedo, A.T., Ribeiro, C.M., Sena, G.J.D., Chaves, A.A., Neto, L.L.S., Moretti, A.C.: Solving the 3D Container Ship Loading Planning Problem by Representation by Rules and Beam Search. ;In ICORES, pp.132-141, 2012.

Deitel, H. M.; Deitel, P. J. Java: How to Programm. 6. ed. Bookman, 2006.

Della Croce, F.; T'kindt, V., A Recovering Beam Search Algorithm for the One-Machine Dynamic Total Completion Time Scheduling Problem, *Journal of* *the Operational Research Society*, vol 54, 2002, pp. 1275-1280.

Gunawan, A., Ng, K.M., Poh, K.L., Solving the Teacher Assignment-Course Scheduling Problem by a Hybrid Algorithm, CiteSeer. Available in: < http://130.203.133.150/viewdoc/download?doi=10.1.1.193.3646&rep=rep1&type=pdf>. Access: 21 jun. 2013.

Java™ , Sun Microsystems, Platform, Standard Edition 6 API Specification. Available in: <http://java.sun.com/javase/6/docs/api/>. Access: 20 jan. 2012.

Michael, W.C., Laporte, G., Recent Developments in Practical Course Timetabling, *Selected papers from the Second International Conference on Practice and Theory of Automated Timetabling II*, Springer-Verlag, London, UK, pp. 3-19, 1998.

Ribeiro, C.M., Azevedo, A.T., Teixeira, R.F.,Problem of assignment cells to switches in a cellular mobile network via Beam Search Method, WSEAS Transactions on Communications, Vol. 9(1): pp.11-21, 2009.

Schaerf, A. A Survey of Automated Timetabling. Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", 1999. Available in: <http://www. diegm.uniud.it/satt/papers/Scha99.pdf>. Access: 16 jun. 2011.

Valente, J. M. S; Alves, R. A. F. S., Filtered and Recovering Beam Search Algorithm for the Early/Tardy Scheduling Problem with No Idle Time, *Computers & Industrial Engineering*, vol. 48, 2005, pp. 363-375.

Willenmen, R. J. School timetable construction: algorithms and complexity. Technische Universiteit Eindhoven, 2002. Available in: < http://alexandria.tue.nl/extra2/200211248.pdf >. Access: 10 jul.2011.

AUTHORS BIOGRAPHY

Anibal Tavares de Azevedo. PhD in Engineering. Dr. Azevedo teaches at Universidade Estadual de Campinas – UNICAMP (http://www.unicamp.br/), in Brazil. Previously, he worked as a researcher and as a teacher at Universidade Estadual Paulista – UNESP (http://www.unesp.br/), Brazil. Dr. Azevedo graduated in Applied and Computational Mathematics from UNICAMP (1999), holds a Master's degree in Electrical Engineering from UNICAMP (2002) and received his Ph.D. degree in Electrical Engineering from UNICAMP (2006). He has experience in software development and in mathematical modeling for Production Engineering and Planning, for Scheduling of Power System Operation and for Education. His research has an emphasis on Linear Programming, Nonlinear Programming and Mixed Dynamics in the following topics: interior point methods, planning and production control of manufacturing, flows in networks, linear programming, graph generalized combinatorial optimization, allocation of cells to central offices, loading and unloading of containers on ships 2D and 3D, genetic algorithms, beam search and simulated annealing. < http://lattes.cnpq.br/9760457138748737 >.

Alexander Kameyama. Engineer. Previously worked as a researcher at Universidade Estadual Paulista – UNESP (http://www.unesp.br/), Brazil.

Joni A. Amorim. PhD in Engineering. Postdoctoral Fellow at the University of Skövde, or Högskolan i Skövde – HiS (http://www.his.se/english/), in Sweden, in collaboration with SAAB Training and Simulation (http://www.saabgroup.com/en/training-and-simulation/). The University of Skövde offers first-class programs and competitive research, which attracts research scientists and students from all over the world. The University of Skövde is one of the most specialized universities in Sweden and its research is focused on the development and use of advanced information technology systems and models. Dr. Amorim previously worked as a researcher and as a teacher at Universidade Estadual de Campinas – UNICAMP (http://www.unicamp.br/), in Brazil. Dr. Amorim collaborates with researchers at UNICAMP, a university with more than 3,600 original scientific publications published in 2009, 78% of which in journals indexed in the ISI/Web of Science. Dr. Amorim collaborates with researchers at Universidade de São Paulo – USP (http://www.usp.br/), the major institution of higher learning and research in Brazil. His research has an emphasis on multimedia production management, project portfolio management, distance education and training based on serious games and simulations. < http://lattes.cnpq.br/3278489088705449 >.

Per M. Gustavsson. PhD in Computer Science. Dr. Gustavsson works as a Research Scientist at Saab Group (http://www.saabgroup.com/). Saab serves the global market with products, services and solutions ranging from military defence to civil security. Dr. Gustavsson also works at the Swedish National Defence College – SNDC (http://www.fhs.se/en/), in Sweden. At SNDC, research is carried out in diverse, but inter-related subject areas and subsequently disseminated to other interested sectors of society both nationally and internationally; the College trains and educates military and civilian personnel in leading positions, both nationally and internationally as part of the contribution to the management of crisis situations and security issues. < http://se.linkedin.com/in/pergu >.

MODELING AUTONOMIC MONITORING AND CONTROL FOR IOT SYSTEMS

Zenon Chaczko[(a)], Ryszard Klempous[(b)], Jan Nikodem[(c)]

[(a)] Faculty of Engineering and Information Technology,University of Technology, Sydney,
15 Broadway, Ultimo, NSW, Australia, 2007.
[(b), (c)]Institute of Computer Engineering, Control and Robotics, Wroclaw University of Technology,
11/17 Janiszewskiego Street, 50-372 Wroclaw, Poland

[(a)]Zenon.Chaczko@uts.edu.au, [(b)]Ryszard.klempous@pwr.wroc.pl, [(c)]Jan.Nikodem@pwr.wroc.pl

ABSTRACT

This paper introduces a bio-inspired approach for development of collective intelligence based computational models. These models are suitable for autonomous sensing, monitoring and control strategies for ambient systems based on Internet of Things technologies. Authors discuss issues and challenges related to modelling, design and implementation of a large scale, Internet of Things based smart system infrastructure such as smart office building, airport, public transport, etc. Additionally, various autonomous management strategies that anticipate variable levels of resource usage in Internet of Things systems are being presented

Keywords: bio-inspired modelling, autonomics, Internet of Things, large-scale infrastructure

1. INTRODUCTION

The management of complex, heterogeneous and distributed (network based) system composed of collaborating, sensors, actuators and robotic devices is a challenging task; and unless done effectively can significantly reduce the overall efficiency; degrade capacity to perform various functions as well as limit access to available resources in remote, dynamic and often hostile environment. This is particularly true in heterogeneous, network based environments as the actual structure of the network based system can change depending upon such disparate factors as the application tasks, communication links, hardware, topology and geography and environmental conditions. The software intensive, autonomous (Ishida 1997) and collective intelligence (Por 1995, Brown & Lauder 2000, Kennedy 2006, Kaiser et al. 2010 infrastructure for IoT system aims to lay the foundations for developing service-oriented and real-time system. These systems can be used for monitoring and control applications where containment of dangerous events or re-collection of available resources is critical. Apart of the on-board computer(s) to support the system's high functions, the system's' infrastructure is built as a highly reconfigurable network of sensors, actuators and robotic devices. The ability to adapt to the changing environments requires a step change in the design approaches. The key research challenge is to provide flexible resource management and data access solutions that are effective in a large-scale, heterogeneous system network. The outcomes of the initial design will enhance the position of the AI group to become not only an advisory body but to ensure a sustainable vision for future development of advanced engineering and co-evolution of and open hardware and software platforms. The modelling and simulation of software intensive systems (Bruzzone and Longo 2005) need to consider the high level decision making and system wide functions as well as individual (autonomous) computational and communication facilities (i.e., localisation, navigation, control and communication) that reside at the systems' lower levels.

The system requires the global information management and application software development facilities that are implemented in higher level programming languages. The development of a prototype required a set of high performance management solutions; middleware and software component frameworks that are able to facilitate the development of a network based, infrastructure oriented and embedded software for swarms (Bonabeau et al. 1999) of sensors, actuators and robotic devices. This paper is aimed at addressing the problem of dynamically managing the network of sensors, actuators, robots and various other associated resources so that specified communication links, data rates and priorities as defined by the real-time management system can be achieved (Das et al. 2004). This entails a development of infrastructure oriented software and algorithms for construction and simulation of real-time, mission critical solutions in resource constrained and possibly ambient environments.

2. AUTONOMICS

The concept of autonomics can be perceived as a capability of software systems to perform and manage their operations completely by themselves or only with a minimal level of human intervention. Autonomic systems, including Internet of Things (IoT) systems by adopting the concepts of autonomics are capable of self-managing all its elements and data communication links. In autonomic scenarios, wireless communication is complex and requires designers to consider a number of problems related to sensor and actuator localisation, clustering, routing, energy management as well as various constraint conditions related to transmission collisions, multipath interference, obstructions that adversely impact the data throughput of high bandwidth communications robustness, reliability and scalability (Loureiro and Ruiz 2007). In order to face the challenges related to implementation of autonomic functions in IoT we propose a model of biomimetic IoT system that is characterised by the following fundamental properties:

2.1. Self-organisation

The phenomenon of self-organisation is pervasive in natural systems (Kauffman 1995), (Bak 1996). It can be defined as a system's tendency to evolve into a more organised state in the absence of external factors as a response to changes in the system's environment; It can be also perceived as a collective and coordinated process in which system's components achieve a higher level of organisation while interacting with other elements as well as and with the environment. Software systems incorporating a collective made of a number of components communicating, interacting among each other and with the environment. Collectively these elements can perform various tasks as a group, coordinating their tasks and activities to obtain a higher level of efficiency. The sum of dynamic behaviour arising due to interactions between different parts of the IoT based system could result in a coherent behaviour of the system as a whole. The IoT sensitivity relies on capability of software services to perform changes even if the value of an observed or control parameters is modified by a small value only. Since by principle, the self-organisation properties cannot be predetermined, the IoT system facilitated by software infrastructure can evolve to a new configuration that is compliant with the global system functions and environmental constraints. The robustness of the self-organising system can be then measured by a rate at which the system in its newer configuration is able to detect and handle its faults.

2.2. Self–shaping

A system's self-shaping property can be defined as allometric and scale-invariant (power-law scaling) characteristics of a system that addresses various aspects of self-adaptation and self-optimisation (McMahon & Bonner 1983), (Niklas 1994), (Phillips 2006). It can be interpreted as the capability of the system to alter or adjust its structure, size and rates of metabolic processes according to its varying internal and external stimulations (or events). In resource constrained IoT networks there is a requirement for software to be able to self-modify the network shape, adapt to variable levels of available resources and changes in the environment. In order to promote the system emergent properties such as robustness, or survivability of the IoT, this needs to occur by respecting scale-invariant relations.

In order for IoT software system infrastructure to support management of the IoT according to varying levels of available resources, tasks and changes in the environment we need to model, design and then implement the self-shaping (i.e. self-modification of the network topology) function requirement. It is suggested that this new type of the system property can be ensured if we follow allometric laws (Darveau et al. 2002, Chaczko 2009) that are often pertinent to living systems (McMahon and Bonner 1983), (Calder 1984), (Bejan 2002).

2.3. Self-adaptation and self-healing

The self-adaptation property can be seen as a capability of a system to modify/alter or adjust its internal structure or/and behaviour according to varying conditions in its surrounding.

The self-healing is to be perceived as a capability that allows automatically detecting, localizing, diagnosing and repairing failures. The process of self-healing is adaptive, fault-tolerant and inter-dependent with the mechanism of self-monitoring.

2.4. Robustness and resiliency

Robustness (fault tolerance) can be perceived as a capability of the system to resist or tolerate noise, disturbances, faults, stress, modification in system architecture (both structure and behaviour) or changes in the ecosystem without negatively affecting the system's functions or having long term effects on its structure and behaviour.

The property of resiliency can viewed as capability to absorb and even utilise (frequently with advantageous results) noise, disturbances and changes that attain them, in order to sustain and persist without any qualitative changes in the architecture of the system.

IoT for resource constrained systems need to be flexible to change and resistant to damage or faults; the systems should be able to self-modify their past behaviour and adapt to newly allocated tasks, changes in levels of available resources or changes in environment. IoT based systems can be perceived as being robust, if they would be able to tolerate failures, non-cooperative behaviour or conflict relation among its components. This can be achieved by including software functions that apply genetic mutation and reproduction to seed autonomic properties in IoT.

2.5. Cooperativeness

Characteristics of cooperativeness are perceived as a system's capability to stimulate collective and cooperative interactions among its components. Components can perform various tasks in teams, coordinating their activities to obtain an optimal efficiency. In reality, there are various degrees of cooperation/competitive (in resource constrained situations) behaviour at place. The sum of dynamic behaviour arising from cooperative interactions between different parts of the system could decide about coherent behaviour and robustness of the whole system. Thus robustness of a system can be measured as a rate at which its components (resources) are repaired or replaced against the rate they diminish.

3. APPROACH AND METHODOLOGY

The software methodologies and tools are core aspects of all networked oriented system infrastructure software. These infrastructure oriented software are implemented to efficiently combine, manage data and control robotic swarms. At some stage, infrastructure oriented software is to play of enormous significance in such areas as: ambient management system's infrastructure management, environment monitoring, adaptation/articulation of all major parts, sensing and actuation, security and safety, education as well as many other areas that depend critically upon software technology. However, building application that make best use of AI and IoT technology in terms of practicality and economics (including time to deploy) cannot be fully realised without a consensus by majority of application developers on adequate methodologies and tools. In the context of remote management of infrastructures, such methodologies and tools for robotic network systems can significantly improve development life cycle the value of embedded devices and sensor infrastructure, reduce the cost of information management and offer technically and economically significant as well as viable implementations to many participating institutions.

3.1. Scope

The scope for cross disciplinary knowledge advancements when discussing hardware and software of architecture is significant. First and foremost is the advancement associated with the application of autonomics and information AI to improve the engineering methods of analysis, simulation and prediction. Second is the advancement associated with the development of cost effective and robust network architecture for local and remote operations. Third is the advancement associated with the application of biomimetic and AI paradigm that will enable the sharing of resources for multiple infrastructure oriented software concurrently. Owing to the richness of the field and the number of open problems in the domain, there is significant scope for several serendipitous advancements in the knowledgebase of several disciplines. The following new methodologies and technologies were being developed in the course of the discussed project:

1. New design paradigms for infrastructure networks with distributed processing and AI for autonomic robotic network infrastructure management.
2. New and open Service Oriented Architectures that support access to the fused/processed remote sensor/actuator information by possibly multiple applications.
3. Advanced infrastructure oriented software tools for development and integration of real time, context sensitive and proactive software Infrastructure for mission critical robotic networks/infrastructures.

3.2. Approach

The presented research approach involves the following stages:

1. Development of a working definition of Service Oriented Architecture-like infrastructure software for IoT based and embedded robotic systems. The output of this stage might be a document describing the architecture and how it applies to the needs of ambient management system. The document will include such high level design components as: Service Configuration, Service Activation, Fault Data Collection, Performance Data Collection and Usage Data Collection modules.

2. Identification of the mechanisms, policies and possible parameters for enforcing control and management of individual robotic drones and their swarms. Outcome of this stage is a design document identifying the policies, algorithms and equipment parameters that can be used for controlling and managing ambient management system.

3. Modelling, design and development of algorithms for management execution of real-time functions of ambient management system architecture in the test-bed environment. During this stage a prototype of performance management software for the equipment currently available in the lab has been developed.

4. Development of suitable algorithms for task allocation within individual embedded devices/motes and across groups of these devices. A small scale simulation environment has been used to test different resource allocation techniques. In this phase, simple search techniques were applied. More sophisticated and more scalable resource allocation techniques will be the focus of the future work.

The modelling and the development of the IoT based demonstrator allows for testing and validation of various autonomic behaviours that could be applied in various user environments. The main objectives is to demonstrate dynamic/adaptive modifications in device resource settings to meet a given control priority setting. This includes the management and decision-making software for the current ambient management environment. The demonstrator can be used to show how the autonomic management and decision making

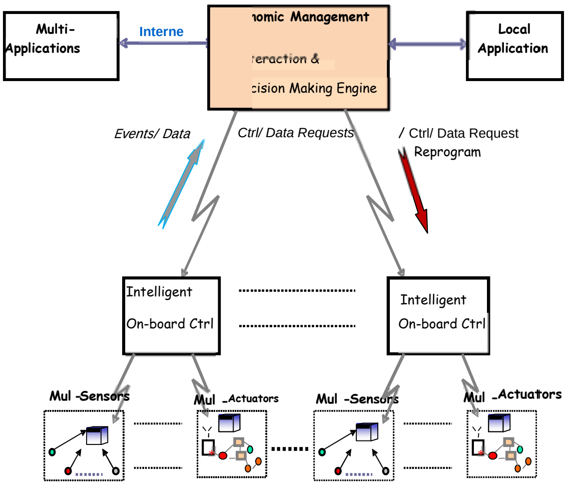


Figure 1: Conceptual architectural model of IoT infrastructure

system works in dynamic environment as well as to test the performance of such a system.

4. MODELLING SYSTEMS BASED ON COLLECTIVE INTELLIGENCE

4.1. The architecture for ambient systems

The main driver requirement for modelling systems based on collective intelligence was to develop an open source, cost effective and reliable architectures which would support a rapid development, validation and test of various autonomic concepts and user applications. Infrastructure hardware and software is to make the best use of emerging computer and network device platforms in order to facilitate information processing at the individual IoT motes, local level controllers (intelligent on-board computers) and the central computer(s). One of the key advantages of the proposed biomimetic model approach is that by processing information locally, the system can reduce the amount of data that needs to go on the networks. More importantly, there would be no need to send all that irrelevant or redundant data through the network and thus burdening the transaction requirements of the systems. In line with this view, the proposed architectural model allows the processing to move from wherever it is to a point closer to the sensors and actuators. This has a potential to fundamentally transform the efficiencies associated the computing and network support infrastructure. The proposed system architecture (hardware and software) has been developed for multi-sensor/actuator, multi-application environment of the swarm-like, human nervous modeled system (Fig. 1 and Fig. 2.)

Although most of the processing may be done closer to source, still one needs an efficient Internet and Web Services like and possibly Cloud Computing (grid middleware) access. The proposed remote information management approach is illustrated in Fig 2. Web user logs into the web system to request services that include the main functionalities of the web system such as view, analysis and control of required information. The web browser transfers the user request to the web server. The web server sends the service request to the information server for information. The information is retrieved from the data storage server. In principle the data can be embedded in the intelligent on-board controller (IOBC). The data server processes the information using heuristic techniques, returns the results to the data storage server, and then returns the data to the web server for the display on the web pages. The web user interacts with the web pages for the next service request.

4.2. Biomimetic model for IoT

In IoT base systems, low cost sensors and actuators with embedded computing and communication capabilities are enabling a new paradigm where sensors and actuators can share resources, just like the way computers are able to form a grid to share the computing power. The discussed ambient management system architecture, at presentation and business logic layers would still require a significant computational power. Hence, the system's higher level functions need to be provided with a dedicated on-board CPU(s) (Linux box). While, the ambient management system can rely on Web services and Cloud computing infrastructure (grid middleware). The vision for the model is to push the advancement of a modern IoT solution by enabling sensors and actuators to form a grid and deliver, autonomous services to various user applications. As Infrastructure software will expand to incorporate multi-sensor/actuator feeds such systems are subject to severe bandwidth loading and potentially may require large amounts of computing power and storage. As we move to consider broader multi-sensor/actuator installations, these limitations might be

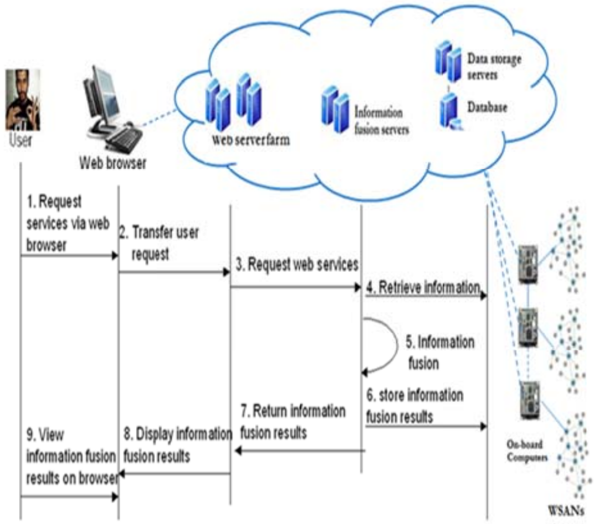

User
Web browser
Web serverfarm
Information fusion servers
Data storage servers
Database
1. Request services via web browser
2. Transfer user request
3. Request web services
4. Retrieve information
5. Information fusion
6. store information fusion results
7. Return information fusion results
8. Display information fusion results
9. View information fusion results on browser
On-board Computers
WSANs

Figure 2: Web information management sequence

further exacerbated. Biology inspired, alternative approach would be able to dramatically reduce the bandwidth, computing and storage requirements while allowing multi-sensor/actuator information and control to be shared between many diverse infrastructure software solutions. The human body is an excellent example of a true multi-sensor/actuator system being used for a multitude of diverse infrastructure software. As can be seen in the Fig. 3, much of the bodies data is being constantly processed by the brain at the *Unconscious* and the *Semiconscious* levels in the background without the direct involvement of the reasoning or cognitive part of the brain. In this way, the brain is capable to obtain information from the Semiconscious parts, when required and then integrate, via data/ctrl AI, with the *Conscious* world. If the human brain was not compartmentalised in this specific manner, the *Consciou*s part of the brain would be completely overloaded with just trying to process the data or events required for the control of *Unconscious* body functions such as respiration, blood circulation, heart rate, temperature control not to mention immune, endocrine or cutaneous system functions. The only time that this Unconscious world communicates with the *Conscious* part of the brain is when serious anomalies are detected, i.e. a high temperature or low sugar levels is signalled to the *Conscious* world where reasoning then takes place followed by *Conscious* actions - take a pill, eat, call the doctor etc.

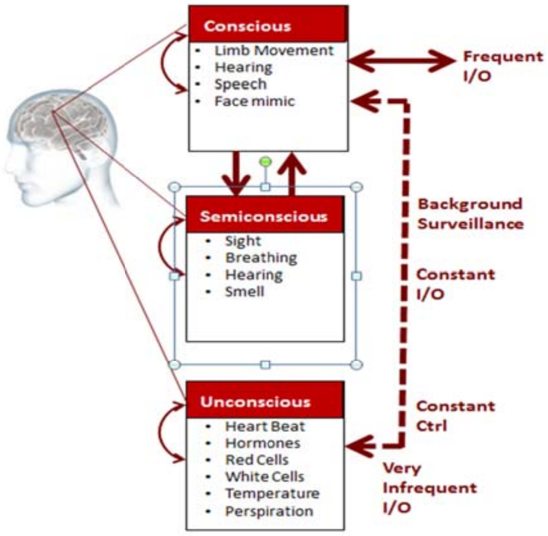


Figure 3: A model of IoT with autonomic feedback & Data/Ctrl

However, the breakdown of computing and communication for optimal data/ctrl AI is a non-trivial problem. The fundamental question is: what processing or control should be executed at the sensor/actuator level, if it has some computing power and what needs to be done at the intelligent controller (Linux box) level; And ultimately what computations need to be carried out at the ambient management level (application layer). One of the suggested approaches might be to first break down the information AI problem in terms of its distribution of communication and computing load in a highly flexible manner and then mapping the information AI on to the proposed architecture. It is proposed to use the emerging IEEE 1451 TEDS standards for effective discovery and maintenance methodologies of transducer (sensor and actuator) infrastructure.

4.3. Software tools and technologies for IoT

Wireless sensors, actuators, various computing devices and technologies play a critical role in the infrastructure (Fig. 4) of IoT base solutions. Advanced software development tools are not only to enable rapid modelling, design, implementation and integration of infrastructure oriented software but also to facilitate transaction processing, manage demands for computing resources, support flexible software composition, provide data from all sensors for the comprehensive assessment of processing conditions and allow software constructs which can support combination of real-time decisions or perceptions from multiple sources. With the advent of wireless sensor and actuators, MEM devices, application developers need to be in a better position to overcome the traditionally over-conservative, less-transparent, labour intensive and costly approach in terms of design development and maintenance of software.

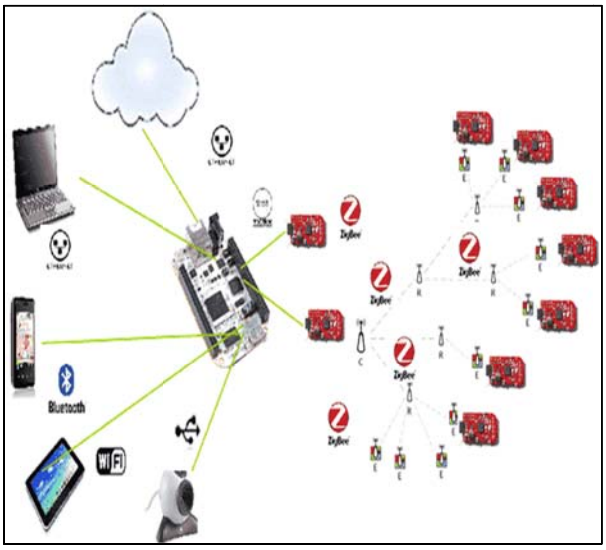
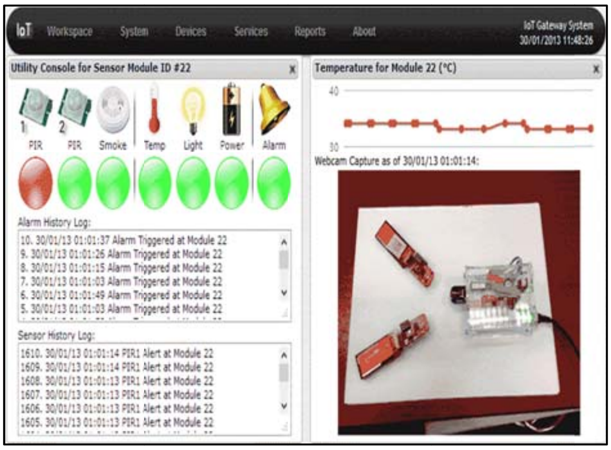


Figure 4: Wireless sensors and actuators technology in a prototype of IoT system



Alarm History Log:

```
10. 30/01/13 01:01:37 Alarm Triggered at Module 22
9. 30/01/13 01:01:26 Alarm Triggered at Module 22
8. 30/01/13 01:01:15 Alarm Triggered at Module 22
7. 30/01/13 01:01:03 Alarm Triggered at Module 22
6. 30/01/13 01:01:49 Alarm Triggered at Module 22
5. 30/01/13 01:01:03 Alarm Triggered at Module 22
```

Sensor History Log:

```
1610. 30/01/13 01:01:14 PIR1 Alert at Module 22
1609. 30/01/13 01:01:14 PIR1 Alert at Module 22
1608. 30/01/13 01:01:13 PIR1 Alert at Module 22
1607. 30/01/13 01:01:13 PIR1 Alert at Module 22
1606. 30/01/13 01:01:13 PIR1 Alert at Module 22
1605. 30/01/13 01:01:13 PIR1 Alert at Module 22
```

Figure 5: Rapid modelling, design, implementation using open source IoT hardware and software tools.

5. DISCUSSION

There are several challenges that modelling of the autonomic monitoring and control functions for IoT systems need to address. These challenges related to such important questions as:

How much intelligence needs to be local and how much can be group intelligence?

It is expected that group intelligence would bring certain tangible benefits. However each individual intelligent mote should have sufficient competence to complete a limited number of tasks on its own taking into account the special, and non-trivial, situation of autonomous execution of various tasks. It is considered that the interplay of these complimentary goals will be beneficial to the development process.

How to design intelligent behaviour (cooperation) in the IoT collective?

It is suggested that three classical rules (Reynolds 1987) should be followed:

1. Separation: do not move too close to nearby members of the collective (drones),
2. Alignment: move into the general direction of the collective (swarm) , and
3. Cohesion: steer towards the general centre of the collective.

It would be interesting to explore automated survival schema that run on MCU's for the situation where the CPU has failed and a limited set of MCU's remain to recover the vehicle. Some evolutionary algorithms for generating and prioritizing scenarios that are subsequently installed on MCU's to provide an adaptive way of fine tuning these schema. It would be possible and that it may even be possible to place adaptive seeds on an MCU allowing them to be somewhat adaptive once the CPU has failed and a vehicle wishes to return to a safe haven.

Other ideas related to cooperation within a swarm may involve exploration of decay properties of pheromone-like or hormone-like (endocrine system) paradigms that could involve encapsulation of weighting system which could allow some more subtlety of control. The decaying positional marker is stigmergy information left by some individuals of a given species for other members of the same species that could be accessed or modified when they come close or pass that point. An information token, or a tuple of information could encapsulate anything, including relations or behaviour of the vehicle (function/computation components). In addition to varying in the time domain through a timed decay, we can modify the relevant information for the individual drone/swarm in the light of changing external conditions to share with the new or old arrivals various information including the relative time sequence or even provide some serendipitous data/info for future use. Additionally, even computational components (whole algorithms) not only some proven parameters or artefacts i.e. images, sounds, graphs, text, etc. can be shared or exchanged. These shared elements can be used to environmental monitoring and forecasting as well as can be used for the decision making in IoT collective of various network and robotic devices.

6. CONCLUSION

The discussed modelling and simulation approaches involve the cutting edge technology of software infrastructure, technologies as well as application of heuristic (AI) techniques in IoT based systems. IoT systems, can be highly dependent on tight coupling between the user applications, infrastructure oriented software, the AI solutions, on-board computer(s). At the low level, IoT system relies on a large number of sensors and actuators devices which might be difficult to deploy and manage. Additionally, the autonomic management paradigm may be limited by actuating and sensing capability of proprietary technologies and devices in IoT systems. Collective intelligence and communication in IoT brings in a new set of challenges and complexities that can be addressed using bio-inspired ambient computational models as well as open hardware/software solutions.

REFERENCES

Bruzzone, A.G., Longo, F., 2005. Modeling & Simulation applied to Security Systems. *Proceedings of Summer Computer Simulation Conference*, pp. 183-188. July 24-28, Philadelphia (Pennsylvania, USA).

Reynolds, C.W., 1987. "Flocks, herds and schools: A distributed behavioral model". *Computer Graphics* **21** (4): 25–34.

Bak, P., 1996. *How Nature Works: The Science of Self-Organized Criticality*. Springer Verlag, New York.

Bejan A., 2002. Constructal Theory of Organization in Nature: Dendritic Flows, Allometric Laws and Flight", *Design and Nature*, Edited by: C.A. Brebbia & L.J. Sucharov, Wessex Institute of Technology, UK and P. Pascolo, Universita degli di Udine, Italy, Transactio**n:** *Ecology and the Environment* vol. 57.

Bonabeau, E., Dorigo, M. and Theraulaz, G., 1999. Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press,

Calder, W.A., 1984. Size, function and life history. Harvard University Press, Cambridge, Mass.

Chaczko, Z. (2007) Autopoietics of Biomimetic Middleware System, private correspondence, November.

Darveau, C. A., Suarez, R. K., Andrews, R. D., & Hochachka, P. W. 2002. Allometric cascade as a unifying principle of body mass effects on metabolism, *Nature*, 417:166-170.

Das, S. K. , Banerjee, N. and Roy, A., 2004. Solving Optimization Problems in Wireless Networks using Genetic Algorithms, Handbook of Bioinspired Algorithms.

Ishida, Y., 1997. The immune system as a prototype of autonomous decentralized systems: an overview," *In proceedings of 3rd International Symposium on Autonomous Decentralized Systems* (ISADS 97).

Kaiser, C., Kröckel, J. and Bodendorf, F., 2010. Swarm Intelligence for Analyzing Opinions in Online Communities. Proceedings of the 43rd Hawaii International Conference on System Sciences, pp. 1–9.

Kaufmann, S.A., 1995. At Home in the Universe: The Search for the Laws of Self-Organization and Complexity. Oxford University Press, New York.

Kennedy, J., 2006. Swarm Intelligence, in *Handbook of Nature-Inspired & Innovative Computing,* Editor: A. Zomaya, Springer Verlag, New York, pp.187-221.

Loureiro, A.A.F. and Ruiz, L.B., 2007. Autonomic Wireless Networks in Smart Environments, In Proceedings of the 5th Annual Conference on Communication Networks and Services Research, CNSR '07. Fredericton, New Brunswick, Canada.

McMahon, T. A. and Bonner, J. T., 1983. On Size and Life. Scientific American.

Niklas, K. J. 1994. Plant allometry: The scaling of form and process. University of Chicago Press, Chicago.

Phillips M.L., 2006. "Study challenges metabolic scaling law," *The Scientist*, January 26. http://www.the-scientist.com/news/display/23012/

Brown, P., Lauder, H., 2000. "Collective intelligence". In S. Baron, J. Field & T Schuller. *Social Capital: Critical Perspectives*. New York: Oxford University Press.

Noubel, J-F, 2004. "Collective Intelligence: the Invisible Revolution", rev. 2007.

Por, George (1995). "The Quest for Collective intelligence". In K. Gozdz. *Community Building: Renewing Spirit and Learning in Business*. San Francisco: New Leaders Press.

THE FAMOUS TRANSPORTATION OF THAILAND AND JAPAN LOGISTICS

Nollaphan FOONGKIATCHAROEN [(a)], Fumio AKAGI [(b)]

Wajiro 3-30-1, Higashi-ku, Fukuoka, 811-0295 Japan

[(a)]bluekiwi_toey@hotmail.com, [(b)]Akagi@fit.ac.jp

ABSTRACT

Transportation is firstly considered to support a variety of trade activity. Moreover, it can measure economic growth rate of each country. Hence, transportation is not only about the development of specific automobiles but it also thinks over the process of systematic management of transportation. For instance, in terms of personal transportation it needs to have public transportation or a shared passenger transport service i.e. buses, trams, rapid transit and trolleybuses. On the other hand, in terms of transporting goods, logistic methodologies are also important to be used in managing transportation system.

Japan is one of the potential export markets that related to Thailand. At the same time, Thailand's Small and Medium Enterprises are an important part being developed in order to support ASEAN Economic Community: (AEC) in 2015. The development plan will build a stronger future economy, especially in terms of investment. Therefore, if we get some knowledge and experiences about business strategies, it will help increase the chance for my own business.

This paper will present about my own export business or Small and Medium Enterprise (SME) in the coming future. For example, if we actually have our own business, we will export those stuffs to a target market such as, Thai fruits, clothes (from Thailand to Japan). In order to understand the types of transportation in the correct category, so they are determined into 5 transportation modes such as Water/Ship, Air, Truck, Rail, and Pipeline Transportation.

Keywords: Thai fruits, transportation, logistics, export, import, send abroad, distribution management, supply chain management, intermarket segmentation, attitude toward foreign products, value chain, target cost

1. INTRODUCTION

Everyone who comes to Thailand, however, will find that domestic fruits are so plentiful, so diversified, so inexpensive and so delicious. These advantages lead to an unexpected reward.

Factors that help to increase productivity advantage that is geographical position. Due to Thailand is a tropical climate country [1]. Hence, Thailand can produce so many different kinds of high quality fruits. The tropical climate is certainly effect on the growth of vegetation. However, there are other factors that have contributed to product benefits – the fertile soil, continuous efforts to improve fruit quality by scientific methods, and the comparative length of Thai territory, which extends right into the subtropical zone, making it possible to grow native fruits in higher latitudes [2].

Thailand has a better chance for exporting fruits into Asian countries due to most Asian countries's potential is much more than other continents. Japan is considered to have a potential market for Thailand because Japanese people have the same way of eating fruits as Thai people. In addition, the main issues of distance and logistic system are also suitable for exporting Thai fruit. Seven countries such as Japan, South Korea, Singapore, Indonesia, Hong Kong, Taiwan and India, will be the most important exporters in the near future [3].

However, exporters and government need to cooperate in order to improve the quality of Thai fruit. The Ministry of Agriculture has launched the future direction for fruit quality improvement into 4 issues [4].

(1) Supporting the whole farmers in the country by making agriculture, harvesting and packing products. Moreover, they (exporters and government) should focus on environmentally friendly manufacturing as well.

(2) Supporting the research and fruit production to meet people's needs.

(3) Developing and using logistic system that suitable for Thai fruit export.

(4) Advertising and suggesting Thai fruit into oversea market.

Nowadays, Thailand's logistics problem is a high cost because there is no freight to other modes that has a low cost [5]. Moreover, there have some problems regarding government regulations throughout the year. In the past, the submitted documentation would be transmitted through various procedures. As a result, all steps spent a lot of time sending documentation from one place to another. In the present time, state agency has tried to gather all steps into one point that can be called "single window" or "one stop service: oss".

The Customs Department is an agency of the federal government that collects customs duties and performs other selected border security duties. The customs service have recently developed "ELECTRIC SINGLE WINDOW". Private sector can hand in document and contact customs department via internet system. Therefore, officers don't need to travel by themselves. Moreover, there are other types of electronic systems i.e. e-Declaration, e-Payment, e-Manifest and e-container. However, it has to be accepted that these developments are still limited for e-logistics. The developed system is just “e-customs” which is not connected with agency that create the certificate and other licenses. To ensure that the implementation of federal government response to the needs of private sector, so government should establish an independent federal agency.

Thailand’s overall exports of goods need to rely on logistic systems, which compose of 5 transportation modes such as road/water/rail/air and pipeline transportations. The most popular mode of transportation is road with 87 percent of total transportation modes. Water transportation becomes the second position that has the amount of transportation about 5 percent. Meanwhile, the proportion of rail transportation is approximately 2 percent. On the other hand, the amount of air transportation is not popular modes compared to other transportation modes, with less than 0.5 percent. Apart from 5 transportation modes, there is a significant number of pipeline transportation, especially in gas or petro transport [6].

Using road transportation system has many advantages. For instance, most vehicles can easily drive into every area. At the same time, advantage of loading goods is more diverse in terms of size and vehicle types. However, there has some transport limitation. For example, the cost of road transportation appears more expensive than any other mode of transportation due to the fact that fuel shows continuously high costs while the number of trucks is going to increase. This problem can affect traffic systems, particularly in urban area, leading to the impact of high logistics costs. In terms of transporting commodities, Thailand should do strategic planning to support direct and indirect transportations, leading to the reduction of usage of road transportation. The strategic planning will be accomplished by developing the other modes of transportation and by using tax to reduce the cost of transporting goods.

2. TRANSPORTATION

The first period of trading transportation might be a cart (the wheeled vehicle). It is used to drag some objects, whereas its vehicle is also driven by some of animals i.e. elephant, horse, donkey, camel, deer and dog. Afterwards, these innovations are developed into the system of rail transport that carries (people or goods) from one place to another. In the present time, ship/water/air/truck and pipeline transportations are being extensively accepted by commercial sector.

Nevertheless, there are some kinds of high technology i.e. Internet, mobile phone, computer and etc, which have more advantage than the former transportation system. In order to understand the types of transportation in one direction, so they are determined into 5 modes [6].

2.1. Ship/Water transportation

Water transportation is like the oldest characteristic compared to the whole transportation modes. It is vastly utilized in the past, present and future. This is because water/boat transport can support large amounts of stuff, while they need only small amount of investments.

There are advantages and disadvantages between domestic and international transport. For example, even though water transportation is suitable for business types, on the contrary timing of receiving commodities still spends a lot of time as well. Therefore, people who involve with importation and exportation should carefully study and try to understand all processes before determining.

2.2. Air transportation

Due to time is a major obstacle for remote trading, especially in terms of boat and road transportation. Some goods spend time travelling from one place to another, which is necessary to be controlled. At the same time the choice of transportation mode also needs to be carefully considered particularly in fragile products or special temperature controls, i.e. flowers, fruits, etc. They have to choose time-based competition and reduce the damaged goods caused from transportation. So air transportation will be the best way compared with other types of transportation.

2.3. Truck transportation

Truck transportation is a major heart of road transport. Government has a mega plan to construct and extend the road system. These strategies are determined in order to make an international cooperation. As a result, commercial international investment would be more convenient to transporting.

2.4. Rail transportation

Rail transportation has some advantages. This mode can transport large amount of stuffs, whereas their payments is quite low compared to other modes. Most exportation used rail transportation system often have low cost and heavy weight such as coal, cement, petroleum, rice, sugar and mineral. However, rail transportation also has some drawbacks such as time because of rail transportation always stops at every rail stations.

2.5. Pipeline transportation

Pipeline transportation has a specific characteristic due to transporting goods is in the form of fluid. Therefore, the pipeline should have low slope area so that runoff will not flow back through the pipe. Popular products for pipeline transportation consist of crude oil, petroleum and natural gas.

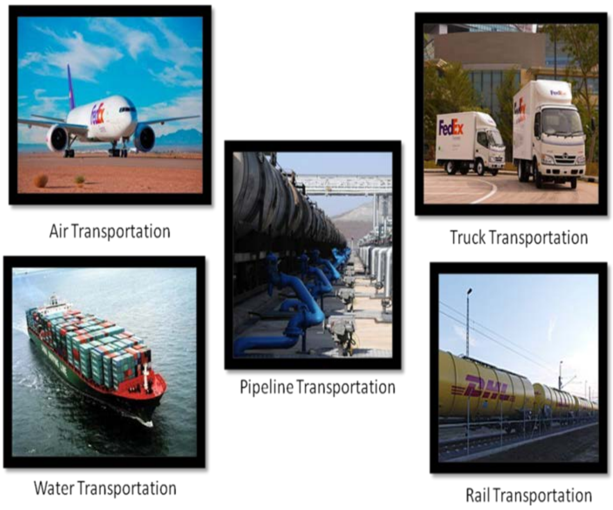


Figure 1: The overall transport systems

Japan is the second country which has the great potential close to USA and Europe. International commercial investment between Thailand and Japan has an expansion rate approximately 11% at the end of 2011. The technological and innovative developments are growing in high level, particularly in automobile and electronic industry. Most vehicles produced in Japan are sold in overseas. There are different brand name vehicles; such as Toyota, Honda, Nissan and so on.

Exportation and importation is like a breath of country. From the past year until now, the opening of trade way as well as free trade has brought about the economic development throughout the country. As a result, most countries can merchandise raw materials, together with other equipment in a proper way [7].

3. ECONOMIC, TRADE AND INVESTMENT RELATIONSHIP BETWEEN THAILAND AND JAPAN

(1) Regarding trade between Thailand and Japan in 2007, the proportion of trade in Thailand's imports and exports was 20.5% and 12.0%, respectively. Based on the Board of Investment (BOI)'s data, Japan-Foreign direct investment accounts for approximately 49 %of the total foreign investment to Thailand. The Japanese Government's bilateral official development assistance (ODA) accounts for 70 – 80% of total assistance to Thailand from the developed countries or the Development Assistance Committee (DAC) member countries.

(2) Thailand is interested not only in trade and investment support, but also in the export. Presently, Thailand is significantly interested in the export to Japan in order to solve economic crisis.

3.1. Trade

(1) In Thai sentiment, Japan is the number one exporter to Thailand and number two import after the United States. Conversely, in Japanese sentiment, the proportion of Thailand's import and export to Japan was 2.9% and 3.6% respectively. Thailand is the number ten importers to Japan, and the number six exporter in 2007

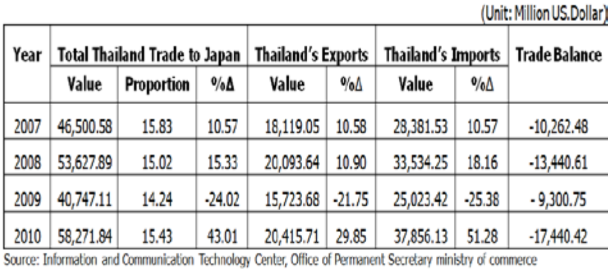
(2) Overall, Japan has continually favorable balance of trade after 1980 (1985-1989). As there are an significant increase in the recovery of Thai economy, and Japan direct investment, the import in production and semi-finished products is increased. This leads to rising trade imbalance.

(3) Japan mostly exports machines, metal products and chemicals (the above-mentioned products account for 90% of total exports) to Thailand while Thai mainly exports foods such as frozen shrimps, boneless chicken, sugar and raw materials to Japan. However, Japan exports of the machinery products to Thailand have significantly increased.

3.2. Investment

In 2007, Japan investment in Thailand was about 3.6 % of the total Japanese foreign investment, and Japan is the number two importer in Asia after China. Conversely, Japan is the number one imports in Thailand, considering investor's request for investment support to the BOI. The percentage of Japan's imports was about 29% of the total foreign investment in Thailand ahead of the United State (20%), of Singapore (20%) and of Nederland (18%).

Table 1: TradingValuebetween Thailand and Japan 2010-d from 2007starte



(Unit: Million US.Dollar)

Year	Total Thailand Trade to Japan			Thailand's Exports		Thailand's Imports		Trade Balance
	Value	Proportion	%Δ	Value	%Δ	Value	%Δ	
2007	46,500.58	15.83	10.57	18,119.05	10.58	28,381.53	10.57	-10,262.48
2008	53,627.89	15.02	15.33	20,093.64	10.90	33,534.25	18.16	-13,440.61
2009	40,747.11	14.24	-24.02	15,723.68	-21.75	25,023.42	-25.38	- 9,300.75
2010	58,271.84	15.43	43.01	20,415.71	29.85	37,856.13	51.28	-17,440.42

Source: Information and Communication Technology Center, Office of Permanent Secretary ministry of commerce

3.3. Trade Balance

(1) Total trade:

Japan is the important trading partner of Thailand. In 2009, Japan became the largest trading partners of Thailand in East Asia. Both of Japan and Thailand have the average trading volumes, which were about 46,961.33 million US. Dollars in the past three years (2007-2009). After that, in 2010 (Jan - Dec) the average trading volumes of two countries was approximately 58,271.84 million US. Dollars. This amount was higher than last year (2009) that was equal to 43.01 percent.

(2) Export:

In 2009 Japan turned into the third exporting market of Thailand, whereasthe UnitedStatesand

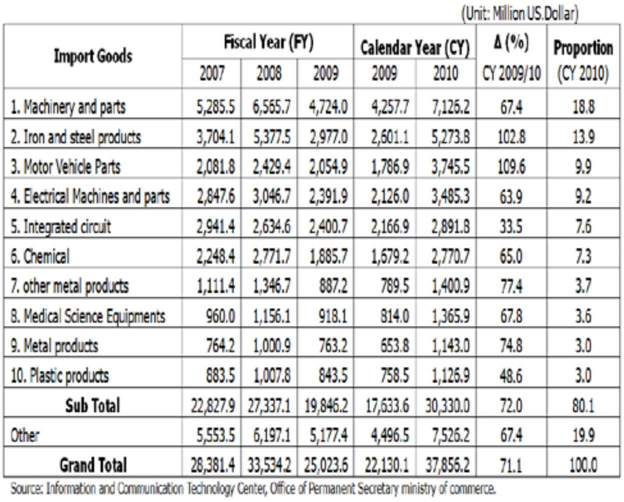
China have lower exportscomparing with Japan. Thailand's exports has the total quantity of 17,981.58million US. Dollarshree for the past t Dec) Exportion -In 2010 (Jan .(2009-years (2007 between Thailand and Japan was about 20,415.71 million dollars increasing from 2009 which has the average yearly trading volume of 15,723.68million US. DollarsTherefore, the percentage difference . .een 2010 and 2009 was about 29.84betw

(3) Import:

In 2009, Japan was the number one source of merchandise imports to Thailand .In the past 3 years (2007-2009), importing goods from Japan to Thailand has the average value of total goods, with about 28,979.78 million million US. Dollars. In 2010, there was the number of importing goods from Japan to Thailand, with about 34,573.78 million US. Dollars. According to the table, the growth rate of Thailand's imports in 2009 was about 25,023 million US. Dollars, which was equal to 51.28 percent.

There were some goods exported from Japan such as, machinery, other components regarding vehicles, electrical machines and etc.

Table 2: Japan Import Statistic



(Unit: Million US.Dollar)

Import Goods	Fiscal Year (FY)			Calendar Year (CY)		Δ (%)	Proportion
	2007	2008	2009	2009	2010	CY 2009/10	(CY 2010)
1. Machinery and parts	5,285.5	6,565.7	4,724.0	4,257.7	7,126.2	67.4	18.8
2. Iron and steel products	3,704.1	5,377.5	2,977.0	2,601.1	5,273.8	102.8	13.9
3. Motor Vehicle Parts	2,081.8	2,429.4	2,054.9	1,786.9	3,745.5	109.6	9.9
4. Electrical Machines and parts	2,847.6	3,046.7	2,391.9	2,126.0	3,485.3	63.9	9.2
5. Integrated circuit	2,941.4	2,634.6	2,400.7	2,166.9	2,891.8	33.5	7.6
6. Chemical	2,248.4	2,771.7	1,885.7	1,679.2	2,770.7	65.0	7.3
7. other metal products	1,111.4	1,346.7	887.2	789.5	1,400.9	77.4	3.7
8. Medical Science Equipments	960.0	1,156.1	918.1	814.0	1,365.9	67.8	3.6
9. Metal products	764.2	1,000.9	763.2	653.8	1,143.0	74.8	3.0
10. Plastic products	883.5	1,007.8	843.5	758.5	1,126.9	48.6	3.0
Sub Total	22,827.9	27,337.1	19,846.2	17,633.6	30,330.0	72.0	80.1
Other	5,553.5	6,197.1	5,177.4	4,496.5	7,526.2	67.4	19.9
Grand Total	28,381.4	33,534.2	25,023.6	22,130.1	37,856.2	71.1	100.0

Source: Information and Communication Technology Center, Office of Permanent Secretary ministry of commerce.

(4) Trade deficit:

From the past 3-year of bilateral trading between Thailand and Japan, the annual trade deficit incurred for Thailand was 10,262.48 million US. Dollar by which the deficit was 9,300.75 and 17,440.42 million US. Dollars at the year-end of 2009 and 2010 respectively.

4. TRADING PROBLEMS BETWEEN THAILAND AND JAPAN

4.1. Export

(1) The regulation for agricultural products imposed by the Japan has been contributing to limitation to Thai export. For this regulation, Thai agricultural products to pass the hazardous residue testing conducted by Department of Agriculture prior to the export. This trade barrier impacts variety of product included;

(a) Plantation: Mangos teen, durian, mango (varieties except Public A), grapefruit, tamarind, lychee, Longan, maize, okra, asparagus, ginger, pepper

(b) Other vegetables including celeries, cilantros, sweet basils, basils, finger grasses, tree basils, Kitchen Mint, herbs, gotu kola, peas, cabbages, Cha Om, acacia pennata, kaffir lime leaves, water mimosa, lemongrass and okra/lady's finger.

(2) The tariff has been imposed, as a barrier to Thai export, for such agricultural product as cassava flour, rice, and canned pineapple as well as such industrial product as synthetic rubber, silk woven-fabric, and shoes.

(3) In the 4th phrase of GSP-tax exempt contract (from May 1st, 2001 to March 31st, 2011), Japan degrades this tax exempt for most of industrial products making many exported industrial products to bear higher GSP-tax. This includes Insulated electrical wire, jewelry, plastic, glutamic acid, leather ware, Sorbitol dextrin, Wood decoration, and synthetic fiber.

(4) The complexity and stringency of the rule for "source of origin" under GSP regulation was not matched to the production platens this day.

5. THAILAND EXPORT/IMPORT FRUIT'S LAWS

In Thailand, law creators have got continuously developed both of form and regulation used in commerce international exportation. However, new trading system of Europe and the United States of America (USA) are now becoming the biggest market of the world. Hence, Thai exportation should rely on western market and use western market as a case study to improve the domestic economy.

Thailand is considered to be reserved supply for fruit production. Fruits, such as durian, mango and etc., provide large amount of products not only from one province but also other province in Thailand. Hence, Thai people can consume these products throughout the year. Simultaneously, there are also a number of products enough to export to abroad. However, the largest fruits consumer markets of Thailand has several countries where are located in both Asia and Southeast Asia e.g. China, Japan, Hong Kong and Taiwan. There are the most five popular fruit of Thailand consists of mango, pomelo, banana, mangosteen, and durian. These fruits are widespread accepted by international consumers [8].

Thai fresh fruits have only 6 types permitted by Japan importers such as mango, mangosteen, durian, banana, young coconut and pineapple [8]. Thailand will have a better chance of exports if Thailand signs free trade agreement. It is called "Japan-Thailand Economic Partnership Agreement (JTEPA)". As a result of making contract, Japan consents to reduce export tax of fruits e.g. mango, mangosteen, durian and coconut. [8].

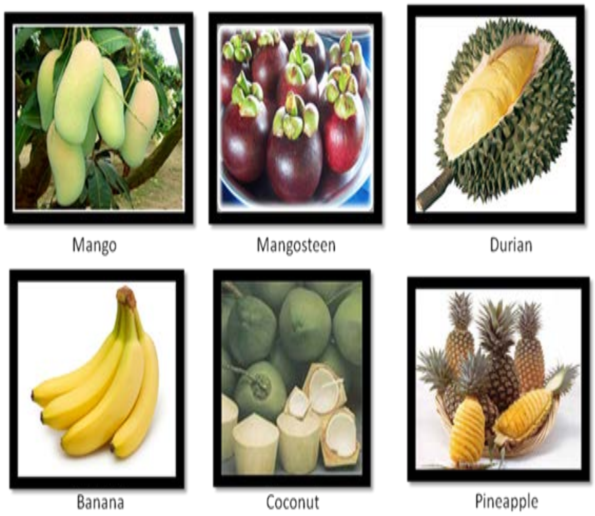


Figure 2: Six types of Thai fresh fruits are permitted by Japan importers

In recent year, Japan has imported large amount of fresh fruits from oversea market. The total amount is 200,000 million yen. More importantly, Japanese consumers have always been sensitive to the standards on food safety. Therefore, fruits should not have remaining toxicity during manufacturing process.

6. PLANNING

The most standpoints of Japanese businessman think that Thailand can be the industrial base to producing Japanese stuffs. There are some aspects of Japanese businessman, which can be used to pick up Thailand as an industrial base;

(1) Japan has chosen Thailand as an industrial base because Thailand is like a potential hurb of region. Moreover, efficient infrastructure of Thailand is also considered in terms of investment, whereas other countries (such as Laos, Cambodia) do not have those good characteristics.

(2) Japan has chosen Thailand because Japan industrial base don't want to rely on the products produced from only one country (such as China, Taiwan, Hong Kong and so on).

(3) Herb used to make medicine and raw materials used in cooking are useful things for the enrichment of human body. For example, garlic can reduce coronary heart disease.

(4) There are some ideas about the importation of products produced from Thailand to Japan. Some Thai productions like fruits, which can be transformed into products used for different parts of the body. That is because Japan has an interest in the care of mental and physical health, by using natural products. Onsen and Spa is an obvious example in Japan tradition. It (onsen & spa), can reduce people's stress. Products that help to reduce stress is taken from natural water. In case of Thailand, there are some products derived from nature such as the coconut. It is extracted from coconut juice in order to anoint on user's body. The advantage of coconut juice is the refreshment on skin. Moreover, it can be extracted into hair care products. Sometimes user can use natural products as diet pills (weight loss pills). Finally, products like cream and soap made from (fruit and herb) can maintain different parts of user's body as well.

REFERENCES

[1] Sorachai P.: *Maketing Research Methodologies*, June Plub Riching (2007), *(In Thai)*

[2] Nechpanna Y.: *Modern management*, 7th ed, Tripple group (2010), *(In Thai)*

[3] Hosokawa H.: *Dai Kyosou Jidai No Tsushou Senryaku*, The foundation for the promotion of social sciences and humanties textbooks project (1999), *(In Thai)*

[4] Teerasorn T.: *Marketing Communication Perspective*, 2nd ed,, Chulalongkorn University (2009), *(In Thai)*

[5] Alan R, Phil C, and Peter B.: *The Handbook of Logistics and Distribution Management*, 3rd ed, Asia Press (1989)

[6] Kasame C.: *Transportation Systems and Operations*, Chulalongkorn University (2012), *(In Thai)*

[7] Aaron C.T. Smith.: *Introduction to sport marketing*, Elsevier Ltd, (2008)

[8] Kumnai A.: *Transport management*, 3rd ed, Focus and plublishing (2011), *(In Thai)*

IMPLEMENTATION OF KANBAN TECHNIQUE WITHIN THE TOTAL FLOW MANAGEMENT MODEL

Mosè Gallo[(a)], Teresa Murino[(b)], Emanuele Guadalupi[(c)], Liberatina C. Santillo[(d)]

[(a) (b) (c) (d)] Department of Chemical, Materials and Industrial Engineering University of Naples Federico II P.le Tecchio, Napoli, 80125, ITALY

[(a)]mose.gallo@unina.it , [(b)]murino@unina.it, [(c)]emanuele.guadalupi@unina.it [(d)] santillo@unina.it

ABSTRACT

This work is focused on presenting the implementation of the kanban tool in the more general context of the Total Flow Management model. This imposes to radically restructure, in a lean perspective. internal and external logistics processes, considering the whole supply chain, through which it is possible to get a global optimization of enterprise's operations.

The Total Flow Management model will be explained, analyzing the pillars that sustain it. A simulation based implementation of this model in a typical automotive production plant has been presented.

Keywords: lean production, Kanban, Total Flow Management, KPI, flexible production,.

1. INTRODUCTION

The instability of competitive markets and the strong competitive dynamics represent tough challenges for companies that want to survive and generate the decisive push proposals to new systems of production and management, which often completely overturn the existing reality.

Comes the need to abandon the "old model" to allow management to cut costs and focus on levers of competitiveness, emphasizing the role of the consumer. Change in a clear way the focus management, no longer solely directed to the productivity but also on the quality and punctuality. Western companies began to turn their eyes to innovative production models, born in Japan after World War II when the Japanese industry conditions was exhausted from the devastation of World War II

In this new approach is the customer who defines the volume of production and the timing, the requirements are then determined downstream and "pull" the entire production system. With a view to reducing costs and increasing the delivery precision, one aims to break down any waste and to act in a continuous manner. It then tries to minimize the stocks of both raw materials and finished goods, reduce lead time by coordinating the entire production system, paying particular attention to elements of reliability and quality of both the production process and the product or service to the customer .

The areas of competitiveness to be investigated can be multiple and dependent also on the type of organization concerned. Certainly, in the current context, it is possible to think that the majority of businesses, goods or services, must in any case pay particular attention to these areas:

- Customer satisfaction and on-time delivery
- Efficiency and quality of processes and product
- Innovation and Growth
- Human resources, employee satisfaction, communication
- Economic and financial results

For each "area" defines the Critical Success Factors (CSF, Critical Success Factors) and then extrapolated a number of Key Performance Indicators (KPI Key Performance Indicator). You can then assume different levels of detail related to specific levels of management:

1. Critical Success Factors: are defined critical areas of business and strategic indicators specified at strategic level
2. Key Performance Indicators: identifies the performance-critical business process-oriented control of operations operational level

Through the CSF calibrates the entire measurement system, focusing mainly on all those strategic variables able to determine business success. Identify the strategic variables you need to build the benchmarks that need to make comprehensive information and comply with certain qualitative and quantitative characteristics. Ultimately, a model of analysis of business performance effectively will have to consider, understand and dynamically analyze all the factors that participate directly or indirectly in the performance of the company to provide an interpretative model of primary and secondary processes and the interactions between them, thus ensuring a monitoring of activities. This model also provides a valuable tool to identify the right cause-effect relationships between the performance recorded as supporting the activities of process improvement and customer satisfaction. In this sense, the measurements are therefore the basis for any attempt to improve the company, especially in optical Kaizen Continuous Improvement, because they offer us a clear idea of the current state and, through the definition of the target, allow you to show the direction to a optimal management, aimed at achieving the objectives.

The goals may be set for each organizational level and monitored through measurements, therefore, can show different levels of detail, from more or less simple (related to operational processes) to measurements consolidated and / or aggregated.

The remainder of this paper is organized as follows. In Section 2 the total flow management model is presented together with a perspective on lean production concepts. In Section 3 the implementation case study is introduced. In Section 4 an analysis on the project's results is conducted. Finally, Section 5 concludes this work.

2. LEAN PRODUCTION FLOW AND PULL

The Lean is recognized as one of the most popular management tools to deal with the strong competitiveness in manufacturing companies or logistics.

In this context, the primary need was to be able to get the maximum output with minimum use of resources, so the goal is to maximize the efficiency of processes, operations. For this to be achieved it is necessary to effectively eliminate all types of waste (muda) and focus on the processes that create value for the customer. It is only the latter that should be concentrated efforts because they are the moments when you generate wealth and competitiveness for the company. The activity of elimination of waste is carried out in the whole production process and follows the principles of Kaizen, a word derived from the Japanese words Kai and Zen (renewal and off, respectively) mutated in the West in "continuous improvement. The strength of this approach lies in the search for simple solutions that can be implemented quickly without the need to resort to strong structural or technological investment.

The focal point of the new organizational model is the creation of a flow that is "pulled" by the actual customer orders and continuously improved. In order to be effectively managed such a system must be complied with the principles:

1. Quality first
2. Elimination of waste
3. Orientation to the Gemba
4. People development
5. Visual standards
6. Observe the processes and results
7. Lean Thinking

Quality management is one of the elements that have more benefits and universally accepted. It was the statistical Deming, Juran and Ishikawa to emphasize how a strategy based on quality could be turned into a competitive advantage and efficiency. The quality has to be defined according to what the market demands, the goal is customer satisfaction which requires certain services at competitive prices. This approach uses a variety of analytical tools based on objective data, so you can better measure and understand the phenomena to be evaluated, and extensive use of statistical tools such as correlation analysis or control charts and introduce, from the methodological point of view, concepts such as cause and effect diagrams and cycles of Deming (or PDCA, plan-do-check-act), which will form the basis of continuous improvement.

The real fundamental concept of Lean to achieve competitiveness and excellence is the elimination of waste, the muda, the causes of variability, walls, and losses, walls, through the Kaizen approach. Excessive volatility is determined by the lack of stability and reliability, resulting in complications for the process control and implementation of timely intervention to limit the damage. For walls are the difficulties that determine loss of time and energy, for example, loss of time can be achieved if the workstations are not ergonomic and therefore cause operators to make unnatural movements may be, energy losses due to lack of satisfactory yields . It is usually placed particular attention to the walls because they may present risks of injury.

The third principle pertains to the Gemba, a Japanese term for the shop floor, or the workshop, in the Kaizen approach the place where the improvements are made. The strength of continuous improvement is to be found in the concrete activities carried out in the field, directly addressing the difficulties identified without deleterious loss of time. A typical resolution of a problem involves the direct observation of the reality on the gemba, the collection of data and information on observed evidence and a search for solutions through the brainstorming group, formed mainly by staff directly employed in criticism. In this way, the improvement comes from below, is not imposed, it is therefore more understandable, shared and achievable. Hence the emphasis on the involvement of individuals in kaizen activities. Working in a team, and invest in the training of people increases the quality of service and reduce costs. The first step to changing habits is enable people to create and manage an improvement, making them aware and conscious of waste and inefficiency where it takes place. Therefore, any person within the company, from top management to the shop floor, must be fully involved in order to obtain a truly efficient.

The visual standard is expressed in the concept that "one picture is worth a thousand words and a standard is the most efficient way to accomplish a task". If an activity is not standardized is exposed to variability, depending on the different ways to carry it out by the various operators. Create a standard with visual support is more immediate and simple to understand, compared to a standard descriptive. The definition of a standard work involves a study of the methods of conducting business, through a careful analysis of the processes and the establishment of targets for improvement, and represents an optimization of the actions of workers, according to a certain cycle time and line with the flow of materials.

Lean thinking is the real principle at issue. Pull Flow means to organize the whole supply chain (or for simplicity the internal logistic and manufacturing flows) with a view to optimize the flow of material and information. To achieve this, the emphasis must surely be placed on the elimination of muda or the expectations and demands of the customer. The idea of the flow of material, ideally a one-piece flow, is often counteracted by the conviction of economy of the production batch. It may seem absurd to even work on individual orders, consumption, to produce individual units, because they might be too small and synonyms, erroneously, of inefficiency. The experience of the Toyota Production System demonstrates what can be winning the introduction, in a radical way, this new method of organization of operations, in terms of ongoing results.

Lean and Kaizen methodologies learned the issue is to define the sustainability of the model identified. For this to happen its implementation must proceed through a single overview of the entire supply chain, from suppliers to customers, keeping in mind what is the direction and purpose which you want to strive

2.1. Total Flow Management (TFM)

Total Flow Management is a management model for integrated logistics, Lean in optics, applied to the entire supply chain of a given company is manufacturing and services. Applying the model in the first instance you want to create an internal pull flow in maximum operational efficiency and free of non-value added (NVAA), then consider the extension to the valley (the delivery side, ie customers) and upstream (the source side, ie the suppliers). The flow of materials can be considered as a repeated sequence of four types of transaction, namely transport, inspection, waiting and transformation, the only real value added activities. The main objective is the reduction of total lead time, as measured coverage of stocks, eliminating muda process, creating benefits in terms of cost reduction and working capital, increased productivity and quality in order to achieve a higher level of service provided to customers and to improve, therefore, the satisfaction. You reach these objectives through the redesign of logistics and implementation of the one-piece flow loop, driven by actual customer orders, which define the so-called customer takt, quantifying the need for restocking of inventories and production volumes. Therefore orders the production or distribution are no longer based on predictive models, which are still useful for capacity management.

2.2. Logistic Loops

The Total Flow Management model divides the supply chain a loop a series of logistic (logistic pull loop, LL), information and material, by the final consumer to the supplier. Based on a simple supply chain, including manufacturing company with a client and a distribution center warehouse supplier of raw materials, it is possible to identify three main types of logistic pull loop:

- Withdrawal finished products
- Production
- Collection of components and raw materials

The first LL is generated when a customer buys a product and inventory management models of the distribution center identified the need to send a replenishment order by generating a flow of information, which will be sent to the production process, and having tried, and a material flow (control, selection, collection and transport of the product) from the finished goods warehouse. The management objective for this loop is the creation of a flow that eliminates the stages in which the material and information remain pending. The major constraint is the availability of the product at the time of sampling and frequency of transport.

At this point it started the production of LL that can respond to a need to restore inventory levels (make to stock, MTS) or the creation of an ordered product does not exist (make to order, MTO). Processing the information available are reworded as production orders, scheduled times and needs, considering the small batches of production and the BOM and sequencing the operations on the first line pacemaker (the line bottleneck that affects the capacity of production of establishment). Meanwhile the client is meeting their demands by resorting to the warehouse stock. In accordance with the schedules the transport of materials or components and the production must take place within the time and in the right places, and it is at this point that concentrate most of the opportunities for improvement aimed at the pull flow.

The third type of loop logistic concerning the sampling of materials or components, i.e. the supply of production processes. The primary objective is to pick up and supply lines in a synchronized way, restoring the inventory levels at the point of use, using kanban systems, physical or electronic call. For the need of just in time several problems arise when increases the number of different components and materials required. The total flow management model aims perfect synchronization using physical systems supply and standard supply that can provide immediate answers.

An additional loop logistics, intermediate, could be defined by the production of supplied pre-assembled in the correct sequence and in the right time at the final assembly lines.

Analyzing the supply chain as a set of logistic loops you can easily identify the difficulties in managing the system. For example in the automotive industry a greater complexity is revealed in the final assembly (i.e. the loop of production) and in the collection of materials and components, evidently due to the large abundance of these components and the need of synchronization thrust, given the nature often automated lines. Every industrial has still endemic in dependence of the complexity as well as the variety of materials managed also by the type of distribution flows.

2.3. The Total Flow Management Pillars

The flow of information and materials identified through the analysis of logistic loop can be grouped into three macro flows, the pillars of TFM (Figure 2):

1. Production Flow
2. Internal logistics flow
3. External logistics flow

In a Total Flow Management optics each of these flows must be analyzed and optimized by applying the principles of Lean and Kaizen tools.

The first step is to improve the production flow, which ranks as objectives the implementation of the one-piece flow, increased flexibility by adjusting the setup and an increase in operational efficiency and supply. The improvement actions are divided into these categories: Redesign of the layout and lines in optical one-piece flow; redesign of the perimeter of the line for the efficiency of supply; definition of standard work for operational efficiency; SMED technique for flexibility in setup, low-cost automation to reduce the walls.

The second step is the optimization of internal logistics flows, including all the movements of small containers inside the plant but also the related information flows. In this case the improvement actions are divided into the following areas: supermarkets to increase the efficiency of the material withdrawals; mizusumashi system to simplify and streamline internal transport of the material to the point of use; synchronization as a co-ordination between supply and production leveling productivity of the lines and equipment in relation to the takt time, pull production planning in accordance with the actual customer orders.

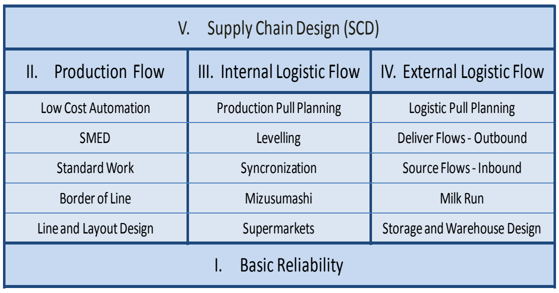
Finally, the third step is the optimization of external logistics flows, i.e. the handling of materials and products, generally parceled out in pallet from the factory to customers and suppliers to the plant. In this case, the categories of intervention are: redesign of the stores and warehouses; creation of the milk run, i.e. an external flow of transport; physical flows in (inbound) and outgoing (outbound) through small containers and pallets; planning pull the logistics to handle the material withdrawals according to the royal orders of consumers.

To complete the Total Flow Management model there are two basic pillars:

- Basic reliability
- Supply chain design

The first one is related to the concept also developed in the Toyota Production System of basic stability, for which the creation of a stream cannot do without a certain level of stability in terms of 4M (Labor, Equipment, Materials and Methods). The basic reliability analyzes the reliability of the available resources and how effectively they are compatible with the realization of the total flow. For efficiency is evident the importance of conducting that analysis beforehand, so as to have a basis on which to develop the other pillars of the TFM model.

The supply chain design is the design instead of the entire logistics system, through the standard tools such as the visual stream mapping, which is useful to represent flows of material and information, or spaghetti chart, which is a graphical representation of the physical flows.



V. Supply Chain Design (SCD)		
II. Production Flow	III. Internal Logistic Flow	IV. External Logistic Flow
Low Cost Automation	Production Pull Planning	Logistic Pull Planning
SMED	Levelling	Deliver Flows - Outbound
Standard Work	Syncronization	Source Flows - Inbound
Border of Line	Mizusumashi	Milk Run
Line and Layout Design	Supermarkets	Storage and Warehouse Design
I. Basic Reliability		

Figure 2: The Total Flow Management model

2.3.1. The Basic Reliability Pillar

The first pillar of Total Flow Management Model is the basic reliability you need, as already mentioned, to develop a stability in terms of manpower, machines, materials and methods (4M). To get a basic reliability is necessary to develop a sustainable capacity for change, the company must initiate a culture that is focused on continuous improvement through the implementation of growth-oriented solutions at the factory (Gemba), in accordance with the principles Kaizen, must i.e. to become a learning organization, or a learning organization. This concept, defined as learning by doing, is expressed in the fact that to get good results is necessary to develop proactive attitude in Gemba in order to try new ideas without bias on the results and judge them only after they have been put into practice, making it valid through 'implementation, so on tangible results and tested. To this end it is essential to the involvement of staff so that the drive for improvement has many sources as possible and often the suggestions of those who are directly involved in the process can be a winner. Before you try to change or improve any activity or process is therefore a clear need to change the corporate culture in a new "mentality Kaizen" and spread awareness of the effectiveness of the tools used.

The first step in the change capability is the development of skills related to a widespread recognition of muda and identify projects, assigned to functional groups, committed to the elimination of muda critical variables, i.e. the occurrences that define waste or inefficiencies in the various processes.

The critical variables muda for the basic reliability can be divided into four groups (labor, machinery, materials and methods), each of which must be analyzed in order to find the major causes that can block the flow. Among the most common causes, however, that may vary from company to company, can be summarized as follows: punctuality and absenteeism, machine availability, percentage of stock out, percentage of defects, resistance to change.

For the basic manpower reliability aspects are focal punctuality and absenteeism. It is important to monitor these critical variables regularly and pay attention to what's going on, because they contrast sharply the creation of a stream. The lack of punctuality, such as delay in the beginning of the activities of 5 or 10 minutes or long pauses, can pose serious problems in the case of flows of material to a minimum waiting time. Absenteeism do completely unexpected, however, can skip a task flow, so some reliability measures need to be taken into account. Surely the improvement activities of these critical variables must involve supervisors and human resources, so that they can directly identify the root causes of absenteeism, define targets for the reduction and measures to achieve them.

One of the most important tools for improving the reliability of the workforce is the SDCA (Standardize, Do, Check and Act).

First you have to create the standard that goes to solve a particular problem. It is important, fundamental, create the standard listening to the views of employees and operators, because they are the people who are most in contact with the business, who know better. Then you have to transfer the standard (phase do) through the training, lessons, in line with the principle of learning by doing, that it is always better that the activity is taught directly in the field, so that the new habit is established directly . New habits take time to be fully digested, so there is this need to reinforce the theory through practice. To ensure that a new standard is actually implemented is necessary to activate a control stage (check), directly into the Gemba, doing audits at regular intervals so as to monitor the implementation and maybe, where there are deviations, establish actions corrective action. If everything is working you can actually start the task according to the standard defined (stage act) and find the possibility of extension and restart the cycle SDCA with a view to Continuous Improvement (Figure 3).

The use of the standard is also important to create a share of professionalism that are conferred on the farm and as a reference in case they found new problems. When we identify a problem are interviewed supervisors and operators, to verify the existence or not of a standard decisive. If the supervisor and the operator confirm the existence of the standard for the resolution of the problem means that the standard is ineffective and needs to be redone, because despite the fact that it is not applied. If the supervisor confirmation but instead the operator is not aware of the standard you have to start a Coaching. In the case in which the supervisor is not to be aware of the standard, while the operator applies, then it must simply formalize. If both ignore the standard must achieve it.

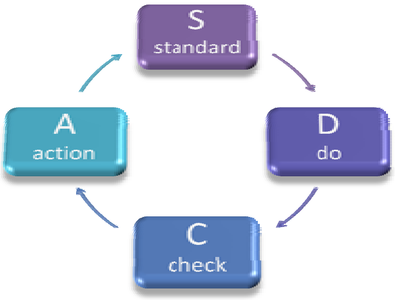


Figure 3: The SDCA cycle

Ultimately, the standard is the first step for the improvement and is the most efficient way to perform an activity.

But when the OEE (Overall efficiency effectiveness, global measure of effectiveness) is very low, it is desirable to carry out projects of basic machine reliability. In this context it is necessary to first define the Equipment Operating Time, or the processing time in which planned and it is expected that the resource work in a day, and is the basis of the takt time (defined as the ratio of operating time / demand daily). The operating time is divided into two aliquots: the time of actual use and losses.

Losses due to the availability are the main causes of inefficiency and may be due to breakage, training, maintenance and cleaning (planned), repairs (unplanned), changeover and adjustments. Certainly if the resource has many stops for breaks or for unexpected stops will have a negative effect on the basic flow reliability. According to the TFM model in order to start a project for Total Flow must start from an availability of 80%, as an average, associated with a low standard deviation.

Other types of detectable inefficiencies are due to leaks for performance, i.e., cycle times slower than in the takt, micro stops (<10 minutes), start up and shut down, or losses for quality, i.e. errors and defects, scrap and rework.

To increase the OEE of the plant is recommended going to act first on that resource pacemaker, or the resource with its capacity is capable of affecting the ability of the entire flow, production or logistics. In optical Kaizen to improve the basic reliability of a machine a quick way is to go through a workshop focused Gemba. We must first identify a list of 10 losses for the machine more problematic and begin to attack the list from the first point. If a significant cause of loss is a specific type of failure must go back to the root cause through the technique of "5 WHY?" Or use an Ishikawa diagram. Identified and including the reasons you can search for countermeasures through a process PDCA (Plan, Do, Check Act), with the support of a group of people who have full awareness of the problem (the variable muda) and he will work until it is resolved. For faults is suitable search for persons of the group between the employees of the maintenance function, while for the defects will be referred to employees of the function of the production and quality control. In short we must identify professionalism appropriate to the resolution of the problem identified. It can well understand the usefulness of this method, which allows the resolution of big problems without resorting to onerous investments for the renewal or even the replacement of the equipment available.

Problems related to basic material reliability occur if there is lack of supply of components or raw materials needed to begin production activities. The problem may be due to an external supplier, unreliable or late, or a disservice internal logistics due to the unavailability of material where it requires the use. An immediate resolution might be to create the buffer I guarantee a pull flow, however, the buffer can be seen as muda, returning to the general principles of Lean. Very often the problems of procurement of materials are related to deficiencies in internal or external logistics processes that are resolved with the various tools of the TFM dedicated, however it is always useful to seek ties in order to achieve systemic solutions. The various projects Kaizen, based on the top ten list and the PDCA cycle, which are open to attack losses in this area are then assigned to groups that, by the principle of competence, should be formed mostly by staff function and Logistics Planning, directly employed in the management of orders, storage, handling and transportation of.

The last class of the critical variables muda is attacked with projects of basic reliability method, which is closely linked to all the losses that may hinder or stop the flow of material or information especially when you try to remove safety stock or buffer, or when you want to change the methods of managing the flow or replace them. The occurrence of losses by methods usually, however, the problems have to do with constraints on quality or time. A usual loss, severe, linked to quality is due to the high variability that cannot be used for a specific machine (check through the OEE) but random problems which have their foundation in the lack of solidity of the methods adopted. For the validation of the methods we resort to the focused group, with specific skills in a stable phase of the process, otherwise you run the risk of contamination of the data due to the normal fluctuations of the transient.

2.3.2. Production Flow pillar

The second pillar of the TFM model is the production flow that has as its objective:

1. Creation of one-piece flow, ideally one piece at a time a production flow, from raw materials to finished product
2. Minimization of muda of motion of the operators, through a restructuring of the perimeter line and standards work
3. High customization of products, thus high flexibility in relation to small batches using tools such as the SMED
4. Simplification of processes and automation

Achieve the goal of one-piece flow production means redesigning the layout, organizing the resources in such a way as to generate a continuous movement from raw material to finished product, then follow the correct sequence of operation by eliminating all activities that do not add value, NVAA (not value added activities).

The TFM model for the realization of the production flow identifies five areas for improvement: the design of lines and layout; redefinition of the scope of the line; utilization of standard work, SMED implementation, use of low-cost automation.

2.3.3. Internal Logistic Flow pillar

This third pillar of the TFM has the objective of optimizing the flow of transport of the containers inside the plant, such as organization of supply line all the parts needed for production according to the cycle time, linked to the takt time. At the same time must be optimized information flow starting from the actual customer orders, which must be converted into production orders as quickly as possible. The reactivity of the system can only be obtained through an integration of production and logistics flows, in order to proceed in a synchronized manner and fulfill the requirements just in time.

The traditional method of organizing the logistics, the push flow, is based, in summary, on the minimization of transport that involves an internal supply lines with lots of large quantities, the use of forklifts to move pallets and planning of large production orders to minimize changeover time (Figure 4).

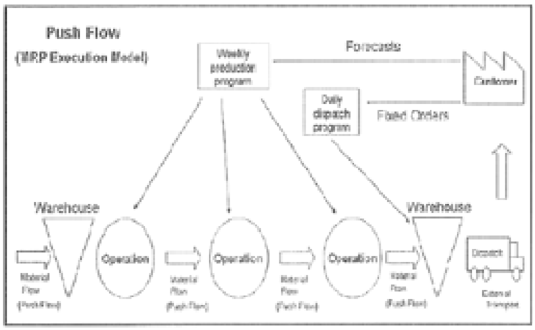


Figure 4: A schematic representation of a Push Flow

By contrast, the method of the pull flow instead aims at supplying the lines of containers of the right size to maximize the efficiency and flexibility, organize withdrawal areas of materials in an efficient manner, using standardized transport routing with fixed times and plan production according to customer orders (Figure 5). The implementation of this pillar includes five areas: the creation of simple supermarket for the organization of the material withdrawals; mizusumashi that optimizes the transport of containers in the establishment; synchronization between production and supply; leveling as scheduling of production according to the process bottleneck, and finally the pull production planning to set the right production capacity in relation to the needs determined by the customers.

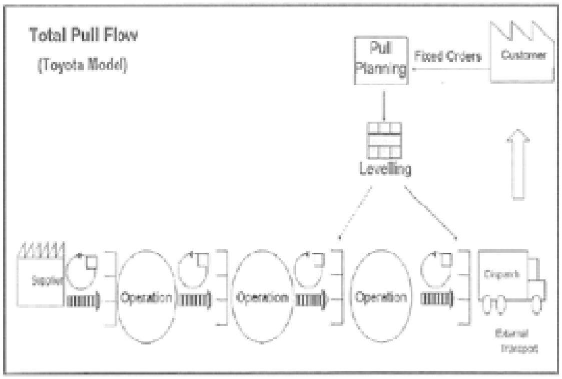


Figure 5: A schematic representation of a Pull Flow

2.3.4 External Logistic Flow pillar

After creating the conditions necessary to pull flow and you have optimized the internal logistics processes will start the fourth pillar of the TFM model focuses on the organization Lean logistics processes external, i.e. the loop logistics delivery flow (cash flow from the factory to customers) and source flow (from the suppliers to the company). In the previous pillar it has been discussed the need to recreate the flows in unique containers along the internal supply chain of supermarkets realizing and managing the movements and orders through the tools described, in order to eliminate muda. In order to integrate the processes of internal and external logistics supermarket there is a final, falling within the defined routing for mizusumashi as an interface with distribution warehouses, from which the withdrawal is made from the customers. A similar argument can be made for upstream suppliers.

The External Logistic Flow in essence is oriented to maximize the delivery precision bringing to 100% the level of service provided to customers and at the same time minimize the amount of stock throughout the supply chain. The traditional approach defines the orders of algorithms based on logical push, ie using MRP schedules, based mostly on assumptions and estimates, and setting high standards for safety stocks. The approach of the TFM model instead complete such a system, however, by placing the emphasis on the importance of physical supermarket, orders and fixed real consumption, optimization of the flow, using standard work reliable.

The proposed objectives are the minimization of stocks and expectations, achieving a delivery on time, on spec and total, complete elimination of muda of movement and handling and the minimization of the total cost of logistics. These objectives can be pursued by acting on the five domains of intervention of external logistics: planning of warehouses and stores; realization of the milk run; resource flows; delivery flows; pull of external logistics planning (figure 6).

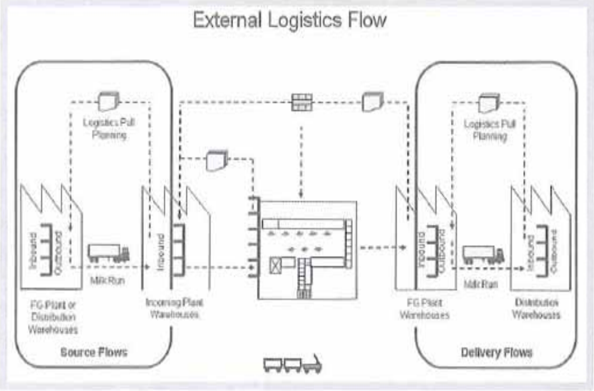


Figure 6: A schematic Logistic Loop.

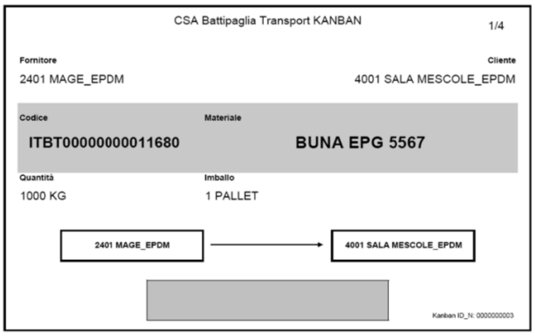
3. CASE STUDY - IMPLEMENTATION OF KANBAN IN TFM VIEW AT CSA BATTIPAGLIA PLANT

The production company (CSA - production and extrusion polymers for automotive) is framed as Repetitive Manufacturing, i.e. a mass production of a large number of identical units, characterized by families, in a continuous flow of lines and work centers standardized.

Defined the area of supply (supply area), the supply strategy and the cycle control (Control Cycle). In optics Total Flow Management was necessary to redesign the layout of the department in a more functional approach to the new, then create the conditions for an optimal management of the flow of materials. The size of mobile containers is equal to the amount contained in a kanban and the location has been chosen as a result of a proposed Workplace Organization to optimize dosing operations.

Figure 7: A picture of storage area before and after the change

Defined in the storage area of the supply source, one must locate supply strategies most suitable to be adopted and for the control of the cycle are defined tabs of handling, i.e. the kanban sampling, through which you authorize the movement of material between the two logistics centers. The card is associated with a standard container and includes a variety of information: the component code; capacity of the container; card number to keep track of the material; identification of "supply source" and "demand source" cycle, identification of the location in the supermarket.



CSA Battipaglia Transport KANBAN 1/4

Fornitore 2401 MAGE_EPDM	Cliente 4001 SALA MESCOLE_EPDM
Codice ITBT00000000011680	Materiale BUNA EPG 5567
Quantità 1000 KG	Imballo 1 PALLET

2401 MAGE_EPDM → 4001 SALA MESCOLE_EPDM

Kanban ID_N: 0000000003

Figure 8: An example of “kanban” label.

The flows of material of this type have therefore to cover such daily production and can be considered constants beyond the processing schedules. Optimizing the application of the consignment stock.si two classes of materials handled according to the classic and the kanban kanban event . The classic kanban is used for small ingredients, divided by the constant and repetitive consumption during production campaigns. In view of this condition, the number of yellow kanban for each code is assumed proportional to the amount consumed in particular in two shifts. The other type of raw materials are managed in a mixing room polimeridi class defined as bulky. Unlike small ingredients depend on the product code, therefore, do not lend themselves to a management with kanban classic, but on them is prepared based on a strategy of replenishment kanban event. The consumption of such materials is not constant and therefore can not be defined control cycles in fixed quantities. The picking list of kanban is sent to the warehouse raw materials through SAP and employees shall supply the supermarket relative Mixing Room.

In the time of withdrawal in order MAGE on tabelliera virtual labels are created to wait for the material (gray) and only when the employee actually kanban delivery in supermarket proceed with the discharge of materials into SAP, which are at this point also accounted for in line with the consignment stock, the label changes to "kanban available" indicated by the green color. When the operator picks up the Banbury Kanban will change the state, passing the empty kanban label in red.

In real time the integrated system controls the actual consumption of the material by the mixer, comparing the amount remaining in the kanban opened with the requirements necessary to shut down the production of the compound produced. If the proposed limit is not respected is sent a warning that immediately initiate a transfer of a kanban to avoid further production blocks. Fixed control loops were defined standard work for picking activities and supply of kanban, fundamental both for the reduction of the time of performance of activities for both the proper operation of the cycles themselves. The strength of Kanban is self-control that determines the requirements according to actual consumption, so the transfer order is sent only upon notice, which in the classical case is represented by the kanban card removed from the container and place it on tabelliera physics, in the case of Electronic Kanban, implemented in CSA, is a virtual input transferred between terminals in the different areas of the plant, according to common protocols. The replenishment process goes through four sequential phases:

1. the replenishment list is transposed by the operator in stock and raw materials is determined by the kanban classic labeled as empty (red dots on the virtual tabelliera) and kanban event labeled as Waiting (gray on tabelliera)
2. the operator picks up the material from the shelves, with the default position and previously studied. According to the principles total each material flow management is located in fixed positions over time, favoring the FIFO and united by frequency of sampling, to minimize movement by the operator
3. the kanban containers are carrying with the aid of forklifts to supermarket kanban, in suitable locations marked with colored labels specific to each code.
4. the attendant after depositing the containers in their rooms at the supermarket pertinent to the interface position Kanban-SAP cult labels gray and red in green, confirming the system the transfer of material from the MAGE Mixing Room. In this phase has the accounting of the materials that are actually sold.
5. For now, the compounds produced are defined by the royal orders scheduled daily by means of MRP II. In the future it is planned to extend the Kanban to downstream processes, in particular a Production Kanban loop through which the extrusion department will regulate the production in mixing room.

4.1 Case study implementation results

The implementation of the Kanban system responds to the need to reorganize the processes of procurement of materials handling and syncing them to the actual production requirements. This transformation aims to apply the principles of the Total Flow Management as the application of Lean methodologies in favor of logistics.

The reorganization of the supply system has allowed a concrete manage orders in Just in Time, fully realizing the expected benefits. Polymeric materials, bulky, are fed in a synchronized manner to the production orders daily scheduled in mixing room, while the feeding of the chemical additives, in quantities less and less expensive, is realized according to cycles kanban for full vacuum. The transfer of materials "pulled" by the orders production allows a reduction of stocks in mixing room, reducing the amount to the size defined by the supermarket planned, calculated on the coverage needs of two rounds. It was therefore moved from a warehouse management of raw materials at a JIT management, with withdrawal from suppliers exclusively in the quantities actually used and consequent economic benefits arising from the type of supply contract in place, namely the consignment stock.

This result translates into a reduction of 57% of the economic value of the stock in the warehouse, by comparing the average value in the range prior to the introduction of the Kanban with the current period corresponding to the first months of activity of the new management. With the old approach was carried in inventory unused material then it would be stuck in storage awaiting consumption and meanwhile other codes had to be replenished more frequently, resulting in an entry and inventory levels undoubtedly higher. With the intervention of the Kanban instead in mixing room comes only what you need and will be consumed at the latest within the next two turns. The resulting material deposited as a stock represents the average amount of kanban in the small tabelliera ingredients out with regard to safety and the amount of WIP EPDM circulating in the mixing room.

The evaluation is made considering the average stock inventory levels before replenishment which, it is recalled, is made one hour and a half before the beginning of the second round daily and covers two rounds, namely the daily production.

The immediate control of inventory levels is possible, as already mentioned, through the use of integrated information systems to SAP. This procedure allows you to undo the need to take inventory, allowing the so-called Zero Patrolling, which translates in lower costs of labor and employment, which in fact can be used for other value-added activities. Other losses of manpower identified and removed or reduced through the adoption of Kanban affecting the movements inside the mixing room and also the activities of picking and warehousing, which has been optimized using Standard Work results from workshops organized on the Gemba. Indirect benefits are the reduction of levy in May MP of the quantity requested is not directly influenced by Kanban, although further improvement by providing a compilation of early material is in the pipeline, but can only be evaluated in the process stabilized. It is good to reiterate, however, that a project of efficient picking was already being realized by identifying the cells logistical grouping of codes for families. The transport of materials for storage in the mixing room is always done in 60 minutes on average, being more or less constant the absolute amount of material to be transported (vary the quantities for the individual codes), and takes place by means of forklifts with a capacity of about 3500 kg. Transport was not able to attend for reasons of layout of the two collection points, which does not allow the use of mechanized systems that could further reduce the time. Before Kanban time of filing and unloading materials employed an additional 60 minutes, on average, during which the products left in the department were deposited on the shelves and then had to be downloaded, and through systems manually before infrared signals as a result, databases MAGE in order to initiate the payment on consignment.

All this has resulted in a reduction of replenishment of 66% and in proportion of labor costs also affected by the activities mentioned above. Ultimately, at the operational level and control the optimization made through the Kanban has enabled better management of flows and control of materials along the loop logistics of supply, by means of inventories more reliable, continuous, in less time, and allowing complete traceability codes handled. Consequently, it is the immediate possibility of an effective FIFO management that eliminates all of the problems such as rework of compounds expired or the difference of the compounds due to the raw materials that do not comply. The stock reduction also results in a reduction of the physical space occupied by the deposited material with consequent improvement to the activities carried out in the mixing room, especially for logistics handling and retrieval of materials. In this direction push the solutions of Visual and Material Management Standard Work adopted, that simplify storage and picking limiting the uncertainty to the timing of implementation and eliminating possible causes of errors. The creation and implementation of Kanban has been achieved without substantial capital investment, in line with the principles Lean and Kaizen methodologies for which the first improvements are those made simply by making changes in the current state attacking and eliminating muda because of all the waste . The application of Total Flow Management for the optimization of the loop logistics supply in Mixing Room, beyond the basic improvements in operational terms then for a broader management Pull-wide establishment, provision would have resulted in a relationship Benefits / Estimated cost approximately equal to 2. Remember that the methodology TFM projects to be considered valid and consistent must propose a relationship between benefits and costs, estimated in the first year of intervention, at least greater than unity. The effective management materialized implementing the project has procured in the early months of the results are more than satisfactory, whose projection on the year portends a benefit / cost ratio of about 2.17, in line with the provisions.

It was decided to simulate the behavior of the system under assessment and implementation through the use of Rockwell Arena simulation software to validate the hypothesis and to make further changes and implementations. The results obtained confirm the validity of the theory of the pillars of sistema. Lo implemented and simulation tool, in particular for the Kanban software Discrete Event Arena, it is useful to be able to test the changes of process and choose the best solution before operating the system in the real world, with a great saving of time and resources. The application of model built has shown, moreover, the reliability of the system and how the environment Arena is reliable and is, therefore, conceivable its use to simulate the entire production process, evaluate the results of the extension of Kanban and optimize the parameters, prior to engage in future improvements of the system.

5. CONCLUSION

In this work we have explored the concept of Lean and its application in logistics. The Lean was born and has grown over the years as the basis for achieving excellence in operations, in particular in the production, attacking waste and eliminating any probable cause of inefficiency, integrating and articulating in a systematic methodologies and "philosophies industrial "rated the quality and efficiency of the processes. Over the years the evolution of the markets in directions increasingly competitive, primarily as a result of globalization, has prompted companies to tighten the focus of more and more businesses around its core business, leading to wastage consider everything that is not directly related to it, thus favoring the use of outsourcing.

This situation has led to the emergence of companies that specialize in providing highly productive services and not, but also created the conditions for a real transformation of productive no longer focused on the company, as an organization of people and vehicles aimed at the production and distribution of goods, but on a network of companies that collaborate and, in a sense, mutually contribute to success. The focus at maximum efficiency therefore extends from internal processes to external relations, namely the processes of coordination, integrated management, with an emphasis on logistics activities that regulate the connections between companies. The Total Flow Management approach fits into this context and through the concepts Lean aims to create Lean Supply Chain, aiming at the optimization of the loop logistic identified within the individual company (such as logistics) and the company to suppliers or customers (outbound logistics). The methodology TFM offers the opportunity of a common "language" for the various actors in the supply chain in order to extend the production flow Lean, based on takt time, the entire distribution chain, from end user to the first supplier of raw materials or semi-finished products. The elimination of the discontinuity in the value chain is needed in order to keep the costs under control and reduce them so as to be competitive on the market and then winning.

REFERENCES

Euclides A. Coimbra (2012), Total Flow Management – Kaizen per l'eccellenza nella supply chain - *KAIZEN INSTITUTE.*

Taichii Ohno (1988). Toyota Production System: Beyond Large-Scale Production, *Productivity Press*

Masaaki Imai (1986). Kaizen: Japanese spirit of improvement,*Japanese Productivity Press*

Frolick, M. N. & Thilini, R. (Š.D.). Business Performance Management: one truth. *Management Business Magazine.*

AA.VV, Criteria for Performance Excellence (2004), *Foundation for the Malcolm Baldrige National Quality Award*

Kaplan, R. & Norton, D. (1996). The Balanced Scorecard: Translating Strategy into Action, *Productivity Press*

Rockart, J. F. (1986). A Primer on Critical Success Factors, *MIT Sloan School of Management*

Womack, J. P., & Jones, D. T. (1991). The Machine That Changed the World: The Story of Lean Production, *United Mechanical Journal*

Maasaki Imai (1986). Gemba Kaizen: A Commonsense, Low-Cost Approach to Management, *Japanese Production Center*

Shigeo Shingo (1989). The Study of the TPS from an Industrial Point of View, *Japan Management Associate*

Taiichi Ohno (1988). Workplace Management, *Productivity Press*

Shigeo Shingo (1985). A Revolution in Manufacturing,*Japan Management Associate*

Rother, M., & Shook, J. (1998). Learning to See: Value Stream Mapping to Create Value and Eliminate Muda, *Lean Enterprise Institute*

Guizzi, G., Chiocca, D., Romano, E.(2012) System dynamics approach to model a hybrid manufacturing system- *Frontiers in Artificial Intelligence and Applications*, 246,pp. 499-517.

Chiocca, D., Guizzi, G., Murino, T., Revetria, R., Romano, E. (2012) A methodology for supporting lean healthcare *Studies in Computational Intelligence*, 431, pp. 93-99.

Gallo, M., Aveta, P., Converso, G., Santillo, L.C.(2012) Planning of supply risks in a make-to-stock context through a system dynamics approach- *Frontiers in Artificial Intelligence and Applications*, 246, pp. 475-496

Converso, G., De Carlini, R., Guerra, L., Naviglio, G. (2012) Market strategy planning for banking sector: an operational model *Advances in Computer Science*, WSEAS Press ISBN:978-1-61804-126-5 , pp. 430-435

USING DATA MINING AND MACHINE LEARNING METHODS FOR SERVER OUTAGE DETECTION – MODELLING NORMALITY AND ANOMALIES

Matthias Wastian[a], Dr. Felix Breitenecker[b], Michael Landsiedl[c]

[a]dwh GmbH, Neustiftgasse 57-59, 1070 Vienna, Austria
[b]Vienna University of Technology, Institute for Analysis and Scientific Computing, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria
[c] dwh GmbH, Neustiftgasse 57-59, 1070 Vienna, Austria

[a]matthias.wastian@dwh.at, [b]felix.breitenecker@tuwien.ac.at, [c]michael.landsiedl@dwh.at

ABSTRACT

This paper will discuss several approaches to detect abnormal events, which are considered to be worth further investigation by the modeler, in a time series of frequently collected data as early as possible and – wherever applicable – to predict them. The approaches to this task use various methods originating in the field of data mining, machine learning and soft computing in a hybrid manner. After a basic introduction including several areas of application, the paper will focus on the modular parts of the proposed methodology, starting with a discussion about different approaches to predict time series. After the presentation of several algorithms for outlier detection, which are applicable not only for time series, but also a chain of events, the results of the simulation gained in a project to detect server outages as early as possible are put up for discussion. The text ends with an outlook for possible future work.

Keywords: abnormal event detection, prediction, data mining, machine learning

1. INTRODUCTION, APPLICATIONS AND STATE OF THE ART

Definition 1 (Event): An event shall be defined as an occurrence happening at a determinable time and place with a certain duration. It may be a part of a chain of occurrences as an effect of a preceding occurrence and as the cause of a succeeding occurrence. It is possible that more than one event occurs at the same time and/or place.

Definition 2 (Abnormal Event): An abnormal event shall be defined as an outlier in a chain of events, an event that deviates so much from the other events as to arouse suspicion that it was caused by something that does not follow the usual behavior of the considered system and that it could change the entire system behavior.

Applications of abnormal event detection can be found in a broad variety of areas, almost all of them following the idea to guarantee a certain level of safety for the system considered. Examples are the prediction or detection of server outages, of natural catastrophes like flooding, hurricanes or earthquakes, of stock market breakdowns and of network intrusions. In the area of audio and video surveillance crowd behavior or traffic might be analyzed, but abnormal event detection also plays an important role in ambient assisted living.

Various approaches have been suggested for abnormal event detection. This paper is going to focus on time series forecasting with artificial neural networks (ANN) and outlier detection of the prediction errors with one-class support vector machines (OC-SVM). OC-SVMs were proposed (among others) by Heller, Svore, Keromytis and Stolfo (2003), by Evangelista, Bonnisone, Embrechts and Szymanski (2005) who additionally propose the use of fuzzy ROC curves, by Zhang, Zhang, Lan and Jiang (2008), Dreiseitl, Osl, Scheibböck and Binder (2010) as well as by Lecomte, Lengellé, Richard, Capman and Ravera (2011). Not all of them take into account the factor time. Other applied methods in the field of abnormal event detection are listed below:

- sparse reconstruction cost (Cong, Yuan and Liu 2011)
- wavelet decomposition (Suzuki and Ihara 2008)
- clustering based abnormal event detection (Jiang, Wu and Katsaggelos 2008)
- statistical methods
  - change point detection (Guralnik and Srivastava 1999)
  - explicit descriptors statistical model
  - bayes estimation
  - maximum likelihood
  - correlation analysis
  - principal component analysis (PCA).

2. DATA GENERATION AND DATA PREPROCESSING

2.1. Data Generation

Server monitoring is rampant nowadays. Server monitoring software allows to measure lots of features of a server that somehow describe its status. For our simulations, we had a total of up to 1439 features per server which were measured at a sampling rate from about one per fifteen minutes up to one per minute.

Besides historic data sets of several servers that were logged in the past, a software tool was used to generate artificial data sets. The capacity-planning tool was used to run tests, also called scripts and workloads, against a targeted server to measure its server capacity and response metrics. During these tests, each client generated a simulated user load of transactions against the server under test, which reported server statistics back to the client.

2.2. Data Preprocessing

First of all, the size of the recorded data set is rather large. All the simulations for a rapid server alert system have to be carried out at least nearly online. Thus a reduction of the original data set is indispensable. We used expert knowledge and did a feature selection by categorizing the features into four groups of different priorities, resulting in up to 14 features of the highest priority 0 and up to 73 features of the two most important priorities 0 and 1. Most simulation runs were implemented using the data labelled with these two priorities.

As the model intends to recognize the actual and future status of a server, those features that accumulate values (e.g., number of mails sent since the start of the server monitoring) were transformed into their differences.

Wrong measurements are also an issue that has to be dealt with for the server outage detection model. Especially features that have something to do with the queue lengths of hard disks delivers impossible values in a few cases. These values were substituted by their predecessors (if those were possible values) during the learning process. Of course, this substitution is also possible during on-line simulation runs. Another possibility is to delete those wrong values like it needs to be done, when a measurement cannot be carried out correctly due to any reason and the feature at this time is NaN. The distribution of these NaNs can be investigated separately, the algorithms proposed in the following sections are not able to deal with NaNs.

The ranges of the features considered in the model differ a lot. To make them comparable, the whole data set needs to be normalized. When using the neuro-predictor for the rapid server alert model, is seems best to use the following minmax-mapping to normalize the data:

$$f(x) = y_{min} + \frac{(y_{max}-y_{min})(x-x_{min})}{(x_{max}-x_{min})} \quad (1)$$

This is an affine transformation from $[x_{min}, x_{max}]$ to $[y_{min}, y_{max}]$.

3. PREDICTOR

Given any process that is checked for abnormal events, usually some features of this process can be measured at a constant sampling rate. Let m be the number of observed features. This results in m univariate time series. Given some past values and the actual value x_n of a certain feature, it is possible to predict the next observation x_{n+1} with a predictor and to calculate the prediction error as soon as the true new value x_{n+1} is measured.

Besides the classic ARIMA models that can be used for time series prediction, a certain kind of ANNs has proven to be an efficient predictor. Both models are going to be introduced in the following subsections. A multivariate approach is not recommended based on the simulation results for the server outage prediction as well as based on the results of various other authors. If a multivariate approach is desired nevertheless, we suggest to cluster the features first into several groups and to use an own multivariate predictor for each group.

The basic idea for any predictor of the abnormal event detection model is that the predictions are very good, if there are no abnormal events, i.e., the system's status is normal. The predictions become worse and do not originate from the usual distribution at least at the beginning of an abnormal event.

From a time series point of view, the most difficult task for the predictor is to consider the seasonality of the time series of some features. For example, the number of logged in users of a company on a certain Monday at 9:00 a.m. will probably strongly depend on the number of logged in users on Monday one week before at the same time. Feasts and holidays can cause problems for such models.

3.1. Neuro-Predictor

ANNs are non-linear and data-driven by nature and therefore at least theoretically very well suited to model seasonality interacting with other components.

Palit and Popovic (2005) refer to Simon Haykin, who suggests choosing the number of training patterns based on

$$N = \frac{W}{\varepsilon}. \quad (2)$$

W shall be the number of weights used in the ANN, ε shall be the error the training examples should be classified with and N shall be the number of patterns in the training set in this context.

When using ANNs to forecast time series, data normalization is a key issue. Various normalization methods can be applied; logarithmic or exponential scaling can be used if problems with non-linearities are expected during the network training. Linear normalizations like (1) can be used to meet the requirements of the network input layer, as the input range must not be too wide.

Significant patterns as seasonality and trends should be removed, if possible, to make the ANN time series model easier. To be able to use the concept of cross-validation, appropriate training, test and validation data sets need to be chosen. For our simulations the training data includes 70%, the test and the validation set includes 15% of the preprocessed data each.

The tasks of structuring the data and choosing the number of input nodes n_i of the ANN predominantly depend on the number d of lagged values to be used for forecasting of the next value in the standard case of a one-step-ahead prediction. Thus the function to be modeled by the ANN is of the type

$$x_{n+1} = f(x_n, x_{n-1}, \dots, x_{n-d+1}) \tag{3}$$

This function can also be alternated to

$$x_{n+1} = f(x_n, x_{n-1}, \dots, x_{n-d+1}, x_{n-s}, \dots, x_{n-2s}, \dots) \tag{4}$$

for a seasonality s. If the seasonality was not removed and the data preprocessing produces suitable input data blocks, seasonality can thus be modeled in an explicit way by the neuro-predictor.

The number of output neurons n_o directly corresponds to the forecasting horizon, i.e. in the case of a one-step-ahead forecast there is only one output neuron. Usually only one hidden layer is used. The number of the neurons in the hidden layer n_h was chosen according to the geometric pyramid rule:

$$n_h = \alpha\sqrt{n_i n_o},\ \ \alpha \in [0.5,2] \tag{5}$$

Choosing the number of hidden neurons as well as the data normalization involves trial-and-error experimentation.

We used the hyperbolic tangent as activation function in the hidden layer (the sigmoid function is also possible) and the linear activation function for the output layer. According to Zhang and Kline (2007), a non-linear activation function in the output layer is only needed, if the time series shows a significant trend even after the data preprocessing.

For the training of such neuro-predictors we use the Levenberg-Marquardt algorithm. The training sets are presented to the ANNs in several epochs. The supervised learning stops as soon as one of the following three break conditions is met:

1. The number of training epochs exceeds the value of a chosen tuning parameter.
2. The number of back-to-back epochs, which the error function of the validation set increases in, exceeds the value of a chosen tuning parameter.
3. The error value of the test data set falls below some minimal error value (e.g. 10^{-6}).

If there are several ANN models that we can finally choose from, an adapted version of the AIC can be applied:

$$AIC = Nn_o \ln(\sigma^2) + 2k \tag{6}$$

The model with the smallest AIC shall be preferred.

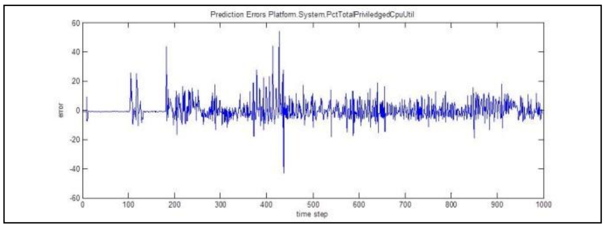


Figure 1: Prediction errors of a certain server feature, using a neuro-predictor

3.2. SARIMA Models

B being the backshift operator, autoregressive integrated moving average models with parameters $p, d$and q for a time series $\{x_t\}$ with error terms $\{\varepsilon_t\}$ are given by

$$\phi(B)x_t = \theta(B)\varepsilon_t \tag{7}$$

with

$$\phi(B) = \left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1-B)^d \tag{8}$$

and

$$\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i. \tag{9}$$

If the time series exhibits a strong seasonality, the model is adapted to a seasonal autoregressive integrated moving average model with parameters $(p,d,q) \times (P,D,Q)_s$, which is given by

$$\Phi(B^s)\varphi(B)\nabla_s^D \nabla^d x_t = \Theta(B^s)\theta(B)\varepsilon_t \tag{10}$$

with ∇ being the differencing operator, D the number of seasonal differences, Φ a polynomial of degree P, Θ a polynomial of degree Q and

$$\varphi(B) = \left(1 - \sum_{i=1}^{p} \phi_i B^i\right). \tag{11}$$

First of all the orders of differencing have to be identified to attain a stationary time series, several transformations like the logarithmic one might be useful. By looking at the plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) - they are in fact bar charts - of the differenced series, the numbers of AR and/or MA terms that are needed can tentatively be identified, for example following the advices that can be found at the course of Nau (2005).

3.3. Comparison Between Neuro-Predictors and SARIMA Models

When using ANNs for prediction, the results obtained by various authors differ widely in quality: Some suggest that ANNs are better than other forecasting models, others contradict them. Some have seemed to obtain better results with seasonally adjusted data, others think that ANNs are able to directly model seasonality in an implicit way, without any seasonal adjustments on the input data. Detailed research results are presented in Zhang and Kline (2007).

In 1991 Sharda, Patil and Tang identified a number of facts that determine which method is superior, by experiments:

- For time series with long memory, both approaches deliver similar results.
- For time series with short memory, ANNs outperform the traditional Box-Jenkins approach in some experiments by more than 100%.
- For time series of various complexities, the optimally tuned neural network topologies are of higher efficiency than the corresponding traditional algorithms. (Palit and Popovic 2005)

A hybrid combination of neural networks and traditional approaches – maybe also including GARCH models – seems very promising.

For the server outage detection model, some time series involved might have a long memory, others a short one. All in all, it seems reasonable that it is less inexact to choose the same parameters for all the feature predictors, if the neuro-predictors are used. Choosing the same parameters for all the predictors simplifies the model a lot.

4. ANOMALY DETECTOR

An analysis of prediction errors is the basis for the anomaly detector. The anomaly detector decides in a multivariate way, whether the prediction errors of all the features belong to the class ‚normal' or not. We did not only let the anomaly detector decide upon the most recent prediction error, but we also made him judge upon a moving average of the prediction errors, which increases the tolerance against weaknesses within the prediction models.

Depending on the number of features predicted, the dimension of the prediction error vector is a key issue for choosing a good anomaly detector. For increasing dimension the relevance of distance converges against 0.

Hodge and Austin (2004) distinct three fundamental approaches to detect outliers:

1. Model neither normality nor abnormality. Determine the outliers with no prior knowledge of the data. This is essentially a learning approach analogous to unsupervised clustering.
2. Model both normality and abnormality. This approach is analogous to supervised classification and requires pre-labeled data, tagged as normal or abnormal.
3. Model only normality; maybe tolerate abnormality in very few cases. Authors generally name this technique novelty detection or novelty recognition, especially if only normal data is given. It is analogous to a semi-supervised recognition or detection task. Only the normal class is taught but the algorithm learns to recognize abnormality. The approach needs pre-classified data but only learns data marked normal.

4.1. Threshold

For lower dimensions a simple threshold for a prediction error norm like the Euclidean norm can be sufficient to detect anomalies (assuming that all the features have been transformed to similar ranges during the preprocessing). If the predictions of several features are as bad as the ones on the outside margin of the Gaussian bell of figure 2, they will be detected by simple threshold.

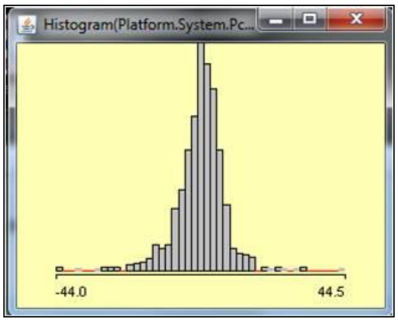


Figure 2: A Typical Histogram of the Prediction Errors of a Single Server Feature: A Gaussian Bell and a Few Outliers Clearly Visible on the Outside Margin

4.2. Angle-Based Outlier Detection

Angles are more stable than distances in high-dimensional spaces, which suggests the use of angles instead of distances for high-dimensional data. In fact, the situation is contrary for low-dimensional data. The angle-based outlier detection (ABOD) method alleviates the effects of the notorious curse of dimensionality compared to purely distance-based methods.

Following the idea of the algorithm developed by Kriegel, Schubert and Zimek (2008), a point is considered as an outlier, if most other points are located in a similar direction, and a point is considered as an inlier, if many other points are located in varying directions. The broadness of the spectrum of the angles between a certain point A and all pairs of the other points is a score for the outlierness of A: The smaller the score, the greater is the point's outlierness. The idea of the algorithm is illustrated for two dimensions in figure 3.

The angles in the so-called angle-based outlier factor are weighted by the squared inverse of the corresponding distances to avoid bigger problems with low-dimensional data sets.

$$ABOF(A) = VAR_{B,C\,\epsilon\,D}\left(\frac{\langle AB,AC\rangle}{\|AB\|^2\|AC\|^2}\right) \qquad (12)$$

A possibility to approximate the computationally expensive ABOF is to calculate the variance of the angles only of the pairs of points which belong to the k nearest neighbors of A, since these are the ones with the largest weights in the formula (12). Pham and Pagh (2012) provide further details on this issue.

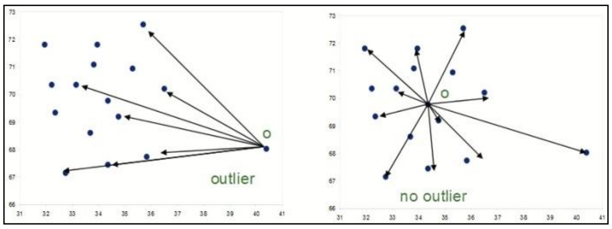


Figure 3: Idea of Angle-Based Outlier Detection

4.3. One-Class Support Vector Machine

In general, one-class support vector machines (OC-SVMs) are designed for the certain type of a $(1+x)$-class learning task. This is a model with an unknown number of classes, but the modeler is only interested in one specific class. Typical examples for these kinds of tasks are content-based image retrieval or document-retrieval in general. Making research for this paper on the internet can be seen as such a task: Papers which treat relevant topics are alike, they represent the class the modeler is interested in. These are the positive examples and it is easy to find some good representatives of this class. The negative examples are simply the rest of the web pages or papers, and they originate from an unknown number of different negative classes.

It is daunting and wrong to try to characterize the distribution of the negatives in such cases; they could belong to any negative class, and the modeler is not even interested which exact negative classes they might belong to. Each negative example is negative in its own way, but as the positive ones are alike, it is possible to model their distribution. According to this the OC-SVM is a typical example of a model of normality, matching the third approach described at the beginning of section 4.

The OC-SVM tries to fit a tight hypersphere W to include most, but not all positive examples. If it is attempted to fit all positive examples, this would lead to overfitting. In fact, the OC-SVM searches for the maximal margin hyperplane

$$\omega x + b = 0 \tag{13}$$

with a normal vector ω and a bias b which separates the training data from the origin in the best way. It may be interpreted as a regular two-class SVM, where almost all the training data lies in the first class and the origin is the only member of the second class.

If the one class the modeler is interested in is considered as the regular data, resulting from normality, the negative examples detected by the OC-SVM can be considered as outliers of a different nature resulting from anomaly. This makes the OC-SVM an effective outlier detection tool.

Let $\{x_1, \dots, x_n\}, x_i \in X \subseteq \mathrm{R}^m$ be a training set of $n \in \mathrm{N}$ observations that belong to a single class. The OC-SVM aims to define the minimum volume region enclosing $(1-\nu)n$ observations. The parameter $\nu \in [0,1]$ thus controls the fraction of observations that are allowed to be outliers. K shall be a kernel with a mapping function φ. ξ_i shall be the slack variables for observations on the wrong side; non-zero slack variables correspond to the tolerated outliers. The OC-SVM algorithm results in the following minimization problem:

$$\min_{\omega,\xi,b} \frac{1}{2}\|\omega\|^2 - b + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i \tag{14}$$

subject to

$$\omega^T \varphi(x_i) - b \geq \xi_i \geq 0 \tag{15}$$

Solving the OC-SVM optimization problem is equivalent to a dual quadratic programming problem with Lagrangian multipliers α_i that can be solved with standard methods:

$$\max_{\alpha_i} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \tag{16}$$

subject to

$$\sum_{i=1}^{n} \alpha_i = 1,\ \ 0 \leq \alpha_i \leq \frac{1}{\nu n} \tag{17}$$

Those patterns with corresponding $\alpha_i > 0$ are the support vectors. By using the Karush-Kuhn-Tucker conditions ω and b can be obtained as:

$$\omega = \sum_{i=1}^{n} \alpha_i x_i \tag{18}$$

$$b = \sum_{i=1}^{n} \alpha_i x_i^T x_j \tag{19}$$

for any support vector x_j.

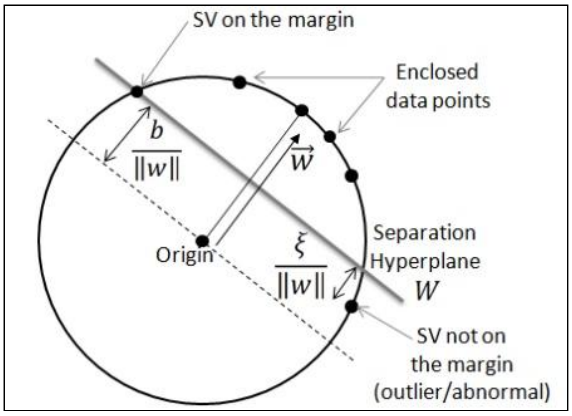


Figure 4: One-Class Support Vector Machine (Lecomte et al. 2011)

A new observation x is labeled by the OC-SVM via the decision function

$$f(x) = \omega^T \varphi(x) - b \tag{20}$$

which is positive for inliers and negative for outliers.

According to Lecomte et al. (2011), it is easily possible to define a family of decision rules introducing a threshold $\gamma \in \mathrm{R}$ by using an adaption of (20) and dividing inliers and outliers along γ, not along 0. This formulation allows controlling the trade-off between the probability to miss outliers and the probability to falsely declare an observation an outlier.

4.4. Combined Detector

As all the proposed outlier detector methods return an outlierness score for a feature vector, they could be used in a hybrid way. Then a weighted sum of the outlierness scores of each method is the final outlierness score of an observation.

5. RESULTS AND OUTLOOK

First of all, it has to be stated that it is almost impossible to precisely define the term server outage, wherefore a definition is not given in this paper. Any limitation to the normal operation of a server is unwanted. Many times only a certain kind of tasks is delayed or cannot be executed at all. The severity of this limitation also depends on the fact whether users can carry out other tasks in the mean time. The only possibilities to give the modeler an idea about the severity of an outage are the total downtime minutes or downtime minutes per user. Thus the basic idea of this model is to be able to provide the administrator of a server with the detection/prediction of irregularities, of anomalies which differ from the usual server operation. A classification of outages would be very useful, but requires outage data to learn from. This data should be labeled with the outage cause by experts. This classification remains future work.

Within the proposed model, the numbers of lagged time series elements that are relevant for the univariate prediction models for each server feature are not very easy to determine and the optimal number probably varies for each variable. Also the seasonality of the feature time series is not easy to diagnose. Nevertheless, the prediction models with global parameters for all the predictors worked very well during a normal operation of servers and seem to be sufficient for an online server outage detection model.

During several test runs, the anomaly detectors easily detected when the servers changed their status from idle to busy and vice versa (see figure 5). They also detected abnormal events within the gas price time series which was used as a benchmark data set (see figure 6). For this time series, an abnormal event is for example the oil crisis of 1979, which was caused by the Islamic revolution in Iran and the first gulf war, i.e. by external events. For the server outage detection model, the verification is rather difficult and there will be done further research on this topic: Besides the difficulty to define a server outage, the model needs to be tested in a real-life scenario which is planned in near future. So far, the detectors worked well with the test data sets.

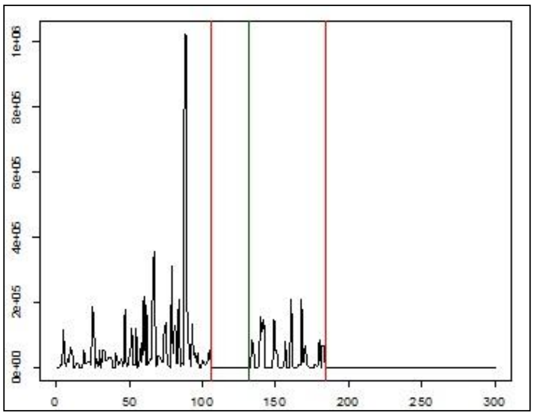


Figure 5: Angle-Based Outlier Detector Detecting the Server Change from Idle to Busy (Green) and Busy to Idle (Red)

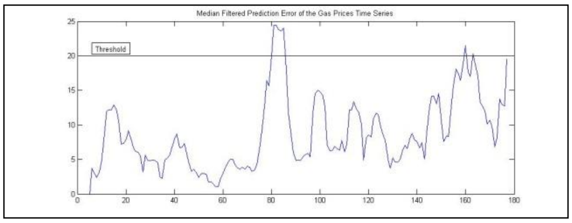


Figure 6: Median-Filtered Prediction Error of the Gas Prices Time Series Using a Neuro-Predictor with a Delay of 3 Months, 10 Hidden Neurons and a Threshold for Abnormal Event Detection. The Median Was Calculated Over 6 Months. The First Peak Above the Threshold 20 Corresponds to the 1979 Oil Crisis.

Of course, a server outage prediction software has a cold start: During the training some internal model parameters that are required to run the model need to be adjusted, before an expert can adjust several tuning parameters to control the alert sensitivity of the software. The most important tuning parameters are part of the anomaly detector. One could say that the server outage detection model needs to get to know the server that the outages shall be predicted of. As parts of the model are able to learn from the past, the software will highly improve its performance after several days.

An important question that still remains unanswered is when the neuro-predictors should be retrained or when the ARIMA models should be updated. Certainly, if the way the server is used changes considerably, a re-start of the model is necessary.

ACKNOWLEDGMENTS

The authors would like to thank the economic partners who helped a lot to acquire tons of data.

REFERENCES

Cong, Y., Yuan, J., Liu, J., 2011. Sparse Reconstruction Cost for Abnormal Event Detection. Proceedings of the 24th *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3449-3456. Colorado Springs, Colorado.

Dreiseitl, S., Osl, M., Scheibböck, C., Binder, M., 2010. Outlier Detection with One-Class SVMs: An

Application to Melanoma Prognosis. Proceedings of the *AMIA Annual Symposium 2010*, pp. 172-176. Washington, D.C.

Evangelista, P., Bonnisone, P., Embrechts, M., Szymanski, B., 2005. Fuzzy ROC Curves for the 1 Class SVM: Application to Intrusion Detection. Proceedings of the *13th European Symposium on Artificial Neural Networks*, pp. 345-350. Bruges, Belgium.

Guralnik, V., Srivastava, J., 1999. Event Detection from Time Series Data, Proceedings of the *5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 33-42. New York, USA.

Heller, K., Svore, K., Keromytis, A., Stolfo, S., 2003. One Class Support Vector Machines for Detecting Anomalous Windows Registry Accesses. Proceedings of the *Workshop on Data Mining for Computer Security in conjunction with the IEEE International Conference on Data Mining 2003*, pp. 2-9. Melbourne, Florida.

Hodge, V., Austin, J., 2004. A Survey of Outlier Detection Methodologies. *The Artificial Intelligence Review* Issue 2 of Volume 22: pp. 85-126.

Jiang, F., Wu, Y., Katsaggelos, A., 2008. Abnormal Event Detection Based on Trajectory Clustering by 2-Depth Greedy Search, Proceedings of the *IEEE International Conference on Acoustics, Speech and Signal Processing 2008*, pp. 2129-2132. Las Vegas, Nevada.

Kriegel, H.-P., Schubert, M., Zimek, A., 2008. Angle-Based Outlier Detection in High-Dimensional Data. Proceedings of the *14th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 444-452. Las Vegas, Nevada.

Lecomte, S., Lengellé, C., Richard, C., Capman, F., Ravera, B., 2011. Abnormal Events Detection using Unsupervised One-Class SVM – Application to Audio Surveillance and Evaluation. Proceedings of the *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pp. 124-129. Klagenfurt, Austria.

Nau, R., 2005, *Forecasting - Decision 411, Online Course*. Available from: http://people.duke.edu/~rnau/Decision411CoursePage.htm [Accessed 13.05.2013]

Palit, A., Popovic, D., 2005. *Computational Intelligence in Time Series Forecasting – Theory and Engineering Applications*. London: Springer Verlag.

Pham, N., Pagh, R., 2012. A Near-Linear Time Approximation Algorithm for Angle-Based Outlier Detection in High-Dimensional Data. Proceedings of the *18th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 877-885. New York, USA.

Suzuki, M., Ihara, H., 2008. Development of Safeguards System Simulator Composed of Multi-Functional Cores. *Journal of Power and Energy Systems* Number 2 of Volume 2: pp. 899-907.

Zhang, G.P., Kline, D., 2007. Quarterly Time-Series Forecasting With Neural Networks. *IEEE Transactions on Neural Networks* Number 6 Volume 8: pp. 1800-1814.

Zhang, R., Zhang, S., Lan, Y., Jiang. J.. Network Anomaly Detection Using One Class Support Vector Machine. Volume 1 of the Proceedings of the *MultiConference of Engineers and Computer Scientists 2008*. Hong Kong.

AUTHORS BIOGRAPHY

Matthias Wastian, born 27.12.1983 in Carinthia, Austria, studies technical mathematics at the Vienna University of Technology and has been writing his diploma thesis about abnormal event detection. He is employed at the dwh, Neustiftgasse 57-59, 1070 Vienna. Apart from that he is the captain of the Austrian wheelchair basketball national team.

THE ROLE OF INNOVATION IN INDUSTRIAL DEVELOPMENT SYSTEM. A SIMULATION APPROACH FOR SUSTAINABILITY OF GLOBAL SUPPLY CHAIN NETWORK

Liberatina Santillo [(a)], Giuseppe Converso [(b)], Luigi De Vito [(c)]

[(a)]DICMAPI University of Naples Federico II Naples, p.le Tecchio 80, ITALY
[(b)]DICMAPI University of Naples Federico II Naples, p.le Tecchio 80, ITALY
[(c)] CNR – IMCB Naples, ITALY

[(a)] santillo@unina.it, [(b)] giuseppe.converso@unina.it, [(c)] ludevito@unina.it

ABSTRACT

This paper shows the relationship between Supply Chain Management and sustainability at global level. As we shall see, it will be shown that relationship using a Causal Loop Diagram then we will move on a Stock and Flow Diagram, developing an ad hoc model in System Dynamics. The model will show how the whole system will crash in total absence of Innovation, then we will figures two scenarios, each with different Innovation levels, and see how the Innovation will compensate the depletion of all primary resources. We will also consider financial sustainability as endless, leaving financial matters to the economists.

Keywords: Innovation, Global Supply Chain Network

1. INTRODUCTION

Innovation has a double impact on the environment. While it contributes to impoverish the environmental resources of the planet, it has the ability to compensate that decrease. Since natural and primary resources are the first link of any supply chain, especially at the global level, it is therefore legitimate to deal with Sustainable Supply Chain Management (SSCM).

SSCM is a set of managerial behaviors that include a low negative or (even positive) environmental impact, also through a multidisciplinary approach of such practices at all levels of the production cycle.

It is assumed that such behavior, arising from the normal process of business decision-making, rather than an imposition by the government.

In this paper we will demonstrate the fundamental role of Innovation in balancing the massive depletion of primary resources, using a System Dynamics simulation model.

System Dynamics is an approach used to model the behavior of complex systems in a certain time range. Developed at MIT back in 1950, by J.W. Forrester first and then by P. Sange, SD is mostly used to describe enterprises behaviors, even if we maintain SD perfectly describes strategic situations.

SD make use of two main tools: Causal Loop Diagram and Stock and Flow Diagram. The first one consist graphic maps showing relationships between all the single elements of the system. It has two pros: it's a first graphic approach to the problem, then it shows all the feedbacks between each quantity. The cons instead are that it is only a qualitative approach, since there's any quantitative information. That limit is overcome with the Stock and Flow Diagram. Its basic elements are: stock variables, shown as boxes that describes the state of the system, flow variables shown as valves that fill or decrease the levels, links that transport the information from levels to flows, the functions that rule the way levels are used to let the flow work.

2. STATE OF THE ART

The SSCM brag a significant literature, developed massively since the mid-90s onwards. It examines the various environmental implications of several business activities such as product design, production cycle,inventory management. Early research on SSCM concerns the management, from collection to rework the returned products.

In 1995, Greenberg highlights the importance of mathematical models for environmental control. Fleischmann analyzes the quantitative models, dividing his work into three areas: distribution planning, production planning and inventory control.

Kleindorfer et al. (2005) use the term sustainability in a diffuse manner, referring to the environmental management,talking about the Closed Loop Supply Chain, integrating profit, people and planet in the corporate culture.

Toktay and Ferguson (2006) have developed models to support a manufacturer's recovery strategy in the face of a competitive threat in the market of the product manufacturing.

Atasuand van Wassenhove (2008) examine the environment rework from the perspective of marketing, focusing on important aspects of remanufactured products, such as low cost, parts reuseand constraints on the supply chain.

Ketzenberg, Van der Laan, and Teunter (2006) treat the value of information (VOI), when a company faces the uncertainty of demand, product returns and repairs. The objective is to evaluate the VOI by reducing the factors of uncertainty one by one, thus evaluating the economic savings.

Therelationshipsamong the actors of a supply chain network, cover every aspect of business, for example through the sub-contracting (outsourcing of activities mainly related to the production, taking advantage of the expertise of the supplier), outsourcing (of one or more activities of the value chain. Involvement greater than the sub-supply involves the expenditure of energy and resources in mate selection and implementation of specific investments to manage the relationship with the latter), licensing (a contract by which a company licensor grants to the licensee company economic exploitation of a patented intellectual property, trademarks and patents and not, or know-how, maintaining the property. licensee agrees to pay royalties relate to the economic results), franchising, venture capital and private equity (respectively forms of participation in the venture capital firm emerging or already).

We may thus apply the concepts of SSCM on a global scale, if we think that the relationship between companies tend not to have more geographical limits.

2.1. Subheadings

Initial Caps, bold, flush left. Use Times New Roman Font and 10 points in size. Start the text on the next line. Please use the “Heading 2**”** style.

2.1.1. Secondary Subheadings

Initial Caps, bold, indented of 0.7 cm. Use Times New Roman Font and 10 points in size. Start the text on the next time. Please use the “Heading 3” style.

3. INNOVATION AND SUSTAINABLE SCMC

From the macroeconomic point of view, research and innovation in particular will have a positive impact on the availability of raw materials, the first ring of each supply chain. The need to model the process of innovation is illustrated by the fact that in addition to the obvious advantages already mentioned, the use of technology may have a negative impact on the environment. The relationship between technology and environment is twofold and moves along two paradoxically opposite. On the one hand the intensive use of technology involves a high consumption of raw materials, with consequent environmental stress, on the other hand the use of technology, in terms of research and innovation, can lead to a better and more efficient use of raw materials , with consequent environmental stress less. In other words, we talk about sustainable technological development.

In any macroeconomic model, a production cycle of goods and services on the one hand has a positive influence on job opportunities, and inevitably the supply of goods and services, on the other hand has a strong negative impact on the availability of raw materials, seen both as materials needed for the production and as energy resources (for example, fuel oil, natural gas and fossil). At the same time, the increase in labor income and therefore inevitably generates consumption, which together with the offer and as a result of solid macroeconomic laws, tends to bring the market back into balance.

One of the tasks of the Supply Chain Risk Management (SCRM) is the analysis and the treatment of all possible threats to the chain. The overall objective of SCRM is to ensure that the chain will continue to operate as expected, with a constant flow and continuous materials from suppliers to end customers. This is possible by increasing their ability to withstand adverse events. If you know the seriousness of the risk, then you know how to deal with it. The response varies depending on the type of risk and consequences associated with it.

We see this behavior in terms of Causal Loop Diagram (CLD, cfr. Fig. 1):

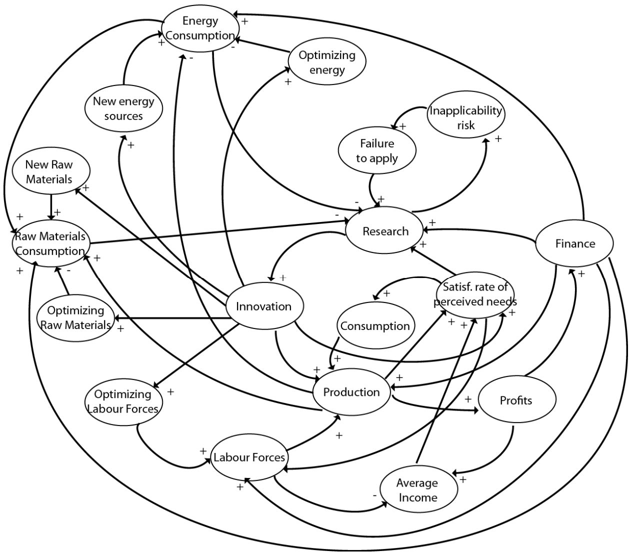


Figure n. 1: Causal Loop Diagram

Let's then suppose the financial sustainability as infinite, and let's take the assumption that, since 20% of the population consumes an average of 80% of the available resources of the planet, what can happen if the remaining 80% of the population increase its own life standards up to or comparable to the 20% mentioned.

In this case, the global supply chain will crash, because when the number of employees increases, and then exceeds the threshold of poverty and access to consumption, increase their disposable income, and therefore will increase the demand first and then consumption. This increase will reverberate positively on industrial production, stimulated to meet the perceived needs of the population and then in turn it will increase as well. In the CLD we already considered implicitly the processes of recycling of raw materials and energy. Of course, an increase in production is given by an increase in labor resources involved, but also of the energy and raw materials, the increase in production will lead to increased profits, they'll be back in a circle fueling monetary reserves, generating hence welfare. We know that the majority of raw materials and non-renewable resources are limited, so naturally tend to run out, for example, despite the continuing search for new sources of raw materials. The impact can therefore be increasing the number of people accessing the consumption does not only accelerate the process suddenly runs out of resources.

A course like this is naturally parasitic and inevitably requires a strong balancing process, or at least of a slowdown in the process of exhaustion, with the ultimate goal of improving the availability of energy resources and raw materials, optimizing consumption (process innovation) but also discovering new (product innovation). The centerpiece of this cycle is precisely balancing innovation and research. Clearly, not everything that is produced by research becomes innovation, In other words, when a research ends, it is not said that the result can be directly implemented, so it is easy to incur the risk of inapplicability which leads to the failure innovation. Such non-innovation in compensation can still be a source of stimulus for subsequent research, in order to successfully transform them into new innovation. Innovation can then act on an optimized consumption of raw materials and / or energy sources or even produce new ones, in any case going to increase their availability, and then to counteract the deleterious effects of massif exploitation. In a nutshell innovation in this case is very strong weapon in defense of the technological sustainability of the global supply chain.

3.1. The Stock and Flow diagram

Based on the process outlined in the causal loop diagram, in the compilation of the model the following quantities have been identified and classified (cfr. Fig. 2):

Level variables:

- Recycled Energy Consumed
- Recycled Energy
- Energy Scrapped
- Energy Used
- MP consumed recycled
- MP dropped
- MP extracted
- MP recycled
- New energy
- Newmaterials
- Production consumed
- Production dropped
- Total production
- Reserve consumed
- Reserve available:
- Reserve power
- Reserve MP

Auxiliary variables:

- Innovation:
- Percentage increase in population
- Population
- Productivity energy
- Productivity matter
- Research
- Expenses for the purchase of materials and energy

Constants:

- Cx Innovation
- Energy Price
- Price MP
- New energy production Price
- Price new production MP

Flow variables:

- Consumption
- New energy added
- New MP added
- Recycling Energy
- Recycling Energy consumed
- Recycling MP consumed
- Recycling MP
- Recycling MP input
- Recycling Energy cons
- Recycling energy input
- Recycling MP cons.
- Scrap energy
- Scrap MP
- Scrap production
- Consumption Rate
- Extraction Rate MP
- New energy Rate
- New MP Rate
- Production rate
- Energy use Rate
- Gains

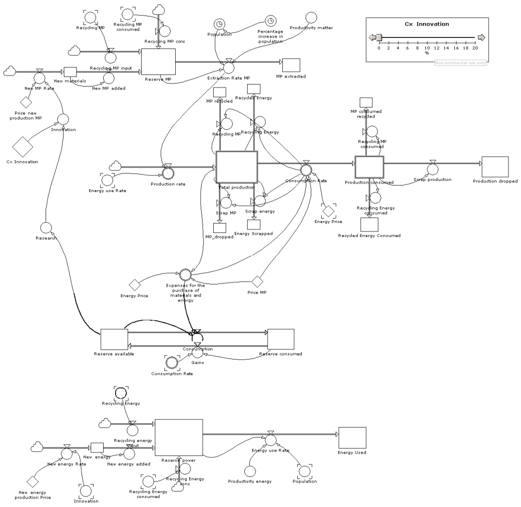


Figure 2: Stock and Flow Diagram

3.2. Validating the model

In our model, a slider was applied on the Coefficient of Innovation (Fig. 2), that allows us to vary at will, its value within a range defined by us between 0-10%.

To validate the model, we started the simulation by resetting the innovation level on the slider. The result, shown in Fig. 3, is in line with those who say, as the Italian ASPO (Association for the Study of Peak Oil) and the maximum of Hubbert, i.e. the peak extraction of raw material for single nation after which the available reserves tend to decrease in a rather sudden, is more or less reached, and then in little more than fifty years the availability of oil tend to run out. The plastic instead will end with a slight delay due to the recycling process. Both raw materials and renewable energy will be depleted since the steady increase in population leads to increasing in consumption.

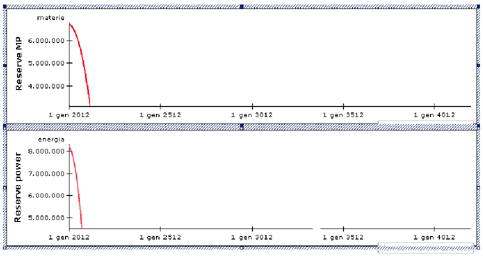


Figure n. 3: Innovation = 0

3.3. First scenario

Therefore we hypothesized a first scenario, introducing a 1% innovation value in the slider. In this case we found a similar pattern for both monitored processes, but with a slower decay, which showed that a low rate of innovation does not solve the problem of scarcity and lack of materials, but move only away the time which all the materials run out and then the system crashes.

The signal given by the model should however be taken positively, either because the innovation, while not completely solving the problem, provides further autonomy, cfr Fig. 4, and encourages us to try new simulations with higher values of innovation, as well as in the second scenario.

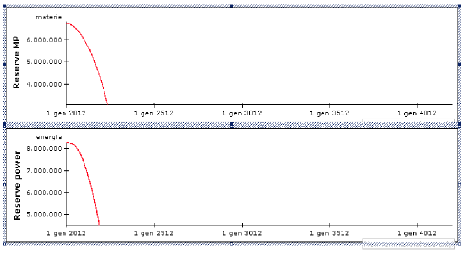


Figure n. 4.: First Scenario , Innovation =1

3.4. Second scenario

In a second scenario, we increased the innovation level on the slider up to 6%. The results here are completely subverted, as shown in Fig. 5; while in the first scenario, the system crash occurs more slowly, materials and energy resources won't deplete, and the system will not crash. In this scenario the innovation plays an important role, both in the optimization of resources and materials, both in the identification of substitute products that are likely to be even more easily recyclable, increasing each own stock.

The trend, as you can imagine, and as you can see from the graphics, is growing.

As already mentioned, one of the key points of this model is the financial sustainability endless. This simulation includes a gold reserve generator, for which at the time when the population increases, a fraction of its richness in the process returns, is able to re-feed the respective initial reservations.

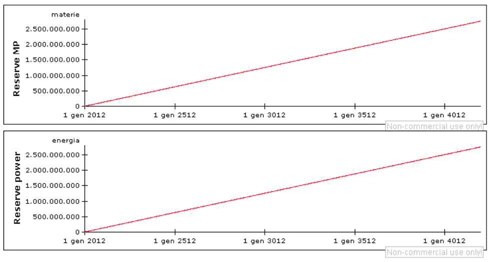


Figure n. 5: Second Scenario Innovation = 6

3.5. Future developments

This particular model, although still in prototype stage, has a strong significant potential. Infact, with a more accurate and complete data collection, it can be vectorialized, creating arrays of simulations, one for each most important raw material, such as steel and cast iron, clay, silver, plastic, and the main sources of energy, such as oil, coal , wood, etc.. In this way it is possible to better understand the dynamics and interrelationships of the exploitation of primary resources, understand the trends and see which resource crashes first, to evaluate the Hubbert peak, possible influences of the limiting factors.

The aim of this work is also laying the groundwork for the creation of a team and the scheduling of work for the implementation and programming of such a model..

ACKNOWLEDGMENTS

If the paper requires an acknowledgements section it can be placed after the main body of the text.

REFERENCES

Greenberg, H. J. (1995), Mathematical programming models for environmental quality control. *Operations Research*, 43(4), pp. 578-622,

Fleischmann, M., Runwaard, J. B. M., Dekker, R., Laan, E., Nnunen, J. A. E. E., & van Wassenhove,

L. (1997), Quantitative models for reverse logistics: a review. *European Journal of Operational Research*, 103(1), pp. 1-17,

Kleindorfer, P. R., Singhal, K., & van Wassenhove, L. N. (2005). Sustainable operations management.*Production and Operation Management*, 14(4), pp. 482-492

Ferguson, M. E., &Toktay, L. B. (2006). The effect of competition on recovery strategies.*Production and Operation Management*, 15(3), pp. 351-368

Atasu, A., Van Wassenhove, L. N. (2008). Remanufacturing as a marketing strategy. *Management Science*, 54(10), pp. 1731-1746.

Ketzenberg, M. E., van der Laan, E., &Teunter, R. H. (2006). Value of information in closed loop supply chain. *Production and Operation Management*, 15(3), pp. 393-406.

Gallo M., Murino T., Santillo L. C. *A lean redesign for a manufacturing process through computer simulation* (2012) Advances in Computer Science, pp 416-422

Bessant J., Lamming R. (2003). Putting supply chain lerning into practice.http://eprints.uk/79/1

Harland R., Zheng J., Johnsen T., Lamming. R. (2004). *A Conceptual Model for Researching the Creation and Operation of Supply Networks*. British Journal of Management, Vol. 15, pp. 1–21

Gallo M., Aveta P., Converso G., Santillo L. C. *Planning of supply chain risks in a make to order context through a System Dynamics approach.*. NEW TRENDS IN SOFTWARE METHOLOGIES, TOOLS AND TECHNIQUES. p. 475-496, IOS Press, ISBN: 9781614991243

Guizzi, G., Chiocca, D., Romano, E., *System dynamics approach to model a hybrid manufacturing system* (2012) Frontiers in Artificial Intelligence and Applications, 246, pp. 499-517

Chiocca, D., Guizzi, G., Murino, T., Revetria, R., Romano, E., *A methodology for supporting lean healthcare* (2012) Studies in Computational Intelligence, 431, pp. 93-99.

Murino, T., De Carlini, R., Naviglio, G., *An economic order policy assessment model based on a customized ahp* (2012) Frontiers in Artificial Intelligence and Applications 246 , pp. 445-456

Guizzi, G., Murino, T., Romano, E., *An innovative approach to environmental issues: The growth of a green market modeled by system dynamics* (2012) Frontiers in Artificial Intelligence and Applications 246 , pp. 538-557

Murino, T., Romano, E., Santillo, L.C., *Supply chain performance sustainability through resilience function* (2012) Proceedings - Winter Simulation Conference , art. no. 6147877 , pp. 1600-1611

Gallo, M., Grisi, R.M., Guizzi, G., *A vendor rating model resulting from AHP and the linear model* (2010) International conference on System Science and Simulation in Engineering - Proceedings, pp. 370-377.

Gallo, M., Guerra, L., Guizzi, G., *Some considerations on inventory-based capacity scalability policies in RMSs* (2010) International conference on System Science and Simulation in Engineering - Proceedings, pp. 342-347.

Di Franco, R., Gallo, M., Guizzi, G., Zoppoli, P., *Project risk management: A quantitative approach through simulation tecniques* (2009) Proceedings of the 8th WSEAS International Conference on System Science and Simulation in Engineering, ICOSSSE '09, pp. 326-333.

AUTHORS BIOGRAPHY

LIBERATINA C. SANTILLO is a Full Professor - at the Department DICMAPI - University of Naples - Federico II -. At the same University Department (Degree in Mechanical Engineering, Industrial Engineering, Materials Engineering) she teaches “Goods and Services Production System” and “Work Safety”. She is Professor of Industrial Plants at Neptune Consortium University. Her research interests cover both the management of the industrial production of goods and services, and the safety procedures of work activities. She is President of a major training institution and a member of several Italian clubs with scientific, social and charity purposes. Her email address is santillo@unina.it.

GIUSEPPE CONVERSO received his Ph.D. in Technology and Production Systems from Department of Materials and Production Engineering - University of Naples - Federico II - Faculty of Engineering - in 2006 and has since collaborated with the DICMAPI - University of Naples - Federico II -. His research interests include optimization via simulation, process reengineering, simulation output analysis, and applications of simulation methods to plant design system. His e-mail and is giuseppe.converso@unina.it.

LUIGI DE VITO holds PhD in Production System Tehnology at University of Naples Federico II, and work for C.N.R., National Research Council. His email address is ludevito@unina.it.

A MACHINE LEARNING APPROACH FOR MODELING AND ITS APPLICATIONS

Dr Shuxiang Xu[(a)], Dr Yunling Liu[(b)], Dr Byeong-Ho Kang[(c)], Dr Wanlin Gao[(d)]

[(a)]School of Computing and Information Systems, University of Tasmania, Launceston, Tasmania 7250, Australia
[(b)](Corresponding Author) College of Information and Electrical Engineering, China Agricultural University, Beijing, China
[(c)]School of Computing and Information Systems, University of Tasmania, Hobart, Tasmania 7001, Australia
[(d)]College of Information and Electrical Engineering, China Agricultural University, Beijing, China

[(a)]Shuxiang.Xu@utas.edu.au, [(b)]lyunling@163.com, [(c)]Byeong.Kang@utas.edu.au, [(d)]gaowlin@cau.edu.cn

ABSTRACT

This paper proposes a new learning algorithm for Higher Order Neural Networks for the purpose of modelling and applies it in three benchmark problems. Higher Order Neural Networks (HONNs) are Artificial Neural Networks (ANNs) in which the net input to a computational neuron is a weighted sum of its inputs and products of its inputs (rather than just a weighted sum of its inputs as in traditional ANNs). It was well known that HONNs can implement invariant pattern recognition. The new learning algorithm proposed is an Extreme Learning Machine (ELM) algorithm. ELM randomly chooses hidden neurons and analytically determines the output weights. With ELM algorithm only the connection weights between hidden layer and output layer are adjusted. This paper proposes an ELM algorithm for HONN models and applies it in an image processing problem, a medical problem, and an energy efficiency problem. The experimental results demonstrate the advantages of HONN models with the ELM algorithm in such aspects as significantly faster learning and improved generalization abilities (in comparison with standard HONN and traditional ANN models).

KEYWORDS: Artificial Neural Network, Extreme Learning Machine, Feedforward Neural Network, Higher Order Neural Network, Machine Learning.

1. INTRODUCTION

An actively researched machine learning algorithm, Artificial Neural Networks (ANNs) have been widely used as powerful information processing tools for modeling a diverse range of applications. HONNs (Higher Order Neural Networks) (Lee et al, 1986) are networks in which the net input to a computational neuron is a weighted sum of its inputs and products of its inputs (see Figure 1.1 for an example of a second order HONN). Such neuron is called a Higher-order Processing Unit (HPU) (Lippman, 1989). It was known that HONN's can implement invariant pattern recognition (Psaltis et al, 1988 ; Reid et al, 1989 ; Wood et al, 1996). In (Giles et al, 1987) it was shown that HONN's have impressive computational, storage and learning capabilities. In (Redding et al, 1993), HONN's were proved to be at least as powerful as any other FNN (Feedforward Neural Network) architecture when the orders of the networks are the same. Kosmatopoulos et al (1995) studied the approximation and learning properties of one class of recurrent HONNs and applied these architectures to the identification of dynamical systems. Thimm et al (1997) proposed a suitable initialization method for HONN's and compared this method to weight initialization techniques for FNNs. More recently, In Alanis et al (2007) an application of HONN was proposed to successfully solve the tracking problem for a class of nonlinear systems in discrete time using backstepping technique. HONNs were employed in Xu (2010a, 2010b) for several data mining tasks and achieved significant results which outperformed conventional ANNs. A HONN was investigated in Dunis (2011) for forecasting and trading EUR/USA exchange rates, with outstanding results when compared against other ANN architectures as well as traditional statistical approaches. In Fallahnezhad et al (2011) a hybrid HONN model was developed for handling several benchmark classification problems, resulted in significant improvements of accuracy (in comparison with the best accuracy obtained from other methods).

Unlike traditional ANN learning algorithms (such as back-propagation), Extreme Learning Machine (ELM) algorithm randomly chooses hidden neurons and analytically determines the output weights (Huang et al 2005, 2006, 2008). With ELM algorithm, only the connection weights between hidden layer and output layer are adjusted. Many types of hidden nodes including additive nodes, RBF (radial basis function) nodes, multiplicative nodes, and other non-neural alike nodes can be used as long as they are piecewise nonlinear. ELM algorithm tends to generalize better at very fast learning speed: it can learn thousands of times faster than conventional popular learning algorithms (Huang et al 2006).

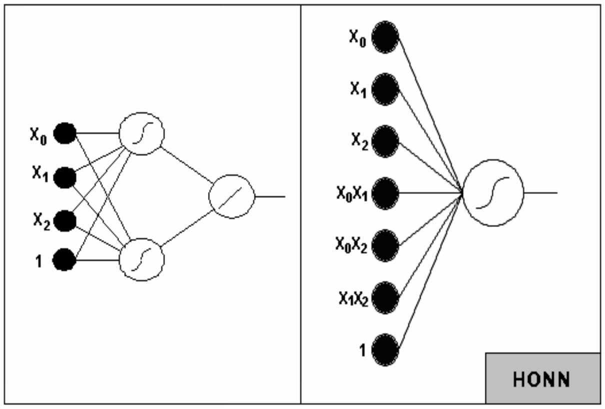


Figure 1.1, Left, FNN with three inputs and two hidden nodes; Right, second order HONN with three inputs

This paper develops an ELM algorithm for HONN models and applies it in an image processing problem, a medical problem, and an energy efficiency problem. Following the introduction, Section 2 introduces the ELM algorithm for HONN. Section 3 describes three HONN modelling experiments to demonstrate the advantages of ELM HONN (against standard HONN and traditional ANNs such as MLP and RBF networks). Results are given and discussed. Section 4 summarises this report.

2. HONN MODELS WITH ELM ALGORITHM

Based on a modified version of a two-dimensional (second order) HONN defined in Zhang et al (2002), this paper proposes the following ELM algorithm for HONN. The main idea of ELM lies in the random selection of hidden neuron activation functions (must be infinitely differentiable) with random initialization of the SFNN (single-hidden-layer feedforward neural network) weights and biases. Then, the input weights and biases do not need to be adjusted during training, only the output weights are learned. In this work, an adaptive neuron activation function (infinitely differentiable) has been used. The free parameters in the adaptive activation function are adjusted in a way similar to how the output weights are tuned.

Consider a set of s distinct training samples (x_i, y_i) with $x_i \in R^n$ and $y_i \in R^m$, where i = 1, 2, …, s, n and m are positive integers, where n represents the dimension of an input space, and m the dimension of an output space. Then an SFNN with N hidden neurons can be mathematically represented by

$$\sum_{i=1}^{N} o_i\, f(W_i \cdot X + b_i) \qquad (2.1)$$

with f being the randomly selected neuron activation function, W_i the input to hidden layer weight vector, b_i the biases, o_i the output weights, and X the input vector: $X = [x_1\ \ x_2\ \ \dots\ x_s]^T$.

In this work, the following adaptive neuron activation function is used:

$$f(x) = A1 \cdot e^{-B1 \cdot x^2} + \frac{A2}{1 + e^{-B2 \cdot x}} \qquad (2.2)$$

where $A1$, $A2$, $B1$, and $B2$ are real variables which will be adjusted during training (in the same way as connection weights, see the end of the current section)). A justification of the use of free parameters in a neuron activation function can be found in Zhang et al (2002).

In case of two-dimensional (second order) HONN with a single hidden layer, equation (2.1) becomes

$$\sum_{i=1}^{N} o_i\, f(W_i \begin{bmatrix} X \\ H(X) \end{bmatrix} + b_i) \qquad (2.3)$$

where

$$H(X) = [x_1 x_2\ \ x_1 x_3\ \ \cdots\ \ x_{s-1} x_s]^T \qquad (2.4)$$

Equations (2.3) and (2.4) show that for a two-dimensional HONN, the number of input neurons is defined by

$$n + C_n^2 = n + \frac{n(n-1)}{2} \qquad (2.5)$$

Assume that the single layer HONN approximates the training samples perfectly, then the errors between the estimated outputs and the actual outputs are zero, which means

$$\sum_{i=1}^{N} o_i\, f(W_i \begin{bmatrix} X \\ H(X) \end{bmatrix} + b_i) = Y \qquad (2.6)$$

where Y is the output vector: $Y = [y_1\ \ y_2\ \ \dots\ y_s]^T$.

Equation (2.6) can be rewritten as $H \cdot O = Y$, with

$$H = \begin{bmatrix} f(w_1 x_1 + b_1) & \cdots & f(w_N x_1 + b_N) \\ \vdots & \vdots & \vdots \\ f(w_1 x_s + b_1) & \ddots & f(w_N x_s + b_N) \\ f(w_1 x_1 x_2 + b_1) & & f(w_N x_1 x_2 + b_N) \\ \vdots & \vdots & \vdots \\ f(w_1 x_{s-1} x_s + b_1) & \dots & f(w_N x_{s-1} x_s + b_N) \end{bmatrix} \qquad (2.7)$$

and O as the hidden layer to output layer weight vector: $O = [o_1\ \ o_2\ \ \dots\ o_N]^T$.

Then the idea of ELM algorithm, when applies to a HONN, states that with randomly initialized input weights and biases, and with the condition that the randomly selected neuron activation function is infinitely differentiable, the output weights can be determined so that the single layer HONN provides an approximation of the sample values to any degree of accuracy. The way to calculate the output weights from the hidden layer output matrix and the target values is proposed with the use of a Moore-Penrose generalized inverse (Rao et al 1972) of the matrix (2.7), denoted as H^{-1}.

Overall, the improved ELM algorithm for HONNs is proposed as follows.

Given a set of *s* distinct training samples (x_i, y_i) with $x_i \in R^n$ and $y_i \in R^m$, an neuron activation function $f: R \rightarrow R$ which is infinitely differentiable, and the number of hidden layer neurons *N*:

Step 1. Randomly assign hidden layer parameters (input weights and biases);

Step 2. Calculate the hidden layer to output layer weight matrix *H*;

Step 3. Calculate the output weights matrix $O = H^{-1}Y$.

3. HONN MODELLING EXAMPLES

In this section, the ELM HONN modelling has been used for an image processing problem, a medical problem, and an energy efficiency problem. The algorithm has been implemented based on a HONN implementation in Matlab version R2011a, run on a standard Windows 7 operating system with a 4-core CPU speed of 2.70GHz and a RAM of 8GB.

To discover the advantages and disadvantages of the HONN model (with the Sigmoid activation function and one hidden layer), the following ANNs have also been applied onto the datasets of the problems for comparison studies: a conventional standard HONN model (with the Sigmoid activation function and one hidden layer); a Multi-Layer Perceptron (MLP) (with the Sigmoid activation function and one hidden layer); An RBF Neural Network (with the Gaussian activation function and one hidden layer). The standard MLP and RBF algorithms offered within the Neural Network Toolbox of Matlab version R2011b are used to train these traditional ANNs. The learning algorithm for the conventional HONN model is from Zhang et al (2002). For all of these ANNs the number of hidden layer neurons has been determined using an approach from Xu et al (2008).

3.1 Skin Segmentation

One of the significant topics in human face image recognition is to automatically determine skin and non-skin areas of a face image. A skin and non-skin dataset has been generated using skin textures which come from face images of people of different ages, genders, and races (Bache et al 2013). The dataset has been collected by sampling RGB (Red, Green, Blue) values of the face images. There are 4 attributes in the dataset with three of them representing the input attributes (the RGB values) and the last one representing the class attribute (skin or non-skin). There is a total of 245057 instances, however, for the purpose of this modelling only half of them have been used for faster learning and testing.

For this experiment, the dataset is divided into a training/learning set made of 80% of the original set and a test set made of 10% of the original set. The final 10% is used for evaluating the model's generalisation abilities.

For ELM HONN and Standard HONN (second order), the number of input neurons is calculated as follows:

$$3 + C_3^2 = 3 + \frac{3(3-1)}{2} = 6$$

For MLP and RBF networks, the number of input neurons is 3 (each for an input attribute).

The experimental results are displayed in Table 3.1. It can be seen that the ELM HONN is considerably faster than standard HONN, and in this case it produces an accuracy which is higher by 7.1%. For standard HONN model, training usually takes longer time because of the increased number of input neurons (compared with conventional MLP and RBF neural networks): in this experiment, the number of input layer neurons is 6 for the HONNs while 3 for the MLP and RBF network. However, ELM HONN is significantly faster. We can also see that the correctness rates (accuracy) produced by the conventional ANNs (MLP and RBF) are lower.

Table 3.1. Comparing ELM HONN against standard HONN, MLP, RBF neural networks. HL: Hidden Layer, TT: Training Time

ANN	Dataset	HL nodes #	TT (secs)	Correctness or accuracy
ELM HONN	Skin	9	10.8	84.2%
Standard HONN	Skin	9	38.7	77.1%
MLP	Skin	22	19.3	71.4%
RBF	Skin	22	17.4	73.5%

3.2 Freezing of Gait Problem

Freezing of Gait (FoG, inability to step) is a typical problem suffered by people with Parkinson's disease. A Daphnet Freezing of Gait dataset has been recorded to recognize gait freeze from wearable acceleration sensors placed on legs and hip of the patients who have volunteered to participate in the study (Bache et al 2013). There is a total of 237 instances, with 9 input attributes and 3 class attributes in the dataset. The input attributes are

- Ankle (shank) acceleration - horizontal forward acceleration
- Ankle (shank) acceleration - vertical
- Ankle (shank) acceleration - horizontal lateral
- Upper leg (thigh) acceleration - horizontal forward acceleration

- Upper leg (thigh) acceleration - vertical
- Upper leg (thigh) acceleration - horizontal lateral
- Trunk acceleration - horizontal forward acceleration
- Trunk acceleration - vertical
- Trunk acceleration - horizontal lateral

The 3 class attributes are:

- 0 (irrelevant movement)
- 1: no freeze (can be any of stand, walk, turn)
- 2: freeze

The challenge is to model this problem to determine what movements would cause a gait freeze (for patients with Parkinson's disease).

For this experiment, the dataset is divided into a training/learning set made of 75% of the original set and a test set made of 15% of the original set. The final 10% is used for evaluating the model's generalisation abilities.

For ELM HONN and Standard HONN (second order), the number of input neurons is calculated as follows:

$$9+C_9^2=9+\frac{9(9-1)}{2}=45$$

For MLP and RBF networks, the number of input neurons is 9 (each for an input attribute).

The experimental results are displayed in Table 3.2. It appears that for this experiment the ELM HONN produces similar accuracy as standard HONN, however, ELM HONN is significantly faster than standard HONN (as well as MLP and RBF networks). Additionally, both HONN modes produce higher accuracy than MLP and RBF neural networks.

Table 3.2. Comparing ELM HONN against standard HONN, MLP, RBF neural networks. HL: Hidden Layer, TT: Training Time

ANN	Dataset	HL nodes #	TT (secs)	Correctness or accuracy
ELM HONN	FoG	11	4.9	81.5%
Standard HONN	FoG	11	21.4	80.3%
MLP	FoG	25	11.4	70.2%
RBF	FoG	25	10.4	68.1%

3.3 Modelling Energy Efficiency

The third experiment deals with modelling energy efficiency of buildings: assessing the heating load and cooling load requirements of buildings based on 8 building parameters (factors) (Bache et al 2013). The buildings differ from each other with respect to the glazing area, the glazing area distribution, and the orientation, and others. There is a total of 768 instances in the dataset, with 8 input attributes (each representing a building feature). The 8 features are

- Relative Compactness
- Surface Area
- Wall Area
- Roof Area
- Overall Height
- Orientation
- Glazing Area
- Glazing Area Distribution

There are 2 class attributes:

- Heating Load
- Cooling Load

The challenge is to learn the relationships between the 8 features and its heating load and cooling load.

For this experiment, the dataset is divided into a training/learning set made of 70% of the original set and a test set made of 15% of the original set. The final 15% is used for evaluating the model's generalisation abilities.

For ELM HONN and Standard HONN (second order), the number of input neurons is calculated as follows:

$$8+C_8^2=8+\frac{8(8-1)}{2}=36$$

For MLP and RBF networks, the number of input neurons is 8 (each for an input attribute).

The experimental results are displayed in Table 3.3. Unsurprisingly the ELM HONN is considerably faster than standard HONN (as well as MLP and RBF), and in this experiment it produces an accuracy which is higher by 10.2% than standard HONN. We can also see that the simulation accuracies produced by the HONN models are significantly higher than the traditional ANN models (MLP and RBF).

Table 3.3. Comparing ELM HONN against standard HONN, MLP, RBF neural networks. HL: Hidden Layer, TT: Training Time

ANN	Dataset	HL nodes #	TT (secs)	Correctness or accuracy
ELM HONN	Energy	13	7.9	87.6%
Standard HONN	Energy	13	25.5	77.4%
MLP	Energy	33	15.3	68.3%
RBF	Energy	33	16.4	70.1%

4. CONCLUSIONS

This paper proposes an ELM algorithm for HONN models and applies it in an image processing problem, a medical problem, and an energy efficiency problem. An obvious outcome is that HONN model with ELM algorithm is significantly faster than standard HONN model as well as traditional ANNs such as MLP and RBF, due to the nature of ELM. It appears that, generally speaking, ELM HONN produces higher accuracy than standard HONN, as demonstrated in the first and third experiments, although occasionally (as in the second experiment) both models produce similar correctness rates. It can be seen that HONN models produce higher accuracies than MLP and

RBF neural networks. This paper is a report of work in progress. In the future, more experiments involving larger datasets will be conducted to further test the new ELM algorithm for HONN. Another direction for future research would be the use of an ensemble of HONN models for modelling and simulation.

ACKNOWLEDGMENTS

This project is supported by Chinese Universities Scientific Fund, Project number: 2013QJ051, and National Key Technology R&D Program of China during the 12th Five-Year Plan Period (Project numbers: 2012BAJ18B07, 2012BAD35B02).

REFERENCES

Alanis AY, Sanchez EN, Loukianov AG, Discrete-Time Adaptive Backstepping Nonlinear Control via High-Order Neural Networks, IEEE TRANSACTIONS ON NEURAL NETWORKS, 18 (4) (2007) 1185-1195.

Bache, K. & Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Dunis, CL, Laws J, and Sermpinis G, Higher order and recurrent neural architectures for trading the EUR/USD exchange rate, Quantitative Finance, 11 (4) (2011) 615–629.

Fallahnezhad M, Moradi, MH, Zaferanlouei S, A Hybrid Higher Order Neural Classifier for handling classification problems, Expert Systems with Applications, 38 (2011) 386–393.

Giles, C.L., Maxwell, T. (1987) Learning, invariance, and generalization in higher order neural networks, *Applied Optics*, 26(23), 4972-4978.

Huang, GB, Siew, CK, “Extreme Learning Machine with Randomly Assigned RBF Kernels,” International Journal of Information Technology, vol. 11, no. 1, pp. 16—24, 2005.

Huang, GB, Zhu, QY, Siew, CK, “Extreme Learning Machine: Theory and Applications”, Neurocomputing, vol. 70, pp. 489-501, 2006.

Huang, GB, Li, MB, Chen, L, and Siew, CK, “Incremental Extreme Learning Machine With Fully Complex Hidden Nodes,” Neurocomputing, vol. 71, pp. 576-583, 2008.

Kosmatopoulos, E.B., Polycarpou, M.M., Christodoulou, M.A., Ioannou, P.A. (1995). High-order neural network structures for identification of dynamical systems, *IEEE Transactions on Neural Networks*, 6(2), 422-431.

Lee, Y.C., Doolen, G., Chen, H., Sun, G., Maxwell, T., Lee, H., Giles, C.L. (1986) Machine learning using a higher order correlation network, *Physica D: Nonlinear Phenomena*, 22, 276-306.

Lippman, R.P. (1989) Pattern classification using neural networks, *IEEE Commun. Mag*., 27, 47-64.

Psaltis, D., Park, C.H., Hong, J. (1988) Higher order associative memories and their optical implementations, *Neural Networks*, 1, 149-163.

Rao C R, Mitra S K, (1972). Generalized Inverse of Matrices and Its Applications, New York: Wiley

Redding, N., Kowalczyk A. and Downs, T., (1993). “Constructive high-order network algorithm that is polynomial time”, *Neural Networks*, Vol.6, pp.997-1010.

Reid, M.B., Spirkovska, L., Ochoa, E. (1989). Simultaneous position, scale, rotation invariant pattern classification using third-order neural networks, *Int. J. Neural Networks*, 1, 154-159.

Thimm, G., Fiesler, E. (1997). High-order and multilayer perceptron initialization, *IEEE Transactions on Neural Networks*, 8(2), 349-359.

Wood, J., Shawe-Taylor, J. (1996). A unifying framework for invariant pattern recognition, *Pattern Recognition Letters*, 17, 1415-1422.

Xu S (2010a), Adaptive Higher Order Neural Network Models for Data Mining, Artificial Higher Order Neural Networks for Computer Science and Engineering: Trends for Emerging Applications, Information Science Reference, Ming Zhang (ed), Hershey, United States, 86-98. ISBN 978-1-61520-711-4.

Xu S (2010b), Data Mining Using Higher Order Neural Network Models With Adaptive Neuron Activation Functions, International Journal of Advancements in Computing Technology, 2 (4) 168 - 177.

Xu, S and Chen, L (2008), A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining, Proceedings The 5th International Conference on Information Technology and Applications, 23-26 June 2008, Carins, Qld, pp. 683-686.

Zhang, M., Xu, S., Fulcher, J., (2002). Neuron-Adaptive Higher Order Neural-Network Models for Automated Financial Data Modeling, *IEEE Transactions on Neural Networks*, Vol 13, No. 1.

EXPERIMENTAL COMPARISON OF IMAGE THRESHOLDING METHODS FOR DEFECT DETECTION IN THE PAPERMAKING PROCESS

Luca Ceccarelli[(a)], Francesco Bianconi[(b)], Stefano A. Saetta[(c)], Antonio Fernández [(d)] and Valentina Caldarelli [(e)]

[(a,b,c,e)] Department of Industrial Engineering
Università degli Studi di Perugia
Via G. Duranti, 67 – 06125 Perugia (ITALY)

[(d)] Department of Engineering Design
Universidade de Vigo
Campus Universitario – 36310 Vigo (SPAIN)

[(a)] lucap3600@gmail.com, [(b)] bianco@ieee.org, [(c)] stefano.saetta@unipg.it, [(d)] antfdez@uvigo.es,
[(e)] vale.caldarelli85@alice.it;

ABSTRACT

Automatic detection and assessment of dirt particles in pulp and paper plays a pivotal role in the papermaking industry. Traditional visual inspection by human operators is giving the way to machine vision, which provides many potential advantages in terms of speed, accuracy and repeatability. Such systems make use of image processing algorithms which aim at separating paper and pulp impurities from the background. The most common approach is based on image thresholding, which consists of determining a set of intensity values that split an image into one or more classes, each representing either the background (i.e.: an area with no defects) or an area with some types of contraries. In this paper we present a quantitative experimental evaluation of four image thresholding methods (i.e.: Otsu's, Kapur's, Kittler's and Yen's) for dirt analysis in paper. The results show that Kittler's method is the most stable and reliable for this task.

Keywords: machine vision, image thresholding, paper, quality assessment

1. INTRODUCTION

Product and process control through machine vision has been receiving increasing attention during the last years. Applications in the industry now cover many produts, such as textile (Carfagni *et al*. 2005), wood (Bianconi *et al*. 2013), ceramics (Kukkonen *et al*. 2001), natural stone (Bianconi *et al*. 2012), food (Furferi *et al*. 2010) and vehicles (Furferi *et al*. 2013) – to cite some.

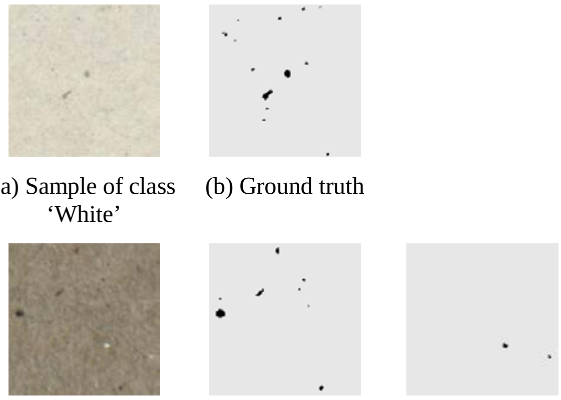
In the papermaking industry, machine vision proved effective in a number of problems, including printability analysis (Kalviainen et al. 2003); control of stripes and holes (Navarrete et al. 2003); assessment of the coating layer (Prykary et al. 2010); curl estimation (Synnergren *et al*. 2001), analysis of microstructural changes (Sjödahl and Larsson 2004) and automatic segregation of waste paper for recycling (Rahman *et al*. 2011). Among them, dirt inspection has always played a central role, due to the strong effect that such defects have on the quality of the final product. An excessive presence of contraries and impurities may cause the pulp or paper to be off-specification, with negative consequences for the producer. The detection and characterization of contraries is also a crucial step to track down and remove (or at least reduce) the source of impurities in the production process. The potential advantages are: a more efficient use of materials and energy, and a reduction of chemicals in the bleaching phase, with beneficial effects on the environment.

Various prototypes and systems for automatic dirt analysis and counting have been described in the literature – for an overview of methods see the works of Torniainen *et al.* (1999); Corscadden and Trepanier (2006) and Ricard *et al*. 2012. From a technical standpoint, the detection of whatever type of particles in pulp and paper can be viewed as an image segmentation process aiming at separating the contraries (foreground) from the rest of the product (background). Most commonly, defects are dark spots on a bright area; but in some types of paper they may well be both brighter and darker than the background. In the paper recycling process, for instance, we expect to find not only traces of toner and wood particles – which tend to be darker than the background – but also stickies – which are likely to be brighter than the background. In either case the segmentation process requires determining one or more intensity values (thresholds) for separating whatever type of defects from the background in the correct way. In this paper we present a quantitative experimental evaluation of four image thresholding methods that can be used for this task. Of each method we consider both the standard single-threshold version, which can be used when defects are all darker or brighter than the background, and the more challenging double-threshold version, which is required when defects are either darker or brighter than the background. To assess the accuracy of the methods in a quantitative way, we compare the results of automatic segmentation against a 'ground truth' of contraries manually generated and cross-validated by two human experts.

In the remainder of the paper we first give an account of the materials and image acquisition devices used in our study, followed by a description of the thresholding methods included in the comparison. Then we outline the experimental set-up, summarize the main results of the study and conclude the paper with some final considerations.

2. MATERIALS

We considered two different classes of recycled paper. According to their appearance, we conventionally refer to the two classes as 'White' and 'Brown' (see Fig. 1).



(a) Sample of class 'White'

(b) Ground truth

(c) Sample of class 'Brown'

(d) Ground truth – defects darker than background

(e) Ground truth – defects brighter than background

Figure 1 (a) Sample of class 'White' and (b) the corresponding ground truth; (c) sample of class 'Brown', and the corresponding ground truth for defects (d) darker and (e) brighter than the background.

Each class includes three sub-classes of different density. The characteristics of each class are reported in Tab. 1.

Table 1 Summary table of the materials used in the experiments.

Class	Sub-class	Density (g/m^2)	No. of samples	Image resolution
White	W1	137	20	400 × 400
	W2	154		
	W3	174		
Brown	B1	154	20	400 × 400
	B2	137		
	B3	137		

For each class we obtained a set of 20 specimens and acquired them through the imaging system described in Sec 2.1. Samples of class White present only defects that are darker than the background. We therefore used this set of samples to test the single-threshold version of the algorithms. By contrast, samples of class Brown show defects that are either brighter or darker than the background (see Fig. 1). Their analysis therefore requires the two-threshold version. The 'true' location and extension of the defects of each sample ('ground truth') have been manually determined and cross-validated by two skilled operators.

2.1. Image acquisition

The imaging system used in the experiments (Fig. 2) is composed of the following parts: one dome illuminator (Monster Dome Light 18.25"), one industrial CMOS camera, one support for the camera, one base and one slot to accommodate the paper specimen. The imaging apparatus can operate either by transmitted or reflected light. The lens can be selected to suit the specific application needs. In this activity we used a 12 mm fixed focal length objective (Pentax H1214-M). The whole imaging system provides a spatial resolution of approximately 370 dpi. The acquisition was carried out in reflected light mode.

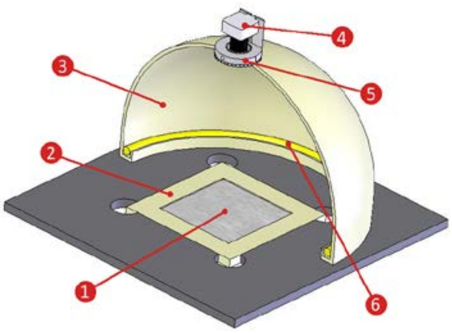


Figure 2 The image acquisition system: 1) paper sample; 2) slot; 3) hemispherical Lambertian surface; 4) camera, 5) rotatable support and 6) illumination ring.

3. METHODS

The problem of segmenting the image of a paper specimen through thresholding consists of determining a set of intensity values $G = \{G_0,\ldots,G_C\}$ that splits the image into a set of C classes, each corresponding to intensity values $i \in [G_{c-1},G_c[$. One of these classes will represent the background of the product; the others different classes of impurities. The case $C = 2$ is the most common, and occurs whenever we need to detect dark/blackish particles on a bright background (but the reverse may also occur). The cases $C > 2$ represent more complex scenarios, in which we have to look for more than one class of impurities. As we mentioned in the preceding sections, we limit our investigation to the cases *C = 2* and *C = 3*. The determination of a proper set of thresholds for a given image (thresholding) has been studied extensively in literature, and several methods exist – for a comprehensive review on the subject see the work of Sezgin and Sankur (2004). Nonetheless, no quantitative data are available, in the literature, as for the effectiveness of the methods for dirt analysis in pulp and paper. In this study we considered four parameter-free, computationally light and easy to implement methods. They are: Kapur's, Kittler-Illingworth's, Otsu's and Yen's. Here below we summarize the basics of each method. References are provided for the benefit of readers interested in the technicalities. All methods take as input the first-order probability distribution (histogram) of gray-levels; we therefore assume that the original images are converted to grayscale before processing. In Equations 1-4 we preliminarily define the weight, mean, standard deviation, entropy and correlation of each *c*-th class:

Weight
$$\omega_c = \sum_{i=G_c}^{G_{c+1}-1} p_i \qquad (1)$$

Mean
$$\mu_c = \sum_{i=G_c}^{G_{c+1}-1} \frac{ip_i}{\omega_c} \qquad (2)$$

Variance
$$\sigma_c = \sum_{i=G_c}^{G_{c+1}-1} (1-\mu_c)^2 \frac{p_i}{\omega_c} \qquad (3)$$

Entropy
$$E_n = \sum_{i=G_c}^{G_{c+1}-1} p_i \log_2 \frac{1}{p_i} \qquad (4)$$

Correlation
$$CR_n = \log_2 \sum_{i=G_c}^{G_{c+1}-1} \left(\frac{\omega_c}{p_i} \right)^2 \qquad (5)$$

where p_i is the probability of the *i*-th grey-value.

3.1. Kapur

In Kapur's method (Kapur *et al.* 1985) the set of optimal thresholds, indicated as $\overline{\mathbf{G}}$ in the following equations, are the intensity levels that maximize the sum of the entropy of each class (Eq. 6). For this reason the procedure is also referred to as *maximum entropy criterion*.

$$\overline{\mathbf{G}}_{\text{Kapur}} = \underset{\mathbf{G}}{\arg\max} \left(\sum_{c=1}^{C} E_c \right) \qquad (6)$$

3.2. Kittler-Illingworth

This approach assumes that the gray-scale histogram of the whole image can be approximated through a mixture of *N* Gaussian distributions, one for each class. Optimal thresholds are the values that minimize the error between the original histogram and the mixture of the approximating distributions (Kittler and Illingworth 1986). In formulas we have:

$$\overline{\mathbf{G}}_{\text{Kittler}} = \underset{\mathbf{G}}{\arg\max} \left[\sum_{c=1}^{C} \omega_c \ln \left(\frac{\omega_c}{\sigma_n} \right) \right] \qquad (7)$$

3.3. Yen

Yen's method (Yen *et al.* 1995) is formally very similar to Kapur's, but instead of maximizing the sum of the entropy of each class, it sets the optimal thresholds at the values that maximize the sum of the correlation of each class (Eq. 8). Therefore the method is also known as *maximum correlation criterion*.

$$\overline{\mathbf{G}}_{\text{Yen}} = \underset{\mathbf{G}}{\arg\max} \left(\sum_{c=1}^{C} CR_c \right) \qquad (8)$$

3.4. Otsu

Otsu's method determines the set of thresholds that maximizes the between-class variance. Originally designed for two level thresholding (Otsu 1979), it has been later extended to the multi-class domain (Liao *et al.* 2001). Mathematically, the method can be formalized as follows:

$$\overline{\mathbf{G}}_{\text{Otsu}} = \underset{\mathbf{G}}{\arg\max} \left[\sum_{c=1}^{C} \omega_c (\mu_c - M)^2 \right] \qquad (9)$$

where *M* is the average intensity of the whole image.

4. EXPERIMENTS AND RESULTS

We carried out a set of experiments to quantitatively evaluate the goodness of the thresholding methods at separating paper impurities from the background. To assess the effectiveness of each method we considered the following parameters: overall accuracy, normalized number of false positives and normalized number of false negatives.

4.1. Overall accuracy

The overall accuracy is the sum of the percentage of foreground pixels (i.e.: defects) correctly classified as foreground and that of background pixels (i.e.: non-defects) correctly classified as background. This parameter gives an overall estimate of the effectiveness of the segmentation process. In formulas we have:

$$A = \frac{|B \cap B_T|}{|I|} + \frac{|F \cap F_T|}{|I|} \qquad (10)$$

where *A* is the overall accuracy; *I* the whole image; *B* and *F* the background and foreground produced by the thresholding method; B_T and F_T the 'true' background and foreground, which have been manually established beforehand. Symbol '| |' stands for 'the number pixels of'.

4.2. False positives

False positives represent 'type I errors': a false positive occurs each time a background pixel (i.e.: non-defect) is incorrectly classified as foreground (i.e.: defect). The normalized number of false positives can be expressed as follows:

$$FP = \frac{|F \cap B_T|}{|I|} \tag{11}$$

4.3. Fale negatives

False negatives are also referred to as 'type II errors'. A false negative arises each time a foreground pixel (i.e.: defect) is incorrectly classified as background (i.e. non-defect). In formulas we have:

$$FN = \frac{|B \cap F_T|}{|I|} \tag{12}$$

4.4. Results

Tables 2-4 summarize the performance of the image thresholding methods considered in the experiment.

Table 2 Overall results of the single-threshold (two-class) experiment.

	Kapur			**Kittler-Illingworth**			**Otsu**			**Yen**		
Dataset	FN	FP	A	FN	FP	A	FN	FP	A	FN	FP	A
W1	0,132	0,004	99,864	0,046	0,095	99,860	0,000	42,546	57,453	0,148	0,002	99,850
W2	0,225	0,003	99,773	0,097	0,128	99,774	0,000	40,619	59,381	0,249	0,001	99,750
W3	0,284	0,001	99,715	0,163	0,066	99,770	0,000	42,689	57,311	0,297	0,000	99,703
Avg	**0,214**	**0,003**	**99,784**	**0,102**	**0,096**	**99,801**	**0,000**	**41,951**	**58,048**	**0,231**	**0,001**	**99,768**

Table 3 Overall results of the double-threshold (three-class) experiment – defects brighter than the background.

	Kapur			**Kittler-Illingworth**			**Otsu**			**Yen**		
Data set	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC
B1 (w)	0,053	0,027	99,920	0,052	0,023	99,925	0,150	0,000	99,850	0,073	0,016	99,911
B2 (w)	0,121	4,980	94,899	0,126	0,001	99,872	0,237	0,000	99,763	0,141	9,935	89,924
B3 (w)	0,129	0,008	99,863	0,166	0,000	99,833	0,292	0,000	99,708	0,164	0,003	99,833
Avg	**0,101**	**1,672**	**98,227**	**0,115**	**0,008**	**99,877**	**0,226**	**0,000**	**99,774**	**0,126**	**3,318**	**96,556**

Table 4 Overall results of the double-threshold (three-class) experiment – defects darker than the background.

	Kapur			Kittler-Illingworth			**Otsu**			**Yen**		
Data sets	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC
B1 (b)	0,069	0,028	99,903	0,073	0,015	99,913	0,000	50,253	49,747	0,078	5,004	94,918
B2 (b)	0,260	9,947	89,793	0,276	0,016	99,708	0,000	47,171	52,829	0,324	9,937	89,739
B3 (b)	0,200	0,019	99,781	0,175	0,033	99,791	0,000	45,604	54,396	0,229	0,010	99,761
Avg	**0,176**	**3,332**	**96,492**	**0,175**	**0,021**	**99,804**	**0,00**	**47,676**	**52,324**	**0,210**	**4,984**	**94,806**

In the single-threshold experiment (Tab. 2) Kapur's, Kittler's and Yen's methods all showed good accuracy with comparable figures. By contrast, the performance of Otsu's algorithm was largely unsatisfactory. Among the first three approaches, Yen's and Kapur's produced less false positives, whereas Kittlers' produced less false negatives.

In the double-threshold experiment (Tab. 3-4), Kittler's method appreciably outperformed the others in terms of overall accuracy. This trend is even more evident when it comes to determining defects that are darker than the background (Tab. 4). Otsu's approach proved rather unreliable in this case too, with an overall accuracy far lower than the other methods. Kittler's method also produced fewer false positives in this experiment, whereas the number of false negatives is similar to that produced by the other methods.

5. CONCLUSIONS

Automatic dirt detection and analysis through machine vision plays a central role in the papermaking industry. A fundamental issue in this process is the problem of separating dirt particles from the background through suitable image processing methods. The typical strategy consists of determining a set of intensity values (thresholds) capable of separating the impurities from the background. In this context we have evaluated, experimentally, the performance of four thresholding methods on a dirt detection experiment. Among the four strategies considered here, the method proposed by Kittler and Illingworth (Kittler and Illingworth 1986) proved the most stable and reliable for dirt analysis.

ACKNOWLEDGMENTS

This work was supported by the European Union under project no. Life09-ENV/FI/000568: VOCless pulping waste waters.

REFERENCES

Carfagni, M., Furferi, R., and Governi, L. 2005, A real-time machine-vision system for monitoring the textile raising process. *Computers in Industry*, 56 (8-9), 831–842

Bianconi, F., Fernández, A., González, E. and Saetta, S.A., 2013, Performance analysis of colour descriptors for parquet sorting; *Expert Systems with Applications*, 40 (5), 1636-1644

Kukkonen, S., Kälviaïnen, K., and Parkkinen, J., 2001, Color features for quality control in ceramic tile industry. *Optical Engineering*, 40 (2), 170–177, 2001.

Bianconi, F., González, E., Fernández, A. and Saetta, S.A., 2012, Automatic classification of granite tiles through colour and texture features, *Expert Systems with Applications*, 39 (12), 11212–11218

Furferi, R., Governi, L., Volpe, Y., ANN-based method for olive ripening index automatic prediction, 2010, *Journal of Food Engineering*, 101 (3), 318-328

Furferi, R., Governi, L., Volpe, Y., Carfagni, M., 2013, Design and assessment of a machine vision system for automatic vehicle wheel alignment, *International Journal of Advanced Robotic Systems*, 10, art. no. 242

Kälviäinen, H., Saarinen, P., Salmela, P., Sadovkinov, A. and Drobchenko, A., Visual inspection on paper by machine vision, 2003, *Proc. of. SPIE Intelligent Robots and Computer Vision XXI: Algorithms, Techniques and Active Vision*. Vol. 5267, pp. 321-332

Navarrete, H., Cadevall, C., Bouydain, M., Antó, J., Pladellorens, J.M., Colom, J.F., and Tosas, A., System for off-line optical paper inspection and quality control, 2003, *In Optical Measurement Systems for Industrial Inspection III, volume 5144 of Proceedings of SPIE*, pp. 774–782

Prykäri, T., Czajkowski, J., Alarousu E. and Myllylä, R., 2010, Optical Coherence Tomography as an Accurate Inspection and Quality Evaluation Technique in Paper Industry, *Optical Review*, 17 (3), 218-222

Synnergren, P., Berglund, T. and Söderkvist, I., Estimation of curl in paper using a combination of shape measurement and least-squares modeling, 2001, *Optics and Lasers in Engineering*, 35 (2), 105-120

Sjödahl, P. and Larsson, L., 2004, Monitoring microstructural material changes in paper through microscopic speckle correlation rate measurement, Optics and Lasers in Engineering, 42 (2), 193-201

Rahman, M. O., Hussain, A., Scavino, E., Basri, H. and Hannan, M. A., 2011, Intelligent computer vision system for segregating recyclable waste papers, *Expert Systems with Applications*, 38 (8), 10398-10407

Torniainen, J.E., Soderhjelm, L.S.A. and Youd G., 1999, Results of automatic dirt counting using transmitted light, *TAPPI Journal*, 82 (1), 194-197

Corscadden, K.W., Trepanier, R.J., 2006, Online measurement of dirt specks in sheets, *Proc. of TAPPI Engineering, Pulping, and Environmental Conference*

Ricard, M., Dorris, G., Gendron, S., Pagé, N., Filion, D. and Castro, C., 2012, A new online image analyzer for macrocontaminants in recycled pulps, *TAPPI Journal*, 11 (2), 19-28

Sezgin, M., Sankur, B., 2004, Survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic Imaging*, 13 (1), 146-165

Kapur, J., Sahoo, P. and Wong, A., 1985, A new method for gray-level picture thresholding using the entropy of the histogram, *Computer Vision, Graphics and Image Processing*, 29 (3), 273-285

Kittler, J. and Illingworth, J., 1986, Minimum error thresholding, *Pattern Recognition*, 19 (1), 41-47,

Yen, J.C., Chang, F. and Chang, S., 1995, A new criterion for automatic multilevel thresholding, *IEEE Transactions on Image Processing*, 4 (3), 370-378

Otsu, N., A Threshold selection method from gray-Level histograms, 1979, *IEEE Transactions on Systems, Man and Cybernetics*, 9 (1), 62-66

Liao, P., Chen, T. and Chung, P., A fast algorithm for multilevel thresholding, 2001, *Journal of Information Science and Engineering*, 17 (7), 13-727,

AUTHORS' BIOGRAPHIES

Luca Ceccarelli received a B.Sc. and M.Sc. in Mechanical Engineering from the University of Perugia, Italy. He is currently Research Assistant within the Department of Industrial Engineering of the same University. His research interests include computer vision and intelligent systems for industrial applications.

Francesco Bianconi received the M.Eng. degree in Mechanical Engineering in 1997 from the University of Perugia (Italy) and the Ph.D. in computer-aided design in 2001 from a consortium of Italian universities. He has been visiting research fellow at the University of

Vigo (Spain) and the University of East Anglia (UK). Currently, he is Lecturer within the Faculty of Engineering of the University of Perugia. His research interests include computer vision, image processing and pattern recognition, with a special focus on texture and color analysis. He is IEEE Senior Member.

Stefano A. Saetta is Associate Professor of Industrial Plants within the Department of Industrial Engineering of the University of Perugia, Italy. His research interests cover: modeling and simulation of logistics and production processes, life cycle assessment, discrete-event simulation, decision methods, lean production, networks of enterprises and environmentally friendly production systems. He has been visiting professor at Rutgers University (USA), the University of Arizona (USA) and the University of Göttingen (Germany). He has directed several national and international research projects supported by private and public companies. He authored/co-authored more than 80 scholarly papers in international journals and conferences.

Antonio Fernández received the M.Eng. degree in electrical engineering in 1993 and the Ph.D. degree (with honors) in applied physics in 1998, both from the University of Vigo, Vigo, Spain. He held a research fellowship in the Department of Applied Physics, University of Vigo, during the period 1994 through 1998. He was appointed to the Department of Engineering Design, University of Vigo, in 1999, where he is currently full-time Senior Lecturer in Engineering Drawing. He has worked as a visiting researcher at Centre for Research on Optics (Mexico), University of Perugia (Italy), Dublin City University (Ireland) and Computer Vision Centre (Spain). His research interests are in image processing, pattern recognition and computer vision, with a special focus on image texture analysis.

Valentina Caldarelli received a B.Sc. and M.Sc. in Mechanical Engineering from the University of Perugia, Italy. She is currently Research Assistant within the Department of Industrial Engineering of the same University. She works on the reduction of VOC emissions in papermaking process and on the simulation and analysis of the supply chain the simulation and analysis of the supply chain in the food industry.

A SOFTWARE ARCHITECTURE FOR INTEGRATED LOGISTIC MANAGEMENT SYSTEM

Fabrizio Simeoni*, Srecko Maksimovic*, Walter Geretto* George Georgoulas, Chrysostomos Stylios****

*Teorema Engineering Srl, Area Science Park Basovizza, Trieste, Italy
**Technological Educational Institute of Epirus, GR-47100, Arta, Greece

Fabrizio.Simeoni@teorema.net; Srecko.Maksimovic@teorema.net; Walter.Geretto@teorema georgoul@gmail.com; stylios@teiep.gr

ABSTRACT
This paper presents an integrated software architecture for the management of transfer goods (trucks and containers) between port and dry-port facilities. This system is a large scale system and it has to deal with huge and continuously updated set of information. The required information is gathered from other information systems managing business activities within the involved areas. This is the reason why there was required the development of software components that are able to manage the processes of acquiring and sending information to the other systems. In this paper the technological choices and the main information flows managed are described.

Keywords: .NET - BizTalk - Marketplace – On Board Computer - RFiD tag – SAIL - SOA – SQL Server-Subscription – Visual Studio 2010 – WCF – Web service – XML.

1. INTRODUCTION

In this paper we are presenting the characteristics of the implementation of the software solution developed within the "ICT System addressed to Integrated Logistic Management and Decision Support for Intermodal port and dray port facilities, SAIL project. The project aims at developing an integrated ICT tool able to support efficiently logistic chain of goods flow and all business operations provided in the port and the dry port areas, mainly related to the transfer between the two facilities (Caris et al., 2008; Turban et al., 2010).

Within the SAIL project, we are designing and implementing the planning and scheduling tools for optimizing the transfer of goods and the sizing of the resource capabilities. To achieve this goal, we have to gather information from other systems in order to know how many transfers are foreseen, requested, when they have to be executed, etc., in other words all the operational data needed by our algorithms to provide reliable results.

The present paper contains a description of the architecture of the implemented solution for the central hub, which is the central SAIL module, aimed to provide the functionalities for data exchange among sub-modules. We define the development environments chosen, programming languages, and the technological choice of BizTalk Server 2010 as the integration tool due to its flexibility and modularity. Moreover we describe the details of the implementation of web services and the routing logic of messages. Microsoft technology has been used as it is the core of Teorema activities.

The rest of the paper is structured as follows: Section 2 presents the overall software architecture. The communication flows are described in Section 3 and Section 4 concludes the papers and presents future developments.

2. ARCHITECTURE

2.1 Programming Languages

The chosen programming language is C # and the software development is based on version 4 of .NET Framework (Figure 1) (Troelsen 2012). The framework libraries for WCF technology have been very important to the development of web services that allows achieving a high level of interoperability and communication with other software modules implemented with technologies other than .NET. Another important aspect is the Workflow Foundation that is the technology on which the designer offered by the SDK (Software Development KIT) of BizTalk Server is based: it was used to define the data orchestrations that implement the logic for message routing through the central hub (Rosanova, 2011).

2.2 Development environment

The chosen development environment is Visual Studio 2010 Ultimate Edition (Randolph *et al.*, 2010). This is the latest Integrated Development Environment (IDE) developed by Microsoft for programmers who develop for Windows and .NET Framework 4.0. It allows the utilization of multiple programming languages, including VB.NET, C ++, C# and others. It also offers the ability to create applications and ASP.NET Web Services in C# or VB.NET (Pathak, 2011). BizTalk Server provides developers with an SDK as an extension for the Visual Studio projects using visual tools.

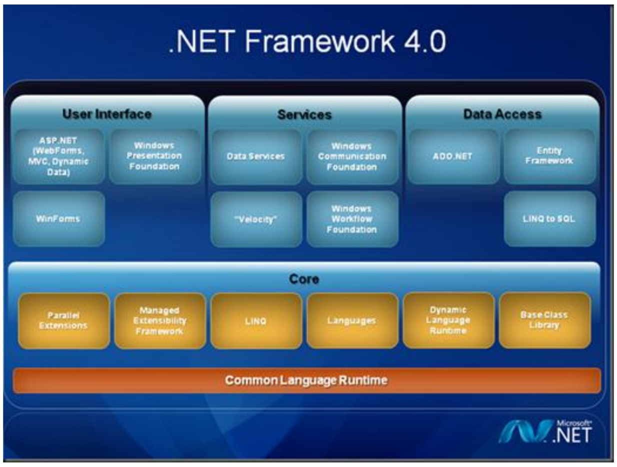


Fig. 1. The .NET Framework 4.0 structure.

2.3 Database server

The engine Database Management System (DBMS) is SQL Server 2008 R2. It is a Relational Database Management System (RDBMS) produced by Microsoft. In early versions it was used for medium-sized databases, but starting from version 2000 it is used also for management of large data bases.

2.4 BizTalk Server

2.4.1 The product

Microsoft BizTalk Server is a middleware (management layer of business logic) that is specializes in the management of business processes and their integration. It is a Business Process Management (BMP) system, i.e. a system that the company uses to create a layer of management of business flows between applications being able to apply existing rules and customizable parameters and can be monitored through a Business Activity Monitoring (BAM) system. In connecting applications across the enterprise, BizTalk Server is an EAI Enterprise Application Integrator (EAI) and also a Message Broker as it converts the native formats of the applications it connects via a system of adapters. BizTalk Server can function as Enterprise Service BUS (ESB) to create Service Oriented Architecture (SOA) infrastructure in which we can grow a forest of services both internal to the company and external - by mechanisms like SaaS (Software As A Service) or Cloud Computing. An ESB manages the flow of information between applications such as exchange of messages, receiving, processing and delivering it, based on metadata associated with the message itself that define the set of operations the message has to go through.

A highly common scenario is that BizTalk is used in companies after having purchased/built systems such as IBM Mainframe, SAP, Navision, web applications, asp.net, jsp/jsf, php, various services, SOAP, etc.. The company faces the need to make these systems coexist, in order to keep data synchronized between the different systems, and even to let a complex operation become feasible on multiple systems simultaneously without writing an application from scratch specifically for such functionality.

One of the requirements of the SAIL system is the interaction between different software modules within the port-dry port system and at the same time to allow integration of both internal and external processes within the business activities to enhance their flexibility and interoperability.

The central hub abstracts the connection between the service and the transport, trying to make it neutral. The solution is actually constituted by a series of processes that can be reused in many other realities, also extremely diversified. Using BizTalk Server provides the infrastructure that enables these processes, and entities related to these data, to be easily inserted into a SOA and they can be used for many other workflows and services within the port.

In the context of SOA, all documents can be transferred safely, while validating the content. Thanks to the reliability of BizTalk, in case of an error, the messages and processes can be completely recovered and the state management is guaranteed.

2.4.2 Operation principles

BizTalk Server architecture is based on the concepts of publication and subscription of messages and is built entirely using XML as a mechanism for data representation. The operation principles of BizTalk are shown in Figure 2 and described in the following paragraphs.

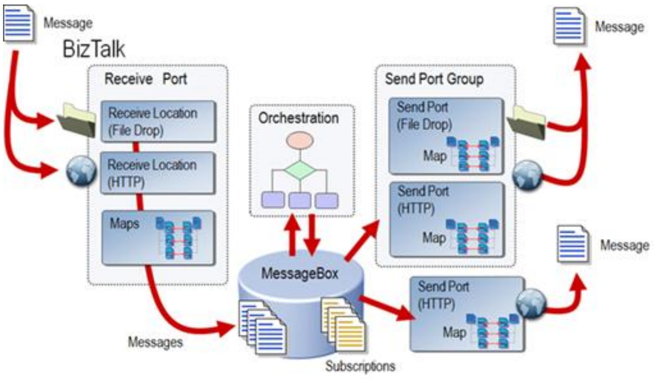


Fig. 2. The operation principle of BizTalk.

Messages are received through the receive port. Each receive port includes more receive locations associated with an adapter that allows communication with a particular type of external entity. The received message goes through a pipeline that takes care of operations like decoding, disassembling, validation and identification of partners. The messages can then be processed in a particular xslt map and then be published in the Message Box. The Message Box is a database for BizTalk internal use: it has a large number of tables many of which are used to store received messages. Each message has associated metadata called Message Context and each element is maintained by a key/value pair called Context Property. Every message that enters in the Message Box is "*published*". There is a rule according to which the posted messages cannot be changed.

The subscription is a mechanism by which doors and orchestrations are able to receive and send messages. Every process running BizTalk Message Agent has a Message Agent that looks for messages responding to subscriptions and occur to subscriptions and routs them to the EPM and sends them where they have to go. EPM is the broker between the Message Box and the pipeline/port/adapter. Subscriptions of orchestration are handled by a different service sub-called XLANG/s. A subscription is a collection of statements of comparison known as predicates that include context property of messages.

Subscriptors who have a valid subscription to a message can receive it and send it to the orchestration or send port. The sending process is similar to the receiving process. Messages sent to a send port can be transformed according to a map and then go through a send pipeline in which the steps are executed for validation, assembling and coding. Eventually, the adapter associated with the send port will transform the xml message in the compatible format, based on the type of adapter.

2.4.3 Monitoring

BizTalk activities can be monitored using the Health and Activity Tracking technology (HAT), and directly through the solutions of the Office suite or Share Point Portal Server 2003 with BAM. The central hub can include a BAM portal ready for use, which allows users to easily examine and configure BAM information. Using the BAM portal, users can select a particular instance of the business processes to be monitored, and then choose a specific BAM view into the process to get a different perspective of key performance indicators monitored. Users could be allowed to receive BAM information such as notifications by e-mail or other communication channels, supporting decision-making processes in real time. You can also expose Web services queries on aggregated data and instances, create alerts and retrieval of BAM configurations. The interface for the Web service can then be used to expose the functionality of the BAM portal interface.

2.4.4 Reliability and scalability

The high reliability of the platform is guaranteed by BizTalk Server that provides this capability natively. The central hub is capable of operating in high reliability and load balancing. Balancing the load network guarantees optimal performances, and the ability to add more servers if necessary and perform maintenance on any server without neither consequences for users nor hardware configuration changes. Through the combination of network load balancing and redundant clusters, administrators can add and remove cluster with minimal impact on users.

2.5 Web services and communication

SOA is a software architecture designed to support the use of services (web services) to meet the requirements of users in order to allow the use of single applications as part of the full business process. As part of a SOA architecture, it is possible to modify, in a relatively simple manner the mode of interaction between services with a very flexible design, making dynamic the combination in which the services are used in the process: in this way it is easier to add new services and change processes to meet specific business and process needs. The flow of information is not bound by a specific platform or by an application but can be considered as a component of a larger process, and then reused or changed.

3. *COMMUNICATION FLOWS*

In this section, flows are described, that are managed through the software architecture described in the previous section.

3.1 The overall system

The system consists of five distinct software modules: 1) Central Hub (HC), 2) Gateway3) Emergency Management Component (EMC), 4) Marketplace and 5) Scheduler

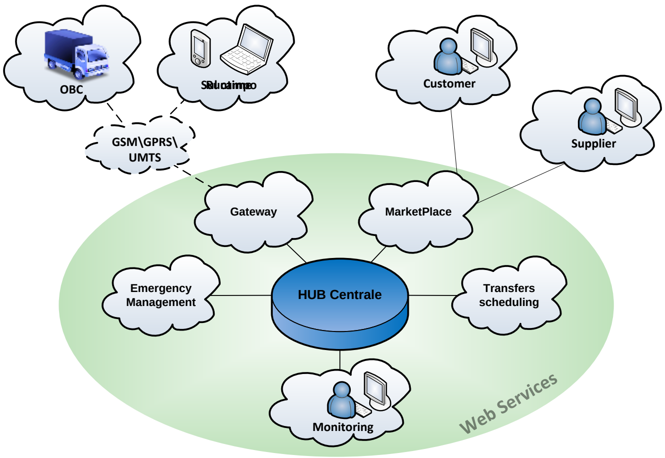


Fig. 3. The overall system.

The communication between the central hub and the other modules is via web service. Web services exposed by the various modules are implemented with different technologies (WCF, asmx, axis2, metro, etc.) and interoperability is guaranteed by compliance with the standard basic profile 1.1.

A global view of the information flows managed within the SAIL project information system is shown in Figure 3, from a logical point of view, and Figure 4, from an architectural point of view. Single scenarios are discussed in the following sections.

3.1 Single scenarios

Central Hub - Gateway

The communication between Central hub and Gateway allows o capture information from the OBC mounted on individual vehicles in transit through the territory and to carry out the functions of group management – a group is formed as a unit consisted of the travel drive and one or two containers.

All communications originating from HC go to the OBC through the Gateway who sends the messages (Figure 5). The Gateway module is in charge to apply the retry policies to the OBC.

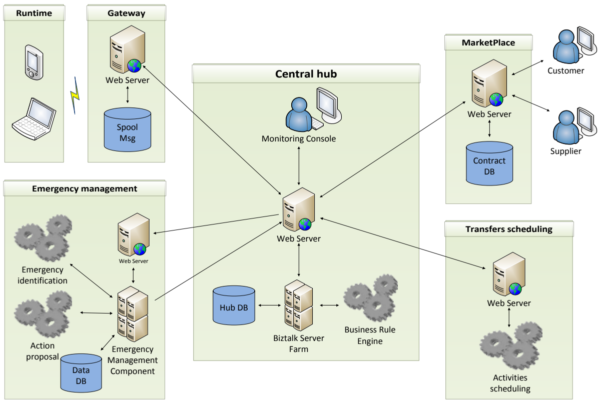


Fig. 4. The modules of the system.

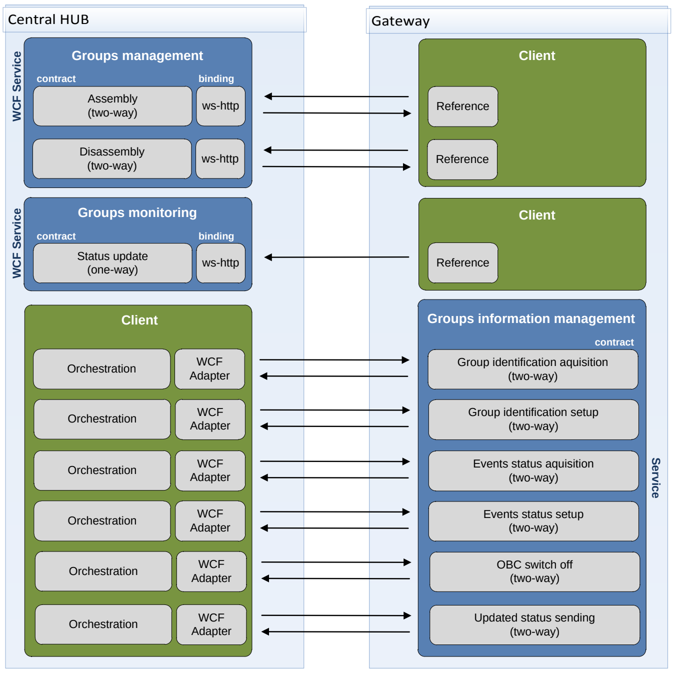


Fig. 5. Exchange of messages between Central Hub and Gateway.

The process of assembly consists of the following steps:

1. the operator installs the RFID tags on containers and on-board computer on the tractor
2. the operator - using a palm device - acquires the identifiers of RFID tags and of the on-board computer. Then it transmits via GSM/GPRS/UMTS to the Gateway module all related information to the equipped vehicle (for example license plate, identifier of the OBC, the identifier of the tag of the container, the container identification, etc)
3. the Gateway, through WCF call, sends this information to HC in synchronous mode
4. the HC generates a new Group ID (id OBC, drive plate, id tags (MAC address), container serial) and responds to the previous call
5. the Gateway responds to the palm with the outcome of the operation and simultaneously sends to the OBC the assembly
6. the operator checks the status LEDs on the OBC and gives the OK to drive, or checks where the procedure failed

The process of disassembly consists of the following steps:

1. EMC sends through a WCF door - exposed in Basic-HTTP – the piece of information that some ID group arrived at the end of its transfer
2. through the Gateway, HC changes the configuration parameters of the OBC in order to turn off periodic sending of messages, but leaves it working on events (e.g. opening of container, etc.): in this way no power is wasted and there are no useless information sent around
3. with the palm device, the operator generates the request of disassembly, reading tags of OBC and containers. He sends this information and waits for the response from the HC that the procedure is successful. Possibly something may be wrong as these tags have never been used for assembly, etc.
4. in the case of a positive outcome of the previous step, HC sends the shutdown message to the OBC

During the transfer, every 30 seconds (configurable parameter), the OBC sends all the information about the state of RFID tags and alarms, through the gateway. Furthermore, for some alarms, like the pressure of the button signalling the presence of an unusual event, the message starts immediately. For this information, HC has a WCF port listening for messages arriving through the gateway. Once arrived, information is written into the database and forwarded to the EMC by a web service exposed by the module itself.

The services offered by Gateway to HC include:

Acquisition of group identification: the HC sends a request that contains the identifier of an OBC. HC receives from gateway the identifier of the group that is currently associated with, or an error message if it no active group is identified

Configuration of Group ID: it is used in the process of disassembly to change the identifier of a group which must be broken and update its value setting it equal to 0

Acquisition of Event Status: given the identifier of an OBC, it returns the status of the events associated with it (eg: battery, GPS connection, GSM connection, etc.)

Configuration of Event Status: it allows changing the status of the configured events on the OBC and in particular the state of sending periodic messages that is disabled when arrival in a safe area is notified by the EMC

OBC switch off: it is used in the process of undressing, to notify the gateway that the OBC must be turned off. The gateway is responsible to send the communication to the OBC

Status Update: it allows the operator to request a status update for a particular group by entering its ID in the request. The request shall be made in case of lack of status for a period larger than expected, or to verify the consistency of the information received from a previous update

Central Hub – Emergency Management Component

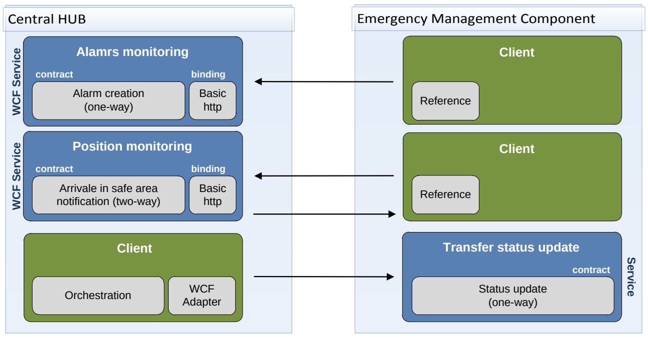


Fig. 6. Interaction between Central Hub and Emergency Management Component.

The services exposed from the HC to EMC include (Figure 6):

Alarm creation: the EMC uses this service to indicate the presence of an alarm identified by an alarm code and accompanied by the description of the action to be taken to manage the emergency

Notification of arrival in safe area: EMC analyzes the status messages and determines when a group reaches the target area, so it is therefore in a safe place. Once the group has arrived at destination, it is no longer needed the transmission of position, state of the tag, etc.. The EMC module sends a message to the HC for notifying the arrival to destination and disable the periodical sending of status messages.

The service exposed by the EMC is (Figure 7):

Status update: HC uses this service to forward to EMC all status messages received by the gateway. All received messages are forwarded without filters. Inspecting the content of individual messages and analyzing their content in relation with archived messages, EMC is able to detect possible alarm situations and notify the HC. The service call does not wait for a response, but it only sends the updated information

Central Hub – Operational level

The communication between the HC and the Operational level is designed to obtain the scheduling of transport activities booked through the Marketplace portal or directly in the traffic center. The scheduler input is a list of activities called units. A unit may represent a container or a vehicle.

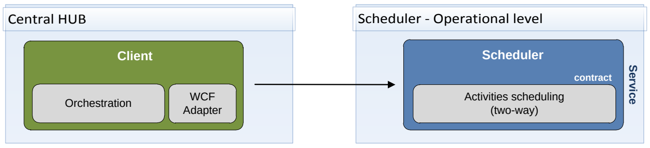


Fig. 7. Connection between Central Hub and Operational level.

Central Hub – Marketplace

Transfer requests are generated within the Marketplace for customers who wish to book the transfer service or from the (dry-)port management system, for vehicles which arrive there without reservation. The (dry-)port staff takes charge of entering into the system orders through an appropriate system outside of SAIL. For the purposes scheduling, both kinds of activities requests are included in the scheduled plan: a) in the first case the description of the activity is obtained from a web service exposed by the marketplace portal, b) in the second case the description of the activity is obtained from the external system that (dry-)port crew uses for order entry. Also this system exposes a web service that the HC can consume to get a list of orders to be taken into account

The Marketplace is considered the master in the management of information on contracts, as it represents the meeting point between supply and demand, where customers meet the suppliers of transport. Only in one case, the Marketplace consumes a service from the HC: this service provides an operation that allows the Marketplace to obtain detailed information on the activities associated with the contracts (Figure 8).

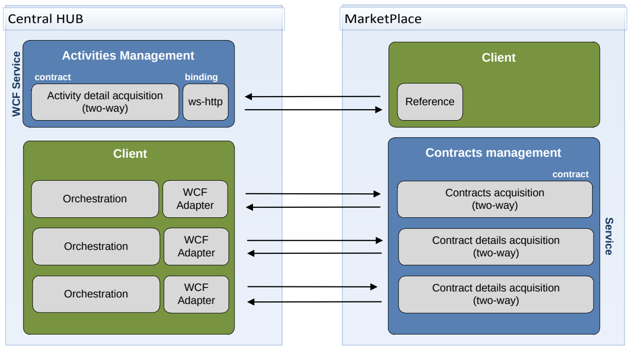


Fig. 8. Central Hub and Marketplace interactions.

Central HUB – Monitoring console

The monitoring console is available to the operator to supervise the movement of vehicles. The console interacts with the central hub in a unidirectional manner using the services offered.

The operations available to the operator are (Figure 9):

Group status request: it allows the operator to ask a group for sending an updated status message. This feature is designed to cover those situations in which a vehicle does not send information about its position and the status of the on-board instrumentation for an interval of time greater than expected. The status update request allows distinguishing situations in which the periodical sending is disabled for some reason arising from situations of potential danger or lack of gsm/gprs signal. The status message obtained has as a sending reason "Request from the center." The call always returns a response, positive or negative, that indicates the assumption by the communication gateway of the communication to the OBC. It may happen that the OBC does not send an updated message, due to lack of gsm/gprs signal or malfunction of the equipment

OBC events acquisition: it allows the user to request the status of the events associated with a group. In response to the request you always get a response that contains the latest information available to the gateway at the moment of the call. The information may not be updated if the on-board computer does not transmit messages for a long time

Group Id acquisition: it obtains the identifier of a group from the identifier of the OBC

Activities re-scheduling: it allows the operator to send a request for a schedule of activities. First, new contracts created within the marketplace portal and new orders entered by the (dry-)port management system are acquired. After updating the list of activities it sends this list to the scheduling module (operational level) that returns a proposal for the execution plan of activities, setting to a higher level of priority those who were already scheduled and confirmed

Contract details acquisition: it allows the operator to obtain detailed information about the customer and the supplier of the shipping service. The information is obtained from the Marketplace that has a master role in the management of this information

Contract cancelling: the operator has the right to cancel a contract if the task cannot be completed on time and as planned. This situation can occur subsequently to the conclusion of the contract. It may be due to network traffic urban or delays in the performance of practical port/customs

4. CONCLUSIONS

This paper presents the architecture of the SAIL project information system and the logical/physical scenarios it is tailored for.

Currently, the system manages the information related to the management of the runtime operations, and their flows. Using an orchestrator allows us to add other information flows just building the adapter to the new information formats and adding the needed maps and orchestrations: the system is open to include any (un)foreseen new set of information.

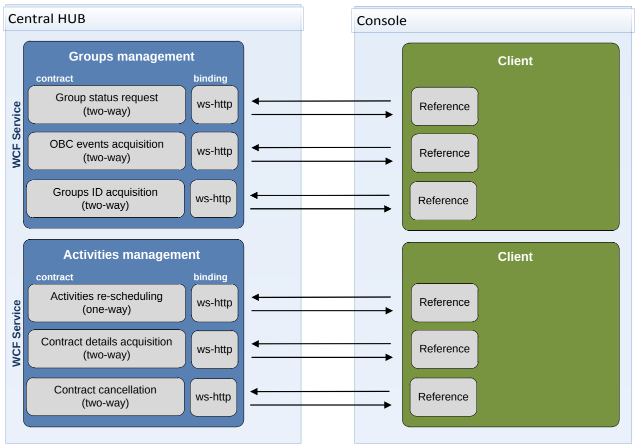


Fig. 9. Central Hub and Console interactions

In an updated version, we will include the modules of tactical and strategic level decisions. Moreover, we are going to include modules for statistical analysis of the gathering data through the connected modules. For the time being, it is not foreseen any connection to other information systems, but in the future the integrations with the (dry-)port information system will be implemented.

ACKNOWLEDGMENT

This work is supported by the E.U FP7-PEOPLE-IAPP-2009, Grant Agreement No 251589, Acronym: SAIL.

REFERENCES

Beckner, M. Goeltz B. and, Gross, B. (2006) BizTalk 2006 Recipes: A Problem-Solution Approach, Apress

Beckner M. (2010), BizTalk 2010 Recipes: A Problem-Solution Approach, Apress

Cibraro, P., Claeys, K., Cozzolino, F., & Grabner, J. (2010). Professional WCF 4: Windows Communication Foundation with. NET 4. Wrox

Dawson, J., Wainwright, J., & Sanders, J. (2009). Pro Mapping in BizTalk Server 2009. Apress

Forum: http://www.biztalkgurus.com/forums/

Klein, S. (2007). Professional WCF programming. John Wiley & Sons.

Lowy, J. (2008). Programming WCF services. O'Reilly Media, Incorporated.

Moukhnitski, S.,Campos, H., Kaufman, S., Kelcey, P., Peterson, D., and, Dunphy G. (2009). Pro BizTalk 2009. Apress, 2009.

A SIMULATION APPROACH FOR SPARE PARTS DEMAND FORECASTING AND INVENTORY MANAGEMENT OPTIMIZATION

Armenzoni Mattia[a], Montanari Roberto[b], Vignali Giuseppe[c], Bottani Eleonora[d], Ferretti Gino[e],Solari Federico[f], Rinaldi Marta[g]

[a]Interdepartmental Centre SITEIA.PARMA,
c/o Department of Industrial Engineering, University of Parma
Parco Area Delle Scienze , 181 - 43124 Parma

[b], [c], [d], [e], [f]Department of Industrial Engineering, University of Parma
Parco Area Delle Scienze , 181 - 43124 Parma

[g]Interdepartmental Centre CIPACK,
c/o Department of Industrial Engineering, University of Parma
Parco Area Delle Scienze , 181 - 43124 Parma

[a]mattia.armenzoni@unipr.it

ABSTRACT

Today, the management of the spare parts is a very important problem for the producers companies of industrial mechanical plant.

In this article, we develop and test a simulation approach, to forecast the demand of spare parts components during the lifetime of a complex product, such as an industrial plant.

The model requires a preliminary phase, where the relevant data of the plant should be collected by the manufacturing company. Then, a specific computational process should be carried out.

In the paper, the model is applied to a case example, referring to a hypothetical manufacturer of industrial plants, to demonstrate the model efficacy and its resolution capacity. Indeed, the application shows that the model proposed is able to identify the optimal level of service the company should set to manage the inventory of spare parts, i.e. the service level that minimizes the cost of spare parts inventory management.

The study finally proposes the future developments of the model and its application on the actual industrial reality.

Keywords: spare parts, simulation, demand forecast, stock management

1. INTRODUCTION

Spare parts (also called service parts, repair parts, or replacement parts) are interchangeable parts used for the repair or replacement of failed parts. Spare parts are an important feature of logistics management and supply chain management, and often require dedicated inventory management policies. Indeed, spare parts inventories need to be available at appropriate points within the supply chain, to provide after-sales services and to guarantee the desired service level (Botter and Fortuin, 2000).

Spare parts inventory management differs from managing the traditional manufacturing inventories, in several ways (Kennedy, et al., 2002). First, the functions of spare parts inventory is different from the traditional manufacturing inventory. For spare parts, service requirements are usually higher, since the effects of stock-outs may be particularly critical and financially remarkable (Huiskonen, 2001). However, the prices of individual parts may be very high, so that the cost of inventory can be relevant. Finally, the demand for parts is extremely sporadic and difficult to forecast, and, under some circumstances, could depend upon the maintenance strategy adopted.

At the same time, however, spare parts are essential in a supply chain. Indeed, they are required anytime there are material buffers in production systems, or maintenance operations are carried out, or the material flow is particularly high. Moreover, spare parts are essential to companies manufacturing complex systems, such as plant manufacturers or companies operating in the aviation industry (Fritzsche & Lasch, 2012), or to maintenance organization companies (Driessen et al., 2010).

On the basis of the scenario described above, it is not surprising that spare parts inventory management is a widely debated issue in literature. Researchers, in particular, are seeking for the optimal management of spare parts inventory, i.e. to the definition of an inventory policy leading to improving the quality of spare parts, at the same time reducing inventories and related costs. To this purpose, the problems addressed by spare parts literature focus primarily on those topics (Bacchetti and Saccani, 2012): (1) modeling the demand of spare parts; (2) classifying spare parts and determining criticality; (3)

developing stock-control techniques, integrating spare parts classification and demand forecasting.

Concerning the first topic, there is an extensive body of literature addressing spare-part inventory systems with random demand under different assumptions on the demand probability distributions. Hausman and Scudder (1982), Pyke (1990), Verrijdt et al.(1998), Perlman et al.(2001), Sleptchenko et al. (2002), (2005), Perlman and Kaspi (2007), and Adan et al.(2009) study stochastic multi-echelon inventory systems with several repair modes. Less studies have been carried out concerning spare-part inventories with deterministic demand (e.g., Prager,1956; Gass, 2003; Abdul-Jalbar et al., 2006; Federgruen et al.,2007; Perlman and Levner, 2010).

On the basis of the demand modeling, the corresponding inventory techniques are generally formulated as a cost minimization problem, with a cost function comprising the holding cost, ordering/setup cost, and either an explicit penalty cost or a specified service level constraint. For instance, Al-Rifai and Rossetti (2007) present an analytic inventory model for a two-echelon non-repairable spare parts system, that consists of one warehouse and several identical retailers. The model is grounded on the reorder point policy, and the inventory control problem is formulated to minimize the total annual inventory investment of the system, subject to a defined annual ordering frequency and an expected number of backorder constraints. Simao and Powell (2009) develop a model to determine the optimal inventory levels at each warehouse, for an aircraft manufacturer. The model is solved using approximate dynamic programming, with a proper design, so as to consider the presence of low-frequency observations.

The basic inventory models (such as EOQ or EOI) have been widely applied to the inventory management of spare parts (Liu and Esogbue, 1999). Conversely, there is relatively little evidence of the use of more sophisticated techniques or integrated models (Bacchetti and Saccani, 2012), that start with the inventory classification and ends with the performance assessment of the inventory control policy adopted. Moreover, as regards the inventory management policies, approaches for spare parts management are rarely optimized in terms of overall inventory costs they generate, including not only the holding stock cost, but also stock-out cost and order cost. Therefore, there is no link between the selected policy and the corresponding cost (or performance) of the spare parts management, nor to the evaluation of the practical usefulness of that policy in practical cases.

Starting from those research gaps, in this paper we develop an integrated approach, that, starting from the identification of critical components, provides, as final output, the identification of the optimal inventory management policy for each spare part, i.e. the policy that minimizes the total cost of the inventory management system. Since spare parts are requested on an irregular basis, simulation is exploited to derive an estimate of the components demand, replacing traditional analytic models. To show the practical use of the approach developed, the model is then applied to the case of a food plant manufacturing company.

The remainder of the paper is organized as follows. Section 2 reviews the literature related to spare parts inventory management, with a particular attention to the studies that include an analysis of the cost of the inventory management. In section 3, we present the model developed in this paper. Section 4 provides an application of the model to a numerical example. Finally, section 5 concludes and indicates future research steps.

2. MODELLING

As mentioned above, few studies in literature propose complex models for spare parts management. The aim of this article, therefore, is to develop an analytic method for spare parts management. The starting point of the model is a set of data provided by a real company, manufacturing food industrial plants; as output, the model provides an estimate of the spare parts the company will have to supply during the plant lifetime.

The initial hypotheses of the model developed are the following:

- The industrial plant, and, in general, any kind of complex equipment, can be considered as a group of mechanical components;
- For each mechanical component, some technical serviceability data (e.g., Mean Time to Failure or Mean Time to Repair) are known;
- For each plant, the lifetime data (i.e., the initial time of working and the date of its disposal) are known.
- The unitary inventory management cost (in particular, the cost of holding stock and the stock out cost) is known.

The mathematical model proposed in this paper is developed exploiting a general purpose software, i.e. Microsoft Excel, appropriately programmed with as Visual Basic for Application. This tool has been chosen due to its capacity to perform complex computations in the inventory management context, as demonstrated by Bottani et al., (2012).

To set up the model, the company should provide the amount of industrial plants installed worldwide; such information is used to forecast the demand of internal spare parts.

Typically, the market phases of complex products, such as industrial plants, consists of five phases [Figure 1]:

- Initial phase: the product is new, and therefore unknown to the external market. The company invests much on the marketing of the new technology;
- Development phase: during this phase, the product is well-known and the number of the requested and installed plants increases in time;
- Stabilization phase: in this phase, the number of plants sold during the year remains almost constant in time;

- Decreasing phase: during this phase, competitors can reach and exceed the technology level of the machinery manufactured by the company. Therefore, the number of plants installed decreases gradually;
- Final phase: during this step, the product will be removed from the company's portfolio and therefore that product will no longer be sold to the market..

Up to the decreasing phase, the demand of spare parts the company should fulfill will depend on the plants already installed, as well as on the new installed plants. Conversely, during the final phase, the demand of spare parts will depend only on the plants the company has previously installed. The end-of-life of the product occurs when the last plant stops working.

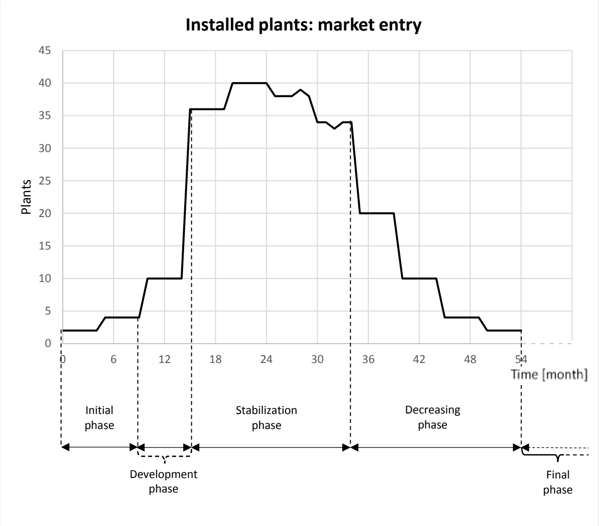


Figure 1 - Different product market phases.

For each plant, it is reasonable that the company knows the number of critical spare parts and its maintenance data. Specifically, for each plant, the model requests these data to be implemented:

- Start operating date: this date is expressed through an absolute temporal system. Conventionally, the first installed plants is sold and begins to work at time $t_0 = 0$;
- End of working time: this date indicates the dismantling of the plant. Since the model is grounded on a simulated approach, the operative life of the machinery could also be hypothesized and varied by means of the simulation. In the numerical case, a particular statistical distribution is used to represent the working life of the plant. Obviously, the model has been developed as a flexible tool, where different statistical distribution can be used to represent the service life of the plant in the specific industrial case.

These two data replicated for each plant analyzed describe the temporal limit of the simulation.

In order to forecast the spare parts demand, the model requires also the maintenance and replace data of those components that can be considered as critical spare parts. In this regard, according to Molenaers et al. (2011), defining the criticality of spare parts is a complex process. In particular, this activity involves several areas of a company, such as logistics, production, business analysis, marketing and more (Braglia et al., 2004).

The relevant criteria for the analysis of critical components depend on the particular industrial application of the method. In general, those criteria can include, among others:

- Logistic characteristics: these factors can be the geometric and dimensional properties. Generally speaking, these factors have a lot of influence on the logistics activities;
- Maintenance and replaceable characteristics: as mentioned above, the higher is the probability of failure, the more critical component is. This point should take into account not only the probability of rupture of the part, but also their corruption, since it can affect the time required for the replenishment of the spare part.

In summary, the proposed model needs several data about the critical component, for which the demand should be estimated. Specifically, to correctly initialize the model, the user must insert the following data about the critical component:

- Mean Time to Failure (MTTF): this is the time elapsing between two subsequent product failures. In the numerical example, this parameter is set to a constant value. Nonetheless, once again, the simulation model has been designed to be flexible, allowing to adopt a different statistical distribution depending on the nature of the component to be examined. Similarly, the MTTF value can change during the lifetime of the component;
- Replenishment Time (RT): the model estimates this data by means of a lognormal distribution. To set up the distribution, the user can insert the average and standard deviation values of the RT in the model. As per the MTTF, a future development of the method can to consider new or other distribution to describe the RT.

Through the data listed above for the critical components, the model is able to compute the rate of the component failure in each plant.

To this extent, the time elapsing from the plant starting date and the first failure is computed according to the classical theory of industrial maintenance, expressed in eq.1:

$$t_I = -\frac{1}{\lambda * \ln(rnd)} + S_t \qquad 1.$$

where:
λ = 1/MTTF
rnd = random number [0;1]
S_t = Start operating time

To estimate the time of the following failures, the models considers the replenishment time as follows:

$$RT \sim logN(\mu, \sigma^2) \qquad 2.$$

$$t_n = -\frac{1}{\lambda * \ln(rnd)} + S_t + RT + t_{n-1} \qquad 3.$$

where:
t_n = time of occurrence the rupture n
S_t = Start Operating Time
RT = Replenishment Time
t_{n-1} = time of occurrence the rupture n-1

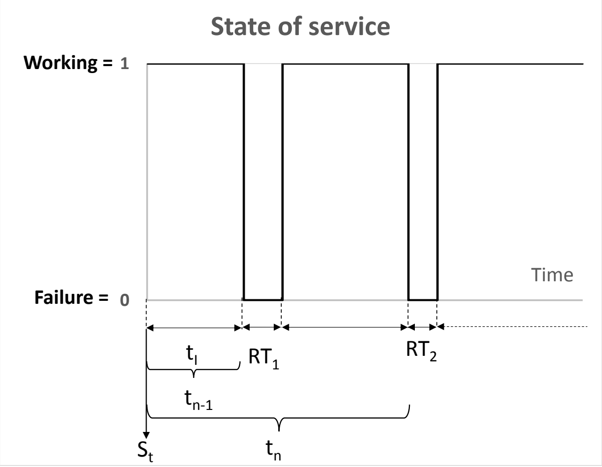


Figure 2 - State of service of a mechanical component.

From this data, the request for components during the lifetime of the plant can be easily computed. This computation should be replicated for each plant sold and installed by the company, to derive the global distribution in time of component failures during the lifetime of the plant.

As a result from the computation of the failure times, the model reproduces the demand of spare parts the company should fulfill, on a selected time unit (e.g., week or month). As an example, the demand of components provided by the model can describe a function as that shown in the figure below:

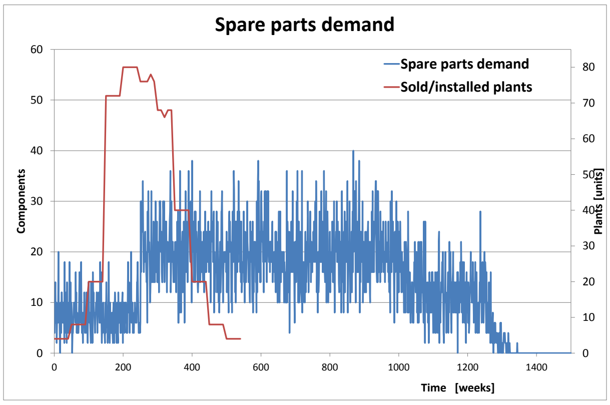


Figure 3 - Spare parts distribution (example).

The model replicates the computation described above, in order to increase the relevance of the statistical analysis related to the lognormal statistical distribution characteristics of the RT and the MTTF. Therefore, for each time step the model does not provide a specific value, but a statistical distribution of the component demand:

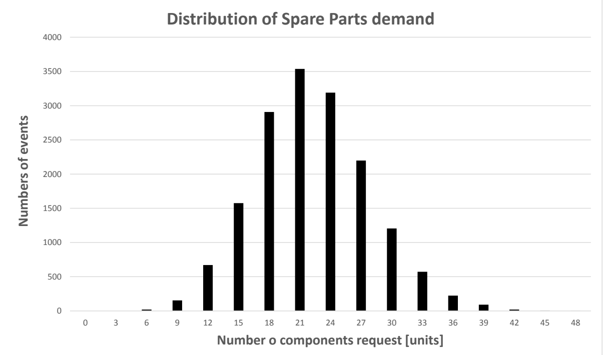


Figure 4 - Distribution of spare parts demand on the model replications for a generic time step.

With these hypotheses, the model tries to identity the optimal level of service through an economic analysis. Specifically, the simulation takes into account the following cost:

- stock-out cost: the company will pay for bad maintenance service. In our model, this costs arises when the required component is not available in the stock;
- Inventory cost: this is the cost that the company pays to maintain the stock of the spare parts.

The simulation system modifies the service level provided by the company, by varying the *k* coefficient, defined by the following equation:

$$k = \frac{number\ of\ service\ events\ soddisfied}{total\ number\ of\ events} \qquad 4.$$

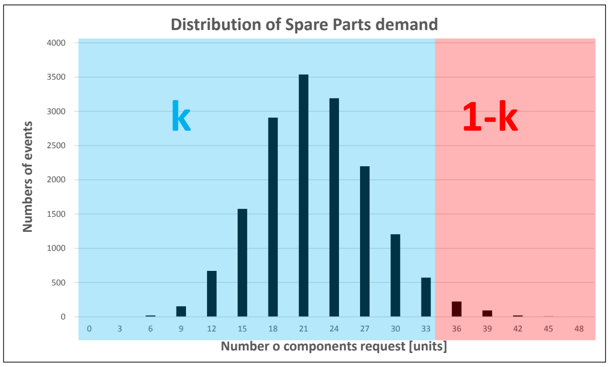


Figure 5 - Service level (k).

The value of *k* varies from 85% to 99.9% with a step of 0.05%; for each value, the model takes into account the distribution of the spare parts demand.

With an integer simulation, the demand value arises with steps of one unit, until the service level reaches the desired value. For each step of each replication, the model computes:

- The expected value of spare parts, as a function of the service level;
- The value of the demand of spare parts.

From these outcomes, for each time step the model will calculate:

- The value of the stock, i.e. the number of components available in the stock. It is assumed that the initial value (at time *t*=1) of the stock accounts for one unit;
- The amount of spare parts to be ordered: by ordering this quantity, the value of the spare parts stock is step by step equal with the expected value.

During the simulation, the model updates the value of the stock and spare parts ordered, as follows:

$$OSP_n = ST_n - EV_{n+1} \qquad 5.$$

$$ST_n = ST_{n-1} - DV_n + OSP_n \qquad 6.$$

where:
OSP_n = spare parts to be ordered at step n;
ST_n = number of component in stock at step n;
EV_n = expected value at step n;
ST_{n-1} = number of component in stock at step n-1;
DV_n = demand value at step n.

The spare parts ordered generates the inventory cost of the time step, as follows:

$$SIC_n = IC * OSP \qquad 7.$$

with:
SIC_n = inventory cost of the temporal step n [€/step];
IC = inventory cost of a unit component [€/unit];
OSP = spare parts to be ordered.

At the same, when the value of the stock is lower than the demand value (ST_n <0), a stock out cost occurs:

$$SCST_n = CST * (-ST_n) \qquad 8.$$

with:
$SCST_n$ = stock out cost at step n [€/step];
CST = stock out cost [€/unit];
ST_n = number of component in stock at step n.

For each service level the total costs of the single replication follow the equations:

$$TIC = \sum_{n=1}^{n} SIC_n \qquad 9.$$

$$TCST = \sum_{n=1}^{n} SCST_n \qquad 10.$$

$$\overline{TIC} = \frac{\sum_{n=1}^{n} SIC_n}{n} \qquad 11.$$

$$\overline{TCST} = \frac{\sum_{n=1}^{n} SCST_n}{n} \qquad 12.$$

TIC = total inventory cost of the replication [€];
TCST = total stock out cost of the replication [€];
$\overline{\text{TIC}}$ = average inventory cost [€/step];
$\overline{\text{TCST}}$ = average stock out cost [€/step].

Finally, the model is able to evaluate the service level cost as the average costs of the replications and the global cost as the sum of $\overline{\text{TIC}}$ and $\overline{\text{TCST}}$. Those cost components could also be combined and compared, to explain the relationship between the cost and the service level. Although the specific cost results may vary depending on the value of IC and CST set into the model, the graph of this relationship is expected to have the following general structure:

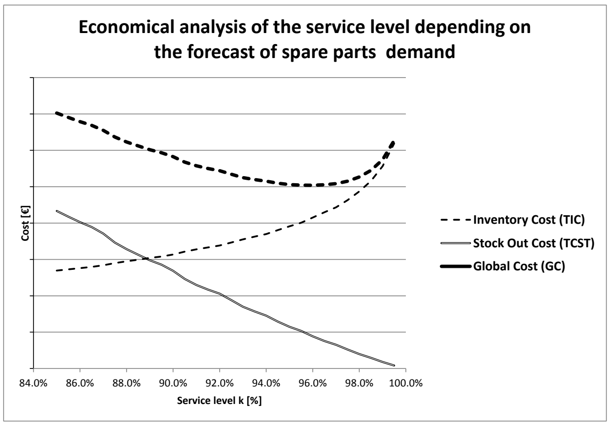


Figure 6 - Final economic analysis (example).

From the economic analysis, the model can calculate the optimal value of k (k_{opt}), i.e. the value that minimize the total costs of inventory.

3. A NUMERICAL EXAMPLE

We then test the model presented in a particular context, reproducing a plant manufacturing company; one of their plants is taken as the example product.

At the beginning, the plants enter in the market with the distribution shown in figure 7.

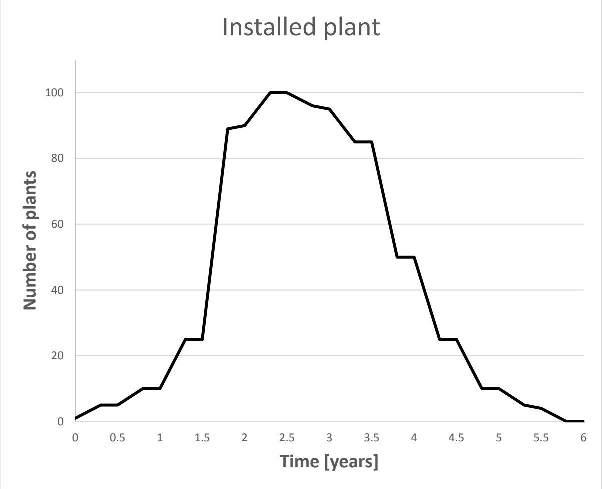


Figure 7 - Trend of installed plants.

According to the graph, the product experiences a rapid increase in its sells between the first and the second year of commercialization; the stabilization phase lasts two years, then the sells decrease. After the sixth years of the commercialization, this product comes out from the catalogue of the company. The graph above describes 1000 plants manufactured by the company and entering the market during the commercialization period.

In this example, the service life (SL) of each plant is modeled as a normal distribution:

- $SL \sim N(\mu, \sigma^2)$
  - μ = 20000 hours;
  - σ = 4000 hours

In this example, we assumed as critical those spare parts that own the following maintenance and serviceability data:

- MTTF = 2000 hours (λ = 0.0005);
- $RT \sim logN(\mu, \sigma^2)$ with:
  - μ = 10 hours;
  - σ = 2 hours.

With these data, the model simulates 16373 replications of the spare part demand, computed using the statistical distribution described above for maintenance and serviceability.

The graph in Figure 8 shows the average of the spare parts demand generated by the characteristic of the plants installed:

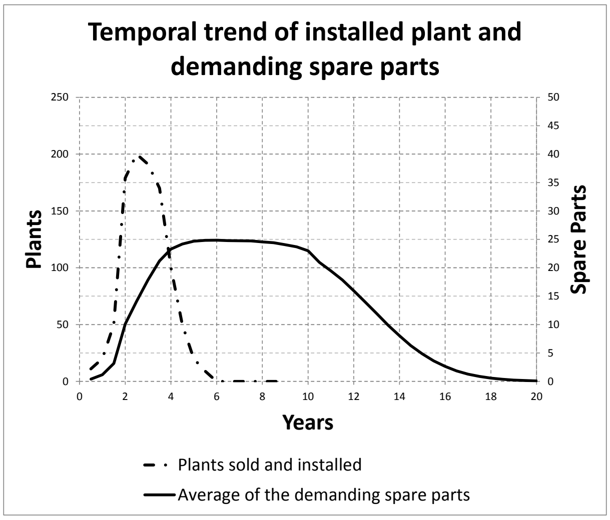


Figure 8 - Trend of the installed plants and relative demand of spare parts.

Then, the model estimates the stock-out distribution, depending on the service level k, whose value varies in the following range:

- minimum service level: 85%;
- maximum service level: 99.9%;
- step of analysis= 0.05%;

In order to identify the optimal value of the k, the costs of stock out and inventory are computed, starting from the following unitary costs:

- Inventory cost (IC) = 75 €/component;
- Cost of stock out (CST) = 3000 €/component.

In our example, the order cost is hypothetically negligible compared to the other cost components, and is, therefore, neglected in the analysis.

With these data, the model can performs an economic analysis, reporting the average weekly costs of stock-out and inventory :

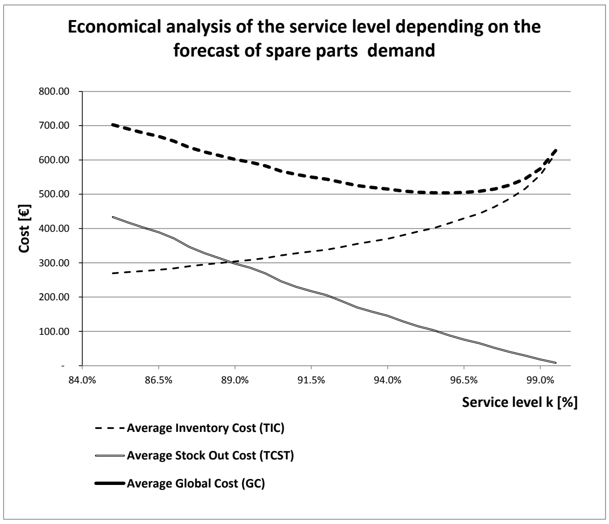


Figure 9 - Economic analysis.

Finally, the model identifies the minimum total cost of the system, in order to find the optimal value of service level. In this example, we found

- $k_{opt} = 96\%$
- $AGC_{min} = 503.79$ *€/week*

4. CONCLUSIONS

This articles has proposed a direct application of a simulative approach to a logistic problem, related to the spare parts inventory management.

In particular, the methodological approach presented in this study allows to define the optimal level of the spare parts stock during the lifetime of a product (i.e., a plant or machinery) containing the component.

The study explains the main steps of the computational process, starting from the distribution in time of the plants manufactured by the company.

The model described represents a preventive and flexible tool for the prediction of spare parts demand. Indeed, a real company can insert the characteristics parameter of its products, in order to adapt the simulative model with its industrial case.

With a numerical example, the articles applies the model on a real case example, based on the theoretical description of the maintenance and the service ability of the mechanical system.

The method involves several activities of the company, basically concerning data regarding the plant, the component and the durability of the product on the global market.

Looking at the results obtained, the natural future develop of this analysis is its application to a real industrial case, with the purpose of testing its applicability in real scenarios. Comparing the proposed approach with other tools in use in a real company could be interesting to test the reliability of the model, as well as to identify potentials for improving its performance and affordability.

ACKNOWLEDGMENTS

This work was partially supported by the Emilia - Romagna Region, with the contribution of the Interdepartmental Centre SITEIA.PARMA, Parma, Italy.

REFERENCES

Bottani, E., Cecconi, M., Vignali, G., Montanari, R., 2012. Optimisation of storage allocation in order picking operations through a genetic algorithm. *International Journal of Logistics: Research and Applications*, 15 (2), 127-146

Braglia, M., Grassi, A., Montanari, R., 2004, Multi-attribute classification method for spare parts inventory management, *Journal of Quality in Maintenance Engineering*, 10 (1), 55-65.

Botter R, Fortuin L., 2000. Stocking strategy for service parts: a case study. *International Journal of Operation and Production Management*, 20, 656-74

Kennedy, W.J., Patterson, J.W., Fredendall, L.D., 2002. An overview of recent literature on spare parts inventories. *International Journal of Production Economics*, 76 (2), 201-215

Huiskonen, J., 2001. Maintenance spare parts logistics: Special characteristics and strategic choices. *International Journal of Production Economics*, 71 (1–3), 125–133

Fritzsche, R., Lasch, R., 2012. An Integrated Logistics Model of Spare Parts Maintenance Planning within the Aviation Industry. *World Academy of Science, Engineering and Technology*, 68. Available at http://www.waset.org/journals/waset/v68/v68-1.pdf (accessed June 2013)Botter, R., Fortuin, L., 2000. Stocking strategy for service parts – a case study. *International Journal of Operations & Production Management*, 20 (6), 656-674

Driessen, M.A., Arts, J.J., van Houtum, G.J., Rustenburg, W.D., Huisman, B., 2010. Maintenance spare parts planning and control: A framework for control and agenda for future research. Working Paper 325 of the Eindhoven University. Available at http://cms.ieis.tue.nl/Beta/Files/WorkingPapers/wp_325.pdf (accessed May 2013)

Bacchetti, A., Saccani, N., 2012. Spare parts classification and demand forecasting for stock control: investigating the gap between research and practice. *Omega*, 40, 722–737

Hausman, W.H. , Scudder, G.D., 1982. Priority Scheduling Rules for Repairable Inventory Systems. *Management Science*, 28 (11), 1215-1232

Pyke, D.F., 1990. Priority repair and dispatch policies for repairable item logistics systems. *Naval Research Logistics*, 37, 1–30.

Verrijdt, J. Adan, I., de Kok, T., 1998. A trade off between repair and inventory investment. *IIE Transactions*, 30, 119–132.

Perlman, Y., Mehrez, A., Kaspi, M., 2001. Setting expediting repair policy in a multi-echelon repairable-item inventory system with limited repair capacity. *Journal of the Operational Research Society*, 52,198–209.

Sleptchenko, A., van der Heijden, M.C., van Harten, A., 2002. Effect of finite repair capacity in multi-echelon multi-indenture service part supply systems. *International Journal of Production Economics*, 79, 209–230.

Sleptchenko, A., van der Heijden, M.C., van Harten, A., 2005. Using repair priorities to reduce stock investment in spare part networks. *European Journal of Operations Research*, 163, 733–750.

Perlman, Y., Kaspi, M., 2007. Centralized decision of internal transfer-prices with congestion externalities for two modes of repair with limited repair capacity. *Journal of the Operational Research Society*, 58, 1178–1184.

Adan, I.J.B.F., Sleptchenko, A., van Houtum, G.J.J.A.N., 2009. Reducing costs of spare parts supply systems via static priorities. *Asia-Pacific Journal of Operational Research*, 26,559–585.

Prager, W., 1956. On the Caterer problem. *Management Science*, 3,16–23.

Gass, S.I., 2003. *Linear Programming: Methods and Applications*. Dover Publications, NewYork.

Federgruen, A., Meissner, J., Tzur,M., 2007. Progressive interval heuristics for multi-item capacitated lot sizing problems. *Operations Research*, 55,490–502.

Perlman, Y., Levner, I., 2010. Modeling multi-echelon multi-supplier repairable inventory systems with backorders. *International Journal of Service Science and Management*, 3, 440–448.

Abdul-Jalbar, B., Gutierrez, J.M., Sicilia, J., 2006. Single cycle policies for the one-warehouse N-retailer inventory/distribution system. *Omega*, 34, 196–208.

Al-Rifai, MH, Rossetti, MD, 2007. An efficient heuristic optimization algorithm for a two echelon (R,Q) inventory system. *International Journal of Production Economics*, 109, 195–213.

Liu, B., Esogbue, A.O., 1999. *Decision Criteria and Optimal Inventory Processes*. Norwell, USA: Kluwer Academic Publishers. ISBN:0792384687.

Simao, H., Powell, W., 2009. Approximate dynamic programming for management of high-value spare parts. *Journal of Manufacturing Technology Management*, 20 (2), 147 - 160

Molenaers, A., Baets, H., Pintelon, L., Waeyenbergh, G., 2011. Criticality classification of spare parts: A case study. *International Journal Production Economics*, 140 (2), 570-578

AUTHORS BIOGRAPHY

Prof. Gino FERRETTI graduated in Mechanical Engineering on February 1974 at the University of Bologna. At the Faculty of Engineering of the same University, he served as assistant professor for the courses of “Machines” and “Mechanical Plants”. In the next years he worked as associate professor at the University of Padua and as full professor at the University of Trento. In 1988, he moved to the Faculty of Engineering, University of Parma, where at present he is full professor of Mechanical Plants. His research activities focus on industrial plants, material handling systems, and food processing plants, and have been published in numerous journal and conference papers.

Prof. Roberto MONTANARI graduated (with distinction) in 1999 in Mechanical Engineering at the University of Parma (Italy), where he is currently employed as Full Professor of “Industrial Plants”. His research activities concern equipment maintenance, power plants, food plants, logistics, supply chain management, supply chain modeling and simulation, inventory management. He has published his research papers in several qualified international journals, as well as in national and international conferences. He acts as a referee for numerous scientific journals and is editorial board member of two international scientific journals**.**

Dr. Eleonora BOTTANI graduated (with distinction) in 2002 in Industrial Engineering and Management at the University of Parma (Italy), where she is lecturer (with tenure) in Mechanical Industrial Plants since 2005. In March 2006, she got her Ph.D. in Industrial Engineering at the same University. She is the author or coauthor of approx. 100 papers in the fields of logistics, supply chain management, supply chain modeling, performance analysis, industrial plants safety and optimization. She acts as a referee for about 60 international scientific journals and for several international conferences. She is editorial board member of three international scientific journals.

Dr. Giuseppe VIGNALI graduated in 2004 in Mechanical Engineering at the University of Parma. In 2009, he received his Ph.D. in Industrial Engineering at the same university, related to the analysis and optimization of food processes. Since August 2007, he works as a Lecturer at the Department of Industrial Engineering of the University of Parma, and, since the employment at the university, he has been teaching materials, technologies and equipment for food packaging to the food industry engineering class. His research activities

concern food processing and packaging issues and quality/safety of industrial plant. Results of his studies related to the above topics have been published in more than 50 scientific papers, some of which appear both in national and international journals, as well in national and international conferences.

Dr. Federico SOLARi is Ph.D. Student in Industrial Engineering at the University of Parma. He got a master degree in Mechanical Engineering of the Food Industry at the same University, discussing a theses related to the design of a plant for the extraction of volatile compounds. He attended several conferences related to food process and plants, and published several papers on the same topics in international conferences.

Dr. Mattia ARMENZONI is scholarship holder in Industrial Engineering at the Interdepartmental Center Siteia.Parma of the University of Parma. He got a master degree in Mechanical Engineering of the Food Industry at the same University, discussing a thesis related to the design of a static dryer for pasta with simulation tools. He attended numerous conferences on the topics of food processing, food plants, modeling and simulation, and published several papers on the same topics in international conferences.

Dr. Marta RINALDI is Ph.D. Student in Industrial Engineering at the University of Parma and she is scholarship holder in Industrial Engineering at the Interdepartmental Center Cipack of the University of Parma. She got a master degree in Management Engineering. She attended numerous conferences on the topics of food processing, logistics process, energy saving, modeling and simulation, and published several papers on the same topics in international conferences.

MODELING CYBER WARFARE IN HETEROGENEOUS NETWORKS FOR PROTECTION OF INFRASTRUCTURES AND OPERATIONS

Agostino G. Bruzzone, Diego Merani
NATO STO CMRE
Email: {bruzzone, merani}@cmre.nato.int
URL: www.cmre.nato.int

Marina Massei, Alberto Tremori
DIME University of Genoa, Italy
Email: {massei, tremor}@itim.unige.it
URL: www.itim.unige.it

Christian Bartolucci, Angelo Ferrando
Simulation Team
Email: {Christian.bartolucci, angelo.ferrando}@simulationteam.com
URL: www.simulationteam.com

ABSTRACT

This paper presents a modeling approach for mapping cyber defense issues with respect to heterogeneous networks; the research is devoted to develop an agent driven simulation environment able to analyze this problem considering different layers including CIS capabilities, operational issues, system architecture, management processes and human factors. The paper analyzes a specific case study to validate and verify the proposed modeling approach; the scenario is focused on an heterogeneous network applied to extended maritime environment including Autonomous Underwater Vehicles (AUV), sensors, platforms, vessels, satellites and relevant military assets and threats. The present document uses this case study as example of System of Systems to be simulated including cyber warfare issues to evaluate their impact on operations.

keywords: Cyber Defense, Interoperable Simulation, Maritime Simulation, Heterogeneous Networks, Autonomous Systems, Modeling & Simulation

1. INTRODUCTION

The research aims at matching the NATO Topology of Heterogeneous Networks with Cyber Defense warfare in order to model the different elements and possible risks (i.e. installation procedures, access methods, training level, networking reliability, data certification, encryption procedures, password management, operator procedure etc.).

Heterogeneous Networks are becoming popular and intensively present in several application areas, since they represent an opportunity and have a big potential, while at the same time introduce new open issues and problems: indeed these systems, whose capability is affected by multiple layers, involve complex phenomena such as data abundance that overpasses the elaboration capabilities, hiding techniques, non-collaborative targets behaviors, environmental conditions, assets reliability, models maturity, agility, node compromised resources, etc. Such increasing popularity is expected to become very common in military field as a consequence of the technology evolution trends with special attention to autonomous systems, robots and sensor networks (Bruzzone et al. 2005; Tether 2009). A great challenge for the near future is related to the possibility to link together lots of light mobile devices in order to have a complete persistent understanding of the battlefield and so get advantages in terms of military results, properly addressing cyber defense issues.

Due to the high level of interactions among the networks and their complexity, this application field requires to be investigated using M&S (Modeling and Simulation); indeed the presence of several stochastic factors affecting the behaviors of the different actors in the scenario needs to be modeled through

Intelligent Agents and properly mapping policies, doctrines as well as new potential threat behaviors.

2. INTEROPERABLE SIMULATION FOR CYBER WARFARE IN HETEROGENOUS NETWORKS

From this point of view, simulation is very important in reproducing heterogeneous networks considering the complexity of the different systems and interactions; in fact, following this approach, it becomes possible to model different devices linked together into a scenario, and to simulate the different layers covering technological, operational and social aspects.

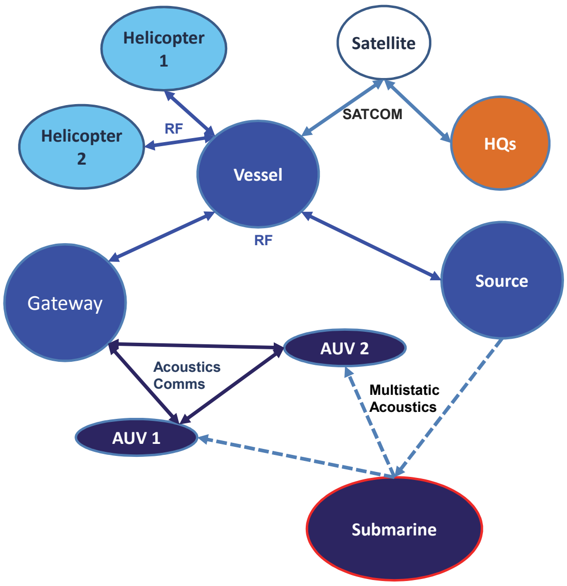


Figure 1: Example of heterogeneous network for an ASW operation involving 2 AUVs

Vice-versa, it could be very difficult to test the effectiveness of this System of Systems in the real context and to identify architectures and actions able to improve it with a convenient cost/benefit rate considering the complexity of the framework, the high concentration of parameters and the number of entities involved. In addition simulation allows users to insert into the scenario new concepts and technologies and to analyze the network performance with respect to introduced threats and considering the interactions among systems and components; by this approach it becomes possible to evaluate in a virtual simulation environment the system capabilities and to finalize new requirements and/or procedures (Hua Guo et al. 2010).

In this context, one of the main goals of this research is to couple topologies and characteristics of heterogeneous networks and cyber defense warfare within conceptual models ready to be federated and simulated; this research aims at defining models reproducing objects, attributes and their behaviors and interactions to reproduce heterogeneous networks immersed in operational scenarios and operated by human social networks.

The authors are currently working on researches related to Marine heterogeneous Network involving autonomous vehicles, satellites, vessels, aircrafts, sensors and emitters as proposed in figure 1 (Bruzzone et al 2013; Wiedemann 2013); these heterogeneous systems could be devoted to conduct different complex missions such as Intelligence, Surveillance and Reconnaissance (ISR). Indeed, these heterogeneous networks result as an aggregation of different assets (i.e. underwater systems, surface drones, ships, helicopters and even satellites) being available for connection on based on their operative status and boundary conditions; this context is a very sensitive environment to cyber warfare issue; in this area the conceptual models are representing the characteristics of nodes, connections, infrastructures of the heterogeneous network; the different models could be implemented in different simulations able to be federated together over the operational scenario by using HLA (Bruzzone et al. 1998; Kuhl et al. 1999; Massei et al. 2013); in addition to these models, even the procedures related to the social layer could be simulated within stochastic discrete event models and synchronized in this federation (Massei and Tremori 2010). These simulated layers federated together represent the environment available to conduct tests and experimentations for high fidelity simulation as well as for preliminary investigations. The authors are currently focusing their attention on the maritime extended framework including multiple domain such as sea surface, underwater, air, space, cyberspace, land and coast (Bruzzone 2013); in this context security issues as well as cyber warfare are critical elements (Longo et al. 2005; Bruzzone, Massei, Tremori, Longo, Madeo, Tarone 2011)

3. CONTEXT OVERVIEW

Cyber security is a major issue in the industrial and business sectors, especially in relationship with emerging contexts such as power grid (Yang et al. 2011) and SCADA systems (Urias et al., 2012). A methodology for analyzing the compromise of a deployed tactical network has been proposed by Asman, B.C. et al. in 2011. Homeland security applications were approached in 2007 by Kotenko, I. in its work on Multi-agent Modelling and Simulation of Cyber-Attacks and Cyber-Defense; a recent work on how to mitigate a cyber-physical attack that disables the transportation network and releases a cloud of chlorine gas has been published by the U.S. Dept. of Homeland Security, whose security analysts developed simulation models and tools to analyze the consequences of complex events on critical U.S. infrastructure and resources (Nabil Adam et al., Communications of the ACM, no.56/6, June 2013); the analysis of ICT infrastructures respect cyber security issues could be addressed even by risk analysis and Monte Carlo Simulation using HPC (High Performance Computing) to solve the computational workload (Baiardi et al, 2011); therefore in most of the cases it could be necessary to include the stochastic components with functional and operational models; for instance this paper addresses the point of combining actions over the real with the cyber battlefield in a coordinated way, indeed this point is supposed to have a major impact on future war operations (Jakobson 2004).

These problems strongly affect the nature of heterogeneous networks that is characterized by dynamic and complex nature (Rumekasten 1994). The use of intelligent agents has been demonstrated very effective for reproducing reactive entities and complex systems; the concept of agent driven simulation where the IAs (Intelligent Agents) are directing objects active within the simulation were experimented in a wide spectrum of applications (Oren and Ylmaz 2009). Indeed the capability to reproduce by agent emergent behaviors in complex system is major benefit of this approach (Thompson and Bossomaier 2006).

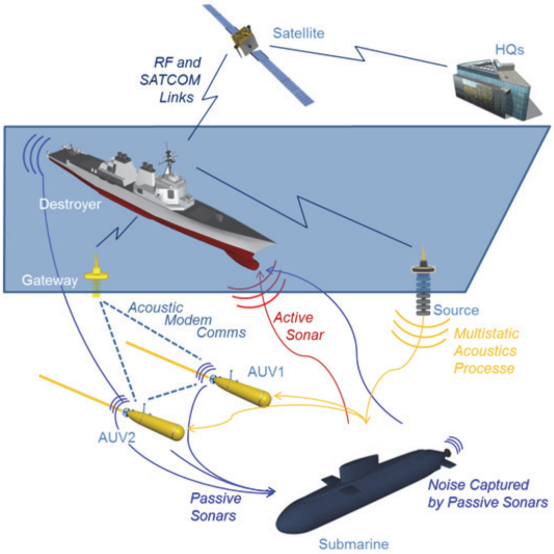


Figure 2: Sensor Network Case Study

The IA allows modeling platforms, humans as well as their interactions (Cornforth et al. 2004; Calfee and Rowe 2004).

The authors have long experience in using intelligent agents for reproducing intelligent reactive behaviors within complex mission environments (Bruzzone 2008; Bruzzone 2010); obviously the intelligent agents could implement different AI (Artificial Intelligence) techniques and methodologies (Bruzzone, Massei, Tarone, Madeo 2011; Affenzeller et al. 2009). Indeed the proposed scenario where AUV operates requires also coordination among Autonomous system that could be directed by the humans just when communication are working over the network; the problem of coordination among UAS (unmanned autonomous systems) has been investigated in order to identify proper approaches and effective control solution (Feddema et al. 2002; Vail and Veloso 2003; Kalra et al. 2010); indeed this problem represents a very good case of heterogeneous network where it is possible to apply different methodologies (Tanner 2007); the problem was even addressed with specific reference to marine mission environments (Merkuriev Y. et al. 1998; Sujit 2009; Martins et al. 2010; Nad et al. 2011; Zini 2012).

4. MODELS & SCENARIO

This research, through its topological approach over heterogeneous networks, is devoted to create a Federation of models based on HLA simulation interoperability standards (i.e. High Level Architecture); such federation should be able to couple the different models and layers related to such kind of networks and to simulate their interactions with respect to operational scenarios. In particular it focuses on cyber defense topological and procedural aspects regarding complex heterogeneous networks; one crucial part it is represented by introducing autonomous and intelligent behavior over the simulated entities, in this case the simulator was adopting Intelligent Agents Computer Generated Forces (IA-CGF) developed by Simulation Team to reproduce threats, behaviors as well as entities reactions (Bruzzone, Tremori, Massei 2011). The specific scenario used as test-case for this research is a surveillance system composed of unattended autonomous underwater sensors (AUV) whose mission is to detect an enemy target through sensing and collaboration via acoustic underwater communication. This system is able to perform different missions (e.g. area clearance or hold-at-risk in specific choke points).

The sensors communicate with a surface node (sink) which acts as a gateway between underwater and above water communications systems. The surface node has the capability to connect with a local naval unit via Line of Sight RF link. The naval unit then provides data fusion and system integration with existing global C2 tactical networks via satellite communication capability (SATCOM) as proposed in figure 2.

The capability of Heterogeneous Networks are combined by the effect of different layers: a CIS layer including user applications, information system equipment, communications equipment, and user applications; but also an operational layer, including business processes (Information Assurance processes, Service Management and Governance processes etc.), information products and other operational capabilities directly connected or derived from the operation or mission type.

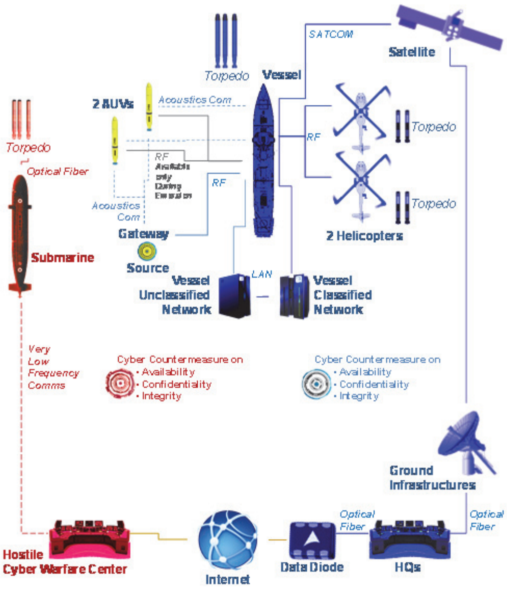


Figure 3: MCWS Objects

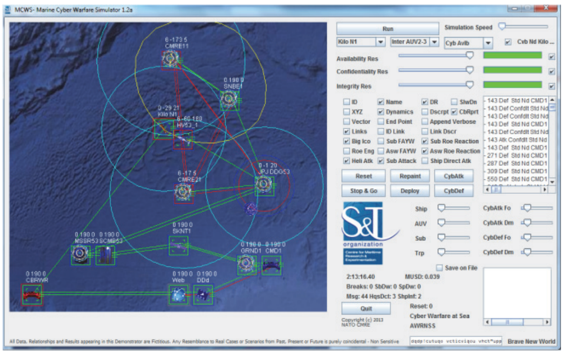
The model considers the distinction between the CIS infrastructure (Computer Information Systems representing the Cyber Infrastructure) and the operational context including the mission and operation type, and all user-related processes that form the operational capability.

Some human behaviors, in fact, could easily compromise a network or its performance due to accidental and/or intentional actions; as regards intentional ones, in some cases terrorists exploit procedural lacks in order to drive their attacks to success. It is interesting to note that even completely automated persistent solutions could quickly degrade their performance due to actions on the human layer. For this reason the Information Assurance components that influence the Operational Context (i.e. information products and the user-related processes) are included in the simulation in order to estimate their impact with respect to the other layers.

In addition to these elements, in defense scenarios key performance indexes and measures of merits are evaluated on the operational layer that is reproduced within the simulation environment: indeed the performance of the heterogeneous network should be tested with respect to the dynamic evolution of the operations since this affects the different alternative Course of Actions (COA).

A first step in the problem analysis is represented by the definition of node security properties through scoring of security properties such as Confidentiality, Integrity, Availability, Non-Repudiation and Authentication. Then criteria for modeling nodes performance in normal conditions, as well as their interaction with other nodes, are defined. A third step is the simulation of the impact that a total or partial break-down of a node, consequence of a successful exploit of a vulnerability, would have on other network nodes, and on the overall performance of the system: such effects are evaluated in percentage points, but even in terms of operational impact (measured in the simulated environment). The choice to restrict the simulation to the vulnerability/threat pairs that determine the most relevant effects in terms of security capability over each node, along the simulation evolution, as well as to the resolution and the most significant details required to reproduce the different layers, is related to the possibility to finalize a relevant demonstration within the research timeframe; such finalization is made possible by the use of simplified meta-models able to approximate the behavior of the objects. Nevertheless the simulation architecture proposed guarantees the possibility to keep the proposed modeling approach open for further developments and even to replace some models in the future by more detailed ones (Zacharewicz et al. 2008).

In general the research includes among different layers the network models as well as cyber architecture objects. The simulated scenario is composed by sensors, entities that collect information from them, units on the field and a Command and Control Network. Not only networks nodes, but also links between them are often physical objects with specific properties (e.g. the LAN, Local Area Network, that connects two computers); the preliminary model was finalized in MCWS (Marine Cyber Warfare Simulator) developed in cooperation among the authors; this simulator does not include any sensitive information even if realistic data were used for the setup; in figure 3 it is proposed the set of the objects representing the assets and connections simulated in the current scenario within MCWS, while its implementation in Java is proposed in figure 4; the HQs in fact is connected to the web adopting proper solutions for guaranteeing the protection of its own infrastructure (Bruzzone et al. 1999)



☐ ID ☑ Name ☑ DR ☐ SlwDn
☐ XYZ ☑ Dynamics ☐ Dscrpt ☑ CbRprt
☐ Vector ☐ End Point ☐ Append Verbose
☑ Links ☐ ID Link ☐ Link Dscr
☑ Big Ico ☐ Sub FAYW ☑ Sub Roe Reaction
☐ Roe Eng ☐ Asw FAYW ☑ Asw Roe Reaction
☑ Heli Atk ☑ Sub Attack ☐ Ship Direct Atk

- 143 Def Std Nd CMD1
- 143 Def Confdlt Std Nd
- 143 Def Confdlt Std Nd
- 143 Def Std Nd CMD1
- 143 Def Confdlt Std Nd
- 143 Atk Confdlt Std Nd
- 143 Def Std Nd CMD1
- 271 Def Std Nd CMD1
- 287 Def Std Nd CMD1
- 309 Def Std Nd CMD1
- 550 Def Std Nd CMD1

☐ Save on File

2:13:16.40 MUSD: 0.039
Breaks: 0 SbDw: 0 SpDw: 0
Msg: 44 HqsDct: 3 ShpInf: 2
Reset: 0
Cyber Warfare at Sea
AWRNSS

Copyright (c) 2013
NATO CMRE

All Data, Relationships and Results appearing in this Demonstrator are Fictitious. Any Resemblance to Real Cases or Scenarios from Past, Present or Future is purely coincidental - Non Sensitive

Figure 4: MCWS Simulator

The characterization of the security properties of the nodes, and the links between them, has been conducted through a process of identification of key properties and behaviors; for example, we assume that the Integrity factor of one network node affects downstream flows of information; or that the confidentiality of a node can be only compromised in its wholeness.

Nevertheless, the simulator can be fed with as many granular security properties or behavior as required, to augment its accuracy. The simulator is able to discriminate cases in which there is a monitoring activity over the nodes with respect to the cases in which such activity is not carried out, and to consider remediation actions in case a cyber-attack is detected.

After defining cyberspace objects with their variables, their attributes and their mutual interactions, input actions such as degrees of freedom as well as threats effects to be are identified: this is the scheme basing on which the model applies stochastic factors and probabilistic rules that reproduce how the exploitation of a node vulnerability influences the others. In a similar way, the operational layer, including user processes, is modeled and simulated even considering stochastic factors affecting the procedure evolution.

Behaviors and Rules of Engagements (ROEs) were implemented by configuring IA-CGF for the specific roles including: cyber warfare actions and operational decisions.

5. MCWS ANALYSIS AND EXPERIMENTATION

The scenario used to validate and verify MCWS was inspired by the case above described with some characteristics.

In the proposed scenario the Blue Force has the role to protect the area from submarines arriving from East; while the submarine (OPFOR) goal is move from West side up to the East borders over a square (20 by 20 nautical miles) in deep waters; however the operations are not limited to the square and could be extended even over it if necessary both by the submarine and the Blue Force. The environmental simulation includes sea current, wind, sea waves, weather conditions, temperature and visibility over day and night. Very simplified public release models have been adopted for sonar detection including:

- Passive Sonar the model are affected by target and sonar platform noise affected by dynamic behaviors as well as by the specific characteristics of the sensor
- Active Sonar Mono-static and Multi-static the model are affected by acoustic target strength and characteristics of the boundary conditions of the assets and sensors dynamically evolving during the simulation.

Blue Force resources include 2 AUVs, 1 Destroyer with 2 helicopters, 1 Buoy able to act as Emitter and gateway, a Satellite Network, a Ground Infrastructure and an HQs with web connection through a data diode; the helicopters and the ship are equipped with ASW (Anti-Submarine Warfare) torpedo and Vessel LAN is simulated as divided between classified Network connecting with HQs and unclassified network connecting AUV.

In this scenario, the Blue Force is not entitled to carry out offensive cyber-attacks, but could adopt preventive and reactive measures to protect their cyber space both in term of nodes and connections.

OPFOR have just a submarine armed with torpedo and a cyber-warfare center transmitting sensitive information through very low communication; in order to simplify the scenario the submarine cyber warfare center communication is consider very reliable over a wide spectrum of operational modes; obviously the reliability and availability of this connection could be consider subjected to all boundary and conditional factors as other connection in the future for more realistic researches.

In this scenario the focus is on the three among the node security properties:

- Availability: this element affects the reliability and throughput of network connections and nodes; if availability is completely disrupted the corresponding resources are not available at all; in this case rerouting is possible, if alternative paths exist.
- Confidentiality: this element measures the capability of Opposing Force (OPFOR) to access data and information present or passing through a network node or link; if this property is compromised and a message including position of an asset (i.e. an AUV or the Destroyer) pass through this cyber resource, the information is transmitted to the submarine who changes its behavior in order to respect the ROE (i.e. avoid contact).
- Integrity: this element measures the accuracy of the content of data (information); if integrity is compromised, the messages going through the compromised entity are disrupted or modified with unuseful or fake information, and cannot be processed; this affects obviously the command chain, therefore the message could be delivered over different paths where they are available, to solve the contingency until the integrity is re-established.

The scenario adopted very simple rules of engagement: ROEs for Blue Force include detecting and discouraging, not use of lethal force, engaging under approval by HQs, reacting to fire, free engagement; while the ROEs for the submarine include hiding, avoiding contact, reacting to fire, engaging at his will; the ROE are directed by the IA-CGF (Bruzzone, Tremori, Massei 2011); the authors conducted V&V (Verification and Validation) on the consistency of ROE application respect different conditions by adopting a testing plan (Bruzzone and Massei 2007; Bruzzone et al. 2002).

The user is entitled to activate the different types of cyber-attacks as well as to change resilience and effectiveness of defensive and offensive actions in cyberspace; in similar way the capabilities of the sensor and assets could be changed by the authors.

The simulation is currently a stochastic hybrid agent driven simulation; stochastic factors include simplified model for communications over the network, failures, success rate and duration of cyber action, detection probabilities and hitting probability, damages, etc.

The communications over the heterogeneous networks are modeled taking into account aspects of reliability and latency affecting both nodes and links, independently from the cyber actions, in order to reproduce the characteristics of the channel (i.e. high latency and disruptions of acoustic underwater comms), but also malfunctions and degenerative operational modes of the ICT (Bruzzone; De Felice et al. 2010); for instance these issues were investigated with non-traditional protocols for underwater communications in heterogeneous networks of AUVs (Merani et al. 2011). It was critical to identify measure of merits in order to compare experimental analysis obtained by MCWS; in this scenario the target functions used to measure the performance have been defined based on desired end state during each single simulation run and are classified in the following 4 classes:

- Sub Success: the submarine successfully passes through the area and reaches East side.
- BF Success: the submarine is detected and tracked successfully and the Blue Force assets reach the condition to be ready to proceed with engagement, blue force stops to act and the submarine resign.
- Sub Down: the submarine is engaged and disabled by Blue Force
- Ship Down: the destroyer is engaged and disabled by the submarine

The scenario was played over the following different hypotheses:

- Limited Scenario: Operations stops when Blue Force is ready to engage the submarine
- Full Operational Scenario: Operations proceed under NATO Art.5 environment
- No Cyber warfare: Cyber Warfare actions are disabled
- Regular Cyber warfare: Cyber Warfare actions are enabled and intensity is set on regular values
- Intense Cyber warfare: Cyber Warfare actions are enabled and intensity is set on high values

It is evident how Cyber Warfare settings are subjected to author hypotheses as well as other parameters; so the experimental results are characterized as relative values (one respect the other ones) much more than as absolute evaluations.

Considering the stochastic nature of the simulator it was necessary to apply ANOVA (Analysis of Variance) in order to estimate the experimental error and confidence bands in the different conditions.

$$\overline{BFS}_{Rate}(k) = \frac{\sum_{i=1}^{k} BE_{ES}(i)}{k} \quad (1)$$

$$MSpE_{BFS_{Rate}}(k) = \frac{\sum_{i=1}^{k}[BE_{ES}(i) - \overline{BFS}_{Rate}(k)]^2}{k-1} \quad (2)$$

$$CF_{BFS_{Rate}}(k) = \tfrac{1}{2} t_{\alpha,k-1}\sqrt{MSpE_{BFS_{Rate}}(k)} \qquad (3)$$

$$BE_{ES}(i) = \begin{cases} 0 & \textit{if end state of } i-th \textit{ run is Ship Down} \\ 0 & \textit{if end state of } i-th \textit{ run is Sub Success} \\ 1 & \textit{if end state of } i-th \textit{ run is Blue Force Success} \\ 1 & \textit{if end state of } i-th \textit{ run is sub down} \end{cases}$$

$\overline{\text{BFS}}_{\text{Rate}}$ (k)	Blue Force Success Rate (BFS Rate) after k replications
BE_{ES}(i)	Blue Force in End State of the i-th run
n	total simulation replications changing pseudo random seed
k	k-th replication among simulation runs
$\text{MSpE}_{\text{BFSRate}}$(k)	Mean Square pure Error of BFS Rate after k replications
$\text{CB}_{\text{BFSRate}}$(k)	Semi amplitude of the Confidence Ban of BFS Rate after k replications
$t_{a,\nu}$	t-Student Distribution with α confidence level and ν degree of freedom

MCWS general architecture is designed in order to federate different models into an interoperable simulation environment; therefore in this case it was implemented within as basic demonstrator and it was used to conduct standalone fast time experiments by using simplified meta-models for sensors and communications.

Such experiments are carried out after defining a specific relevant scenario in order to restrict the range of investigation and test the research most important concepts versus interesting target functions. Indeed, thanks to the experimentation activity, it is possible to evaluate system performance and sensitivity on measure of merits referred to procedures, policies, architectures and technological alternative solution; in the following figures it is proposed only the analysis of the Mean Square pure Error over the different scenario hypotheses (see figure 5, 6, 7 and 8) and a basic comparison of the overall results (figure 9).

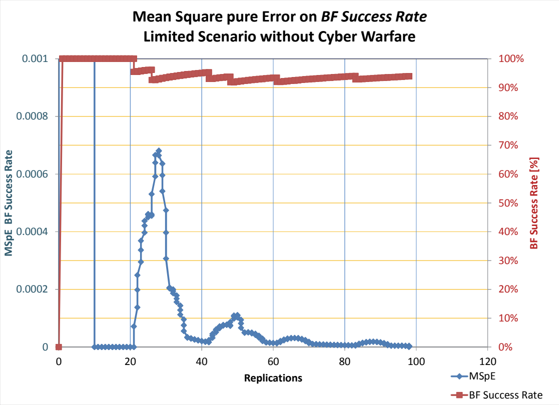


Figure 5: MSpE and Mean BF Success in Limited Scenario without Cyber Warfare

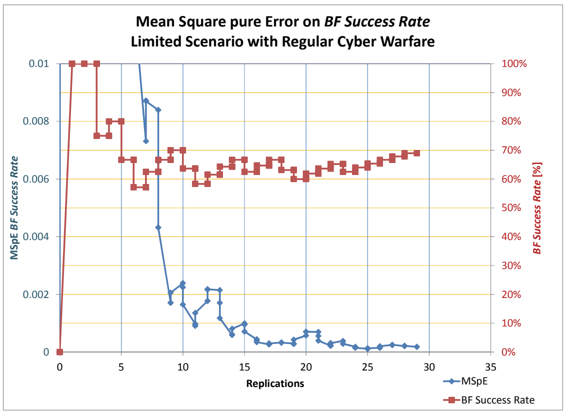


Figure 6: MSpE and Mean BF Success in Limited Scenario with Regular Cyber Warfare

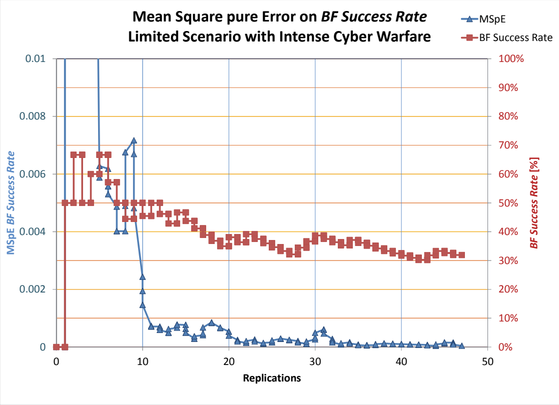


Figure 7: MSpE and Mean BF Success in Limited Scenario with Intense Cyber Warfare

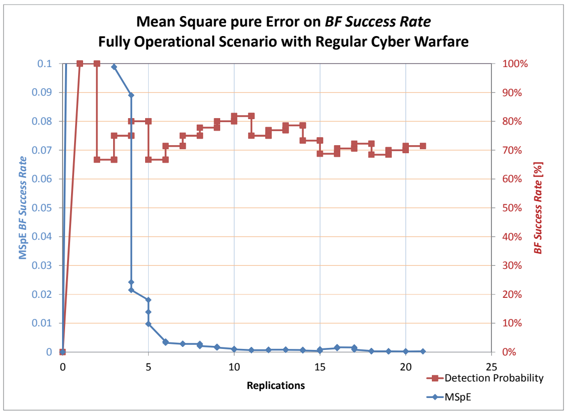


Figure 8: MSpE and Mean BF Success in Fully Operational Scenario with Regular Cyber Warfare

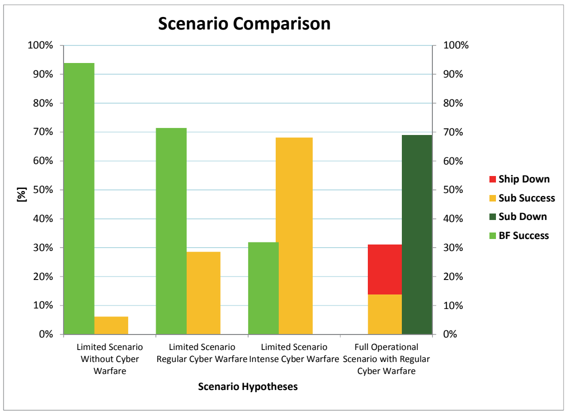


Figure 9: Result comparisons over different hypotheses respect the different end states

From the analysis it emerges the optimal number of replications for each combination of scenario hypotheses; the MSpE and consequently the confidence band results pretty good even with a limited number of replications;

In figure 9 it is proposed the comparison among the different results changing the scenario hypothesis; the analysis confirm the impact of cyber warfare on the Blue Force Success Rate; the *Fully Operational Scenario* produces just a smoothly change respect *Limited Scenario* as expected, considering the additional, even if limited, probability for the submarine to succeed in a confrontation against the Destroyer after successfully being detected and tracked.

6 CONCLUSIONS

The general architecture and conceptual models proposed in the paper were successfully implemented in MCWS simulator focused on a specific basic scenario, inspired to a collaborative ASW mission "hold at risk/secure friendly maneuver area", conducted via autonomous underwater vehicles; therefore even the case study proposed represent a relevant mission environment respect existing research and available models; the results obtained are very interesting and the potential of this approach by the interoperability with other models is very great providing scalable solution to complex scenarios; indeed the described approach is open to be extended and applied to more sophisticated context.

The use of MCWS allows conducting experimental analysis; by this approach, it is possible to use sensitivity analysis in order to evaluate the most influent parameters, the second and high order effects, and to quantify the degree of uncertainty as well as the experimental error (Montgomery 2000); the simulation allows to test criteria to identify emergent behaviors and to estimate risk to violate or to compromise cyber resources; preventive action efficiency, mitigation procedures and reactions are tested and evaluated in terms of their impact on the operational scenario through simulation experiments; indeed the quantitative experimentation proposed in this paper confirms the benefits of the proposed approach and the importance of adopting simulation as investigation aid for cyber warfare within operational frameworks.

REFERENCES

Affenzeller M., S. M. Winkler, S. Wagner, A. Beham (2009) "Genetic Algorithms and Genetic Programming" CRC Press (Taylor & Francis Group)

Baiardi F., Telmon C., Sgandurra G. (2012) "Haruspex—Simulation-driven Risk Analysis for Complex Systems", JournalISACA Volume 3

Bruzzone A.G. (2013) "New Challenges for Modelling & Simulation in Maritime Domain", Keynote Speech at SpringSim2013, San Diego, CA, April

Bruzzone A.G., Berni A., Fontaine J.G., Brizzolara S., Longo F., Poggi S., Dallorto M., Dato L., (2013) "Simulating the Marine Domain as an Extended Framework for Joint Collaboration and Competition among Autonomous Systems ", Proceedings of I3M2013, Athens, September

Bruzzone, A.G., Tremori, A., Massei, M., (2011) "Adding Smart to the Mix," Modeling, Simulation & Training: the International Defence Training Journal, 3, 25-27.

Bruzzone, A.G., Massei, M., Tremori, A., Longo, F., Madeo, F., Tarone, F. (2011) "Maritime security: emerging technologies for asymmetric threats.” In Proceedings of the European Modeling and Simulation Symposium, EMSS2011 (Rome, Italy, September 12-14).

Bruzzone A.G., Massei M., Tarone F., Madeo F. (2011) "Integrating Intelligent Agents & AHP in a Complex System Simulation", Proceedings of the international Symposium on the AHP, Sorrento, Italy, June.

Bruzzone A.G. (2010) "Project Piovra on Intelligent Agents and CGF", Technical Report for A-03-IT-1682 Italy USA M&S Data Exchange Agreement, November 4-5

Bruzzone A.G. (2008) "Intelligent Agents for Computer Generated Forces", Invited Speech at Gesi User Workshop, Wien, Italy, October 16-17

Bruzzone A.G., Massei M. (2007) "Polyfunctional Intelligent Operational Virtual Reality Agent: PIOVRA Final Report", EDA Technical Report

Bruzzone A.G., Frydman C., Junco S., Dauphin-Panguy G. (2005) "International Mediterranean Modelling Multiconfernece - International Conference on Integrated Modelling and Analysis in Applied Control and Automation", LSIS Press, ISBN 2-9520712-5-X (pp 194)

Bruzzone, A.G. et al. (2002) "Simulation -based VV&A methodology for HLA federations: an example from the Aerospace Industry”, Proceedings of 35th Annual Simulation Symposium, vol., no. , pp.80,85, April

Bruzzone A.G., E.Page, A.Uhrmacher (1999) "Web-based Modelling & Simulation", SCS International, San Francisco, ISBN 1-56555-156-7

Bruzzone A.G., Giribone P. (1998) "Decision-Support Systems and Simulation for Logistics: Moving Forward for a Distributed, Real-Time, Interactive Simulation Environment", Proceedings of the 31st Annual Simulation Symposium, Boston MA, April

Calfee, S.H., Rowe, N.C., 2004. “Multi-agent simulation of human behavior in naval air defense,” Naval Engineers Journal, 116, no.4, 53-64.

Cornforth D., Kirley M., Bossomaier T. (2004) "Agent Heterogeneity and Coalition Formation: Investigating Market-Based Cooperative Problem Solving", AAMAS: 556-563

De Felice F., G. Di Bona, D. Falcone, A. Silvestri (2010) “New Reliability allocation methodology: the Integrated Factors Method”, International Journal of Operations & Quantitative Management Volume 16 Number 1, ISSN: 1082-1910

Feddema, J.T.; Lewis, C.; Schoenwald, D.A., "Decentralized control of cooperative robotic vehicles: theory and application, "Robotics and Automation, IEEE Transactions on, vol.18, no.5, pp.852,864, Oct 2002

Hua Guo, Fei Tao, Lin Zhang, Suyi Su, Nan Si (2010) "Correlation-aware web services composition and QoS computation model in virtual enterprise", International Journal of Advanced Manufacturing Technology, 51, 5-8 , 817-827.

Jakobson, G., Lewis, L., Buford, J., Sherman, C.E. (2004) "Importance of Considering Future War as a Convergence of Real and Cyber Battlespaces - Battlespace situation analysis: the dynamic CBR approach", Military Communications Conference, MILCOM IEEE, Vol. 2 Page(s): 941 - 947

Kalra N., D. Ferguson and A. Stentz: A generalized framework for solving tightly-coupled multirobot planning problems, Proc. of the IEEE International Conference on Robotics and Automation, April 2007, pp.3359-3364. Kennedy, K.P., 2010. “Training: the key to keeping your head in a crisis situation,” Naval Engineers Journal, 122, no.3, 73-85.

Kotenko, I. (2007) “Multi-agent Modelling and Simulation of Cyber-Attacks and Cyber-Defense for Homeland Security” Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2007. 4th IEEE Workshop

Kuhl, F., Weatherly, R., Dahmann, J., 1999. Creating Computer Simulation Systems: An Introduction to the High Level Architecture. Prentice Hall, Upper Saddle River, USA.

Longo, F., Bruzzone, A.G. (2005) “Modelling and Simulation applied to Security Systems”. Proceedings of Summer Computer Simulation Conference, pp. 183-188

Martins, R.; de Sousa, J.B.; Afonso, C.C.; Incze, M.L., "REP10 AUV: Shallow water operations with heterogeneous autonomous vehicles," OCEANS, 2011 IEEE - Spain, vol., no., pp.1,6, 6-9 June 2011

Massei, M., Tremori, A., 2010. “Mobile training solutions based on ST_VP: an HLA virtual simulation for training and virtual prototyping within ports.” In Proceedings of the 2010 International Workshop on Applied Modeling and Simulation, WAMS2010 (Buzios, Brazil, May 5-7).

Merkuriev Y., Bruzzone A.G., Novitsky L (1998) "Modelling and Simulation within a Maritime

Environment", SCS Europe, Ghent, Belgium, ISBN 1-56555-132-X

Montgomery D.C. (2000) "Design and Analysis of Experiments", John Wiley & Sons, New York

Nad, D., Miskovic, N., Djapic, V., Vukic, Z. (2011) "Sonar aided navigation and control of small UUVs",Proceeedings of the 19th Mediterranean Conference on Control & Automation (MED), Corfu, Greece, June

Nabil Adam, Randy Stiles, Andrew Zimdars, Ryan Timmons, Jackie Leung, Greg Stachnick, Jeff Merrick, Robert Coop, Vadim Slavin, Tanya Kruglikov, John Galmiche, and Sharad Mehrotra. (2013) "Consequence analysis of complex events on critical U.S. infrastructure". Commun. ACM 56, 6 (June 2013), 83-91.

Massei M., Tremori A., Poggi S., Nicoletti L., (2013) "HLA based real time Distributed Simulation of a Marine Port for Training Purposes", International Journal of Simulation and Process Modeling, Vol.8, No.1, pp.42-51

Merani, D., Berni, A., Potter, J., Martins, R. (2011) "An Underwater Convergence Layer for Disruption Tolerant Networking", Proceeedings of Baltic Congress on Future Internet Communications, Riga Feb. 16-18

Ören, T.I. and L. Yilmaz (2009) "On the Synergy of Simulation and Agents: An Innovation Paradigm Perspective", Special Issue on Agent-Directed Simulation. The International Journal of Intelligent Control and Systems (IJICS), Vol. 14, Nb. 1, March, 4-19.

Rumekasten, M. , (1994) "Simulation of Heterogeneous Networks", Proceedings Winter Simulation Conference, pp. 1264 - 1271

Sujit, P. B.; Sousa, J.; Pereira, F.L., "UAV and AUVs coordination for ocean exploration,"OCEANS 2009 - EUROPE, vol., no., pp.1,7, 11-14 May 2009

Tanner H.G., D.K. Christodoulakis, Decentralized cooperative control of heterogeneous vehicle groups, Robotics and Autonomous Systems 55 (2007) 811–823

Tether, T. (2009) "Darpa Strategic Plan", Technical Report DARPA, May

Thompson J., Bossomaier T. (2006) Agent Based Modelling of Coevolution of Trust between Client and Wealth Managers", CIMCA/IAWTIC, 131

Vail D. and M. Veloso: Dynamic multi-robot coordination, In Multi-Robot Systems: From Swarms to Intelligent Automata, Vol II, 2003, pp. 87-100.

Wiedemann J. (2013) "Naval Forces", Special Issue 2013, Vol.XXXIV ISSN 0722-8880

Zacharewicz G., Frydman C., Giambiasi N. (2008) "G-DEVS/HLA Environment for Distributed Simulations of Workflows", Simulation, Vol.84, N.5, pp 197-213

Zini A. (2012) "Virtual Ship: Dream or Reality", Keynote Speech at Summersim 2012, Genoa, July

"Methodology for analyzing the compromise of a deployed tactical network", Asman, B.C. ; Kim, M.H. ; Moschitto, R.A. ; Stauffer, J.C. ; Huddleston, S.H., Systems and Information Engineering Design Symposium (SIEDS), 2011 IEEE Digital Object Identifier: 10.1109/SIEDS.2011.5876871 Publication Year: 2011 , Page(s): 164 - 169

Urias, V. ; Van Leeuwen, B. ; Richardson, B. (2012) "Supervisory Command and Data Acquisition (SCADA) system cyber security analysis using a live, virtual, and constructive (LVC) testbed" MILITARY COMMUNICATIONS CONFERENCE- MILCOM Digital Object Identifier: 10.1109/MILCOM.2012.6415818 Publication Year: 2012 , Page(s): 1 - 8

WEB REFERENCES

- http://www.cmre.nato.int
- http://www.itim.unige.it
- http://www.liophant.org/projects/ia_cgf_ucoin.html
- http://www.mastsrl.eu/solutions/ia_cgf_t4.html
- http://www.simulationteam.com

COOPERATIVE TRAINING FOR SHIPS AND TUGBOATS PILOTS BASED ON INTEROPERABLE SIMULATION

Francesco Longo[(a)], Letizia Nicoletti[(b)], Alessandro Chiurco[(c)]

[(a)(b)(c)]DIMEG, University of Calabria, Rende (CS), Italy

f.longo@unical.it[(a)], letizia.nicoletti@unical.it[(b)], a.chiurco@unical.it[(c)]

ABSTRACT

This paper presents an advanced training system for ship pilots and tug pilots involved in the last mile of navigation. Such training system relies on the interoperable simulation based approach and it is made up of a federation of simulators that have been integrated according to the High Level Architecture standard (HLA 1516). In particular, the federation includes, among others, the simulator of a container carrier and a tugboat simulator that can interoperate each other and share the same 3D virtual environment. The physical behaviour of the boats (both the containership and tugboat) and their interaction patterns have been recreated by using 6 Degree of Freedom (DOF) mathematical models encoded within the simulators. In addition, the 3D geometric models and the operational scenario (including the port area) have been carefully developed.

Keywords: Ships Pilots Training, Ship Tugboat interactions, Modeling & Simulation, marine ports

1. INTRODUCTION

On May 7th, 2013 the Italian container ship Jolly Nero slammed into a dock as it was exiting from the Port of Genoa striking and toppling the port control tower. The incident caused ten deaths, several injured and huge economic damages. So far, investigations into the causes are still underway but it is clear that:

- an intervention of two tugboats assisting the ship during the manoeuvre would have prevented the accident from happening;
- the lack of dredging of the west exit has required big ships to execute unusual manoeuvres to leave Genoa port.

The Jolly Nero facts, as well as other incidents occurred over the years, show that exit/entry manoeuvres within the port area are very complex and must be regarded as critical operations; in fact owing to the big dead-weight of commercial ships a single mistake, even at low speed, can cause unexpected and tragic consequences not only for the operators involved but also for those operators working in the same area. To this end, a good interaction and communication between the ship and the tugboat is important to prevent accidents; in addition, it is also important to define standard operating procedures and policies for entry/exit manoeuvres aimed at ensuring high security and safety levels (Bruzzone et al., 2012-a). Within this framework, considering also that the most of the ports impose to incoming and outgoing ships to be supported by a certain number of tugboats (according to the ship weight/dimensions), the cooperative training of both ship and tugboat pilots plays a critical role for security and safety levels enhancement as well as for a better productivity (indeed, according to the shipping company goals, ships must leave the port as soon as possible, Longo 2007; Bruzzone et al. 2012-b).

Obviously, training activities on the real system with real equipment would be too expensive and too dangerous, therefore Modeling & Simulation (M&S) based approaches, that allow trainees to act in a synthetic environment learning how to put in practice theoretical concepts/procedures and see the immediate consequences of their actions in a visual manner, can offer substantial advantages (Bruzzone et al., 2010; Bruzzone et al. 2011; Merkuryev and Bikovska, 2012). Hence, this research work is devoted to describe an advanced interoperable simulation based training system for ship pilots and tug pilots involved in the last mile of navigation. Such training system is part of an ongoing research project “HABITAT, Harbor Traffic Optimization System” co-financed by the Italian Ministry of Education, University and Research as part of the “PON Ricerca e Competività ” Program.

The proposed training tool is conceived in order to provide its users with a realistic experience thanks to the possibility of experiencing a joint and cooperative training environment. To this end, interoperability has been one of the main driving requirements that has been achieved integrating the simulation resources according to the High Level Architecture integration standard (HLA 1516). In this way a federation of simulators has been developed. The federation includes three federates:

- the ship federate, that reproduce different types of ships according to the initial settings (containership, tanker, ro-ro ship; in this paper the case of the containership is presented);
- the tugboat federate; such federate runs on a different PC, even distributed over a

LAN/WAN network, and can be used to recreate the operations usually carried out by tugboats in real port environments;

- the control tower federate; this simulator recreates the activities that usually take place within the port control tower including port traffic monitoring and control and communications with incoming and outgoing ships (in this paper only the first two federates are presented; the implementation of the control tower federate is still on-going).

As a result, pilots can exercise their operational and manoeuvring skills, became acquainted with the behaviour of the ship/tugboat and with the effects of their interaction patterns. Moreover, the proposed tool allows the pilots to learn the procedures that are currently adopted in a specific port, and decision makers to design and test new procedures.

To ensure a satisfactory level of realism, the dynamic behaviour of each federate has been implemented and is based on a 6 Degree Of Freedom (DOF) mathematical model. In particular, surge, sway and yaw are modeled using the so called Manoeuvring Mathematical Modeling Group (MMG, 1985) model, that takes its name from the Japanese research group that has implemented it between 1976 and 1980.

The MMG model has been studied extensively in conjunction with the findings of posthumous research works such as Kijima and Nakiri (2003), Lee et al. (2003), Hasegawa et al. (2006), Perez et al. (2006) and Armaoğlu et al. (2009). As a matter of facts, the MMG group defined for the first time a prediction method of ship manoeuvrability while Kijima and Nakiri (2003) proposed the approximate formulas for calculating the hydrodynamic forces, taking into account the effects of the stern shape. Moreover, Rhee et al. (1999) and Lee et al. (2003) have proposed some empirical formulas aimed at finding out the hydrodynamic coefficients for the ship manoeuvrings equations.

Hasegawa et al. (2006) have discussed the course-keeping ability of a pure car carrier in windy condition.

Perez et al. (2006) and Armaoğlu et al. (2009) describe how some parameters and dimensions influence manoeuvrability characteristics.

As for roll, pitch and heave, ship motions have been estimated based on the models of Jensen, (2001), Jensen et al. (2004).

In the sequel, a brief description of the paper organization is given. Section 2 describes the reference scenario, the port of Livorno, reproduced in the 3D virtual environment; in addition, the ship and the tugboat taken as a reference for the respective simulators are described. Section 3 deals with the mathematical models behind the dynamic behaviour of the ship and of the tugboat both when they are kept separate and when they are mutually interacting. In section 4, a tool developed in the C++ programming language and devoted to evaluate and validate such mathematical models is presented. Lastly, in section 5 the integration of the federates into the proposed HLA federation is discussed. The last section summarizes the contribution of the work.

2. THE MARINE PORT SCENARIO

The scenario that has been reproduced within the proposed simulation framework is the port of Livorno (Figure 1), which is located on the Tyrrhenian Sea in the northwestern part of Tuscany. It is one of the largest seaports in Italy and in the whole Mediterranean Sea with an annual traffic capacity of around 30 million tonnes of cargo and 600,000 TEU's. It can handle every kind of goods: bulk, liquid, frozen foods, fruits, cars and most of all containers. Moreover, the port of Livorno is the only in Italy, and the second in Europe, with a liquefied natural gas (LNG) terminal where double-hull gas tankers from North Africa unload liquid gas in artificial caves located below the sea level. Port areas cover about 800,000 square meters; however, surrounding areas include warehouses and yards devoted to support port activities therefore the overall port size is around 2,500,000 square meters. Other features that are worth mentioning include 11 km of quay, 100 docking points, and maximum deep water 40 feet. These features, along with its connection to the railway and to the major roads, make this port a strategic hub for both Italian and European freight traffic. For these reasons, the port of Livorno has been chosen as a reference scenario for the simulators development.

On the other hand, the ship simulator has been developed to recreate the behaviour of different types of ships including one containership, one tanker and one ro-ro ship. In this paper the case of the containership, that is based on the KRISO model (conceived by the Korea Research Institute for Ship and Ocean Engineering, 1997) , is presented. The main characteristics of the ship are given below (no real scale model of the KRISO exists, however the KRISO data have been used for executing flow physics and CFD validation for modern containership).

Figure 1: The port of Livorno

- Hull:
  - Length between perpendiculars: 230.0 m
  - Length water line: 232,5 m
  - Breath: 32.2 m
  - Depth: 19.0 m

- Displacement: 52030 m^3
- Coefficient block 0.651

- Rudder
  - Type semi-balanced horn rudder
  - Surface of rudder:115 m^2
  - Lat. Area: 54.45 m^2
  - Turn rate: 2.32 deg/s

- Propeller
  - Number of blades: 5
  - Diameter: 7.9 m
  - Pitch ratio P/D (0.7R): 0.997
  - Rotation Right hand

The tugboat is characterized as follows:

- Hull
  - Length between perpendiculars: 26.0 m
  - Breath: 8.3 m
  - Depth: 3.7 m
  - Displacement: 710 m^3
  - Coefficient block: 0.6

- Rudder
  - Height:1.38 m
  - Area ratio A$_R$/Ld: 0.02
  - Aspect ratio: 1.4

- Propeller
  - Diameter: 1.1 m
  - Pitch ratio P/D (0.7R): 0.86

3. SHIP AND TUGBOAT EQUATIONS FOR THE MOTION AT SEA

The dynamic behaviour of both the ship and tugboat has been modelled according to a 6 DOF model. The so-called MMG model was adopted for surge, sway and heading whereas for the remaining 3 DOF (roll, pitch and heave) the reference model is given by Jensens (2001) and Jensens et al. (2004).

3.1 MMG model

The MMG model (MMG, 1985) includes two coordinate systems: a global coordinate system where the axes are marked as x_0 and y_0, and the other is a body-fixed coordinate system, where the axis are labelled as x and y and are centred at the ship centre of gravity.

The MMG model is based on the three equations of motion (given in 1, 2 and 3), one for each DOF. Such equations have been found out applying the Newton's second law.

$$(m + m_x)\dot{u} - mvr = X \qquad (1)$$

$$\left(m + m_y\right)\dot{v} + mur = Y \qquad (2)$$

$$(I_{zz} + i_{zz})\dot{r} = N - x_G Y \qquad (3)$$

In equations 1, 2, and 3:

- m is the mass of the ship;
- m_x and m_y are the added mass in the x and y directions respectively;
- I_{zz} is the moment of inertia;
- i_{zz} is the added moment of inertia around z;
- u is the surge speed;
- v is the sway speed;
- r is the rate of turn;
- the point above the variable identifies the derivative of that variable;
- x_G is the distance from amidships to the center of gravity of the ship
- X and Y are respectively the total external surge and sway forces;
- N is the yaw moment;

External forces are generated by the hull resistance, the propeller and the rudder, such forces are identified by the subscript *H, P* and *R* respectively as shown in equations 4, 5 and 6.

$$X = X_H + X_P + X_R \qquad (4)$$

$$Y = Y_H + Y_R \qquad (5)$$

$$N = N_H + N_R \qquad (6)$$

According to the The Specialist Committee on Esso Osaka (2002) it is possible to use the equations 7, 8 and 9, where the variables marked by the primed symbol are non-dimensional variables, to calculate the hull forces.

$$X'_H = -(X'_0 + \left(X'_{vr} - m'_y\right)v'r') \qquad (7)$$

$$Y'_H = Y'_v v' + \left(Y'_r + m'_x\right)r' + Y'_{vvv}v'^3 + Y'_{vvr}v'^2 r' + Y'_{vrr}v'r'^2 + Y'_{rrr}r'^3 \qquad (8)$$

$$N'_H = N'_v v' + \left(N'_r + m'_x\right)r' + N'_{vvv}v'^3 + N'_{vvr}v'^2 r' + N'_{vrr}v'r'^2 + N'_{rrr}r'^3 \qquad (9)$$

The equations 7, 8 and 9 define relations between velocities and hull resistance using hydrodynamic non-dimensional coefficients. These coefficients are normally calculated with thank tests but in Lee et al. (2003) a set of semi-empirical equation is given.

In equation 10 X'_0 is the total non-dimensional resistance, C_T is the total resistance coefficient (obtained from model resistance tests), S is the wetted surface, L is the length between perpendiculars and d is the draft.

$$X'_0 = \frac{C_T S}{Ld} \qquad (10)$$

The propeller force has been evaluated according to Kijima and Nakiri (2003), see equation 11.

$$X_P = (1 - t)\rho n^2 {D_p}^4 K_T \qquad (11)$$

In equation 11, ρ is the density of the water, n is the propeller rate expressed in rounds per minute (RPM), t is the suction coefficient, D_p is the propeller diameter and K_T is the propeller thrust coefficient (that has been calculated according to Kijima and Nakiri, 2003).

In addition, the equations 12, 13 and 14 used to evaluate rudder forces, are taken once again from Kijima and Nakiri (2003).

$$X'_R = -(1 - t_R)F'_N sin\delta \qquad (12)$$
$$Y'_R = -(1 - a_H)F'_N cos\delta \qquad (13)$$
$$N'_R = -(x'_R - a_H x'_H)F'_N cos\delta \qquad (14)$$

When evaluating rudder forces, t_R is the rudder drag coefficient; F'_N is the normal force applied on the rudder; a_h is a coefficient that expresses the interaction between rudder and hull forces; x'_R is the non-dimensional coordinate of the centre of lateral force along the x-axis; x'_H is the non-dimensional coordinate of the centre of additional lateral force along the x-axis; δ is the rudder angle.

3.2 Heave, Pitch and Roll model

The response of a ship to wave induced loads is quite complex. As a matter of facts complex interactions between the ship dynamics and several hydrodinamic forces have to be considered. Basically, the equations of motion can be derived from the Newton's second law. In particular, recalling the linear theory, acting forces and moments can be divided into excitations and radiations forces/moments. Excitation forces are the forces of the waves acting on a restrained ship while radiation forces are caused by ship motions.

Considering heave, roll and pitch denoted by w, θ, and φ, the related coeffiecients labelled with the subscript *3, 4* and *5* respectively, the excitation and radiations forces/moments labeled with the subscripts *EX* and *H* respectively, for sinusoidal uncoupled motions the equations are given in 15, 16 and 17 where derivation with respect to time is denoted by a dot (Lewis, 1989).

$$(\Delta + A_{33})\ddot{w} + B_{33}\dot{w} + C_{33}w = |F_{EX3}\cos(\varpi t + \epsilon_3)| \qquad (15)$$

$$(I_{44} + A_{44})\ddot{\varphi} + B_{44}\dot{\varphi} + C_{44}\varphi = |F_{EX4}\cos(\varpi t + \epsilon_4)| \qquad (16)$$

$$(I_{55} + A_{55})\ddot{\theta} + B_{55}\dot{\theta} + C_{55}\theta = |F_{EX5}\cos(\varpi t + \epsilon_5)| \qquad (17)$$

In equations 15,16 and 17, the coefficients A_{33}, A_{44}, and A_{55} have the dimension of a mass and are called hydrodynamic masses or added masses, the coefficients B_{33}, B_{44}, B_{55} have the dimension of a mass per unit of time and are called damping coefficients; the coefficients C_{33}, C_{44}, C_{55} are the restoring spring coefficients. Interesting theoretical insights about such coefficients and equation of motions can be found in Lewis (1989). In addition, Δ is the displacement, ω is the wave frequency, ϖ is the frequency of encounter and I_{44} and I_{55} are the mass moment of inertia for roll and pitch respectively.

Obviously, in order to obtain numerical values for motion amplitudes the coefficient values and exciting forces amplitudes have to be known. The calculation is easy for the C_{33}, C_{44}, C_{55} coefficients that can be derived from stability calculations, but is very difficult for excitation forces, added mass and damping coefficients since a very complex hydrodynamic problem has to be solved. In most practical application the Strip Theory is applied (Ogilvie and Tuck, 1969). However, it has been found out that the implementation of a numerical code based on the strip theory was out of the scope of this study. As a matter of facts the simulator should work real time and therefore a lightweight computation model is needed even for visualization purposes. To this end, past related works such as Chen and Fu (2007), Ueng et al (2008), Sandaruwan et al. (2009), Sandaruwan et al. (2010), Yeo et al. (2012), have been investigated. These works seek to calculate buoy motions arising from waves in order to achieve realistic visualization results and are based on simple dynamic models. However, ship responses are evaluated just in terms of visualization and the numerical results have been accepted even if in some cases they have proven to be far from real empiric data. Under these considerations a more realistic and accurate model has been selected. This model has been proposed by Jensen, (2001) and Jensen et al (2004) and is based on simplified equations for ship motions in regular waves where the coupling terms are neglected and the sectional added mass is equal to the displaced water.

$$2\frac{kT}{\omega^2}\ddot{\omega} + \frac{A^2}{kB\alpha^3\omega}\dot{\omega} + \omega = aFcos(\varpi t) \qquad (18)$$

$$2\frac{kT}{\omega^2}\ddot{\theta} + \frac{A^2}{kB\alpha^3\omega}\dot{\theta} + \theta = aGsin(\varpi t) \qquad (19)$$

$$\left(\frac{T_N}{2\pi}\right)^2 C_{44}\varphi^{\cdot\cdot} + B_{44}\dot{\varphi} + C_{44}\varphi = Macos(\varpi t) \qquad (20)$$

The equations 18 and 19 refer to heave and pitch respectively; k is the wave number, B is the ship breadth, T is the ship draught, a is the wave amplitude and A is the sectional hydrodynamic damping that can be evaluated according to Yamamoto et al., (1986). Moreover, F and G are the forcing functions whose values can be worked out according to Jensen et al. (2004). As for roll, the equation is given in 20 where, T_N is the natural period for roll, B_{44} is the ship hydrodynamic damping, C_{44} is the restoring moment coefficient and M is the roll excitation moment. The hydrodynamic damping coefficient, B_{44}, can be found by applying the method described in Jensen et al. (2004), the roll excitation moment M can be derived from the Haskind relation (one of the most outstanding results in the ship oscillations theory), while the restoring moment coefficient C_{44} can be expressed as a linear function of the displacement Δ, the transverse metacentric height GM_T and the acceleration of gravity g (see equation 21).

$$C_{44} = gGM_T\Delta \qquad (21)$$

In addition, it is worth noticing that the semi-analytical approach proposed in Jensen et al (2004) is intended to derive frequency response functions of wave-induced motions. Therefore, the model, as it is, is mainly addressed to naval engineering applications. However, the application proposed in this research work is quite new since the model has been solved in the time domain by the Euler's method and is used to provide the simulation and visualization system with real time data.

3.3 Ship and tugboat interactions

A tugboat can support ship manoeuvres in two ways: it can either pull the ship using a rope or push it. As for the interactions between the ship and the tugboat, since in calm water such interactions are not critical for roll, pitch and heave, only their effects on surge, sway and yaw have been modelled. However, before going into the details of the interaction model, it is worth saying that, while interacting, the ship and the tugboat can be considered as a single system that can be described with a single mathematical model. In particular, a new external force needs to be added in the dynamic model of both the vessels; this force is labelled with the subscript *T* and is the force that each ship applies on each other. As a result, the ship-tugboat system can be described as reported on equations 22, 23, 24, 25, 26, 27, and 28.

$$(m_S + m_{x_S})\dot{u}_S - m_S v_S r_S = X_{H_S} + X_{P_S} + X_{R_S} + \text{T}\cos\gamma_S \qquad (22)$$

$$\left(m_S + m_{y_S}\right)\dot{v}_S + m_S u_S r_S = Y_{H_S} + Y_{R_S} + \text{T}\sin\gamma_S \qquad (23)$$

$$(I_{zz_S} + i_{zz_S})\dot{r}_S = N_{H_S} + N_{R_S} - \text{T}\cos\gamma_S A_{Y_S} + \text{T}\sin\gamma_S A_{X_S} - x_{G_S}\left(Y_{H_S} + Y_{R_S} + \text{T}\sin\gamma_S A_{X_S}\right) \qquad (24)$$

$$(m_{TB} + m_{x_{TB}})\dot{u}_{TB} - m_{TB} v_{TB} r_{TB} = X_{H_{TB}} + X_{P_{TB}} + X_{R_{TB}} + \text{T}\cos\gamma_{S\ TB} \qquad (25)$$

$$\left(m_{TB} + m_{y_{TB}}\right)\dot{v}_{TB} + m_{TB} u_{TB} r_{TB} = Y_{H_{TB}} + Y_{R_{TB}} + \text{T}\sin\gamma_{TB} \qquad (26)$$

$$(I_{zz_{TB}} + i_{zz_{TB}})\dot{r}_{TB} = N_{H_{TB}} + N_{R_{TB}} - \text{T}\cos\gamma_S A_{Y_S} + \text{T}\sin\gamma_S A_{X_S} - x_{G_{TB}}\left(Y_{H_{TB}} + Y_{R_{TB}} + \text{T}\sin\gamma_{TB}\right) \qquad (27)$$

$$D_{STB} = l \qquad (28)$$

In equations 22-28:

- The subscript *S* identifies ship-related variables;
- The subscript *TB* identifies tugboat-related variables;
- A_x identifies the y coordinate of the *T* force application point;
- A_y identifies the x coordinate of the *T* force application point;
- l is the length of the rope if the tugboat is pulling and it is 0 if it is pushing;
- *DSTB* is the distance between the application point of the force *T* to the ship and application point of the force *T* to the tugboat; it depends on the ship accelerations, on the tugboat accelerations, and on the force.
- γ is:
  - the angle between the rope and the *axis* of the ship/tugboat when a rope is used (the force T has the same direction as the rope);
  - the angle between the perpendicular to the hull, in the vessels point of contact, and the *x* axis of the ship/tugboat if they are pushing each other.

Such system of differential equations has been solved by the Euler's Method.

4. PRELIMINARY TEST FOR SHIP AND TUGBOAT INTERACTIONS

The aforementioned mathematical model has been evaluated by an ad-hoc tool implemented by the C++ programming language in the Visual Studio 2008 Integrated Development Environment.

This tool allows setting some input parameters such as:

- pushing/pulling;
- points of contact for pushing;
- docking points of the rope for pulling;
- ship/tugboat position and orientation;
- ship/tugboat engine turn rate;
- ship/tugboat rudder angle;
- sea state;

Based on the input parameters, the tool draws a dynamic plot about the trajectory of both the ship and the tugboat. The ship is shown as a segment with a circle on the top (indicating the bow), while the tugboat is depicted as a smaller segment with a circle indicating the bow. In addition some relevant data, as linear and angular velocities, position and orientation, are recorded on a text file.

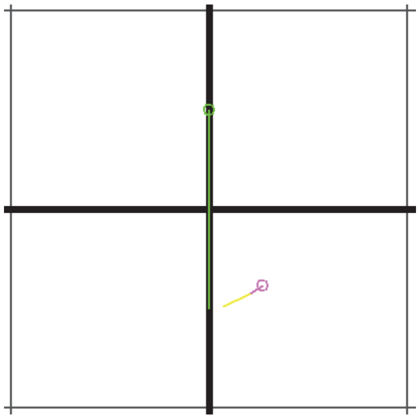


Figure 2: Chart of the ship and tugboat during a pushing experiment at time 0

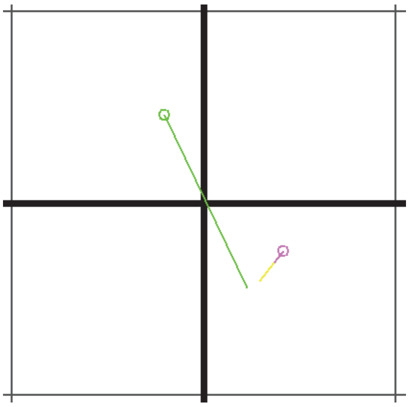


Figure 3- Chart of the ship and tugboat during a pushing experiment after 175 seconds

Figure 2 and Figure 3 depict the plots related to a particular pulling experiment. The ship engine has 0 RPM and the initial speed is 0.0 kn, the tugboat engine has 400 RPM, the initial speed 0.0 kn and the rudder angle is 0°. The green object is the ship, the violet one is the tugboat and the yellow line is the rope (this line becomes red in colour when the rope is not taut). Figure 2 shows the ship and tugboat position at the instant of time 0, while figure 3 shoes their position after 175 seconds. It is worth saying that the system is able to predict the position of the ship and tugboat even if the engine of the ship or the side thrusters are operating. Indeed, most of the times, ship manoeuvrings in the port areas are executed by using both the help of tugboats and the ship propulsion systems (including main engine and bow/stern thrusters). The system is also able to take into account the effects of the wind and marine currents.

The C++ tool developed by authors has been particularly useful to test the interactions between the ship and the tugboats in many different cases; indeed the interactions has been verified and validated with the help of subject matter experts (ship and tugboats pilots). After some preliminary analysis subject matter experts were able to identify errors in the behaviour of ship and tugboat during their interactions (i.e. by executing manoeuvres such those depicted in figures 2 and 3 they noticed errors in the positions of the ship and tugboat). Such errors have been corrected by acting iteratively on the values of the hydrodynamic coefficients; therefore the simulators have been used as system for tuning the model parameters according to subject matter experts' suggestions.

5. THE HLA ARCHITECTURE FOR COOPERATIVE TRAINING

As already mentioned, the ship simulator and the tugboat simulator are fully interoperable. Interoperability has been achieved integrating the simulators according to the HLA 1516 integration standard.

Before going into the details of the simulation architecture, it is worth focusing the attention on some basic definitions and concepts within the HLA framework. Basically, HLA relies on three elements namely the federates, the federation and the Run Time Infrastructure (RTI). A federate is an individual, HLA-compliant simulation application; a federation is a simulation system composed of two or more federates; the RTI is the software that manages the simulation execution and data exchange among the federates (Bruzzone et al., 2008; Massei et al. 2013).

In turn, our federates include three components: the hosting environment that is the virtual environment where the federate is located, the geometric models and the dynamic process module that includes information about the federate configuration, state, dynamical behaviour, etc, including the interfaces with the RTI (the authors have already successfully applied a similar approach to develop interoperable simulators in logistics area, Bruzzone and Longo 2013; Longo, 2012).

Within the proposed simulation architecture, there are three federates: the ship federate, the tugboat federate and the control tower federate (the latter not described in this paper because still under development). As already stated in Section 2, the reference scenario is based on the port of Livorno (Italy). Both the virtual environment and the federates' geometrical models have been developed by the tool Creator provided by Presagis. The figure 4 shows the geometric models of the Livorno Port area as it appears after texturing operations and ready to be imported in the virtual environment.

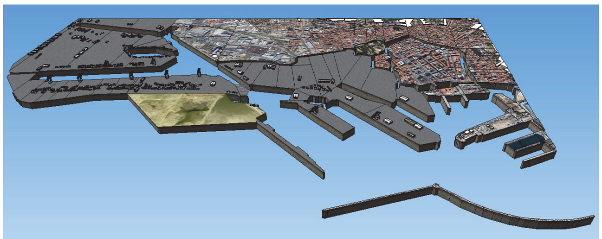


Figure 4: the geometric model of the Livorno port area after texturing operations

On the other hand, the dynamic process modules of both the federates has been developed from scratch by programming code, algorithms and functions written in C++ (most of the programming code is the same used to developed the testing tool presented in the previous section). In particular the code that allows the federates to reproduce the behaviour of the real ship/tugboat, to interact with the virtual environment, to collect data and performance measures is called internal functions. Conversely, there are external functions responsible for data management and information exchanges between the federates; therefore such functions play a crucial role in reproducing the interactions between them. The graphical engine that has been used is Vega Prime from Presagis.

Figure 5 and 6 show a screenshot of the container ship simulator interacting with the tugboat simulator according to two different viewpoints.

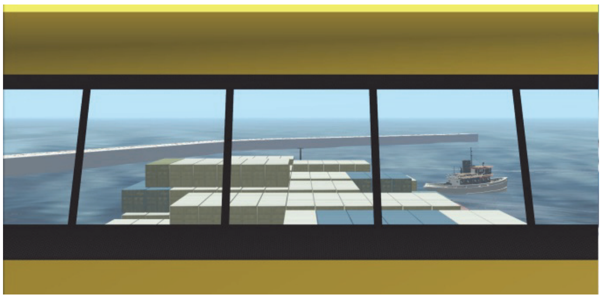


Figure 5: Ship simulator: tugboat pulling the ship (bridge view).

As mentioned in the previous sections great attention has been paid on modelling the interactions among the federates. To this end the simulation architecture has been provided with interoperability capabilities that allows reproducing with satisfactory accuracy the real operational processes taking place in the last mile of navigation.

However, it should be noted that, even if the federates have been integrated into a federation in order to allow the cooperative and joint training of ship pilots and tugboat pilots, the integration via HLA does not prevent each federate from being executed standalone or to be reused in different federations (Longo, 2011). In this perspective, the proposed tool ensures a great flexibility that is very important because it allows the trainees to train alone with the single ship/tugboat when necessary.

6. CONCLUSIONS

Even though several ship simulators are already on the market few of them offer the possibility of setting up combined and integrated training sessions involving both ship pilots and tugboat pilots. In fact, the last mile of navigation has not been subject of particular scientific interest until the recent disaster in the port of Genoa.

However, in standard conditions, manoeuvres within the port area may be the most critical moment of the whole navigation and during these operations, ports regulations, may require that a certain number of tugboats support the ship. Even if this rule facilitates the manoeuvres inside the port, on the other hand it makes the operations even thornier than expected since ships and tugboats have to work in synergy and relatively close to each other; therefore training is crucial in order to carry out such manoeuvres in a safe and efficient way.

Considering these aspects this paper describes the work done (that actually is still on-going) to develop an interoperable simulation framework for training of ship and tugboat pilots and for port traffic controllers with the aim of:

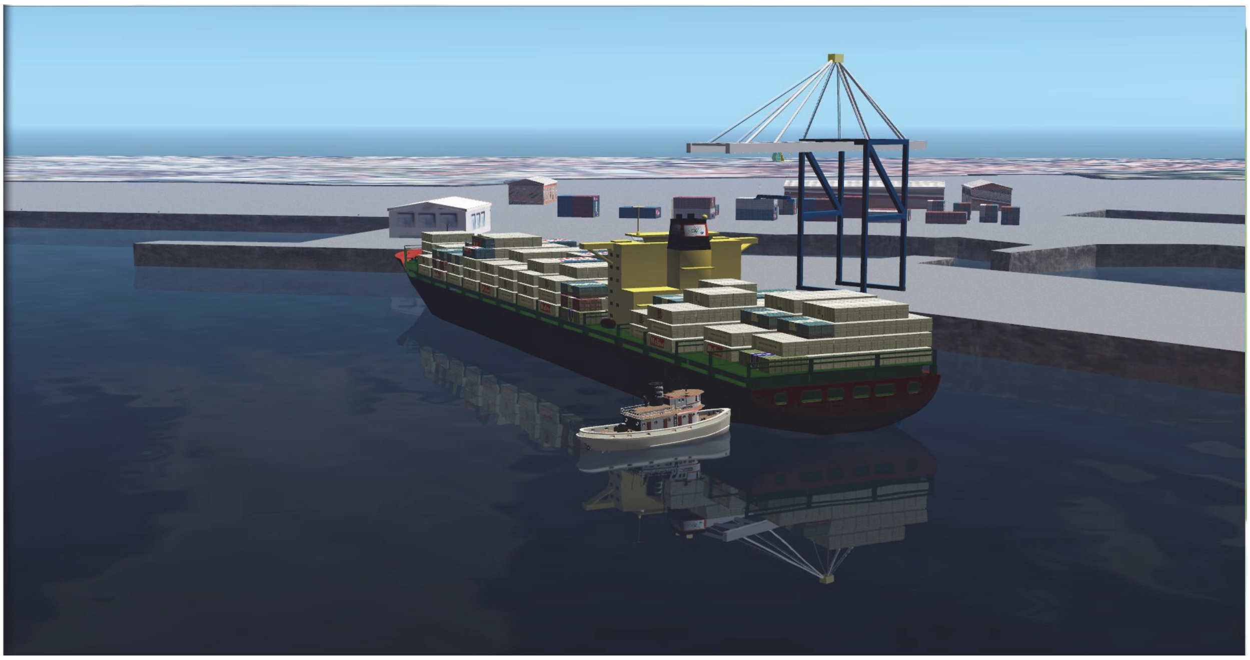


Figure Ship simulator: the tugboat pushing the ship, rear view.

- improving the trainees skills on steering a ship/tugboat;
- learning about the procedures adopted in a certain port (it is important in this sense to underline the possibility to replace the virtual environment easily with another one);
- improving synergy and communication between ship pilots, tugboats pilots and control tower;
- defining new policies and designing new procedures;
- testing the effectiveness of new policy/procedures.

Since HLA allows creating a distributed federation, it is possible to locate each federate (simulator) in a different rooms (or even geographic areas) in order to recreate the conditions that occur in the real word. To this end the research activities are still ongoing with the following objectives:

- tune the effects of the wind and of the side thrusters;
- design and provide each simulator with a reproduction of a real cockpit (external hardware);
- complete the development of the control tower simulator that will be integrated within the federation.

ACKNOWLEDGMENTS

The research work described in this paper is part of an ongoing research project, HABITAT PON01_01936, co-financed by the Italian Ministry of Research and Education (MIUR) under the program "PON Ricerca e Comptetitività".

REFERENCES

Armaoğlu, E., Garcia-Tuñón, J., Alonso, J.R.I., Verdugo, I., Bron, I., 2009. Manoeuvring mathematical model for "ship docking module (sdm™)g. ", *Proceedings of MARSIM '09*, August.

Bruzzone A.G., Poggi, S., Bocca, E., 2008. Framework for interoperable operations in port facilities. *Proceedings of the European Conference on Modeling & Simulation*,vpp. 277-282.

Bruzzone A.G., Fancello, G., Fadda, P., Bocca, E., D'Errico, G., Massei, M., 2010. Virtual World and Biometrics as strongholds for the development of innovative port interoperable simulators for supporting both training and R&D. *International Journal of Simulation and Process Modeling,* 6(1), 396-402.

Bruzzone, A.G., Fadda, P., Fancello, G. Massei, M., Bocca, E., Tremori, A., Tarone, F., D'Errico, G., 2011. Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes. *Simulation*, 87(10), 857-874.

Bruzzone, A.G., Longo, F., Nicoletti, L., Diaz, R., 2012-a. Traffic controllers and ships pilots training in marine ports environments. *Proceedings of the 2012 Symposium on Emerging Applications of M&S in Industry and Academia Symposium*, Article No. 16, March 26-29, FL, USA.

Bruzzone, A.G., Longo, F., Nicoletti, L., Bottani, E., Montanari, R., 2012-b. Simulation, Analysis and Optimization of container terminal processes. *International Journal of Modeling, Simulation and Scientific Computing,* 3(4), art. 1240006.

Bruzzone A.G., Longo F., 2013. 3D simulation as training tool in container terminals: The TRAINSPORT simulator. Journal of Manufacturing systems, 32(1), 85-98.

Chen, C. H., Fu, L. C., 2007. Ships on real-time rendering dynamic ocean applied in 6-DOF platform motion simulator. *Proceedings of the CACS international conference*.

Hasegawa, K., Kang, D., Sano, M., Nagarajan, V., Yamaguchi, M., 2006. A study on improving the course-keeping ability of a pure car carrier in windy conditions, *J Mar SciTechnol*, 11,76–87.

Jensen, J.J., 2001. Loads and global response of ships.*Elsevier Ocean Engineering Book Series*, vol. 4, Elsevier.

Jensen, J.J., Mansour, A.E., Olsen, A.S., 2004. Estimation of ship motions using closed-form expressions, *Ocean Engineering*, 31(1), 61–85

Kijima, K., Nakiri, Y. 2003 On the Practical Prediction Method for Ship Manoeuvring Characteristics, *Transactions of the West-Japan Society of Naval Architects*, 105, 21-31.

Lee, T., Ahn, K.-S., Lee, H.-S., Yum, D.-J., 2003 On an Empirical Prediction of Hydrodynamic Coefficients for Modern Ship Hulls, *Proceedings of MARSIM ´03*, August, Kanazawa Japan.

Lewis, E. V., 1989. *Principles of Naval Architecture-Second Revision, Volume III Motions In Waves And Controllability*. New Jersey, The Society of Naval Architects and Marine Engineers.

Longo, F., 2007. Students Training: integrated models for simulating a container terminals. *Proceedings of the International Mediterranean Modeling Multiconference, I3M*, Bergeggi, Italy.

Longo, F., 2012. Supply chain security: an integrated framework for container terminal facilities. *International Journal of Simulation & Process Modelling*, 7(3), 159-167.

Massei, M., Tremori, A., Poggi, S., Nicoletti, L., 2013. HLA based real time distributed simulation of a marine port for training purposes. *International Journal of Simulation and Process Modeling*, 8(1), 42-51.

Merkuryev, Y., Bikovska, J., 2012. Business Simulation game development for education and training in supply chain management. *Proceedings of the 6th Asia International Conference on Mathematical Modeling and Computer Simulation*, AMS 2012, art. No 6243943, pp. 179-184.

MMG 1985. Prediction of Manoeuvrability of A Ship. In: *Bulletin of the Society of Naval Architects of Japan*.Japan, The Society of Naval Architects of Japan.

Ogilvie, T. F., Tuck, E. O., 1969. A Rational Strip Theory of Ship Motions: Part I. *Naval Architecture & Marine Engineering.*

Perez, F. L., Clemente, J. A., 2006. The influence of some ship parameters on maneuverability studied at the design stage*. Science Direct Ocean Engineering 34*, 518–525.

Rhee, K. P., Kim, K. H., 1999. A new sea trial method for estimating hydrodynamic derivates. *J Ship Ocean Technol* 3(3), 25–44.

Sandaruwan, D., Kodikara, N., Keppitiyagama, C.,Rexy, R., 2010.A Six Degrees of Freedom Ship Simulation System for Maritime Education, *The International Journal on Advances in ICT for Emerging Regions*, 03 (02), 34 – 47.

Sandaruwan, D., Kodikara, N., Rexy, R., Keppitiyagama, C., 2009. Modeling and Simulation of Environmental Disturbances for Six degrees of Freedom Ocean Surface Vehicle.*Sri Lankan Journal of Physics*, 10,39-57.

The Specialist Committee on Esso Osaka, 2002. Final Report and Recommendations to the 23 rd ITTC.*Proceedings of the 23 rd ITTC* – Volume II, 2002

Ueng, S.-K. , Lin D., Liu C.H., 2008. A ship motion simulation system. *Virtual Reality*,12, 65–76, DOI 10.1007/s10055-008-0088-8.

Yamamoto, Y., Sugai, K., Inoue, H., Yoshida, K., Fugino, M., Ohtsubu, H., 1986. Wave loads and response of ships and offshore structures from the viewpoint of hydroelasticity. *In Proceedings of the International Conference on Advances in Marine Structures, Admiralty Research Establishment*, May 20-23, Dunfermline, Scotland.

Yeo, D. J., Moohyun, C., Duhwan, D., (2012) ship and buoy motions arising from ocean waves in a ship handling simulator. *Simulation,* 88 (12), 1407-1418.

AUTHOR BIOGRAPHIES

Francesco Longo received his Ph.D. in Mechanical Engineering from University of Calabria in January 2006. He is currently Assistant Professor at the Mechanical Department of University of Calabria and Director of the Modelling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES). He has published more than 120 papers on international journals and conferences. His research interests include Modeling & Simulation tools for training procedures in complex environment, supply chain management and security. He is Associate Editor of the “Simulation: Transaction of the society for Modeling & Simulation International”. For the same journal he is Guest Editor of the special issue on Advances of Modeling & Simulation in Supply Chain and Industry. He is Guest Editor of the “International Journal of Simulation and Process Modelling”, special issue on Industry and Supply Chain: Technical, Economic and Environmental Sustainability.

Alessandro Chiurco is Researcher at MSC-LES, Department of Mechanical, Energy and Management Engineering, University of Calabria. His research activities concern the development of 3D immersive and interoperable simulators for training based on the HLA standard. He is also used simulation for investigating problems related to supply chain and marine ports security.

Letizia Nicoletti is PhD student at MSC-LES, Department of Mechanical, Energy and Management Engineering, University of Calabria. Her research interests include Modelling & Simulation for training in complex system with particular attention to marine ports and container terminals. She is also using Modelling & Simulation based approaches for inventory management problems in industrial plants and supply chains. From 2010, she actively supports the organization of the I3M Multi-conference where she is co-chairing the Inventory Management Simulation track.

DRIVERS AND PARKERS TRAINING IN CAR TERMINALS

Francesco Longo[(a)], Letizia Nicoletti[(b)], Alessandro Chiurco[(c)], Adriano Solis[(d)]
Francisco Spadafora[(e)]

[(a)(b)(c)(e)] DIMEG, University of Calabria, Italy
[(d)] York University, Canada

[(a)]f.longo@unical.it, [(b)]letizia.nicoletti@unical.it, [(c)]a.chiurco@unical.it,
[(d)]asolis@yorku.ca, [(e)]f.spadafora@msc-les.org,

ABSTRACT

The paper presents a simulation based training framework for drivers and parkers in car terminals along with some preliminary results achieved during the development phase. The proposed framework relies on a modular simulation architecture devoted to reproduce the main processes and activities taking place in a car terminal. To reproduce the car terminal operational processes with accuracy and provide the users with a cooperative training scenario, the simulation architecture includes three interoperable simulation modules; namely the Operator Simulator (the Marshalls), the Ship Simulator and a Vehicle Simulator. In particular, the paper focus is on the Vehicle Simulator, whose development and implementation is discussed in detail. As for the other modules, a general overview is provided.

Keywords: car terminals, drivers training, parkers training, simulation

1. INTRODUCTION

A car terminal is a complex system whose complexity relies on two main elements: the nature of inner daily processes and the great number of operators and vehicles (of different types) involved. Therefore, operators' training is critical to preserve operators', vehicles and equipment safety and security, ensure operational efficiency and avoid economic losses. Each operator has his own role and therefore specific training needs. However, within the whole scenario the most critical roles are taken by drivers (both vehicle and taxi drivers) and marshalls (operators assisting drivers during parking operations) whose errors may cause severe human and economic damage. In fact, drivers have to deal with a variety of working conditions including different levels of risk. The main risk factors depend on the configuration of the ship involved in loading/unloading operations; i.e. steep ramps, sharp bends, narrow aisles, slippery floors, etc. Moreover, for optimization purposes, vehicles loading and unloading are concurrent operations therefore it is likely that opposite flows cross each other increasing even more the risk of accidents.

Needless to say that only well-trained (both in theory and practice) and highly qualified staff can cope with the complexity of such a working environment and carry out its tasks effectively and safely. As mentioned before, improper behaviors, lack of coordination, incorrect procedures may result in losses of human lives and, from an economic point of view, in increased direct and indirect costs (Bruzzone and Longo, 2013).

Traditionally, training activities include frontal classes aimed at illustrating and discussing best practices and operational procedures that should be adopted in standard, unusual and dangerous conditions. Usually, training is not limited to frontal classes but includes also practical training where inexperienced operators are involved in coaching sessions driving in the real system with real vehicles. In particular, drivers and Marshalls courses last between 20 and 40 hours whereas coaching sessions last between 40 and 80 hours. In addition, training does not involve inexperienced operators only; further training is needed to illustrate lessons-learned, successful experiences or even new procedures. Furthermore, training for after-action review may be required in case of accidents and vehicles damage. In this case, training activities aim at understanding why the accident occurred identifying which measures and operational modes can prevent the same situations from happening in the future.

Hence, operators' training is a crucial and critical activity in car terminals; as a consequence there is a continuous search for tools that allow reducing training costs and maximizing training effectiveness.

To this end, Modeling & Simulation (M&S) has proved to be a powerful methodology for dealing with complex systems design, management and even training. As a matter of facts, development and testing real prototypes, even prototype solutions based on simulation, is a relevant training opportunity for its users (i.e. operational testing of weapon systems in the military industry is a clear example).

Simulation allows reproducing a real system and its behavior through an artificial system (the simulation model) therefore, operators involved in simulation-based training activities, while interacting with the simulated environment (that in most of the cases is a 3D virtual environment) can learn how to interact with the real system and the real equipment. As a result this approach can be advantageous especially when using real equipment in the real system is costly and dangerous as it may be in a car terminal. Indeed, simulation provides a safe training environment (the operator interacts with a virtual word reproducing the real system) where human errors have not economic impacts. In other words, operators can even apply wrong procedures to see the consequences of his actions and learn how to handle vehicles and equipment safely to perform their tasks effectively and efficiently.

The main benefits of M&S can be summarized as follows (Cimino et al., 2010):

- practice theoretical concepts and gain awareness of the main consequences related to the undertaken course of action in a very immediate and visual manner;
- provide instructors with a controlled environment where a large amount of data can be recorded and analyzed to evaluate the trainee's evolution and performances;
- avoid hazardous situations that usually occur when inexperienced users manipulate real machines;
- reduce costs associated to training operations;
- provide trainees with the possibility of working in any desired condition (i.e. arbitrary weather conditions).

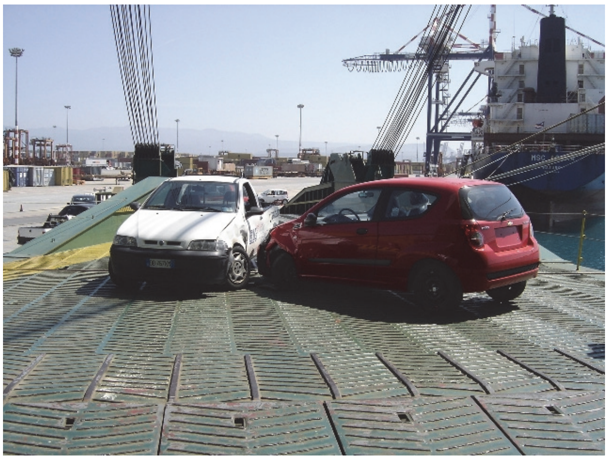


Figure 1: Cars accident during loading/unloading operations in a car terminal

2. RELATED WORKS

Over the years Modeling & Simulation (M&S) has proved to be a very effective problem solving methodology in different areas including Industry, Logistics and Supply Chains (Piera et al. 2004). As far as the marine ports domain is concerned, many are the cases in which M&S has been used for supporting decision making at strategic, tactical and operative level (Longo 2007, Bruzzone et al., 2000; Bruzzone et al. 2012) also including (above after September 11th 2001) security enhancement (Longo, 2010). However, in the same domain, M&S has been also widely applied for supporting operators training (i.e. container terminals) especially for high-risk and costly activities. There are many examples of works and research projects where simulation is profitably used as a training tool. In the following a general survey on past related works is proposed. In particular, the state of the art allows pointing out that simulation has been used extensively for the training of the operators involved in containers handling processes and loading/unloading operations. As follows a brief description of different types of applications is reported. Many are the examples of simulation systems devoted to train quay cranes operators: Wilson et al. (1998) propose a 3D virtual system devoted to simulate crane operations; such system allows reproducing also the feelings and sensations that can be experienced in a crane cockpit. Huang (2003) presented a method to design an interactive visual simulation mobile crane training. Daqaq (2003) developed a virtual simulation for training of ship-mounted cranes operators. Rouvinen et al. (2005) developed a gantry crane simulator intended for container handling operations between yard and ships. Fernandez et al. (2009) present a training simulator for different kinds of operators, namely quay crane, gantry, rubber tired gantry and reach-stacker operators. The simulator includes an automated system devoted to track and monitor operators' skills. In Elazony et al. (2010), attention is focused on the design and implementation of reusable and interactive simulation-based training systems. Similarly, Lau et al. (2007) present a distributed real-time simulation model for container terminal processes. Furthermore specific research works have been developed to support operators' training and procedures design within container terminals. Moreover, several examples of distributed simulation for operators training in container terminals can be found in Merkuriev et al. (1998), Bruzzone et al. 2010, Bruzzone et al. (2011).

In addition it is worth mentioning that a complete survey on the major projects focusing on simulation systems for training of marine operators is one of the main deliverables of the OPTIMUS (Operational Port Training Models Using Simulators, that is financed by the European Union) project.

A careful analysis of the state of the art shows that the most common simulators for training in the port area include:

- ships bridge simulators:
- engine room simulators;
- handling loads simulators

In these simulators, usually the particular attention is paid on the visualization system that consists of a series of screens where the virtual environment (which recreates the real system) is projected. In addition, these simulators are designed for the training of the following kinds of operators:

- ships pilots;
- forklift operators;
- Reach Stacker operators;
- Straddle Carrier operators;
- Gantry Crane operators (STS, RTG, RMG);
- Offshore Crane operators;
- Tower Crane operators;

Some of the most important commercial simulators include:

- Drilling Systems (http://www.drillingsystems.com) whose simulator KraneSim is an advanced tool for simulating a wide range of quay cranes and vehicles.
- Oryx Simulations AB (http://www.oryx.se/), crane simulators that provide the users with different scenarios and different options in terms of cockpit, motion-based system, real-time graphics, background sounds, etc.
- ARI (http://www.ariworld.com/simulation/default.asp) and Total Soft Bank Ltd. (http://www.tsb.co.kr/) simulators for training on different types of cranes QC (Quay cranes), RTG (Rubber Tyre Gantry), RMG (Rail Mounted Gantry), SG (Ship Gantry), PC (Pedestal cranes), SC (Straddle Carriers)
- MPRI Ship Analytics (http://www.mpri.com/esite/), develops crane simulators for training that can play faithfully the operational characteristics of 12 types of cranes.
- STC Group http://www.stc-group.nl has developed simulators for various types of cranes such as containers cranes, bulk cranes and off-shore cranes.
- Simulation Team http://www.simulationteam.com has developed HLA interoperable simulators of different logistical means including gantry cranes, transtrainers, stackers, trucks, etc. These simulators are used for training, performance analysis and biomedical operators, as well as for virtual prototyping. Such simulators are available also in a full motion, immersive cave and containerized solution that can be easily transported where it is needed.

The studies on the effectiveness of M&S applications have pointed out their usefulness for training applications. In fact, simulators are widely used both for the first contact with machines and equipment and for the skill upgrading experienced operators. The effectiveness of simulation-based training is evaluated according to the transfer of the learned concepts to the real world during scheduled sessions where the operator acts in the real world under the supervision of an experienced instructor (Morrison et al 2000).

The literature review allows pointing out that there are many simulation-based applications for the port operators and especially for container terminal operators and for operators handling different types of vehicles (cars, trucks, buses, etc.). However, as far as we know, no research projects about training issues and M&S solutions applied to car terminals exist.

In fact, further analysis of the state of the art show that existing research works on car terminals are focused on transshipment operations using multi-agent systems (Fischer et al 2004), on operations management (Mattfeld et al 2002) and on business processes definition. In addition, the role of such logistic nodes in the supply chain of the automotive sector has been analyzed be Dias et al. 2007. However, the issues related to training and exercise of various professionals using advanced approaches based on M&S and 3D immersive virtual environments, are unexplored yet.

2.1. Contribution of the research work

Even if this paper shows only some preliminary results, the authors are carrying out a research project (called CTSIM) that will provide the following contribution to the current state of the state of the art:

- a training system for drivers of small, medium and large cars operating in car terminals;
- a training system for bus drivers operating in car terminals;
- a training system able to offer multiple scenarios (only the yard, only the ship, ship-yard, etc.), with at least two types of ships (a big ro-ro ship and a feeder ship);
- a training system for carrying out cooperative training of drivers of different vehicles (i.e. cars and buses);
- definition of appropriate performance metrics for evaluating trainees;
- a virtual advanced environment that can be used for performance evaluation also in case of structural lay-out changes in the terminal area;
- development of a business model easily exportable between different car terminals.

3. MAIN PROCESSES AND ACTIVITIES IN A CAR TERMINAL

In this section the main processes and activities that usually take place in a car terminal are described. One of these processes includes vehicles unloading and their placement into the yard. This process occurs after each ship entering the port is towed and moored. However, before unloading operations can start, some preliminary validation activities and macroscopic controls are carried out; in particular during these activities the staff verifies the compliance of each group of vehicles with the information reported in the informative systems and checks whether there are damaged vehicles. In addition, after that, the optimal unloading sequence of vehicles is defined and the yard position assigned to each vehicle is established. At this stage, unloading operations can start.

In the next step, each vehicle is placed in the previously assigned location; after arriving to its assigned location the vehicle is checked for damages (caused during the movement to the interiors or to the bodies). In case of damage a damage report is drawn up.

Another typical process that is carried out in a car terminal includes vehicles transfer from the yard area to a service area devoted to pre-delivery inspection (PDI) activities and a subsequent transfer to the loading area.

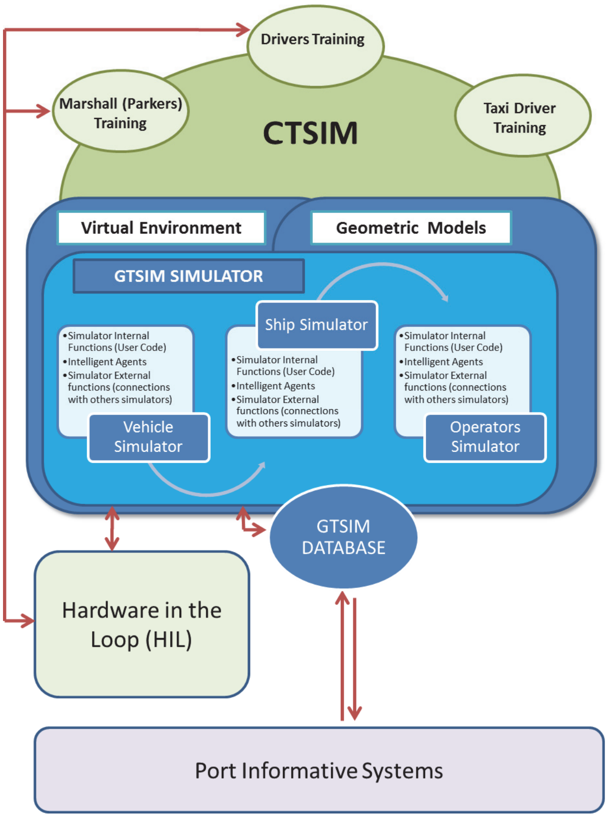
As far as the loading process is concerned, it should be noted that some of the vehicles are loaded on trains or trucks while other vehicles are loaded again on ships. Before the loading operation, further controls are carried out and if damaged vehicles are found these vehicles have to return to the PDI buffer area (to be repaired before leaving the terminal area). The vehicles that have to be loaded on ships are driven from the yard to the boarding ramps and, after an integrity check, are driven to the assigned spot on board.

Even in this case, it is possible that after inspections a damage report is drawn up and is attached to the involved vehicle. The main actors responsible for the aforementioned processes are:

- Drivers: they are in charge of vehicles handling (from the ship to the yard and vice versa) during loading/unloading operations and in the yard during shifting operations. Moreover, drivers have to cooperate with quality checkers and marshalls (parkers) in order to avoid incidents and errors while executing particular manoeuvres.
- Taxi Drivers: they pick drivers up from the yard and move them onto the ship (or vice-versa) during loading/unloading processes. In addition, they work in cooperation with quality checkers and marshalls, to ensure the correct loading/unloading sequence.
- Quality checkers: they verify that operators' behaviours are compliant with the instructions and procedures they must adhere (in particular they execute severe controls at dangerous points and during the vehicles inspections). On the other hand, coordination functions include those activities that are carried out in collaboration with taxi drivers and parkers to choose and communicate the assigned position (on the yard or on board the ship) and to ensure that the established loading/unloading sequence is respected.
- Service persons. The service persons are responsible for the viability on board and on the ramps; moreover, they assign bar codes and they are the first responders in case of accidents.
- Tally Men. The tally men are in charge of bar codes scanning (to get Vehicle Identification Numbers, VINs), of assigning a destination to the vehicles of each row/parking area; particular attention has to be paid in order to avoid scanning (wrongly placed) of vehicles with a destination different from the one assigned to the same row.
- Marshall (Parker). The marhsalls or parkers have to ensure that vehicles are parked according to the required instructions (such as distances between adjacent vehicles, parking on the line, checking handbrake / first gear, etc).

4. THE CTSIM GENERAL ARCHITECTURE

As already mentioned in section 2.1, the authors are carrying out a research project (CTSIM, Car Terminals Simulator) with the aim of developing a simulation framework devoted to car terminals operators training. On overview on the CTSIM architecture is given in figure 2. Basically, the main components of the architecture are three different interoperable simulators: the vehicle simulator, the ship simulator and the operator simulator. The vehicle simulator has been already developed while the development of the other simulators is still on-going. In this section the main characteristics of the ship and operator simulators are briefly described, while in section 5 the vehicle simulator is presented.



Drivers Training
Marshall (Parkers) Training
CTSIM
Taxi Driver Training
Virtual Environment
Geometric Models
GTSIM SIMULATOR
Ship Simulator
•Simulator Internal Functions (User Code)
•Intelligent Agents
•Simulator External functions (connections with others simulators)
Vehicle Simulator
•Simulator Internal Functions (User Code)
•Intelligent Agents
•Simulator External functions (connections with others simulators)
•Simulator Internal Functions (User Code)
•Intelligent Agents
•Simulator External functions (connections with others simulators)
Operators Simulator
GTSIM DATABASE
Hardware in the Loop (HIL)
Port Informative Systems

Figure 2: The CTSIM general architecture

The Ship Simulator allows simulating two types of vessels devoted to transport cars and buses: ro-ro car/truck carrier for long transport routes (international and/or intercontinental) and ro-ro feeder. In addition this simulator will provide the users with the opportunity to select climatic conditions such as wind, visibility, rain and sea state so as to offer the possibility to train over a wide variety of operating scenarios. The two types of ships will include side ramps and a stern ramp. Moreover, for each kind of ship, it will be possible to choose among different configurations in terms of bridges layout, lanes, ramps, etc. Therefore it will be possible to set-up different training scenarios for the operations that occur on board of ships. Since the Ship Simulator will be part of an interoperable architecture that will allow the cooperative training of different operators (i.e. parkers and drivers), the visualizations system will allow changing the viewpoint and therefore different operators can see the ship interiors from different perspectives as it happens during operations on board a ship in a real car terminal. Furthermore user interfaces will be based on Man in the Loop (MIL) solutions that will be integrated within the simulator; for instance basic parameters such as operators' viewpoints, type of display etc will be controlled by a computerized console and some hardware devices. Figure 3 shows the cars parked inside a ro-ro ship as they appears within the Ship Simulator.

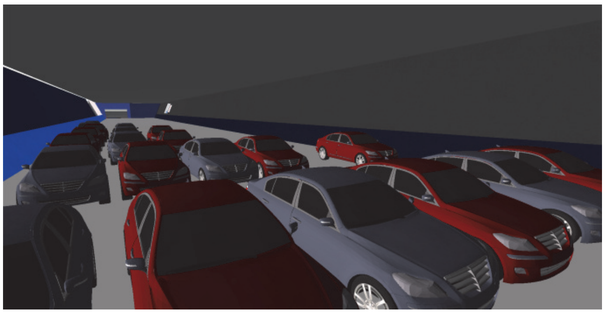


Figure 3: Cars parked inside a Ro-Ro ship

The Operator Simulator will recreate the main tasks of marshalls, tally men and quality checkers. The scope of this simulator is twofold: it can be used to train operators different from drivers and it can also be used to train drivers in being acquainted with the meaning of parkers' gestures (or in general to interact correctly with the other operators). As in the ship simulator, operators will have the possibility to change their point of view based on the needs that arise from the contingent situation they are dealing with. In this way they can act as it happens in the real system and can provide the drivers with accurate instructions. Moreover, even in this case, user interfaces will be based on advanced hardware and MIL solutions (i.e. the avatar of the operator in the virtual environment can be controlled through motion controllers, the virtual environment can be seen through head mounted display). The figure 4 shows the view of an avatar while interacting with a car that is approaching the ship, while figure 5 depicts the real operator controlling the avatar through a motion controller and seeing the virtual environment through a mounted head display.

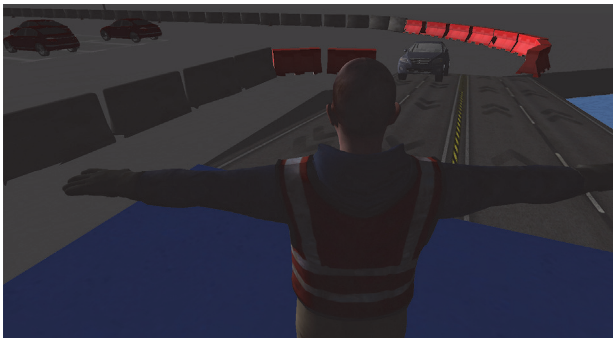


Figure 4: an avatar interacting with a car approaching the ship

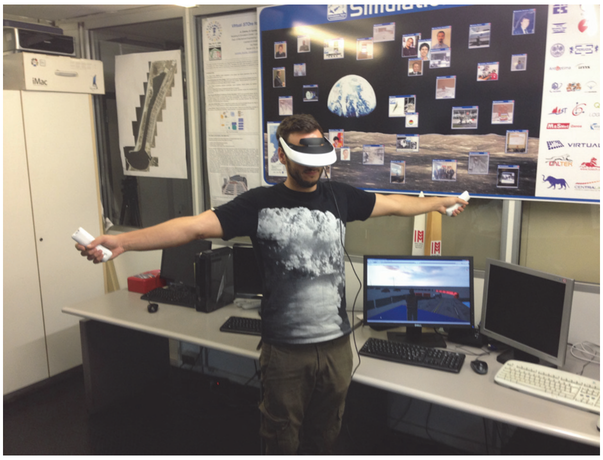


Figure 5: the real operator controlling the avatar through a motion controller and experiencing the virtual environment through a mounted head display

4.1. Performance Measure description

The main objective of any training simulator is to raise the level of personnel qualification as a function of time that elapses from the moment the training is started. One of the main problems in car terminals is related to the time required for an operator to be considered an "expert operator". The main recommendation coming from the navigation lines in ro-ro sector and from the automobile manufacturers suggest 1 year of work as estimated average time to reach an acceptable level of qualification. Indeed the estimated time for an operator to experience at least two times all the possible driving scenarios in a car terminal is 2 years. As already highlighted complex driving situations are characterized by:

- Concurrent operations involving multiple ships and simultaneous loading/unloading operations;
- Simultaneity of operations on ships, trains and yard;

- Simultaneity of operations on the same ship (concurrent vehicles embarkation and disembarkation);
- Operations during the night or during adverse weather conditions.

Therefore, a higher level of qualification obtained in a shorter time would have a direct impact on the following performance measures

- increase of the operators productivity in terms of number of handled vehicles per day;
- reduction of the risk of accident
- reduction of the number of collisions;
- reduction of total number of major damage (total loss) and micro damage with consequent reduction of all direct costs;
- reduction of insurance costs;
- optimization of human resources in terms of operators flexibility in carrying out different types of operations.

5. THE VEHICLE SIMULATOR

As part of the CTSIM general architecture, the vehicle simulator has been already developed (even if additional research activities are still on-going trying to improve some aspects of the simulator). Figure 6 shows a panoramic view of the car terminal area.

The Vehicle Simulator allows recreating the standard operations carried out by drivers while moving vehicles within a car terminal. In particular, the Vehicle Simulator includes three different types of cars (small, medium and large) and a generic model of bus. Figure 7 shows developing and testing activities MSC-LES lab, University of Calabria.

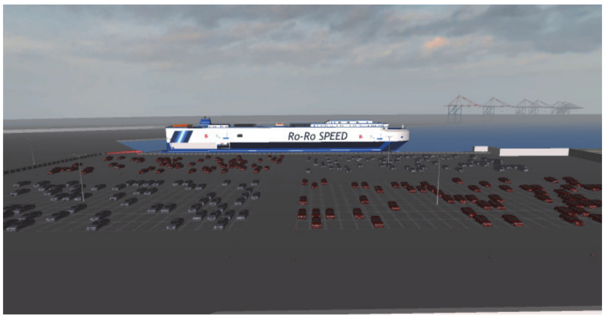


Figure 6: Panoramic view of the car terminal

The Vehicle Simulator can be controlled by specific hardware interfaces (i.e. Steering wheel, pedals, dashboard, etc..) and could be equipped with a 6 Degree of Freedom motion platform. The Vehicle Simulator includes a user interface for the operator based on computerized console. These interfaces (MIL and HIL) allow the handling of the vehicle in accordance to the inputs provided by the driver, creating at the same time, the dynamic behaviour of the real vehicle.

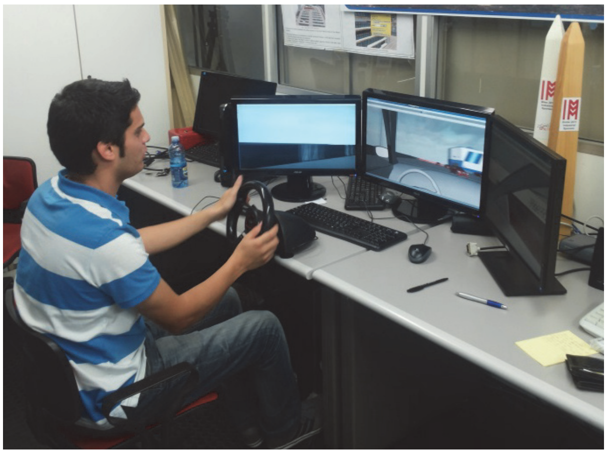


Figure 7: developing and testing the vehicle simulator at MSC-LES, University of Calabria

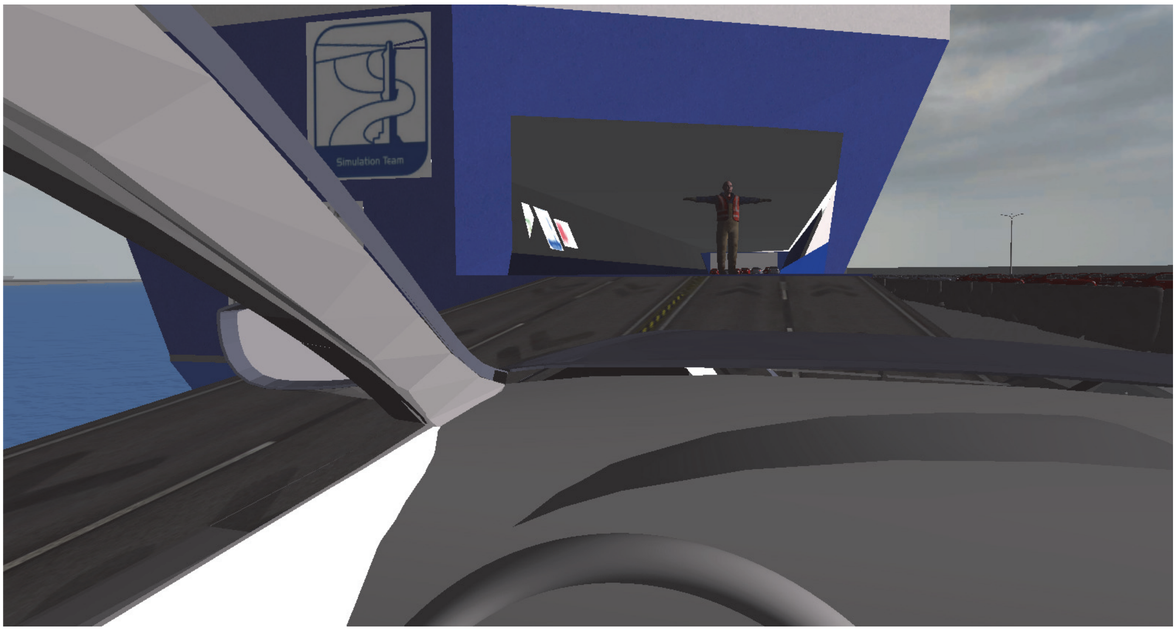


Figure 8: internal view from the car while approaching the ramp for entering the ship

The simulator can run on an immersive visualization system based on multiple screens and an integrated sound system in order to guarantee the view of the external port environment and the feeling of being in the real port. The Vehicle Simulator is also equipped with multiple views that allow the visualization of the vehicle in the virtual car terminal. The figure 8 shows an internal view from the car while approaching the ramp for entering the ship (note the presence of the avatar controlled by the real operator through the Operator Simulator).

Through parameters setting, it is possible to change the yard scenario (i.e. number and types of vehicles parked in the yard). Indeed the Simulator engine includes a method that, once selected the vehicles types, the number of each vehicle type and the parking requirements, it fills the yard with the appropriate number of vehicles arranging them in a casual order (therefore a huge number of possible scenarios are possible) or according to a specific sequence provided by the user. The method loads only one geometric model for each vehicle and it replicates this model to render all the cars. This approach allows the trainer to set the parking conditions just modifying very few parameters and, at the same time, permits the minimization of the GPU workload.

Figure 9 shows one of the possible results of the automated procedure for filling the yard area

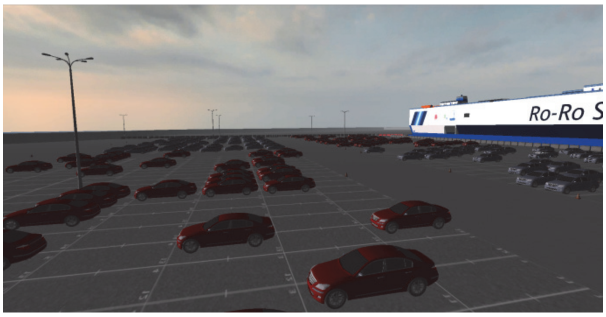


Figure 9: Results of the automated procedure for filling the yard area

6. CONCLUSIONS

The paper presents the general architecture of the CTSIM simulation framework devoted to cooperative training of car terminal operators. In the first part, the paper presents a survey of the current state of the art highlighting that there are many works in the area of container terminal operators training by using simulation while a lot can be done for car terminal operators training.

After having identified and described the main operators usually working in a car terminal and the training needs, the general architecture of the CTSIM framework is presented. CTSIM is a modular simulators system composed by three interoperable simulators: an Operator Simulator, a Ship Simulator and a Vehicle Simulator. A description of the three simulator is provided.

There are research activities still ongoing mainly related to the development of the Ship and Operators simulators, while the Vehicle simulator is currently under testing.

REFERENCES

Anon, 1994. *Virtual Reality and Training*. Government Executive, June.

Anon, Virtual Reality in Industrial Training, *Virtual Reality*, Grindelwald, Vol. 5, No. 4, pp. 31-33.

Banks, J., 1998. *Handbook of Simulation*. J. Wiley & Sons: New York.

Bruzzone A.G., Fadda P., Fancello G., D'Errico G., Bocca E., Massei M. (2010). Virtual world and biometrics as strongholds for the development of innovative port interoperable simulators for supporting both training and R&D. *International Journal of Simulation and Process Modeling*, Vol. 6, Issue 1, pp. 89-102.

Bruzzone, A.G., Fadda, P., Fancello, G. Massei, M., Bocca, E., Tremori, A., Tarone, F., D'Errico, G., 2011. Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes. *Simulation*, 87(10), 857-874.

Bruzzone A.G. Longo, F., Nicoletti, L., Bottani, E., Montanari, R., 2012. Simulation, analysis and optimization of container terminal processes. *International Journal of Modeling, Simulation and Scientific Computing*, 3(4), art. No. 1240006.

Bruzzone A.G., Longo F., 2013. 3D simulation as training tool in container terminals: The TRAINSPORT simulator. Journal of Manufacturing systems, 32(1), 85-98.

Campbell, C. H., Knerr, B.W., Lampton, D.R., 2004. *Virtual Environments for Infantry Soldiers: Virtual Environments for Dismounted Soldier Simulation, Training and Mission Rehearsal*. ARMY research INST for the Behavioral and social sciences, Alexandria May

Carraro, G.U., Cortes, M., Edmark, J.T., Ensor, J.R., 1998. The peloton bicycling simulator. *Proceedings of The Third Symposium On Virtual Reality Modeling Language*, pp. 63-70.

Chin-Teng, L., I-Fang C., and Jiann-Yaw L., 2001. Multipurpose Virtual-Reality-Based Motion Simulator. *Proceedings of the IEEE International Conference on Systems*, Man and Cybernetics, Vol. 5, pp. 2846-2851.

Cimino A., Longo F., Mirabelli G. (2010). Operators training in container terminals by using advanced 3d simulation. In: *Proceedings of the Summer Computer Simulation Conference*. Ottawa, Canada, July 11-15, SAN DIEGO: SCS

Cosby N., Severinghaus R., 2004. Critical Needs for Future Defense Simulations Capabilities Needed for M&S Users. *Proceedings of European Simulation Interoperability Workshop*, Edinburgh, Scotland, 28 June - 1 July

Cramer, J., Kearney, J., and Papelis, Y., 2000. Driving Simulation: Challenges for VR Technology. *IEEE Computer Graphics and Applications*, Vol. 16, No. 12, pp. 1966-1984.

D.C. Mattfeld, H. Kopfer,2002. *Terminal Operations Management In Vehicle Transshipment*

Daqaq Mohammed F., 2003. *Virtual Reality Simulation of Ships and Ship-Mounted Cranes*. Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University.

Ferrazzin, D., Salsedo, F., Bergamasco, M., 1999. The MORIS simulator. *Eighth IEEE International Workshop on Robot and Human Interaction (RO-MAN '99)*, pp. 135–141.

Freund, E., Rossman, J., and Thorsten, H., 2001. Virtual Reality Technologies for the Realistic Simulation of Excavators and Construction Machines: From VR-Training Simulators to Telepresence Systems. *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 358-367.

Furness Z., Tyler J., 2001. Fully Automated Simulation Forces (FAFs): A Grand Challenge for Military Training. *Proceedings of European Simulation Interoperability Workshop*, Stockholm, Sweden, June

Greenberg, J.A., Park, T.J., 1994. Driving Simulation at Ford. *Automotive Engineering*, pp. 37–40.

Henry Lau · Leith Chan · Rocky Wong *A Virtual Container Terminal Simulator For The Design Of Terminal Operation*

Huang, J.-Y., 2003. Modelling and designing a low-cost high-fidelity mobile crane simulator. *International Journal of Human-Computer Studies*, 58(2):151–176.

Ignacio García-Fernandez, Marta Pla-Castells, Miguel A. Gamón And Rafael J. Martínez-Durá Usim: *A Harbor Cranes Training System*

J.C. Quaresma Dias , J.M.F. Calado , M.C. Mendonça, (2007) *The Role Of European «Ro-Ro» Port Terminals In The Automotive Supply Chain Management*

Jiing-Yih, L., Ji-Liang, D., Jiun-Ren H., Ming-Chang, J., and Chung-Yun G., 1997. Development of a Virtual Simulation System for Crane-Operating Training. *Proceedings of ASME*, Paper No. 6p 97-AA-45.

Kallmeier V., Henderson S., McGuinness B., Tuson P., Harper R., Price S. Storr J., 2001. Towards Better Knowledge: A Fusion of Information, Technology, and Human Aspects of Command and Control. *Journal of Battlefield Technology*, Volume 4 Number 1.

Kim, G., 2005. Designing Virtual Reality Systems: *The Structured Approach*. Springer.

Kwon, D.S., et al., 2001. KAIST interactive bicycle simulator. *IEEE International Conference on Robotics and Automation (ICRA)*, Vol. 3, pp. 2313-2318.

Lee, W.S., Kim, J.H., Cho, J.H., 1998. A driving simulator as a virtual reality tool. *IEEE International Conference on Robotics and Automation* 1, 71–76.

Lindheim, R., Swartout, W., 2001. Forging a new simulation technology at the ICT. *IEEE Computer* 34 (1), 72–79.

Longo F. (2007). Students training: integrated models for simulating a container terminal. In: *Proceedings of the International Mediterranean Modelling Multiconference (European Modeling & Simulation Symposium)*. Bergeggi, Italy, October 4-6, GENOA: vol. I, p. 348-355.

Longo F., Mirabelli G, Bocca E., Briano E., Brandolini M. (2006). Container Terminal Scenarios Analysis And Awareness Trough Modeling & Simulation. In: *Proceedomgs of the European Conference on Modeling and Simulation*. Bonn, Germany, May 28th – 31st, vol. I, p. 619-624.

Longo, F. 2010. Design And Integration Of The Containers Inspection Activities In The Container Terminal Operations. *International Journal of Production Economics*, vol. 125(2); p. 272-283.

Maged Elazony, Ahmed Khalifa And Mohamed Alsaied *Design And Implementation Of A Port Simulator Using Formal Graphical Approach (Fga)*

Melnyk, R., 1999. Flight simulators: a look at Linux in the Aerospace Training Industry. *Linux Journal 64*, Article No. 5. Available online at http://www.linuxjournal.com/

Menendez, R.G., Bernard, J.E., 2000. Flight simulation in synthetic environments. *IEEE 19th Proceedings of the Digital Avionics Systems Conferences*, Vol. 1, pp. 2A5/1-2A5/6.

Merkuriev Y., Bruzzone A.G., Novitsky L., 1998. Modelling and Simulation within a Maritime Environment. *SCS Europe*, Ghent, Belgium, ISBN 1-56555-132-X

Morrison, J. E., & Hammon, C. (2000). *On Measuring The Effectiveness Of Large-Scale Training Simulations* (Ida Paper P-3570). Alexandria, Va: Institute For Defense Analysis. (Dtic No. Ada394491).

OPTIMUS Project: *State of the Art Survey on past experiences and on operational port professions* (2009). WP1.1, available online at http://www.optimus-project.eu/

Park, M.K., et al., 2001. Development of the PNU vehicle driving simulator and its performance evaluation. *IEEE International Conference on Robotics and Automation (ICRA),* Vol. 3, pp. 2325-2330.

Piera M.A., Narciso M., Guasch A., Riera D., 2004. Optimization of logistic and manufacturing systems through simulation: a colored Petri net-based methodology. *Simulation,* 80(3), 121-129.

Ray D.P.,2005. Every Soldier Is a Sensor (ES2*)* Simulation: Virtual Simulation Using Game Technology. *Military Intelligence Professional Bullettin.*

Roger D. Smith: Simulation Article, *Encyclopedia Of Computer Science*, Nature Publishing Group, Isbn 0-333-77879-0).

Rouvinen, A., 2005. Container gantry crane simulator for operator training. In Publishing, P. E., editor, Proceedings of the Institution of Mechanical Engineers, Part K: *Journal of Multi-body Dynamics*, volume 219, pages 325–336.

Seròn, F., Lozano, M., Martìnez, R., P´erez, M., Vegara, P., Casillas, J., Martìn, G., Fernàndez, M., Pelechano, J., Brazàlez, A., and Busturia, J., 1999. Simulador de gruas portico portuarias*. In Congreso Espanol de Informàtica Gràfica (CEIG'99).*

Signorile R., Bruzzone A.G., 2003. Harbour Management using Simulation and Genetic Algorithms. *Port Technology International*, Vol.19, pp163-164, ISSN 1358 1759.

Stretton M.L., Hockensmith T.A., Burns J.J., 2002. Using Computer Generated Forces in an Objective-Based Training Environment. *Proceedings of the 11th Conference on Computer Generated Forces and Behavioral Representation*, Orlando, Florida, 7 - 9 May

T. Fischer , H. Gehring, 2004. *Planning Vehicle Transhipment In A Seaport Automobile Terminal Using A Multi-Agent System*

Tam, E.K., et al., 1998. A low-cost PC-oriented virtual environment for operator training. *IEEE Transactions on Power Systems* 13 (3), 829–835.

Torsten Fischer And Hermann Gehring Business Process Support In A Seaport Automobile Terminal – A Multi-Agent Based Approach

Wilson, B., Mourant, R., Li, M., and Xu, W., 1998. A Virtual Environment for Training Overhead Crane Operators: Real-Time Implementation. *IIE Transactions*, Vol. 30, 1998, pp. 589-595.

Yuri Merkuryev, Vladimir Bardatchenko, Andrey Solomennikov, And Fred Kamperman *Simulation Of Logistics Processes At The Baltic Container Terminal: Model Validation And Application*

Zeltzer, D., Pioch, N.J., Aviles, W.A., 1995. Training the officer of the deck. *IEEE Computer Graphics and Applications* 15 (6), 6-9.

MODELLING SWARMS OF UNMANNED AUTONOMOUS FIGHTERS FOR IDENTIFICATION AND ANALYSIS OF NEW STRATEGIES AND TECHNOLOGICAL ENABLERS

Agostino Bruzzone[(a)], Alberto Tremori [(b)]
Simonluca Poggi[(c)], Luciano Dato[(d)], Angelo Ferrando[(e)], Antonio Martella[(f)]

[(a)] [(b)]DIME University of Genoa (ITA)
[(a)]agostino@itim.unige.it, [(b)]tremori@itim.unige.it

[(c)(d)(e)(f)] Simulation Team
[(c)]simonluca.poggi@simulationteam.com, [(d)]luciano.dato@simulationteam.com, [(e)] angelo.ferrando@simulationteam.com
, [(f)] antonio.martella@simulationteam.com

ABSTRACT

This paper presents preliminary results of ongoing researches devoted to use agent based simulation to investigate political, military, economical social infrastructural and informative effects of new underwater tactics based on swarms of autonomous unmanned underwater vehicles attacking cargos and tankers with a single or a group coordinated approach.

The basic idea is to use relatively cheap UAFs (Unmanned Autonomous Fighters) bringing small economic torpedoes or mines to create consistent threats to international commercial trades' routes and to critical infrastructures or to mount a blockade to a hostile country.

To investigate such innovative tactic a discrete event, stochastic agent based simulator has been created and preliminary on going experimentations are summarized.

Keywords: agent based stochastic discrete even simulation, unmanned autonomous vehicles, underwater warfare.

1 INTRODUCTION

During the Second World War two main categories of tactic for submarines were adopted: defensive and offensive. In the first case submarines were used to protect harbors and other critical infrastructure. Offensive tactics can be, as well divided in two main typologies: attack to military or merchant ships. This is the area of interest of the studies described in this paper. With the evolution of submarines with bigger, faster and much more expensive vehicles and torpedoes and with the development of more efficient anti-submarine warfare the usage of modern submarines for attacking merchant ships has almost disappeared. In this paper it is described an innovative tactic based on the usage of unmanned submarine devoted to attack, damage and if possible sink cargos and tankers both sailing alone and on convoy.

This is a preliminary work and the goal is mainly to describe basics of the approach considering historical events that inspired authors. Authors developed a simulator to test hypothesis and concept related both to tactics and vehicles catachrestic and the effectiveness. This work is divided in the following phases:

- Study of historical data about *Rudeltaktik* and current an analysis of economical impacts deriving from such kind of potential threats
- Development of conceptual models of new family of economical underwater unmanned vehicles, torpedoes or mine and related tactics
- Description of the stochastic discrete event simulator developed to provide preliminary experimental results.

2 HISTORY AND PRESENT OF UNDERWATER WARFARE

The submarine warfare in past conflicts introduced new aspects in Naval Warfare and represented a strategic issue; the use of submarine indeed introduce many new aspects including technical, operational, ethical, diplomatic etc.

In this chapter are summarized historical events and tactics used adopted during World War II, when submarine warfare was split into two main geographical theaters areas - the Atlantic and the Pacific. In the Pacific theater US submarines represented a minority of the US fleet, nevertheless sank more than 30 percent of Japan's navy, including eight aircraft carriers. Furthermore US submarine fleet contributed to weaken Japanese economy by sinking almost 5 million tons of ships, about 60 percent of the Japanese merchant marine. Japanese submarines were initially successful, destroying two U.S. fleet aircraft carriers, a cruiser, and several other ships but on the long term they proved to be ineffectual due to a doctrine that concentrated attacks on warships instead on weaker merchant ships.

US submarine tactics were mainly based on lonely ambushes to Japanese cargos, with a maximum of three submarines that operated in the *coordinated attack group.*

Different was the approach of the Kriegsmarine (German Navy) on the Atlantic *the Battle of the Atlantic*. This is considered the longest continuous military campaign in Second World War. Being an island, United Kingdom was highly dependent on imported goods. Britain required more than a million tons of imported (about one million tons per week) to survive and fight. In essence, the Battle of the Atlantic was a tonnage war where Allied struggle to supply Britain and the Germany and Italy attempted to stem merchant shipping flows.

This campaign was based on convoy and submarine warfares based on the so-called Wolf Pack or "Rudeltaktik" that was a tactic used by Kriegsmarine (German Navy), and made famous by Karl Donitz. The Wolf Pack had the goal to defeat the convoy system created by Alleys during the First World War and then extensively re adopted to bring goods across the Atlantic Ocean to support UK and was to have a devastating impact on allied shipping.

U-boat movements requested a high level of coordination by the Befehlshaber der Unterseeboote (BdU; English translation: "Commander of Submarines"). The idea was to form a pack of U-boats (Rudel) and to wait for other U-boats until all boats were in position to conduct a massed attack, so a Rudel consisted of as many U-boats as could reach the scene of the attack with strong possibility to overwhelm the escorts. The first boat to make contact was the "shadower" – whose job was to remain invisible, to maintain contact and to report the convoy's position back to BdU. When enough boats have converged with the convoy, BdU would give order to attack, usually in the night.

This had devastating effects and attacking in groups easily overwhelmed escort ships. When a destroyer detected and attacked one U-boat, another would attack at a different location, bringing confusion and chaos.

With the exception of orders and coordination given by the BdU, when the attack started U-Boat commanders were free to chose the best tactic and the attacked as they saw fit.

Once Wolf Pack operations began U-boats inflicted heavy losses until the allies developed new technology to counter the threat. One of the most famous Wolf Pack attacks took place between the nights from October 16th to the 19th, 1940. Convoy SC7 was attacked by a pack of seven boats, sinking 20 of 34 ships in the convoy. The next night, convoy HX79 was attacked with 14 ships further losses, making a total of 34 ships in 48 hours.

Allies developed countermeasures to turn the U-boat organization against itself. Most notably was the fact that wolf packs required extensive radio communication to coordinate the attacks. This left the U-boats vulnerable to a device called the High Frequency Direction Finder (HF/DF or "Huff-Duff") which allowed Allied naval forces to determine the location of the enemy boats transmitting and attack them. Also, effective air cover, both long-range planes with radar, and escort carriers and blimps, allowed U-boats to be spotted as they shadowed a convoy (waiting for the cover of night to attack).

Wolf packs fell out of use after the Second World War, modern submarines have better weapons and underwater speed than those of Second World War. These boats could operate faster, deeper and had much longer endurance. They could be larger and so became missile launching platforms. Furthermore the cost of each single submarine and modern torpedoes respect the value of a single merchant ships and the capability for escorts to attack and sink it by helicopters has made not economically convenient to attack cargos or tankers. However, the importance of the submarine has shifted to an even more strategic role than the disruption of merchant shipping, with the advent of the nuclear submarine carrying nuclear weapons to provide second-strike capability.

3 ECONOMIC SCENARIOS

It is still pretty controversial the real impact of the U-boat Arm and if it came close to winning the Battle of the Atlantic with the Allies almost defeated, and Britain close to starvation it. At no time during the campaign were supply lines to Britain interrupted and the German never succeeded in mounting a comprehensive blockade.

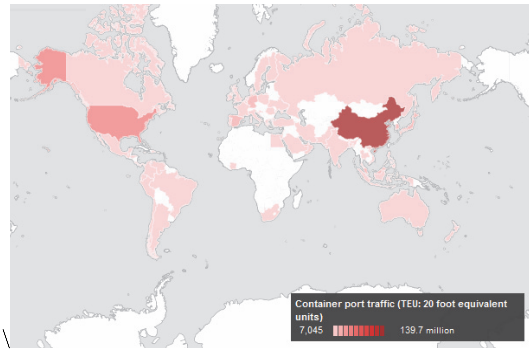
In only four out of the first 27 months of the war did Germany achieve the target of 350,000 tons that was calculated as necessary to isolate Great Britan, while after December 1941, when Britain was joined by the U.S. merchant marine and ship yards the target effectively doubled. Such target was achieved in only one month, November 1942.

By the end of the war, although the U-boat arm had sunk 6,000 ships totalling 21 million grt, the Allies had built over 38 million tons of new shipping.

These are figures of the Battle of Atlantic and the impact of Wolf Pack tactic. Nevertheless, today economical mechanisms that are ruling our world are much more complex and even fragile respect the situation dramatic situation of the Second World War. Furthermore the scenarios we are envisioning and studying are completely different from a global war scenario. It must also be considered that with the US entrance in War forces inequality bocame considerable. The ideas presented in this paper are focusing on a more symmetric scenario with two countries in war. It could be also applied to a completely different scenario with terrorist organization wit the goal to threat global trades but this will be studied in further analysis.

Considering global economical situation we have to focus on some figure related to extended supply chain and maritime transportation. Delocalization of production sites, continuous request of raw materials import and export flows that are continuously changing based on a high dynamic richness allocation are some of the most relevant elements of an extended logistics process that is more and more complex.

In figure 1 and Table 1 are summarized statistics of ports container traffic. These figures measure the flow of containers from land to sea transport nodes, and vice versa, in Milion of twenty-foot equivalent units (TEUs).



\

Figure 1: geographical distribution of container traffic

Data refer to coastal shipping as well as international journeys. Transshipment traffic is counted as two lifts at the intermediate port (once to off-load and again as an outbound lift) and includes empty units. (Years 2006-2011 Source Word Bank).

Table 1: container traffic by nations

Country	2006	2007	2008	2009	2010	2011
China	**85**	**104**	**116**	**109**	**130**	**140**
India	**6**	**7**	**7**	**8**	**10**	**10**
Italy	**10**	**11**	**11**	**10**	**10**	**10**
Japan	**18**	**19**	**20**	**16**	**18**	**19**
Netherlands	**10**	**11**	**11**	**10**	**11**	**12**
UK	**8**	**9**	**8**	**8**	**9**	**9**
USA	**41**	**45**	**42**	**37**	**42**	**43**

In Figure 2 are described shipping routes and it illustrates the relative density of commercial shipping in the world's oceans. For instance in 2011 about 39 million TEUs passed through the Suez Canal.

Table 1 shows that, despite last years international crisis, maritime traffic trend is still positive with a growing number of TEUs shipped from and to the three main areas: China and Far East, Europe and USA as shown in Figure 1.

There are several countries or areas where a naval blockade based on the submarine tactics could be effective. In the Middle East there are countries higly dependant from sea traffic with few ports located in areas that could be easily interdicted. Also in Far East and South Asia there area with a growing trend of more than +150% from 2006 to 2011 in container flows but with coasts, and of course ports located in areas that could be easily controlled by hostile countries. The world is plenty of countries with such conditions and complex international political relations and looking at Figure 2 we could easy imagine several critical areas: in fact such figure depicts the level of concentration of routes in some "hot spots" which represent, with the areas around the main ports in the world, a set of ideal areas for any kind of military or terroristic operations aiming to create threats to international trades and global economy or targeting a specific country as well.

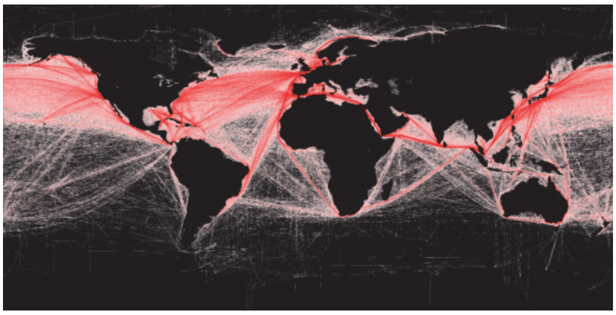


Figure 2: shipping routes and relative density of commercial shipping

4 TACTICS AND CONCEPTUAL MODELS OF UAFS

In the last paragraph of previous chapter elements summarizing some basic concepts underling potential success factors of a submarine tactic based on unmanned vehicles. Indeed a stealth flotilla or swarm of UAFs (Unmanned Autonomous Fighters) bringing torpedoes could create such a threat to international trades able to mount a comprehensive blockade of an enemy country; the models proposed in this paper are devoted to address both the analysis of possible strategies and tactics as well as to identify the technological enablers and requirements that are most sensitive to the impact of using these innovative systems; these kind of analysis are addressing a very complex mission environment affected by many stochastic factors and interactions, so it emerges evident the necessity to use modeling and simulation and the potential of this approach to identify drivers and critical elements of the whole problem.

In this chapter are described the hypothesis considered to define UAFs and are also described tactics and variables to be investigated of this submarine warfare; obviously the use of such autonomous systems will affect many sectors, including regulations as well as diplomacy and are expected to introduce complex operational modes, caveats and strategic decision making related to the rule of engagement, tactics and modes to be applied during the operations.

The UAF should copy some basic elements characterizing Second World War submarines, respect modern ones. They should be relatively cheaper vehicles bringing cheaper torpedoes and these elements should make economically convenient attacking cargos and tankers. But respect II World War submarines UAF are unmanned so smaller, with more autonomy and, consequently more difficult to be detected by anti-submarines systems and ships. Furthermore no men on board means, of course, no cost (direct and indirect) related to human losses.

Such kind of vehicles and the tactics under evaluation could be used both by country, with no capability to deploy a navy fleet or a terrorist organization.

Since our aim is to design an underwater unit having at least the same operational skills of the German U-Boot during the II World War, made more efficient thanks to the modern technologies and weapon systems, in order to represent a real threat for cargo or, based on coordination and swarm artificial intelligence algorithms, even a merchant convoy, itself composed by naval units built with advanced construction techniques, we have to consider different aspects.

First of all the autonomous propulsion underwater vehicle, provided with a propeller and thrusters for the horizontal asset, will get power supply from fuel cells batteries and accumulators fed by solar panels placed on the UAF surface. This represents a limit for the maximum operative depth: to use cells also diving it's important not to exceed 10-15 meters of depth, where the sunlight is about the 10% than over the surface.

Since the UAF can operate autonomously , that is without crew on board, it can have very limited dimensions in relation to conventional submarines. Moreover there are other annexed benefits: no acoustic emissions caused by human activities, no heat produced by heating plants for the crew, fewer necessity to ventilate and cool spaces, no more delays in the execution of offensive and defensive maneuvers other than the time necessary to make decision about actions to be taken.

Limited dimensions permit to hold the needed electronic equipment: sensors for the underwater passive detection (hydrophone), radio for sat communication (VHF or UHF), and active sonar (beacon) for underwater transmission, GPS for position identification, echo sounding.

The payload the UAF must have is strongly related to the methods of use. Since the UAF is quite complex and provided with communication and detection systems, we may imagine the simultaneous employ of UAF s with restricted detection capability, placed in proximity of the Command UAF, the only one with payload and with the required propulsive power to keep the speed to silently follow the target. One kilogram of TNT explosive charge brought near the propellers or the rudders could restrict the ship ability to maneuver. Thus exposing it to assaults on the part of other UAF s, let alone the fact that possible supply vessels would expose themselves to further attacks.

There are other mission typologies that could considered: such as surveillance of sea areas to control surface navigation, sea bottom mapping, identification through acoustic tracking, navigation interdiction, communication between air and water taking advantage of sound canalization phenomena, coordination between UAF s.

In other words this is to present a possible evolution of submarine warfare: less expensive units, compared to submarines, able to seriously damage merchant and military navigation. Supposing the presence of autonomous propulsion mines, with low probability to be detected, behind enemy lines, becomes a nightmare for who has to organize a convoy escort and keep navigation safe.

Among the variables and operational modes that could be investigated by the model concerning vehicles characteristics and different tactics it is worth to mention:

- Autonomy, supplying procedures and maintenance politics of UAFs
- How to deploy in and recover from theater of operations UAFs
- Coordination and Communications among UAFs, in particular to use a wolf pack approach to mass assault convoys or critical infrastructures.

In fact invisibility of UAFs is, probably, the main advantage of such vehicles and tactics and all the above mentioned issues are related to this element. Another area that it is worth to be investigated is focused on Command and Control and to the related concept of agility.

5 SIMULATION AND UUVs, STATE OF ART

Unmanned vehicles is an area of great interest in the defense sector with researches from the beginning of the Twentieth Century.

The US Navy began experimenting with radio-controlled aircraft during the 1930s but battlefield Unmanned aerial vehicles (UAV) would not come into their own until the 1980s. For what concerns underwater the first Autonomous Unmanned Vehicle (AUV) was developed at the Applied Physics Laboratory at the University of Washington in 1957. It was named SPURV that stays for “Special Purpose Underwater Research Vehicle”, or SPURV and was used, as name suggests, to study diffusion, acoustic transmission, and submarine wakes.

Typical military missions are to monitor and an area to detect mines or unidentified objects. AUVs are also employed in anti-submarine warfare.

Long-Term Mine Reconnaissance System (LMRS) and its successor, the Mission Reconfigurable UUV (MRUUV) are systems imagined as torpedo tube-launched and tube-recovered underwater search and survey against mines and are supposed to be roughly the size of a 21-inch torpedo. Futuristic concepts, such as Naval Undersea Warfare Center's MANTA vehicle, would approach, or even exceed, the dimensions of today's Advanced SEAL Delivery System (ASDS) – 65 feet long and 55 tons. Vehicles of that dimension could carry a variety of full-scale weapons – conceptually, MANTA could launch heavyweight torpedoes and – depending on future rules of engagement – might even be unleashed to wield lethal force against enemy ships, submarines, and shore installations. Such kind of platforms could represent a sample, or a starting point of a vehicle used to apply tactics described in this paper.

Typical missions for such kind of unmanned vehicles are reconnaissance, Search and Survey and supporting communications/navigation of submarines. Most of these roles are motivated by the unique advantages of underwater stealth and the need to manage risk, but there are many missions in which using unmanned underwater vehicles could complement crewed platforms providing a significant force-multiplier or simply a more cost-effective approach. The concept of cost-effectives is one of the reference point.

Considering the interest of the military, homeland security and industry communities for such kind of platforms there are several samples of researches and patents in this area. An interesting survey from RAND Corporation (Button et al. 2009) describes most promising missions for unmanned undersea vehicles. Another interesting work from Zhu et al. recently published describes how to control a team of AUVs to reach all appointed target locations for only one time on the premise of workload balance and energy sufficiency while guaranteeing the least total and individual consumption. As an application of simulation in this area we can cite a paper that describes the capability, design and application of the generic underwater warfare simulation environment called ODIN. The model was developed by QinetiQ (Robinson T., 2001)

6 SIMULATOR AND PRELIMINARY EXPERIMENTATIONS

Simulation represents the ideal methodology to test a tactic that is completely new with no real experimental data to be used or analyzed. This is the situation of UAF approach, so a simulator has been developed and, in this chapter are summarized the characteristics of the simulator and preliminary experimental results.

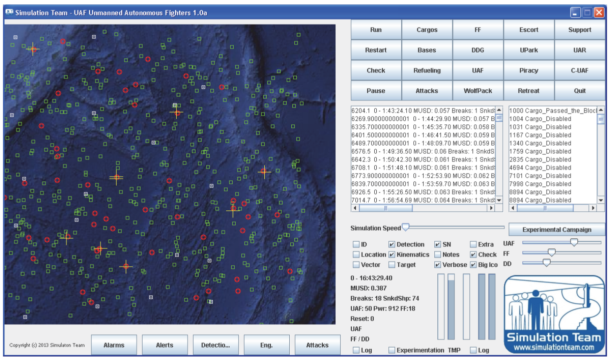
Based on authors' long experience an agent based simulator has been developed. Intelligent agents provide autonomous behavior to the following agents:

- UAFs: for the this experimental scenario the number of UAFs is set to 50; every UAF boards up to a maximum 20 torpedoes
- Military Frigates and Destroyers with helicopters: there are 18 vessels in the area devoted to patrolling and ASW (Anti Submarine Warefare)
- Cargo traffic in the area: there is a daily flow of about 250 cargos per day
- Scenario: it is a square of 1,000 by 1,000 Nautical miles located in Deep Sea

The simulator detects successful attacks (sunken cargos) from UAFS to cargo and from anti-submarine vessels in the area to UAFs. UAFs failures, overall costs and other parameters are also tracked.

UAFS and Frigates/Destroyers operate autonomously and attack their targets based on condition; obviously both need refueling, reloading and support activities based on their use. Cargos navigate in the area following assigned routes.

For this first phase of the analysis a scenario with no convoys has been created, so UAFs operate independently with no particular coordination and C2 with exception to the support, logistics and service operations.



```
6204.1  0 - 1:43:24.10 MUSD: 0.057 Breaks: 1 Snkd
6269.900000000001  0 - 1:44:29.90 MUSD: 0.057 B
6335.700000000001  0 - 1:45:35.70 MUSD: 0.058 B
6401.500000000001  0 - 1:46:41.50 MUSD: 0.059 B
6489.700000000001  0 - 1:48:09.70 MUSD: 0.059 B
6576.5  0 - 1:49:36.50 MUSD: 0.06 Breaks: 1 SnkdS
6642.3  0 - 1:50:42.30 MUSD: 0.061 Breaks: 1 Snkd
6708.1  0 - 1:51:48.10 MUSD: 0.061 Breaks: 1 Snkd
6773.900000000001  0 - 1:52:53.90 MUSD: 0.062 B
6839.700000000001  0 - 1:53:59.70 MUSD: 0.063 B
6926.5  0 - 1:55:26.50 MUSD: 0.063 Breaks: 1 Snkd
7014.7  0 - 1:56:54.69 MUSD: 0.064 Breaks: 1 Snkd
```

```
1000 Cargo_Passed_the_Block
1004 Cargo_Disabled
1031 Cargo_Disabled
1167 Cargo_Disabled
1340 Cargo_Disabled
1759 Cargo_Disabled
2835 Cargo_Disabled
4694 Cargo_Disabled
7101 Cargo_Disabled
7998 Cargo_Disabled
8894 Cargo_Disabled
8894 Cargo_Disabled
```

☐ ID ☑ Detection ☑ SN ☐ Extra
☐ Location ☑ Kinematics ☐ Notes ☑ Check
☐ Vector ☐ Target ☑ Verbose ☑ Big Ico

```
0 - 16:43:29.40
MUSD: 0.387
Breaks: 18 SnkdShp: 74
UAF: 50 Pwr: 912 FF:18
Reset: 0
UAF
FF / DD
```

☐ Log ☐ Experimentation TMP ☐ Log

Figure 3: Simulator GUI

In Figure 4 is represented the Mean Square pure Error (MSpE) analysis to define the simulation duration, that confirms the possibility to obtain after a number of simulation days a pretty reliable estimation of the confidence band for the different target functions.

Preliminary experimentation, mainly to complete the V&V phases, generated some encouraging findings about effectiveness of this approach.

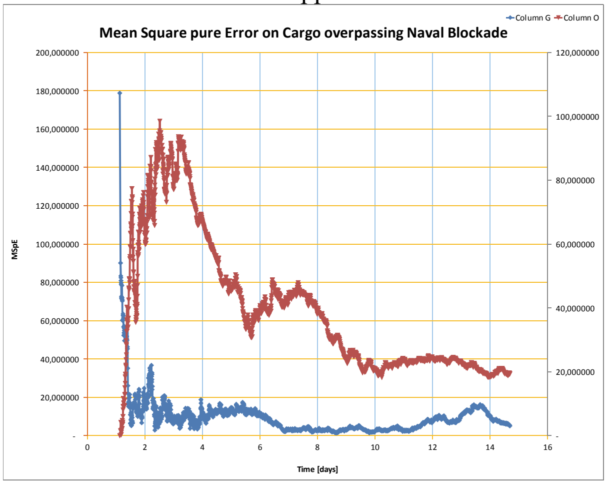


Figure 4: MspE of Cargos passing blockade

In Figure 5 are depicted three observed target functions:

1. Disabled cargos [%]
2. Cargo able to break the Naval Block created by the UAF respect the overall flow [%]
3. Availability of UAFs's weapons respect initial maximum capability [%]

From this figure it is emerging that about 18% of cargos crossing the blockade area are hit and disabled by UAFs' attacks and the UAFs flotilla is able to operate with acceptable losses, but requires proper support and logistics.

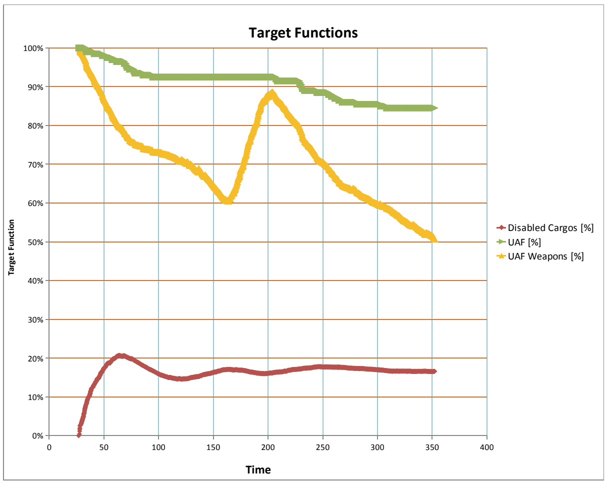


Figure 5: Disabled Cargos and UAFs Fire Power

7 CONCLUSIONS

This paper describes on going researches to define innovative strategies and tactics as well as critical technological aspects related to the development of innovative UAF; the paper address specifically the submarine warfare based on unmanned vehicles and presents some preliminary result from experimentation based on the usage of a agent based discrete event simulator developed ad hoc.

The general idea is to attack merchant routes or mount a blockade to a country or an area by an innovative stealth submarine force based on unmanned vehicles (named Unmanned Autonomous Fighters - UAFs) even if readapt tactics from World Wars. In the analysis and in the development of the conceptual models and scenarios, the authors have considered the economical trade off between results of attacks to cargos and potential losses, considering cost of the vehicles and of mounted weapons.

Preliminary results based on the developed agent based simulator are encouraging about the effectiveness of such approach and provide useful insight about the UAF technical characteristics, requirements and features.

Actually authors are creating an international team of subject matter experts to complete the accreditation of this model and to support future researches and developments. Among these are the study of new tactics with more coordination capabilities, the definition of a detailed plan based on costs and an analysis of direct and indirect impacts of such kind of attacks and further experimentation to define procedures and variables such communications or autonomy of UAFs.

REFERENCES

"Political, military, economic, social, Infrastructure, information (PMESII) Effects forecasting for course of Action (coa) evaluation", AFRL-RI-RS-TR-2009-160 Final Technical Report, June 2009

Benedict, JR (2000) "Future Undersea Warfare Perspectives" Johns Hopkins Apl Technical Digest Volume: 21 Issue: 2 Pages: 269-279

Bocca E., Pierfederici, B.E. (2007) Intelligent agents for moving and operating Computer Generated Forces, Proc. of SCSC, San Diego July

Brooks, John (2007) "Anti-submarine warfare in World War I: British naval aviation and the defeat of the U-boats" Mariners Mirror Volume: 93 Issue: 1 Pages: 115-117, Feb

Brown, Gerald; Kline, Jeff; Thomas, Adam; et al. (2011) "A Game-Theoretic Model for Defense of an Oceanic Bastion Against Submarines", Military Operations Research Volume: 16 Issue: 4 Pages: 25-40

Bruzzone A.G., Cantice G., Morabito G., Mursia A., Sebastiani M., Tremori A. (2009) "CGF for NATO NEC C2 Maturity Model (N2C2M2) Evaluation", Proceedings of I/ITSEC2009, Orlando, November 30-December 4

Bruzzone A.G. Tremori A., Massei M. (2011) "Adding Smart to the Mix", Modeling Simulation & Training: The International Defense Training Journal, 3, 25-27, 2011

Bruzzone A.G., Fadda P, Fancello G., Massei M., Bocca E., Tremori A., Tarone F., D'Errico G. (2011) "Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes", Simulation October 2011, vol. 87 no. 10, p. 857-874, ISSN: 857-874, doi: 10.1177/0037549711418688.

Bruzzone A.G., Massei M. Tremori A., Longo F., Madeo F., Tarone F, (2011) "Maritime Security: Emerging Technologies for Asymmetric Threats", Proceedings of European Modeling and Simulation Symposium, EMSS 2011, Rome, Italy, September 12 -14

Bruzzone A., Tremori A., Longo F., (2012) "Interoperable Simulation for Protecting Port as Critical Infrastructures", Proc. of International Conference on Harbor, Maritime and Multimodal Logistics Modeling & Simulation, HMS 2012, Wien, September 19-21

Bruzzone A.G., Tremori A., Merkuryev Y. (2011) "Asymmetric marine warfare: PANOPEA a piracy simulator for investigating new C2 solutions" Proceedings SCM MEMTS 2011, St.Petersburg June 29-30

Button R. Kamp J. Curtin T.; Dryden J. (2009) "A Survey of Missions for Unmanned Undersea Vehicles), RAND National Defense Research Inst Santa Monica CA ADA503362

Changjun Pan, Guo Yingqing (2013), Design and simulation of ex-range gliding wing of high altitude air-launched autonomous underwater vehicles based on SIMULINK, Chinese Journal Of Aeronautics Volume: 26 Issue: 2 Pages: 319-325

Chu, PC; Perry, MD; Gottshall, EL; et al. (2004) "Satellite Data Assimilation for Improvement of Naval Undersea Capability" Marine Technology Society Journal Volume: 38 Issue: 1 Pages: 12-23

Coy, Peter (2008)"Anti-Submarine Warfare", Mariners Mirror Volume: 94 Issue: 4 Pages: 496-497

Dingman, Roger (2010) "Execute Against Japan: The US Decision to Conduct Unrestricted Submarine Warfare" Pacific Historical Review Volume: 79 Issue: 2 Pages: 308-309

Dönitz K., (1958) Memoirs: Ten Years and Twenty Days, World Publishing Company, NYC

Dorwart, JM (1992) "Diplomatic Ramifications of Unrestricted Submarine Warfare, 1939-1941", Journal of American History Volume: 78 Issue: 4 Pages: 1505-1505

Franklin, George (2006) "The Royal Navy and anti-submarine warfare, 1917-1949", Mariners Mirror Volume: 92 Issue: 4 Pages: 520-522 Nov 2006

Gardner,WJR, Channon, RF (1997) "Anti-Submarine Warfare", Mariners Mirror Volume: 83 Issue: 1 Pages: 120-121

Guo, Rui; Zhao, Xiaozhe; Yu, Hayang (2006) "One application case of artificial life approach in the simulation of surface ship's anti-submarine warfare" Dynamics Of Continuous Discrete And Impulsive Systems-Series B-Applications & Algorithms Volume: 13 pp. 610-614

Hamilton, Michael J.; Kemna, Stephanie; Hughes, David (2010) "Antisubmarine Warfare Applications for Autonomous Underwater Vehicles: The GLINT09 Sea Trial Results", Journal Of Field Robotics Volume: 27 Issue: 6

Higgins, TM; Turriff, AE; Patrone, DM (2002) Simulation-based undersea warfare assessment, Johns Hopkins Apl Technical Digest Volume: 23 Issue: 4 Pages: 396-402

Horcicka, Vaclav (2012) "Austria-Hungary, Unrestricted Submarine Warfare, and the United States' Entrance into the First World War", International History Review Volume: 34 Issue: 2 Pages: 245-269

Jethi, SC; Jain, RK (2004) "Simulation of naval wargames" Defence Science Journal Volume: 54 Issue: 3 Pages: 407-415

Maas P., (1999) The Terrible Hours: The Man Behind the Greatest Submarine Rescue in History, HarperCollins New York

Massei, M., Tremori, A., Poggi, S. and Nicoletti, L. (2013) 'HLA-based real time distributed simulation of a marine port for training purposes', Int. J. Simulation and Process Modelling, Vol. 8, No. 1, pp.42–51

Milner, Marc (2006) "From Nelsonic to Newtonian: The development of anti-submarine warfare in the North Atlantic 1939-45" Mariners Mirror Volume: 92 Issue: 4 Pages: 465-476

Morgan, JG (1998) "Networking ASW systems: Anti-submarine warfare dominance - The direction for US Navy operational primacy - To remain the most formidable ASW force in the world" Sea Technology Volume: 39 Issue: 11 Pages: 19-22

Potter E. B. and Chester W. Nimitz, 1960; Sea Power: A Naval History, N.J.: Prentice-Hall, Englewood Cliffs

Robinson, T. (2001) "ODIN - an underwater warfare simulation environment ", Simulation Conference, 2001. Proceedings of the Winter (Volume:1) ISBN 0-7803-7307-3, Arlington (VA) 09 Dec 2001-12 Dec 2001

Sea Technology, (2010) "Anti-Submarine Warfare", Sea Technology Volume: 51 Issue: 10 Pages: 31-32 Published: Oct 2010

Steffen D. (2004)"The Holtzendorff memorandum of 22 December 1916 and Germany's declaration of unrestricted U-boat warfare", Journal of Military History Volume: 68 Issue: 1 Pages: 215-224

Sturma, Michael (2009) "Atrocities, Conscience, and Unrestricted Warfare US Submarines during the Second World War", War In History Volume: 16 Issue: 4 Pages: 447-468

Todd, Jonathan (2008) "The Royal Navy and anti-submarine warfare, 1917-1949", Journal of Strategic Studies Volume: 31 Issue: 4 Pages: 665-667

Tremori A., Massei M., Madeo F., Reverberi A., (under publication), "Interoperable simulation for asymmetric threats in maritime scenarios: a case based on virtual simulation and intelligent agents" International Journal of Simulation and Process Modelling (IJSPM).

van Vossen, Robbert; Beerens, Peter; van der Spek, Ernest (2011) "Anti-Submarine Warfare With Continuousiy Active Sonar TNO Tests the Principle of Continuously Active Sonar With the Interim Removable Low-Frequency Active Sonar System", Sea Technology Volume: 52 Issue: 11 Pages: 33-35

Vigliotti, V. (1998) "Demonstration of Submarine Control of an Unmanned Aerial Vehicle" Johns Hopkins Apl Technical Digest Volume: 19 Issue: 4 Pages: 501-512

Volkert, Richard; Jackson, Carly; Whitfield, Cecil (2010) "Development of Modular Mission Packages Providing Focused Warfighting Capability for the Littoral Combat Ship", Naval Engineers Journal Volume: 122 Issue: 4 Pages: 75-92

Zhu DQ; Huang H.; Yang, SX (2013) "Dynamic Task Assignment and Path Planning of Multi-AUV System Based on an Improved Self-Organizing Map and Velocity Synthesis Method in Three-Dimensional Underwater Workspace", IEEE TRANSACTIONS ON CYBERNETICS Volume: 43 Issue: 2 Pages: 504-514 DOI: 10.1109/TSMCB.2012.2210212, Published: APR 2013

Author's Index