THE 29TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 18 – 20, 2017
BARCELONA, SPAIN

EMSS
2017

EDITED BY
MICHAEL AFFENZELLER
AGOSTINO G. BRUZZONE
EMILIO JIMÉNEZ
FRANCESCO LONGO
MIQUEL ANGEL PIERA

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2017

ISBN 978-88-97999-93-5 (Paperback)
ISBN 978-88-97999-85-0 (PDF)

ISBN 978-88-97999-93-5 (Paperback)
ISBN 978-88-97999-85-0 (PDF)

THE 29TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 18 – 20, 2017

BARCELONA, SPAIN

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

AUTONOMOUS UNIVERSITY OF BARCELONA

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

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

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

University of Bordeaux

University of Cyprus

Dutch Benelux Simulation Society

I3M 2017 Industrial Sponsors

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2017 Media Partners

Inderscience Publishers – International Journal of Simulation and Process Modeling

Inderscience Publishers – International Journal of Oil, Gas and Coal Technology

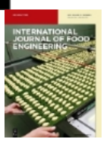
Inderscience Publishers – International Journal of Service and Computing Oriented Manufacturing

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

SAGE
Simulation Transaction of SCS



De Gruyter
International Journal of Food Engineering

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

EDITORS

MICHAEL AFFENZELLER
UPPER AUSTRIA UNIVERSITY OF APPLIED SCIENCES, AUSTRIA
Michael.Affenzeller@fh-hagenberg.at

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

MIQUEL ANGEL PIERA
AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN
MiquelAngel.Piera@uab.cat

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2017

GENERAL CO-CHAIRS

AGOSTINO BRUZZONE, *MITIM DIME, UNIVERSITY OF GENOA, ITALY*
MIQUEL ANGEL PIERA, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*

PROGRAM CO-CHAIRS

FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
YURY MERKURYEV, *RIGA TECHNICAL UNIVERSITY, LATVIA*

THE 29TH EUROPEAN MODELING & SIMULATION SYMPOSIUM, EMSS 2017

GENERAL CO-CHAIR

FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA ITALY*

PROGRAM CO-CHAIR

MICHAEL AFFENZELLER, *UPPER AUSTRIA UNIVERSITY OF APPLIED SCIENCES, AUSTRIA*

EMSS 2017 INTERNATIONAL PROGRAM COMMITTEE

TRACKS AND WORKSHOP CHAIRS

ADVANCED IN DISCRETE AND COMBINED SIMULATION
CHAIRS: 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, MSC-LES, UNIVERSITY OF CALABRIA, ITALY

AGENT DIRECTED SIMULATION (ADS)
CHAIRS: TUNCER ÖREN, UNIVERSITY OF OTTAWA, CANADA; YU ZHANG, TRINITY UNIVERSITY (USA)

PETRI NETS BASED MODELLING & SIMULATION
CHAIRS: EMILIO JIMÉNEZ, UNIVERSITY OF LA RIOJA, SPAIN; JUAN IGNACIO LATORRE, PUBLIC UNIVERSITY OF NAVARRE, SPAIN

WORKSHOP ON CLOUD MANUFACTURING
CHAIRS: PROF. LIN ZHANG, BEIHANG UNIVERSITY, CHINA; PROF. FEI TAO, BEIHANG UNIVERSITY, CHINA; SIMONLUCA POGGI, MAST SRL, ITALY

SIMULATION OPTIMIZATION APPROACHES IN INDUSTRY, AVIATION, SERVICES AND TRANSPORT
CHAIRS: IDALIA FLORES, UNIVERSITY OF MEXICO; MIGUEL MÚJICA MOTA, AMSTERDAM UNIVERSITY OF APPLIED SCIENCES, NETHERLANDS

HUMAN-CENTRED AND HUMAN-FOCUSED MODELLING AND SIMULATION
CHAIRS: AGOSTINO BRUZZONE, DIME, UNIVERSITY OF GENOA, ITALY; PEER OLAF SIEBERS, UNIVERSITY OF NOTTINGHAM, UK

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
CHAIR: 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

RETAIL-ORIENTED SIMULATION
CHAIR: ELEONORA PANTANO, MIDDLESEX UNIVERSITY, UK

COMPUTER SIMULATION FOR SECURITY
CHAIRS: EDGAR ALONSO LOPEZ-ROJAS, BLEKINGE INSTITUTE OF TECHNOLOGY, SWEDEN, JONI A. AMORIM, UNIVERSITY OF SKÖVDE SWEDEN; RICHARD ZALUSKI, CENTER FOR STRATEGIC CYBERSPACE + SECURITY SCIENCE, UK

WORKSHOP ON MODEL ENGINEERING FOR SYSTEM OF SYSTEMS
CHAIRS: LIN ZHANG, BEIHANG UNIVERSITY, CHINA; PROF. BERNARD P. ZEIGLER, RTSYNC CORP, USA; DR. SAURABH MITTAL, DUNIP TECHNOLOGIES, USA

MODELLING, COMPUTER SIMULATION AND VISUALISATION OF HERITAGE
CHAIRS: GALINA MERKURYEVA, RIGA TECHNICAL UNIVERSITY; RICARDO GONÇALVES, UNIVERSIDADE NOVA DE LISBOA (UNINOVA)

MODELING AND SIMULATION IN PHOTONICS SCIENCE AND INDUSTRIAL TECHNOLOGIES

CHAIRS: ALEXANDER S. SIGOV, MOSCOW TECHNOLOGICAL UNIVERSITY (MIREA), (MOSCOW, RUSSIAN FEDERATION); MIKHAIL E. BELKIN, MOSCOW STATE TECHNOLOGICAL UNIVERSITY (MIREA), (MOSCOW, RUSSIAN FEDERATION)

CHAIRS' MESSAGE

WELCOME TO EMSS 2017!

Nowadays, in spite of the financial crises, investments in new simulation applications are becoming increasingly relevant in many sectors recognizing M&S as an enabling technology for break-through innovations. Therefore, M&S plays a key role in many critical areas and is intended to face problems that are even more challenging.

In this perspective, the European Modeling and Simulation Symposium (that this year has reached its 29th edition) presents innovative applications of M&S. From year to year, EMSS provides the ideal framework for presenting the best research results, new emerging research directions, problem solutions and insights in M&S applications allowing the creation of networks of academicians, technicians, subject matter experts and practitioners.

Largely considered as the most important appointment in Europe for M&S professionals, EMSS proposes a multi-disciplinary perspective on research macro-areas such as Education, Engineering, Industry, Simulation Applications, Risk Management, Technology Transition and other emerging applications.

Special thanks go to the authors for their high quality contributions. Their research works are in line with the meaning and the scope of EMSS 2017 and allow giving evidence of real and concrete developments in the field of M&S. Papers presented in EMSS 2017 have been subject to a meticulous review process aiming both at selecting high quality works and at providing the authors with feedbacks on how to further increase the scholarly and scientific relevance of their works. These papers provide evidence on the potentials of the simulation approach applied to a wide spectrum of practical problems.

We wish therefore that EMSS 2017 would offer a fruitful and enjoying experience to all its attendees in the breathtaking scenario of Barcelona.

Francesco Longo
DIMEG, University of Calabria
Italy

Michael Affenzeller
Upper Austria University of Applied Sciences
Austria

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

ACKNOWLEDGEMENTS

The EMSS 2017 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 2017 IPC would like to thank all the authors as well as the reviewers for their invaluable work. A special thank goes to Prof. Miquel Angel Piera from Autonomous University of Barcelona, as Local Organizer and to all the organizations, institutions and societies that have supported and technically sponsored the event.

I3M 2017 INTERNAL STAFF

MATTEO AGRESTA, *SIMULATION TEAM, ITALY*
MARGARITA BAGAMANOVA, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
LUIGI BRUNO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
AGOSTINO G. BRUZZONE, *DIME, UNIVERSITY OF GENOA, ITALY*
ALESSANDRO CHIURCO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
RICCARDO DI MATTEO, *SIMULATION TEAM, ITALY*
CATERINA FUSTO, *CAL-TEK SRL, ITALY*
THIMJO KOÇA, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
GIANLUCA MAGLIONE, *SIMULATION TEAM, ITALY*
MARINA MASSEI, *DIME, UNIVERSITY OF GENOA, ITALY*
ROMUALDO MORENO, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
MERCEDES NARCISO, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
ANTONIO PADOVANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
MIQUEL ANGEL PIERA, AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN
MARKO RADANOVIC, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
JUAN JOSE RAMOS, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
CATALDO RUSSO, *CAL-TEK SRL, ITALY*
NINA SCHEFERS, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*
MARCO VETRANO, *CAL-TEK SRL, ITALY*

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

INDEX

A SIMULATION OPTIMIZATION METHOD FOR DESIGN AND CONTROL OF SIGNALIZED INTERSECTIONS IN URBAN AREAS

Yuan Chen[(a)], Ling Li[(b)], Yinghua Shen[(c)], Bingsheng Liu[(d)]

[(a)]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada
[(b),(d)]School of Management and Economics, Tianjin University, Tianjin, 300072, P.R. China
[(c)] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada

[(a)]ychen10@ualberta.ca, [(b)]liling@tju.edu.cn, [(c)]yinghua@ualberta.ca, [(d)]bluesea_boy_1979@163.com

ABSTRACT
The signalized intersections play a significant role in road network management. Bad traffic light control system results in more waste time the drivers spend on waiting and other negative influences like gaseous pollutant emission, more fuel usage, further deteriorating the environment and lowering the overall living standard. However, traditional analytical techniques are difficult to be applied to traffic lights due to its randomness, complexity and nonlinearity. Thus, simulation could be instrumental in designing and controlling signalized intersections and analyzing various situations by evaluating postulated "what-if" scenarios. This paper first utilized the AHP (analytic hierarchy process) method to determine the weights of evaluation indicators of traffic lights, and then applied simulation modelling (Simphony) to optimize the signalized intersection for the objectives of efficient design and control. Then, some comparisons of different lane designs were made to give some suggestions to road construction. The results demonstrated the effectiveness of the methodology.

Keywords: AHP, signalized intersections, simulation, optimization

1. INTRODUCTION
With the rapid development of regional economy, the residents tend to pursue material and spiritual life with high quality when basic living standard is guaranteed. One typical phenomenon is a great many of households could afford at least an automobile. Take China as an instance. There were 0.12 billion private vehicles in 2014 and the annual number had increased at the rate of 117.5% (National Bureau of Statistics, 2014). The automobile provides convenience for people's work and life to large extent, but the massive cars flooring into the urban area bring in serious traffic issue, especially traffic congestion. In terms of the drivers, they are reluctant and bored to waste most of time on waiting. Besides, the time spent idling in traffic wastes fuel, increases emissions, deteriorates the environment, and affects the overall standard of living (Sadoun, 2003). Meanwhile, traffic flow may largely exceed the maximum load capacity of the road compared with its initial setting. Thus, traffic congestion has become the main bottleneck of urban sustainable development, extremely concerned by city planners, engineers, local governments, urban policy makers, legislators, economists, and law enforcement agencies, among others.

The intersection is the important node in road network, and the signal control strategy has a significant effect on the traffic conditions. If one or several intersection signal setting is unreasonable, it will lead to slow traffic in the related roads, more seriously can cause a large number of vehicles stranded and the regional road network paralysis (Wunderlich et al., 2008). Moreover, the number of lanes and various types of roads also influence traffic congestion. Thus, how to design the new signalized intersections and adjust the existing signalized intersections are considered as the most complex task in the management of traffic system, due to its randomness, complexity and nonlinearity. Currently, there are two main traffic light signal control systems, that is, tradition control and adapting control (Sadoun, 2003). The former depends on historical data and prestored timing plan offline, so it's more easy and cost-effective to control traffic light signal system. By contrast, the latter adapts to actual traffic condition and is capable to perform continuous optimal system, but it needs to deploy traffic detectors and surveillance equipment in advance. Actually, traffic conditions could be divided into peak time, general time and slack time in a given day except bad weather conditions, accidents and holidays, and so on. As for the three phases, different scenarios should be set to meet various traffic conditions. Considering the features of the two main methods, they ignore the utilization of resources (e.g. green traffic lights and the lanes) and limit to make the adjustment based on the roads with the fixed number of lanes. Simphony is a Microsoft Windows based computer system developed with the objective of providing a standard, consistent and intelligent environment for both the development and the utilization of construction SPS (special purpose simulation) tools (Hajjar and AbouRizk, 1999). Also, Simphony provides various services that enable the users to easily control different behaviors in the developed tool such as simulation behaviors, graphical representation, statistics, and animation (AbouRizk and Mohamed, 2000). Besides, how to evaluate each simulation result is also an essential part, since traffic light signal control belongs to a multi-index evaluation and the important degree of indicators should balance the needs from different stakeholders on a certain traffic issue. Based on this, this paper combines the AHP (analytic hierarchy process) method to determine the weights of indicators and simulation modelling (Simphony) to optimize the signalized intersection for the objective of efficient design and control.

2. LITERATURE REVIEW

A complete design and control of signalized intersections involves analysis, modelling and simulation to obtain an efficient system. Basically, the process includes three important components: phases and sequence of traffic light, the cycle length and the duration for green light of each phase. According to various setting methods of the three elements, pretime operation, semiactuated operation and fully actuated operation are proposed to control traffic signals (Sadoun, 2003). Pretimed control needs to preset the cycle length, phases and all intervals of each signalized intersection (Clement and Koshi, 1997). In semiactuated operation, detectors are placed on minor approaches to the intersection, while both cycle length and the green-light time for every phase of the intersection can be varied in fully actuated operation based on mounted detectors (Sadoun, 2003).

Pretime operation belongs to the traditional signal control system, which depends on historical data and prestores timing plan offline. Therefore, it's more easy and cost-effective to control traffic light signal system, but not responsive to dynamic traffic demand. Semiactuated operation and fully actuated operation are the adaptive traffic light signal control system. Due to the detectors installed, the system could adapt to actual traffic condition and perform continuous optimization. There is some industry standard of the adaptive control system, such as SCATS, SCOOT, LHOVRA, UTOPIA, PRODYN and OPAC. However, these adaptive signal control systems control traffic on an areawide basis rather than an uncoordinated intersection basis, and are only applied to that specific case (Sadoun, 2003). Besides, large expenditure is spent on detector installation at the early stage and operating in the usage phase.

From the above, it could be seen that the two main signal control systems both have advantages and disadvantages. Simphony proposed by AbouRizk (1999) has been widely used in construction simulation. Simphony provides various services that enable the users to easily control different behaviors in the developed tool such as simulation behaviors, graphical representation, statistics, and animation (AbouRizk and Mohamed, 2000). Also, it could realize the dynamic simulation in a convenient and economical way, and directly reflect the simulation results (e.g. waiting time, queue length and utilization of resources). Thus, it's worthy and meaningful of applying Simphony to facilitating the traffic management, which could overcome the shortages of the signal control systems above and provide the simulation results that the users need.

3. METHODOLOGY

In this part, we firstly introduce some basic knowledge of traffic signal including traffic phases, traffic signal evaluation index and traffic arrival pattern. Based on this, the methodology will be proposed to present how to optimize signalized intersections by the means of the AHP method and Simphony.

3.1. Basic Knowledge of Traffic Signal

3.1.1. Traffic Phases

Traffic phase is the portion of the signal timing cycle that is allocated to one of these sets of movements (Davol, 2001). Typically, each phase consists of three intervals, that is, green, yellow and red. A phase will progress through all its intervals before moving to the next phase in the cycle. In reality, a three-leg signalized intersection and a four-leg signalized intersection are most common in the daily life. The corresponding layout and phase plans of the two forms of traffic intersections could be seen in Figure 1.

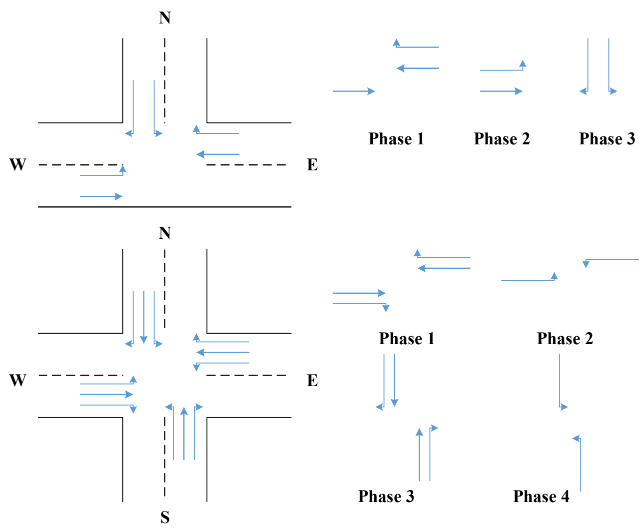


Figure 1: The Layout and Phase Plans of Two Typical Signalized Intersections

3.1.2. Traffic Signal Evaluation Index

Through literature review, traffic signal evaluation index adopted by the scholars is listed in Table 1.

Table 1: Traffic Signal Evaluation Index

Year	Authors	Index
2003	Sadoun	average waiting time, average queue length, average utilization of green-light time
2008	Yang et al.	traffic flow, average delays

2009	Ou et al.	average trip duration, volume of link, total volume, total stop
2013	Černický and Kalašová	delay time, mean speed, number of stop, stop time
2014	Liu et al.	average queque length, maximum queue length, delay time
2014	Wang et al.	average waiting time, average queque length, average number of vegicles, service intensity
2016	Osorio and Chong	average travel time

Due to various research objects, there exists some difference in the selection of traffic signal evaluation index. Ou et al. (2009), and Osorio and Chong (2016) mainly focused on large-scale urban transportation problems with the indicators like average trip duration, while other scholars tended to study signal time change influencing the vehicle behavior in a signalized intersection or multi-signalized intersections. The selection of evaluation index not only facilitates the traffic light optimization in the next stage, but also guarantees simulation results meet the needs from different stakeholders.

3.1.3. Vehicle Arrival Pattern

In the signalized intersection simulation, the first step is to collect data of the inter-arrival time for the vehicles, analyze and conform to the specific distribution. Currently, scholars have proposed three common vehicle arrival distribution models in different situation, that is, Poisson distribution, binomial distribution, and negative binomial distribution (Tohti et al., 2015). Where, Poisson distribution is used to describe the traffic flow which has short counting interval and low traffic density, without much interaction between the vehicles and other external interference (Wu, 2003). The basic form of Poisson distribution is described as below:

$$P(X = x) = e^{-\lambda}\lambda^{x}/x! \quad (x = 0,1,\ldots,n) \tag{1}$$

By contrast, binomial distribution and negative binomial distribution are more suitable for heavy traffic (Yang et al., 2008). The basic forms of two distributions are shown as below, respectively.

$$P(X = x) = \frac{n!}{x!(n-x)!} p^{x}(1-p)^{n-x} \quad (x = 0,1,2,\ldots,n) \tag{2}$$

$$P(X = x) = \frac{\Gamma(k+x)}{x!\Gamma(k)}\left(1+\frac{\mu}{k}\right)^{x}\left(\frac{\mu}{\mu+k}\right)^{n} \quad (x = 0,1,2,\ldots,n) \tag{3}$$

3.2. A Simulation Optimization Methodology for Signalized Intersections

3.2.1. Evaluation of Simulation Results Based on the AHP Method

From Table 1, scholars selected different indicators to evaluate the simulation results from various perspectives. Of them, some index refers to the features of vehicle and drivers' behavior, such as delay time and number of stop. Delay time refers to the difference between the necessary time a vehicle needs to pass by the leading way of crossing entry in a forbidden position and the time a vehicle needs to use while be in no hindering position (Wang, 2006). Average trip duration is selected as the index for the large-scale urban transportation issue. Based on this, two assumptions are given first: (1) all the automobiles as independent individuals; (2) the duration for a vehicle to go through a signalized intersection involves the influences from drivers' behavior and the features of vehicle. This paper will choose average waiting time, average queue length and traffic flow (out) as the traffic signal evaluation indicators. The index is mainly utilized to evaluate a signalized intersection or multi-signalized intersections and could meet the needs from various stakeholders.

As for the multi-index evaluation problems, the importance degree of each indicator should be determined and then weightings and values are integrated to obtain the final evaluation results. Based on this, this paper will introduce the AHP method proposed by Saaty (1977), which has been extensively applied and studied for weighting determination and decision making. The evaluators are from different stakeholder, so it guarantees the values of weightings are suitable for the certain simulation environment.

3.2.2. A Simulation Optimization Methodology

Based on the literature review and research objectives, this paper combines the AHP method and Simphony to propose a simulation optimization methodology, which could be applied to optimize one or more signalized intersections in the phases of design and control. The methodology framework could be seen in Figure 2.

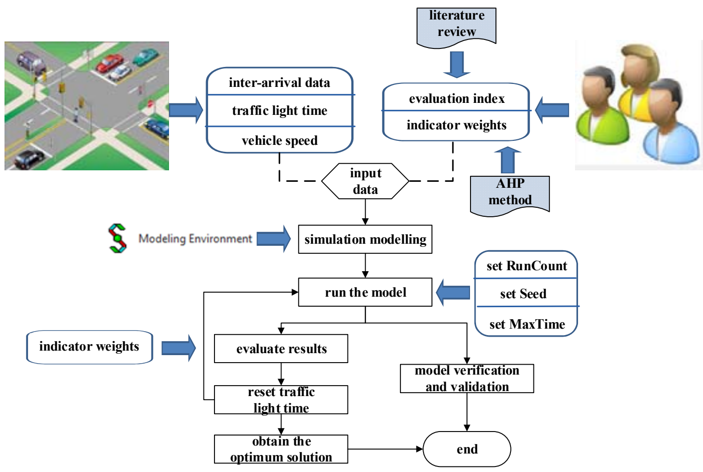


Figure 2: A Simulation Optimization Methodology for the Signalized Intersections

4. SIMULATION OPTIMIZATION FOR A SIGNALIZED INTERSECTION

4.1. Abstract of the Signalized Intersection

This paper utilized a real signalized intersection given by Liu et al. (2014) to present how the methodology proposed to realize the optimization for traffic lights. The signalized intersection of Jianshe Ave and Xin Hua Lu is a typical four-leg intersection located in Wuhan city of China (see Figure 3). The west entrance road and the east entrance road are comprised of three lanes while only two lanes are on the other side of the road. In contrast, the north entrance road and the south entrance road consist of four lanes while only three lanes are on the other side of the road. To be mentioned, the right lanes of all the roads are shared by straight drive and right turn. And two directions of each road have different numbers of lanes and the total lanes of two roads are also various, that is, asymmetry.

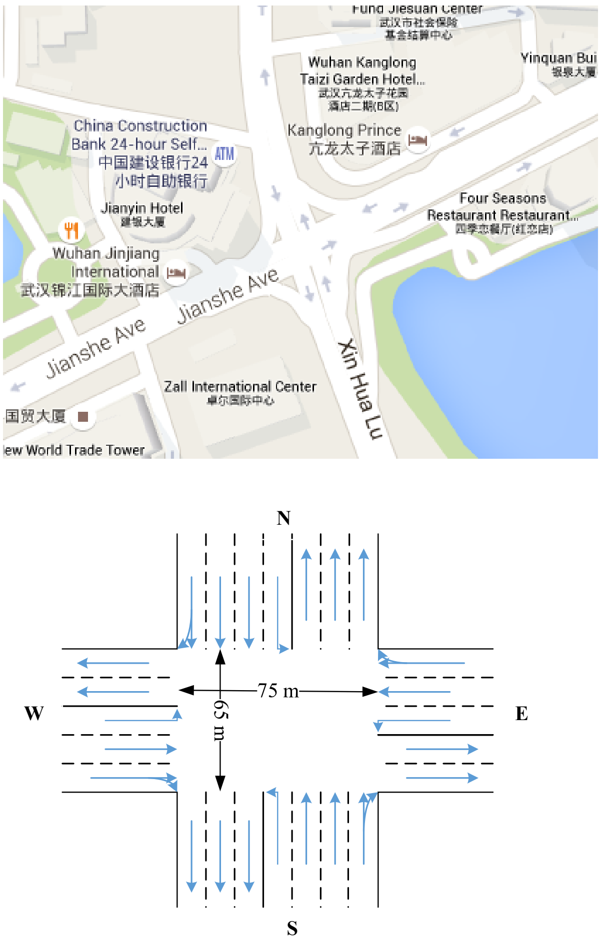


Figure 3: A Signalized Intersection in Wuhan City

4.2. Data for the Signalized Intersection

The traffic flow at the entrance of each road was monitored in Table 2 (Liu et al., 2014). Based on vehicle arrival pattern used by scholars in Section 3.1.3, Poisson distribution was selected to describe the inter-arrival time in the signalized intersection and the means of inter-arrival time could be calculated on the basis of traffic flow in Table 2. The initial duration of the signalized intersection could be summarized in Table 3, which will be applied to the input data for the simulation.

Table 2: Traffic Flow Data of the Signalized Intersection

		Lane	Traffic Flow (pcu/h)
East	Left	1	238
	Straight	2	727
	Right	1	198
South	Left	1	257
	Straight	3	1000
	Right	1	222
West	Left	1	275
	Straight	2	741
	Right	1	177
North	Left	1	240
	Straight	3	1227
	Right	1	200

* pcu indicates passenger car unit.

Table 3: Initial Duration of the Signalized Intersection

		Inter-arrival time (s)	Phase		Duration (s)
East	Left	P(15)	1	south/north straight	U(5,6)
	Straight	P(5)		to end straight	
	Right	P(18)		green light	50
South	Left	P(14)	2	south/north left	U(4,5)
	Straight	P(3.6)		to end left	
	Right	P(16)		green light	30
West	Left	P(13)	3	west/east straight	U(6,7)
	Straight	P(5)		to end straight	
	Right	P(20)		green light	50
North	Left	P(15)	4	west / east left	U(4,5)
	Straight	P(3)		to end left	
	Right	P(18)		green light	30
Yellow light					3

* P indicates Poisson; U refers to Uniform.

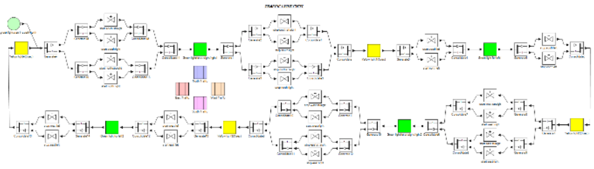
4.3. Simulation Modelling and Evaluation

4.3.1. Simulation Model Layout

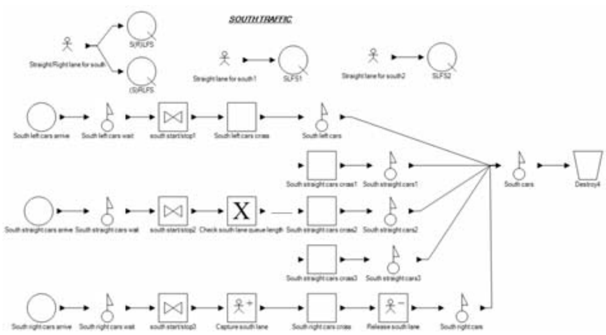
According to the plan view of the signalized intersection in Figure 3 and its input data of duration in Table 3, Simphony.NET was utilized to build the simulation model for the traffic lights in Figure 4. The four composition elements inside the traffic light cycle

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

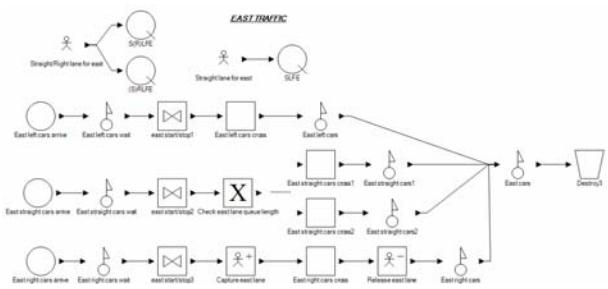
indicate south traffic, north traffic, west traffic and east traffic. As south traffic and north traffic have the similar model structure, only south traffic model was presented. This is identical to west traffic and east traffic. Assume that (1) the peak time for the traffic is one hour, so the simulation model is to be terminated after 3600 seconds; (2) a car arriving at each entrance will select the lane for straight drive based on the shortest queue length principle; (3) the south (the west) and the north (the east) have the same green times for straight drive/ right turn and left turn, respectively.



(a) Traffic Light Cycle



(b) South Traffic



(c) East Traffic

Figure 4: Simulation Model Layout for the Signalized Intersection

4.3.2. Result Analysis

Based on literature review, average waiting time, average queue length and traffic flow were selected as traffic signal evaluation index. After different stakeholders evaluated the importance degrees of the three indicators, the AHP method was utilized to determine their weights. The weights of average waiting time, average queue length and traffic flow were 0.4, 0.4 and 0.2, respectively. Then, through adjusting the green times for straight drive and right turn as well as left turn of each direction, the simulation results with 100 runs and 1000 seeds could be obtained. The optimum solution to control the signalized intersection could be seen in Table 4.

Table 4: The Optimum Traffic Light Time for the Signalized Intersection

	Green time for south/ north straight and right turn (s)	Green time for south/ north left turn (s)	Green time for east/ west straight and right turn (s)	Green time for east/ west left turn (s)
Duration	12	10	10	10
	Waiting time (s)	Queue length	Traffic flow (pcu/s)	Evaluat -ion value
South straight	20.86	4.456	0.274	8.81
South right turn	20.99	1.019	0.062	10.19
South left turn	21.86	1.28	0.072	9.27
North straight	20.92	5.412	0.330	10.60
North right turn	20.94	0.904	0.055	8.75
North left turn	21.69	1.175	0.067	9.16
West straight	21.94	3.573	0.196	10.25
West right turn	21.92	0.91	0.050	9.14
West left turn	22.05	1.389	0.077	9.39
East straight	21.91	3.573	0.197	10.23
East right turn	21.91	1.002	0.055	9.18
East left turn	28.78	0.026	0.068	11.54
Evaluation value	South/North: 9.463		West/East: 9.955	

4.3.3. Validation and Verification

Model validation is usually defined as substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model (Schlesinger, 1979). Model verification is often defined as ensuring that the computer program of the computerized model and its implementation are correct (Sargent, 2004). Sargent (2004) summarized various validation techniques and tests to be used in model verification and validation. Where, the "traces" approach is one of the common methods that the behavior of specific entities in the model are traced (followed) through the model to determine if the model's logic is correct and if the necessary accuracy is obtained.

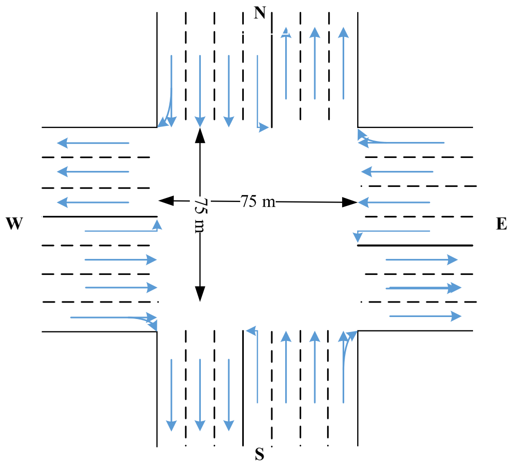
Based on this, to verify the simulation models, the trace functionality provided by Simphony was utilized to prove that the model replicated the logical traffic flow sequence as intended. User-written code was embedded within the model in order that the chronological list of events and "TimeNow" when each car goes across the signalized intersection in the optimum traffic light situation. Part of trace results for this signalized intersection is shown within the trace in Figure 5, where the recorded times for crossing the intersection are found to follow the expected trend; hence the model can be considered reliable.

No.	TimeNow	No.	TimeNow	No.	TimeNow	No.	TimeNow	No.	TimeNow	No.	TimeNow
1	4.435	21	40.878	41	59.056	61	73.702	81	88.998	101	104.940
2	4.558	22	41.042	42	59.711	62	73.996	82	90.732	102	105.968
3	6.904	23	41.046	43	60.175	63	74.371	83	90.766	103	106.101
4	9.702	24	41.498	44	61.204	64	74.866	84	90.818	104	106.741
5	9.872	25	44.933	45	63.323	65	74.999	85	92.468	105	107.497
6	10.116	26	45.666	46	63.622	66	77.445	86	92.676	106	108.777
7	15.462	27	45.995	47	63.903	67	77.747	87	95.099	107	110.047
8	15.849	28	46.373	48	64.204	68	78.469	88	95.420	108	111.229
9	19.003	29	47.537	49	64.215	69	79.553	89	96.036	109	111.922
10	19.613	30	47.602	50	64.226	70	80.123	90	96.586	110	112.974
11	23.537	31	49.226	51	66.955	71	80.311	91	98.676	……	……
12	27.710	32	49.951	52	67.526	72	81.586	92	98.796	5417	3592.679
13	31.722	33	50.573	53	68.434	73	82.225	93	99.106	5418	3595.673
14	32.112	34	52.380	54	68.858	74	82.738	94	99.244	5419	3596.181
15	32.171	35	53.769	55	69.214	75	85.377	95	99.270	5420	3596.380
16	34.182	36	54.383	56	69.332	76	85.406	96	101.616	5421	3597.001
17	34.996	37	54.900	57	69.505	77	85.675	97	102.053	5422	3597.122
18	35.746	38	54.923	58	70.087	78	86.131	98	102.335	5423	3597.424
19	36.415	39	58.404	59	73.263	79	86.295	99	103.259	5424	3598.031
20	36.699	40	58.664	60	73.390	80	86.375	100	103.435	5425	3598.889

Figure 5: Chronological List of Traffic Flow through the Signalized Intersection

4.4. Comparisons with Alternatives

In the signalized intersection, the lanes of the two direction roads are asymmetry. As for city planners, engineers or local governments, they could consider the lane design for the road, which has direct effects on the operation of the signalized intersection. Thus, different alternatives could be compared based on the actual or forecasted traffic information. Then, other two-lane design plans of the signalized intersection in Wuhan city were given in Figure 6.



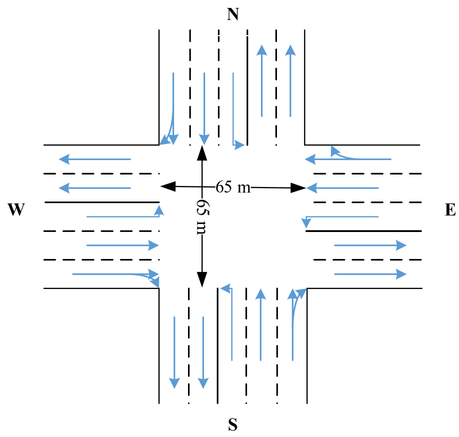


Figure 6: Two Lane Design Plans for the Signalized Intersection

Based on the simulation model of the signalized intersection with the optimum traffic light time, adjust the number of lanes as the above figure and then utilize Simphony to obtain the utilization rate of the lanes in Table 5. Plan 1 is the original lane design; plan 2 and plan 3 are the new design plans.

Table 5: The Comparison of Utilization Rate of the Lanes in Different Design Plans

		Left	Straight	Right
East	Plan 1	30.3 %	78.3 %	74.8 %
	Plan 2	29.9 %	77.5 %	74.8 %
	Plan 3	30.2%	47.0%	59.4%
South	Plan 1	32.2 %	46.0%	85.8 %
	Plan 2	32.1 %	82.2 %	95.6 %
	Plan 3	31.9%	46.6%	86.0%
West	Plan 1	34.4 %	75.7 %	74.6 %
	Plan 2	34.3 %	75.0 %	75.3 %
	Plan 3	34.2%	48.1%	54.5%
North	Plan 1	29.9 %	58.4 %	88.8 %
	Plan 2	29.9 %	96.6 %	98.7 %
	Plan 3	30.3%	58.2%	88.6%

From the table, it could be seen that the utilization rates of the right lanes in all the entrances are large, since the right lanes are shared by the vehicles going straight and turning right. The west and the east in plan 1 and plan 2 have higher usage rates of straight drive than those of plan 3. It shows that increasing the number of lanes could mitigate the road pressure to large degree. The straight lanes of the north and the south in plan 2 bear the tremendous traffic flow, so plan 1 used in the reality should be designed based on the predicted traffic volume. However, the road design (plan 1) was built before a long period, thus resulting currently the straight lanes of the east and the west support relative higher traffic volume. Thus, through the comparison of different lane plans for the roads, city planners could make a scientific decision for the road design based on the predicted lane utilizations in multi-alternatives.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

5. CONCLUSIONS

Traffic congestion is the hot topic concerned by the scholar circles and the experts from the industries, since it has become the main bottleneck of urban sustainable development and brings other bad influences on environment and people's life. This paper firstly chose average waiting time, average queue length and traffic flow (out) as traffic signal evaluation indicators, and determined the weights of the index by the AHP method to meet the needs from different stockholders. Then, Simphony was utilized to simulate the traffic lights in various scenarios and further obtain the optimization solution to control the signalized intersections. Also, the simulation models developed were verified and validated by tracing the chronological list of events and observing relevant statistics (e.g. cycle time for a vehicle across the intersection). Besides, the methodology could also facilitate city planners or the government to select the most suitable lane design by comparing the utilization rates of several alternatives.

Besides the major needs from the drivers, other evaluation indicators such as CO2 emission could be considered during the traffic process. Also, CO2 emission should depend on the whole time spent in the intersection and the features of the vehicles. Moreover, the optimum parameters of traffic lights mainly depend on the adjustment of traffic cycle time manually. In future, the simulation process with Simphony could consider environmental influences and integrate optimization algorithm to obtain the optimum parameters of traffic lights automatically and more efficiently.

REFERENCES

AbouRizk S., Mohamed Y., 2000. Simphony: An Integrated Environment for Construction Simulation. Proceeding of the 2000 Winter Simulation Conference, pp. 1907-1914. December 10-13, Orlando (Florida, USA).

Černický Ľ., Kalašová A., 2013. Microscopic Simulation of the Coordinated Signal Controlled Intersections. In Activities of Transport Telematics: 13th International Conference on Transport Systems Telematics, pp. 15-22. October 23-26, Katowice-Ustron (Poland).

Clement S., Koshi J., 1997. Traffic Signal Determination: The Cabal Model. Proceeding of 1997 Genetic Algorithms in Engineering Systems: Innovations and Applications Conference, pp. 63-68. September 2, London (UK).

Davol A.P., 2001. Modeling of traffic signal control and transit signal priority strategies in a microscopic simulation laboratory. USA: Brown University.

Hajjar D., AbouRizk S., 1999. Simphony: An Environment for Building Special Purpose Construction Simulation Tools. Proceeding of the 1999 Winter Simulation Conference, pp. 998-1006. December 5-8, Phoenix (Arizona, USA).

Liu X., Hong H., Wang C.W., Yang W., 2014. Signal Coordination Control and Simulation for multi Intersections. Proceeding of the 7th IEEE Information Technology and Artificial Intelligence Conference, pp. 464-467. December 20-21, Chongqing (China).

National Bureau of Statistics, 2014. China Statistical Yearbook. Beijing: China Statistical Press.

Osorio C., Chong L., 2016. A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. Transportation Science, 49(3), 623-636.

Ou D.X, Balmer M., Liu H., Shi L.J., 2009. Modeling and Signalizing Intersections for Large-Scale Transportation Simulation Network. Proceeding of the 12th International IEEE Conference on Intelligent Transportation Systems, pp. 609-614. October 4-7, Saint Louis (Missouri, USA).

Saaty T.L., 1977. A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15(3), 234-281.

Sadoun B., 2003. An efficient simulation methodology for the design of traffic lights at intersections in urban areas. The Society for Modelling and Simulation International, 79(4), 243-251.

Sargent R.G. 2004. Validation and Verification of Simulation Models. Proceeding of the 2004 Winter Simulation Conference, pp. 17-28. December 5-8, Washington (D.C., USA).

Schlesinger S., 1979. Terminology for model credibility. Simulation, 32(3), 103-104.

Tohti G., Geni M., Wang B., 2015. Traffic Flow Modeling and Simulation of Signalized Intersections. Proceedings of the 15th COTA International Conference of Transportation Professionals, pp. 2467-2473. July 25-27, Beijing (China).

Wang F.Y., Ye C.M., Zhang Y.N., Li Y., 2014. Simulation analysis and improvement of the vehicle queuing system on intersections based on MATLAB. The Open Cybernetics & Systemics Journal, 8, 217-223.

Wang L., 2006. The urban real-time traffic control (URTC) system: A study of designing the controller and its simulation. Thesis (Master). University of Quebec.

Wu J.G., 2003. Study of traffic flow microcosmic simulation model at signal junction. Transportation Science and Technology, 4, 33-40.

Wunderlich R., Liu C.B., Elhanany I., Urbanik T., 2008. A novel signal scheduling algorithm with quality-of-service provisioning for an isolated intersection. IEEE Transactions on Intelligent Transportation Systems, 9(3), 536-547.

Yang Z.Y., Huang X.Y., Du C.H., Tang M.X., Yang F.X., 2008. Hierarchical Fuzzy Logic Traffic Controller for Urban Signalized Intersections. Proceedings of the 7th World Congress on Intelligent Control and Automation, pp. 5203-5207. June 25-27, Chongqing (China).

USE OF CPANS FOR GROUPING DNA SEQUENCE FRAGMENTS IN THE CONSTRUCTION OF GNOMES

M. Rossainz-López[(a)], Manuel I. Capel[(b)], R. Hernández-Munive[(a)], I. Olmos-Pineda[(a)] , J. A. Olvera-López[(a)]

[(a)] Faculty of Computer Science, Autonomous University of Puebla, San Claudio Avenue and South 14th Street, San Manuel, Puebla, Puebla, 72000, México
[(b)] Software Engineering Department, College of Informatics and Telecommunications ETSIIT, University of Granada, Daniel Saucedo Aranda s/n, Granada 18071, Spain

[(a)]rossainz@cs.buap.mx, [(b)]manuelcapel@ugr.es, [(a)] azhu116@gmail.com, [(a)]iolmos@cs.buap.mx, [(a)]aolvera@cs.buap.mx

ABSTRACT
This paper proposes the representation through the model of the High Level Parallel Compositions or CPANs, of the communication / interaction patterns denominated Pipeline and Farm, and their usefulness in combinatorial optimization problems within the classification of NP-hard problems such as grouping fragments of DNA sequences and assembly of these fragments to construct gnome; through a Structured Parallel Programming approach based on the concept of Parallel Objects. The Pipeline and Farm models are shown as CPANs under the Object Orientation paradigm and with them it is proposed the creation of a new CPAN that combines and uses the previous ones to solve the problem of obtaining and assembling Strings of DNA. Each CPAN proposal contains a set of predefined synchronization constraints between processes (maximum parallelism, mutual exclusion and synchronization of producer-consumer type), as well as the use of synchronous, asynchronous and asynchronous future modes of communication. We show the algorithm that solves the assembly of DNA sequences, their design and implementation as CPAN and the performance metrics in their parallel execution using multicores.

Keywords: CPAN, High Level Parallel Compositions, Parallel Structured Programming, Parallel Objects, Pipeline, Farm, DNA, Contigs, Genome.

1. INTRODUCTION

Currently within the parallel programming one of the open problems of major interest is the lack of acceptance structured parallel programming environments of use to develop applications. Structured parallelism is a type of parallel programming based on communication/interaction patterns (pipelines, farms, trees, etc.) that are predefined among the processes of a user application. Patterns also encapsulate parallel parts of the application, in such way that the user will only program the sequential code of the application. Many proposals of environments exist for the development of applications and structured parallel programs (Bacci, Danelutto, Pelagatti and Vaneschi 1999; Darlington 1993), but until now only a very limited circle of expert programmers use them. In the literature there are several proposals and all agree on the importance of determining a complete set of patterns and try to define a semantics for them (De Simone 1997). At moment, in HPC, a great interest exists in structured-parallel environments research, as the ones previously mentioned. The trend is the use of object-oriented programming approaches. It has been shown that defining parallel objects for the development of new methodological proposals, models and parallel programming communication patterns has generated good results (Corradi and Leonardi 1991; Corradi and Zambonelli 1995). HLPCs or CPANs are parallel patterns defined and logically structured that, once identified in terms of their components and of their communication, can be adopted in the practice and be available as high level abstractions in user applications within an OO-programming environment (Brinch Hansen 1993). The process interconnection structures of most common parallel execution patterns, such as pipelines and farms can be built using CPANs, within the work environment of POs that is the one used to detail the structure of a CPAN implementation. With them, problems like the assembly of DNA strings proposed in this paper can be solved. Finding the solution to these types of problems has become indispensable in research in biology and in many fields such as medical diagnosis, biotechnology, forensic biology, virology, applied biology and bioinformatics among others. The problem of the assembly of DNA strings enters the so-called NP-Hard, because it is a problem of combinatorial optimization in which diverse heuristics and met heuristics have been proposed to assemble sequences of DNA strings and to provide essential information to understand the species and their mechanisms of life including the human species. This work shows the implementation of a grouping algorithm that evaluates a set of DNA sequence fragments as a CPAN. The CPAN represents a Farm where worker processes are themselves Pipeline CPANs. The algorithm determines subgroups of fragments by DNA sequences matching found, which have a high probability of being aligned in an assembly task. Each worker process of CPAN Farm works in parallel with the other worker processes that are generated with a group of fragments of DNA sequences that are internally constructed as graphs represented through the CPAN Pipeline and through an in-depth search the new groups of DNA sequences are generated, which must be processed by some assembly technique to form the contigs of a genome that has been sequenced covering most of its structure but missing a fragment to be completed. Finally the design of an experiment is shown through the use of the new CPAN generated called Cpan GraphADN, with genomes of viruses and bacteria available on the web. The pseudo random synthetic readings created to form contigs are shown and the execution performance of this proposal is obtained for eight genomes with an Intel Core i8 processor, a video accelerator card with 1664 CUDA cores and a clock frequency of 1178 MHz.

2. HIGH LEVEL PARALLEL COMPOSITIONS (CPAN)

HLPCs or CPANs are parallel patterns defined and logically structured that, once identified in terms of their components and of their communication, can be adopted in the practice and be available as high level abstractions in user applications within an OO-programming environment (Rossainz 2005, Rossainz and Capel 2008). The process interconnection structures of most common parallel execution patterns, such as pipelines, farms and trees can be built using CPANs, within the work environment of POs that is the one used to detail the structure of a CPAN implementation, for details see (Rossainz and Capel 2012). A CPAN comes from the composition of a set three object types: an object manager that represents the CPAN itself and makes an encapsulated abstraction out of it that hides the internal structure. The object manager controls a set of objects references, which address the object collector and several stage objects and represent the CPAN components whose parallel execution is coordinated by the object manager (figure 2). The objects stage are objects of a specific purpose, in charge of encapsulating a client-server type interface that settles down between the manager and the slave-objects. These objects do not actively participate in the composition of the CPAN, but are considered external entities that contain the sequential algorithm that constitutes the solution of a given problem. Additionally, they provide the necessary inter-connection to implement the semantics of the communication pattern which definition is sought. In other words, each stage should act a node of the graph representing the pattern that operates in parallel with the other nodes. Depending on the particular pattern that the implemented CPAN follows, any stage of it can be directly connected to the manager and/or to the other component stages. In collector object we can see an object in charge of storing the results received from the stage objects to which is connected, in parallel with other objects of CPAN composition. That is to say, during a service request the control flow within the stages of a CPAN depends on the implemented communication pattern. When the composition finishes its execution, the result does not return to the manager directly, but rather to an instance of the collector class that is in charge of storing these results and sending them to the manager, which will finally send the results to the environment, which in its turn sends them to a collector object as soon as they arrive, without being necessary to wait for all the results that are being obtained (Rossainz and Capel 2012). If we observe the scheme as a black box, the graphic diagram of a CPAN representation would be the one that is shown in Figure 1.

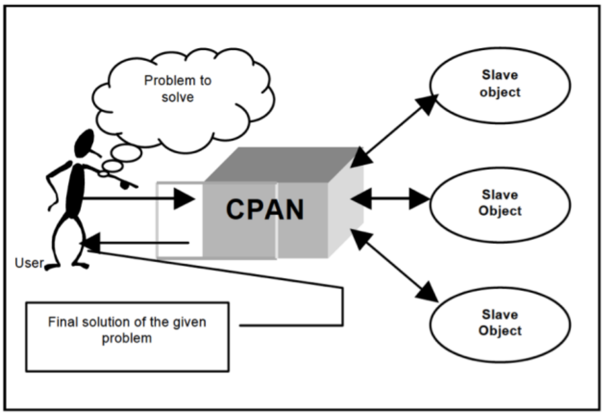


Figure 1: Graphical representation of a CPAN as black-box

In summary, a CPAN is composed of an object manager that represents the CPAN itself, some stage objects and an object of the class Collector, for each petition that should be managed within the CPAN. Also, for each stage, a slave object will be in charge of implementing the necessary functionalities to solve the sequential version of the problem being solved (Figure 2).

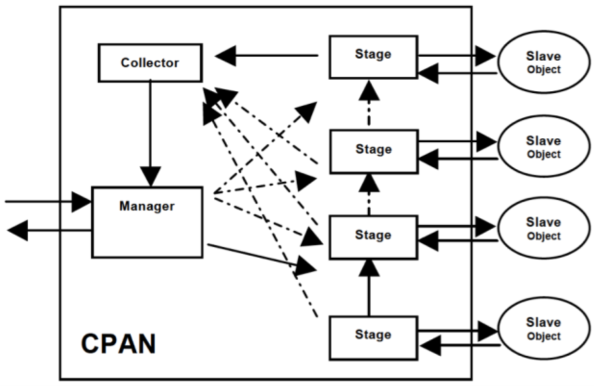


Figure 2: Internal structure of CPAN. Composition of its components

Figure 2 shows the pattern CPAN in general, without defining any explicit parallel communication pattern. The box that includes the components, represents the encapsulated CPAN, internal boxes represent compound objects (collector, manager and objects stages), as long as the circles are the objects slaves associated to the stages. The continuous lines within the CPAN suppose that at least a connection should exist between the manager and some of the component stages. Same thing happens between the stages and the collector. The dotted lines mean more than one connection among components of the CPAN. For details CPAN model, see (Rossainz and Capel 2014).Manager, collector and stages are included in the definition of a Parallel Object (PO), (Corradi 1991). Parallel Objects are active objects, which is equivalent to say that these objects have intrinsic execution capability (Corradi 1991). Applications that deploy the PO pattern can exploit the inter-object parallelism as much as the internal or intra-object parallelism. A PO-instance object has a similar structure to that of an object in Smalltalk, and additionally defines a scheduling politics, previously determined that specifies the way in which one or more operations carried out by the instance synchronize (Danelutto and Orlando 1995; Corradi 1991). Synchronization policies are expressed in terms of restrictions; for instance, mutual exclusion in reader/writer processes or the maximum parallelism allowed for writer processes. Parallel objects support multiple inheritance in the CPAN model. Parallel objects define 3 communication modes: The synchronous communication mode stops the client activity until it receives the answer of its request from the active server object (Andrews 2000), The asynchronous communication does not delay the client activity. The client simply sends the request to the active object server and its execution continues afterwards (Andrews 2000) and the asynchronous future will delay client activity when the method's result is reached in the client's code to evaluate an expression. For details see (Lavander and Kafura 1995).

3. THE CPAN PIPELINE

It represents the aforementioned pipeline technique of parallel processing as a HLPC or CPAN, applicable to a wide range of problems that are partially sequential intrinsically. The CPAN Pipe guarantees the parallelization of sequential code using the pattern PipeLine.

3.1. The technique of the Pipeline

Using the technique of the Pipeline, the idea is to divide the problem in a series of tasks that have to be completed, one after another, see figure 3. In a pipeline each task can be executed by a process, thread or processor for separate (De Simone 1997; Wilkinson and Allen 1999).

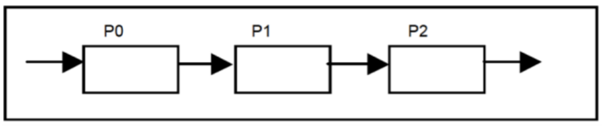


Figure 3: Pipeline

The processes of the pipeline are sometimes called stages of the pipeline (Roosta 1999). Each stage can contribute to the solution of the total problem and it can pass the information that is necessary to the following stage of the pipeline. This type of parallelism is seen many times as a form of functional decomposition. The problem is divided in separate functions that can be executed individually, but with this technique, the functions are executed in succession (Barry and Allen 1999). The technique of parallel processing pipeline is then presented as a High Level Parallel Composition applicable to solving a range of problems that are partially sequential in nature, so that the Pipe CPAN guarantees code parallelization of sequential algorithm using the pattern Pipeline.

3.2. Representation of the Pipeline as a CPAN

The Figure 4 represents the parallel pattern of communication Pipeline as a CPAN. The details of the implementation can be consulted in (Rossainz, Capel and Domínguez 2015).

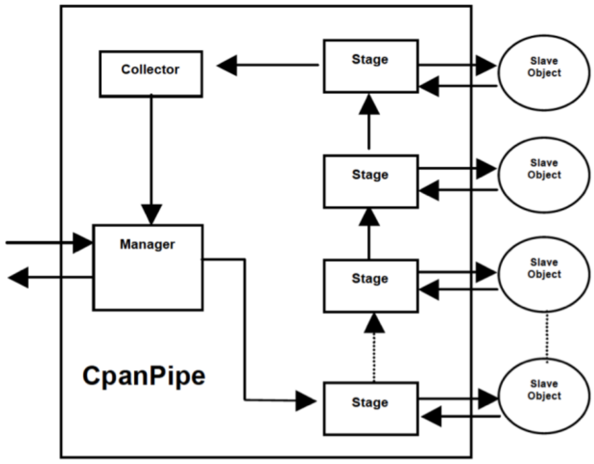


Figure 4: The CPAN of a Pipeline

Once the objects are created and properly connected according to the parallel pattern Pipeline, then you have a CPAN for a specific type of parallel pattern, and can be resolved after the allocation of objects associated with slave stages.

4. THE CPAN FARM

The technique of the parallel processing of the farm as a HLPC or CPAN is shown here. The so named farm parallel pattern of interaction is made up of a set of independent processes, called worker processes, and a process that controls them, called the process controller (Roosta 1999) and (Rossainz and Capel 2008). The worker processes are executed in parallel until all of them reach a common objective. The process controller is in charge of distributing the work and of controlling the progress of the farm until the solution of the problem is found (Barry and Allen 1999). Figure 5 shows the pattern of the farm.

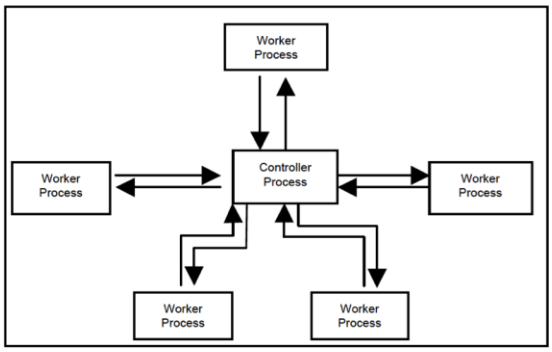


Figure 5: Farm with a controller and five workers

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

4.1. Representation of the Farm as a CPAN

The representation of parallel pattern farm as a CPAN is shown in Figure 6. The details of the implementation can be consulted in (Rossainz and Capel 2012).

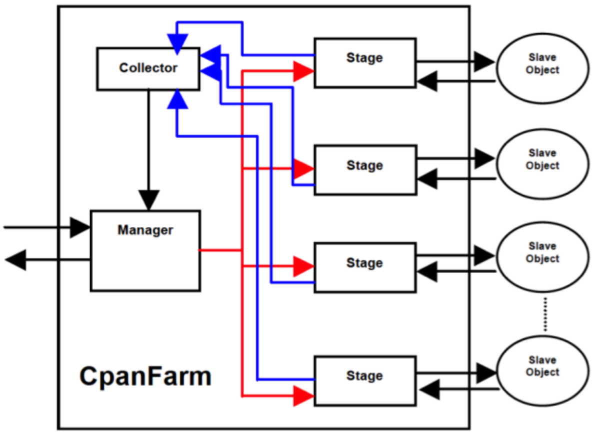


Figure 6: The CPAN of a farm

5. REPRESENTATION OF THE GENOME IN AN ORGANISM

The hereditary information of living beings is stored in a molecule of deoxyribonucleic acid called DNA. This molecule consists of two strands of nucleotides that form a structure, similar to a twisted ladder, called a double helix, see Figure 7 (Dias 2011).

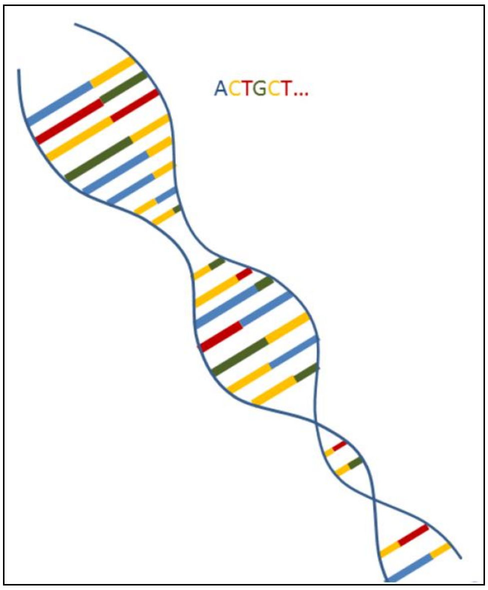


Figure 7: DNA structure, double helix

Each strand of DNA is composed of several nucleotides that are molecules formed by a nitrogenous base, a sugar that contains five molecules of carbon and a phosphoric acid. There are four types of nitrogen bases: adenine (A), guanine (G), cytosine (C) and thymine (T). In the double helix the nitrogenous bases are paired through bridges of hydrogen, giving the circumstance that the A always joins the T and the G to the C (Dias 2011). This is called base pairs (bp).

Chromosomes are filament-like structures that are inside the nucleus of a cell and contain the genetic material of a species, see Figure 8 taken from (Samiksha 2016). Each organism has a certain number of chromosomes per cell. Humans, for example, have 46 chromosomes.

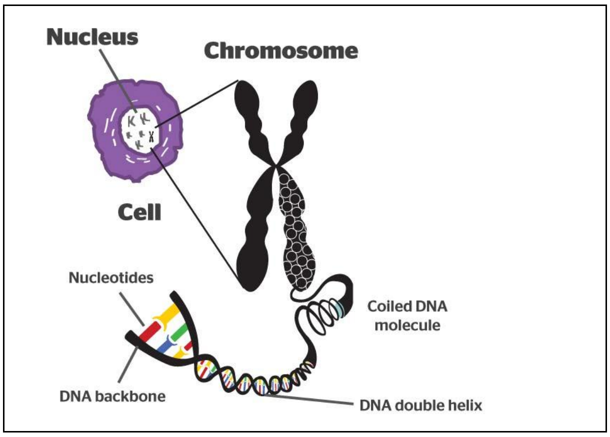


Figure 8: Structure of a chromosome (Samiksha 2016)

Along the DNA strand there are sequences of nitrogenous bases containing genetic information. These sequences are called genes and are responsible for telling cells how, when and where to produce all the necessary structures for life. All cells in the same organism have the same genetic information (Días 2011).

Genome is therefore the complete set of genetic information contained in the chromosome. Both the genome and the DNA sequences belonging to it are measured by counting their number of base pairs (bp). For very long sequences, as for a complete genome, kbp, Mbp and Gbp are used. According to the species, the genome may be composed of hundreds of thousands to billions of base pairs. However, the length of the genome is not related to the degree of complexity of a species. There are similar organisms that may differ in the size of their genomes and more complex organisms that may have smaller genomes than other simpler ones. The number of genes in a genome also has no relation to the size of the genome. But, the number of chromosomes can influence the number of base pairs of a given species. With all this we have that, the exact representation of the genome is formed of 3 parts: the assembly of DNA sequences, correction of error and the formation of contigs and scaffolding.

5.1. The sequence assembly of DNA

One of the initial problems in genome study is the measurement of the similarity of DNA sequences within the same genome but also within different genomes of equal species or within genomes of different species. The set of readings of DNA sequences in the analysis of a gnome is disordered. To obtain the original DNA fragment, all sequences obtained correctly must be fitted. This process is called sequence assembly and aims to determine the correct order of fragments of the sequenced DNA. The output of the assembly is a set of contigs that are formed by the alignment of the input sequences, see Figure 9 taken from (Cañizares, Blanca and Ziarsolo 2015). The number of contigs is well below the number of fragments of entry, since each contig can be formed by several fragments. When assembling a set of fragments it is unlikely that all the fragments will form part of the resulting contigs. These sequences that do not form any contig are called singletons (Días 2011; Hernandez, Olmos and Olvera 2016).

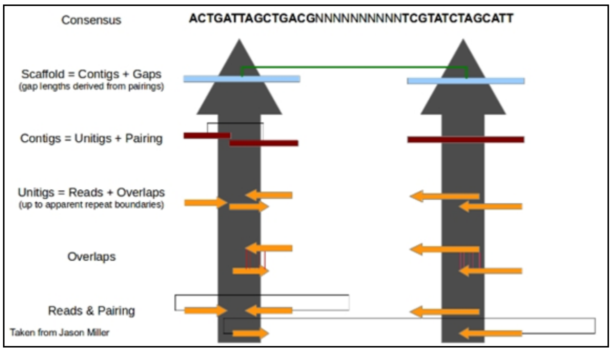


Figure 9: Terminology in the assembly sequence (Cañizares, Blanca and Ziarsolo 2015)

A contig is a set of overlapping DNA segments that together represent a consensus region of DNA. A supercontig is an even longer sequence than a contig, formed by the union of two or more contigs. The union of two or more supercontigs forms a scaffold. Several scaffolds finally create the structure of the chromosome see Figure 10 taken from (Monya 2012).

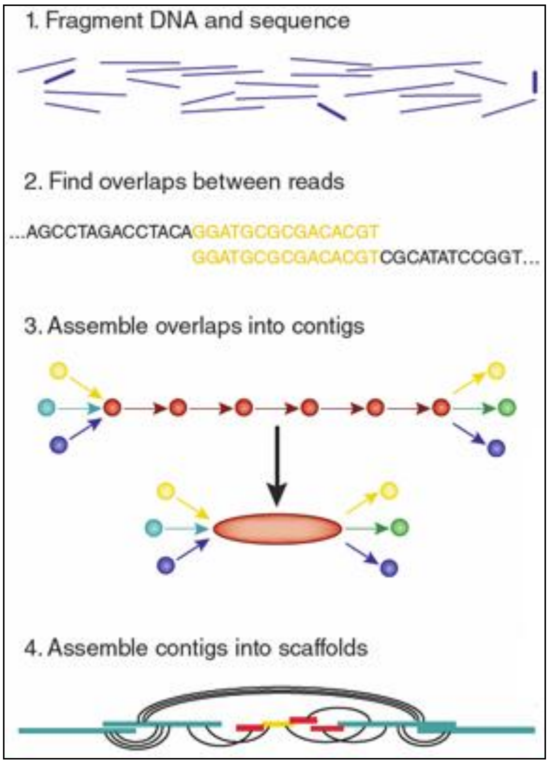


Figure 10: Genome assembly

1. Fragment DNA and sequence: A series of readings are made by the machines called sequencers. Readings are small fragments of DNA.
2. Find overlaps between reads: To assemble the sequence, the splices between the DNA fragments obtained are found. With this information a map is created to assemble longer sequences.
3. Assemble overlaps into contigs: A continuous sequence of DNA is generated that has been assembled through the spliced DNA fragments, and larger fragments of the DNA are being created.
4. Assemble contigs into scaffolds: A sequence formed by the union of one or more contigs is created. In a scaffold it is not required that there is an overlap between contigs (not so for the latter) to assemble longer strings.

6. ASSEMBLY OF DNA SEQUENCES USING CPANS

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases: adenine, guanine, cytosine, and thymine in a strand of DNA (Masoudi-Nejad et al. 2013). The advent of rapid DNA sequencing methods has greatly accelerated biological and medical research and discovery. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, or genomes of numerous types and species of life, including the human genome and other complete DNA sequences of many animal, plant, and microbial species (Pareek et al., 2011). In this regard use of CPANs for grouping DNA sequence fragments from the parallelization of a clustering algorithm to evaluate a set of fragments are made, which have a high probability of being aligned in an assembly task (DanishAli and Farooqui, 2013). The algorithm finds the splices between the fragments using the Myers algorithm and links them in a graph. Then an in-depth search is done in the graph to form the groups and send them as a result (Blelloch and Guy 1996), (The Myers algorithm performs search matching of substrings using a deterministic automaton: Read all the characters of the text one by one and modify in each step some variables that allow to identify the possible occurrences. Some similar algorithms are: Knuth-Morris-Pratt, Shift-Or, Shift-And).
The assembly of DNA strings is proposed as a combinatorial optimization problem and is classified as NP-hard and is based on the paradigm divide-and-conquer using a structure type farm, so that the computational cost of finding the sequence alignments and its splice is substantially reduced with respect to its sequential version. The number of processes required to process the fragments of DNA sequences of a specific genome such as that of a virus or bacteria is determined by the splice of the strings found by the sequential solution algorithm, which looks in parallel for overlaps in the remaining fragments. Two sub-strings of each

fragment are taken for comparison with other fragments; and thus, particular splices are located and associated with the processes (a process for each splice sequence of DNA strings). A splice graph is then generated that shows the relationship between pairs of nodes (processes), as well as the lack of communication among others. The set of nodes (processes) of the graph that are inter-related are grouped together within a worker process pattern farm. Each set of related nodes in the graph are independent and represent the grouping of fragments found.

6.1. Parallelization algorithm for the assembly of DNA sequences by an example

Given a genome which we will assume has been sequenced covering most of the structure, but missing a small fragment:

GTAGTCATGTTGAAAAACTTACGAGTAAATTACGTTGTCGA
GGGCGTGCAAGTAGCGCAACCCGTGACAAGCGCAAATTCG
GAAGTAT<u>**ACGCCAA**</u>TCTACCGCCTCCCGTACCCGCGGAGAC
GTATCAAACCGACGAAGATTACGAGGAAGATGACGGAGG
GTGGGC

Figure 11: Sequenced genome

Figure 11 shows the genome where it is assumed that the reading of the part underlining was not obtained. Assembling the readings of this genome is expected to obtain two contigs by not having been able to read the complete sequence (table 1).

Table 1: Read Contigs

Contig 1	Contig 2
GTAGTCATGTTGAAAAACTT	TCTACCGCCTCCCGTACCCGCG
ACGAGTAAATTACGTTGTCG	GAGACGTATCAAACCGACGAA
AGGGCGTGCAAGTAGCGCA	GATTACGAGGAAGATGACGG
ACCCGTGACAAGCGCAAATT	AGGGTGGGC
CGGAAGTAT	

As a result of the readings we obtain the list of fragments shown in table 2, where the splices are highlighted, the prefixes marked in green, the suffixes in red and in bold the splices that are overlapped by a suffix and prefix of other strings. A prefix and a suffix represent strings of length pb. The unmarked parts have no overlap, which indicates that those sections were read only once. The size of the desired splice is selected, for this example a splice of length 6 is used, the prefix and suffix of each string must be compared to the rest of the complete strings.

Table 2: Fragments Read

Process	Fragments read
1	GTAGTCATGTTGAAAAACTTACGAGT
2	ACGAGTAAATTACGTTGTCGAGGGCG
3	AGGGCGTGCAAGTAGCGCAACCCGTG
4	CCCGTGACAAGCGCAAATTCGGAAG
5	CGTGACAAGCGCAAATTCGGAAGTAT
6	TCTACCGCCTCCCGTACCCGCGGAGAC
7	CCGCCTCCCGTACCCGCGGAGACGTA
8	GACGTATCAAACCGACGAAGATTACGA
9	TTACGAGGAAGATGACGGAGGGTGGGC

The first column of the table shows the number of processes needed to process the fragments in a parallel way in the CPAN, which will be grouped later forming the working processes within the FARM of the CPAN, each of which will be constituted by a pipeline of related processes. Each of the nine processes is mapped to work with a string and look for overlaps in the remaining fragments. Two substrings (prefix and suffix) of each fragment are taken to compare them with the other fragments, thus obtaining the results shown in table 3, using the algorithm Myers.

Table 3: Result of the Myers algorithm by fragment

Fragment	Location of the splice found								
	1	2	3	4	5	6	7	8	9
1	-	**5**	27	25	26	27	26	27	27
2	**24**	-	**5**	25	26	27	26	27	27
3	26	**25**	-	**5**	26	27	26	27	27
4	26	25	**26**	-	**23**	27	26	27	27
5	26	25	27	**7**	-	27	26	27	27
6	26	25	27	25	26	-	**22**	27	27
7	26	25	27	25	26	**9**	-	**5**	27
8	26	25	27	25	26	27	**25**	-	**5**
9	26	25	27	25	26	27	26	**26**	-

The calculation of the overlaps is stored in a list for each process, where the incidence of overlaps is added at the end. The results of the exhaustive search are shown in Table 3. The results where splices in the prefix or suffix are found contain the location of the splicing within the indices of the string. Table 4 shows the list that represents the splicing graph constructed using the Myers algorithm. It can be seen that there is no relationship between the pair of nodes (5,6) and there is no arc between the nodes {1,2,3,4,5} and {6,7,8,9}. The graph obtained is shown in Figure 12 and will be represented in the CPAN with a process farm with worker processes and a controller process (figure 13). Each worker process is a CPAN Pipeline that represents each graph of the figure 12. The slave object associated with each CPAN FARM will perform an in-depth search to output the corresponding strings for each group.

Table 4: Graphs with found splices

Process	List of splices	
1	2	-
2	1	3
3	2	4
4	3	5
5	4	-
6	7	-
7	6	8
8	7	9
9	8	-

This is shown in Figure 12. It shows the number of groups formed (graphs) and the elements that belong to each group. For more details see (Hernández, Olmos and Olvera 2016).

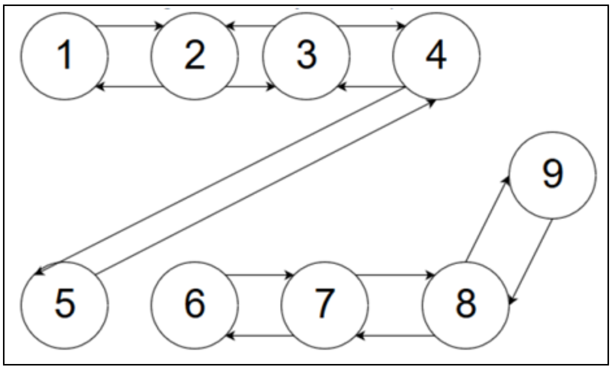


Figure 12: Graph splices

Finally the groups formed by this algorithm are shown in Table 5. This result can be put in a DNA sequence assembler to form the initial contigs of Table 1.

Table 5: Groups obtained by in-depth search

Group 1		Group 2	
1	GTAGTCATGTTGAAAAACTTACGAGT	6	TCTACCGCCTCCCGTACCCGCGGAGAC
2	ACGAGTAAATTACGTTGTCGAGGGCG	7	CCGCCTCCCGTACCCGCGGAGACGTA
3	AGGGCGTGCAAGTAGCGCAACCCGTG	8	GACGTATCAAACCGACGAAGATTACGA
4	CCCGTGACAAGCGCAAATTCGGAAG	9	TTACGAGGAAGATGACGGAGGGTGGGC
5	CGTGACAAGCGCAAATTCGGAAGTAT		

Similarly, the relationships between nodes within each worker FARM process can be represented as the patterns Pipeline. A double communication precisely represents the splices found in the DNA sequences. In Figure 13 is shown the representation of graph-splices as a CPAN.

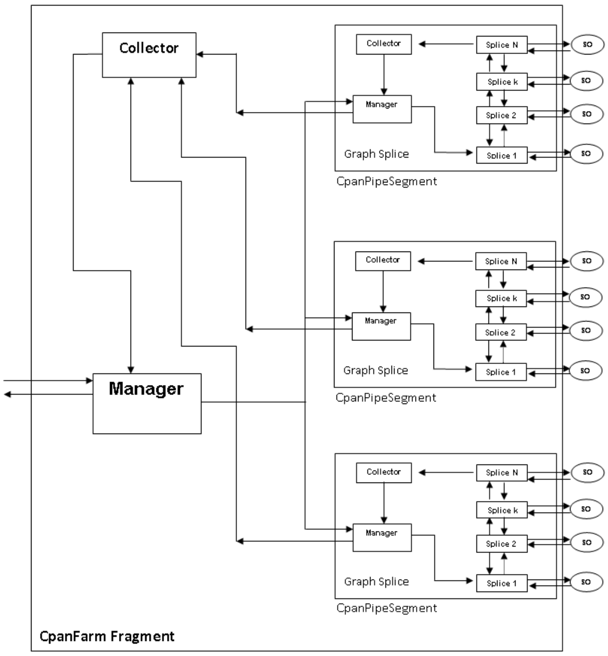


Figure 13: The CpanGraphADN

The new CPAN named CpanGraphADN is structured as a FARM of n-worker processes, i.e., n-fragments of DNA sequences and each worker process is itself a two directions-communication pipeline CPAN formed by m-stages where each stage of CpanPipe represents a splice sequence of DNA strings connected with both, the previous stage as the next stage. The collector object receives the number of formed groups and the elements that belong to each of the formed groups. With the latter information collected, a in-depth search is performed to locate these items and obtain the sequence groups formed by the sequential algorithm assigned to each of the CPAN's slave objects With this result, the user can use an assembly of DNA sequences to try to complete a particular genome or to finish an incomplete sequence of DNA strings of some animal or plant type species. In summary:

1. Two substrings are taken from each fragment of DNA sequences found to compare them with the remaining fragments. The splices are located and are associated with processes within a working process of CPAN Farm, in the new CpanGraphADN of figure 13.
2. The represented graph is processed in a Cpan Farm worker process using a Cpan Pipe and eliminates repetitions in the Cpan Farm Collector process once worker processes send their result to it.
3. It creates the hierarchy contigs, scaffolds, DNA sequence

6.2. CpanGraphADN utility, experiments and results

An experiment was designed by using the CpanGraphADN with genomes of viruses and bacteria available on the web see Table 6, whose data were obtained from European Nucleotide Archive, ENA http://www.ebi.ac.uk. Sequencer readings were simulated to create pseudo-random synthetic readings, so that the number of contigs can be formed, see (Vera and Gonzalez 2014).

Table 6: Genome used by the CpanGraphADN for experimentation

	Extracted characteristics		
Genome	Length	Number of readings	Genome coverage (average)
Avalone herpesvirus victoria	211519	12691	0.9999
Adoxophyes orana granulovirus	99658	5979	0.9997
Adoxophyes orana nucleopolyehedro	111725	6703	0.9996
African swine fever virus benin	182285	10937	0.9999
Feline coronavirus	29215	1752	0.9998
Shrimp white spot syndrome	307288	18437	0.9999
Trichoplusia ni ascovirus 2c	174060	10443	0.9999
Vaccinia virus GLU-1 h68	203058	12183	0.9998

The CPAN used a fragmented genome, and the resulting fragments were pooled ensuring the same number of groups readings and the follower contigs results were obtained, see table 7. However, it should be noticed that the same number of genomes groups on the contigs was not always obtained. Thus, in the last three contigs that read the genome, only 4 groups were obtained by grouping the simulated fragments readings in CPAN. Similarly, for the second genome only two groups were obtained and only one contig was read. For the remaining genomes, the number of expected readings searches (of different lengths) by the CPAN was obtained.

Table 7: DNA sequence groups found by the CPAN

		Joint length (pb)						Contigs
No.	Genome	15	17	23	29	40	56	read
1	Avalone herpesvirus victoria	2	2	2	4	16	33	2
2	Adoxophyes orana granulovirus	2	2	3	5	5	7	1
3	Adoxophyes orana nucleopolyehedro	1	2	2	2	4	9	2
4	African swine fever virus benin	1	2	3	4	8	23	2
5	Feline coronavirus	1	1	1	1	1	3	1
6	Shrimp white spot syndrome	1	1	1	3	8	35	1
7	Trichoplusia ni ascovirus 2c	1	2	2	3	8	13	2
8	Vaccinia virus GLU-1 h68	1	1	2	4	6	25	3

7. PERFORMANCE

Finally, the execution time of CpanGraphADN is shown in Figure 14 and Scalability of the speedup found is shown in figures 15 to 22. The plot shows the number of processes deployed for the calculation of eight genomes in an experiment conducted on a computer with Intel Core i8 processor, and using a video accelerator card with 1,664 CUDA cores and clock frequency 1,178 MHz.

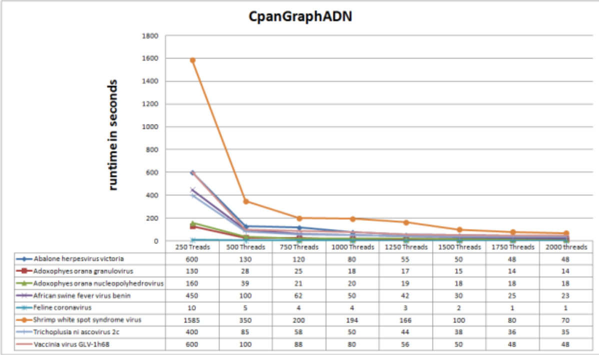


Figure 14: Runtime CpanGraphADN

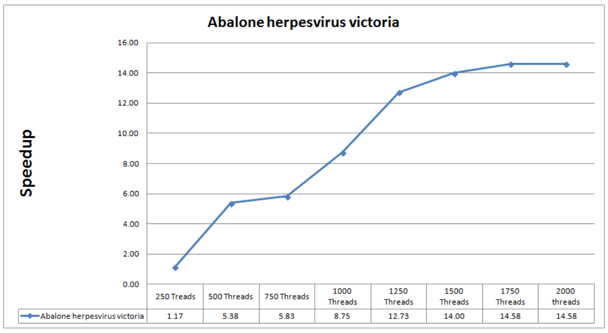


Figure 15: Scalability of the speedup found for the CpanGraphADN for the genome Abalone herpesvirus victoria

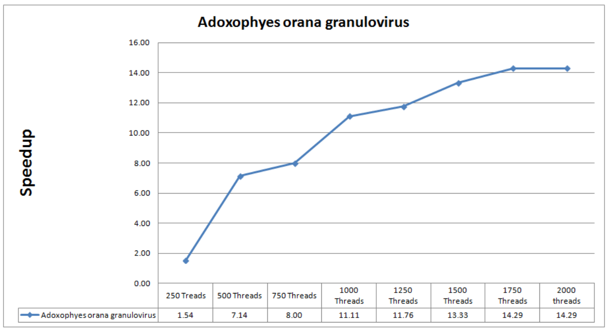


Figure 16: Scalability of the speedup found for the CpanGraphADN for the genome Adoxophyes orana granulovirus

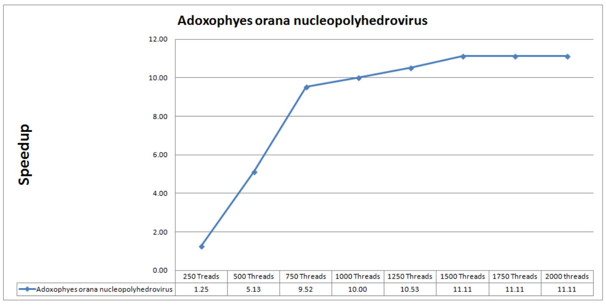


Figure 17: Scalability of the speedup found for the CpanGraphADN for the genome Adoxophyes orana nucleopolyhedrovirus

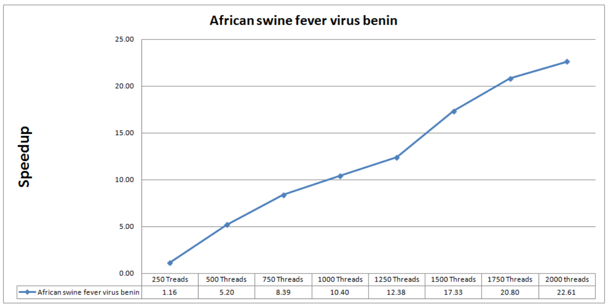


Figure 18: Scalability of the speedup found for the CpanGraphADN for the genome African swine fever virus benin

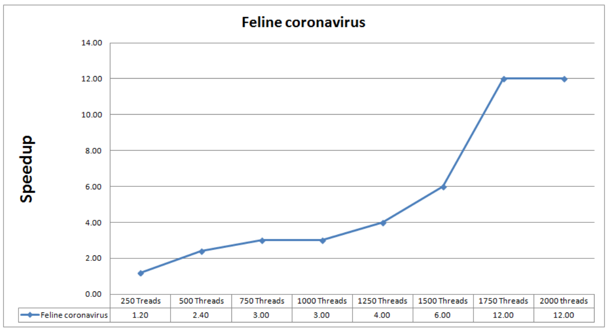


Figure 19: Scalability of the speedup found for the CpanGraphADN for the genome Feline coronavirus

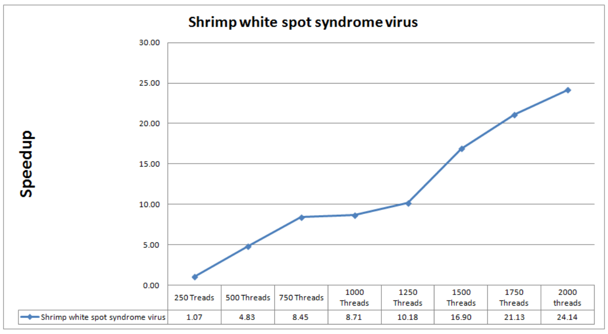


Figure 20: Scalability of the speedup found for the CpanGraphADN for the genome Shirimp white spot syndrome virus

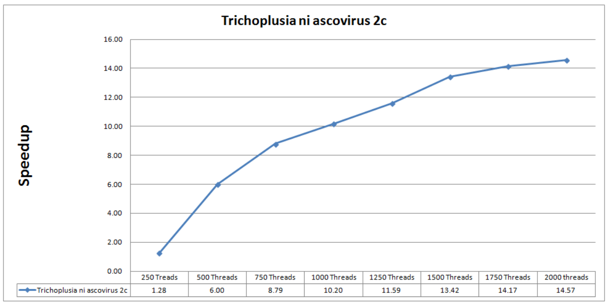


Figure 21: Scalability of the speedup found for the CpanGraphADN for the genome Trichoplusia ni ascovirus 2c

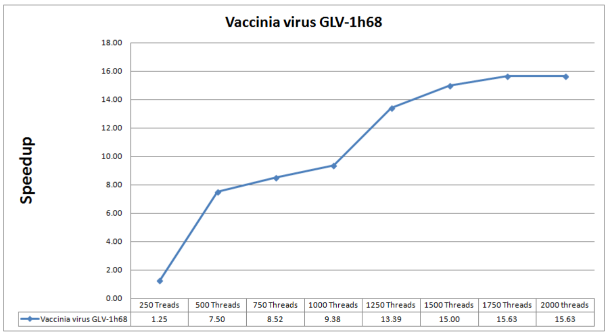


Figure 22: Scalability of the speedup found for the CpanGraphADN for the genome Vaccinia virus GLV-1h68

8. CONCLUSIONS

We have presented a method of design concurrent applications based on the CPAN construction that has been shown susceptible to be used in different platforms (not only with C ++ and POSIX threads). The CPANs have been exercised: Pipe and Farm extracted from a larger library and have shown their utility as communication patterns between concurrent processes, which can be used by programmers with little experience in Parallel Programming. The implemented CPANS can be exploited, thanks to the adoption of the approach oriented to objects. Well-known algorithms that solve sequential problems in algorithms parallelizable have transformed and with them the utility of CPANS has been proven.

We have presented as a case study the parallel calculation of the DNA sequences for 8 genomes and it has been demonstrated efficient, observing Amdahl acceleration, for a restricted range of the number of processors. In the future, more efficient methods will be used to solve the assembly problem: pairing of DNA fragments sequences and voracious algorithms on representation graphs to solve the splices instead of the divide & conquer technique.

REFERENCES

Andrews G.R., 2000. Foundations of Multithreaded, Parallel, and Distributed Programming, *Addison-Wesley*

Bacci, Danelutto, Pelagatti, Vaneschi, 1999. SklE: A Heterogeneous Environment for HPC Applications. *Parallel Computing 25*.

Blelloch, Guy E., 1996. Programming Parallel Algorithms. *Comunications of the ACM*. Volume 39, Number 3.

Brinch Hansen, 1993. Model Programs for Computational Science: A programming methodology for multicomputers, *Concurrency: Practice and Experience*, Volume 5, Number 5.

Barry W., Allen M., 1999. Parallel Programming. Techniques and Applications Using Networked Workstations and Parallel Computers. *Prentice Hall*. ISBN 0-13-671710-1.

Cañizares Sales J., Blanca J., Ziarsolo P. 2015. Ensamblaje y mapeo de secuencias tipo Sanger. *Bioinformatics & Genomics Group. Polytechnic*

University of Valencia. Valencia, España. https://bioinf.comav.upv.es/courses/intro_bioinf/ensamblaje.html
Corradi A., Leonardi L., 1991. PO Constraints as tools to synchronize active objects. *Journal Object Oriented Programming* 10, pp. 42-53.
Corradi A, Leonardo L, Zambonelli F., 1995. Experiences toward an Object-Oriented Approach to Structured Parallel Programming. *DEIS technical* report no. DEIS-LIA-95-007.
Danelutto, M.; Orlando, S; et al., 1995. Parallel Programming Models Based on Restricted Computation Structure *Approach. Technical* Report-Dpt. Informatica. Universitá de Pisa.
DanishAli, S. and Farooqui, Z. 2013. Approximate Multiple Pattern String Matching using Bit Parallelism: *A Review. International Journal of Computer Applications*, Volume 74, No.19: pp.47-51.
Darlington et al., 1993, Parallel Programming Using Skeleton Functions. *Proceedings PARLE'93*, Munich (D).
De Simone, et al. 1997. Designs Patterns for Parallel Programming. *PDPTA International Conference*.
Días J.D. 2011. Estrategia de solución al problema de la anotación de secuencias de ADN mediante la metodología CommonKADS. Disertación Master en Inteligencia Artificial. *Facultad de Informática. Universidad Complutense de Madrid*. Madrid, España.
Hernandez R., Olmos I., Olvera A. 2016. Agrupación de fragmentos de secuencias de ADN a partir de Técnicas paralelas para tareas de ensamblaje de genomas. *Disertacion de Tesis de Licenciatura*. Universidad Autónoma de Puebla. México.
Lavander G.R., Kafura D.G. 1995. A Polimorphic Future and First-class Function Type for Concurrent Object-Oriented Programming. *Journal of Object-Oriented Systems*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.7088&rep=rep1&type=pdf
Masoudi-Nejad, A., Narimani, Z. and Hosseinkhan, N. 2013. Next Generation Sequencing and Sequence Assembly. *SpringerBriefs in Systems Biology*.
Monya Baker, 2012. De novo genome assembly: what every biologist should know. *Nature America Inc*, Volume 9. No. 4: Pp. 333-337. ISSN: 1548-7091. http://www.nature.com/nmeth/journal/v9/n4/full/nmeth.1935.html
Pareek, C., Smoczynski, R. and Tretyn, A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, Volume 52, No.4: pp.413-435.
Roosta, Séller, 1999. Parallel Processing and Parallel Algorithms. *Theory and Computation. Springer*.
Rossainz, M., 2005. Una Metodología de Programación Basada en Composiciones Paralelas de Alto Nivel (CPANs). *Universidad de Granada, PhD dissertation*, 02/25/2005.
Rossainz, M., Capel M., 2008. A Parallel Programming Methodology using Communication Patterns named CPANS or Composition of Parallel Object. *20TH European Modeling & Simulation* Symposium.Campora S. Giovanni. Italy.
Rossainz, M., Capel M., 2012. Compositions of Parallel Objects to Implement Communication Patterns. *XXIII Jornadas de Paralelismo. SARTECO 2012*. Septiembre de 2012. Elche, España.
Rossainz M., Capel M., 2014. Approach class library of high level parallel compositions to implements communication patterns using structured parallel programming. *26TH European Modeling & Simulation* Symposium.Campora Bordeaux, France.
Rossainz M, Capel M., Domínguez P., 2015. Pipeline as high level parallel composition for the implementation of a sorting algorithm. *27TH European Modeling & Simulation* Symposium.Campora Bergeggi, Italy.
Samiksha S., 2016. Structure of Chromosome: Size and Share.*YourArticleLibrary.com*. http://www.yourarticlelibrary.com/biology/structure-of-chromosome-size-and-share-474-words/6648/
Vera, F. and González, B. 2014. LPS: una estrategia de ensambles de secuencias cortas de ADN, Disertacion de examen de master. *Instituto Nacional de Astrofsica, Óptica y Electrónica*, Tonantzintla Pue., Mexico.
Wilkinson B., Allen M., 1999. Parallel Programming Techniques and Applications Using Networked Workstations and Parallel Computers. *Prentice-Hall*. USA.

FRACTIONAL POSITIVE AND STABLE TIME-VARYING CONTINUOUS-TIME LINEAR ELECTRICAL CIRCUITS

Tadeusz Kaczorek

Bialystok University of Technology
Faculty of Electrical Engineering
Wiejska 45D, 15-351 Białystok
kaczorek@ee.pw.edu.pl

ABSTRACT
The positivity and asymptotic stability of fractional time-varying continuous-time linear electrical circuits is addressed. Necessary and sufficient conditions for the positivity and sufficient conditions for asymptotic stability of the electrical circuits are established. New definitions of the capacitance and inductance of fractional time-varying linear electrical circuits are proposed. It is shown that there exists a large class of fractional positive and stable linear electrical circuits with time-varying parameters. Examples of positive and stable linear electrical circuits are presented.

Keywords: linear circuits, time varying circuits, positive circuits, stability

1. INTRODUCTION

A dynamical system is called positive if its trajectory starting from any nonnegative initial state remains forever in the positive orthant for all nonnegative inputs. An overview of state of the art in positive theory is given in the monographs (Farina and Rinaldi 2000, Kaczorek 2002). Variety of models having positive behavior can be found in engineering, economics, social sciences, biology and medicine, etc..

The Lyapunov, Bohl and Perron exponents and stability of time-varying discrete-time linear systems have been investigated in (Czornik, Newrat, Niezabitowski, and Szyda 2012; Czornik and Niezabitowski 2013a, 2013b, 2013c; Czornik, Newrat, and Niezabitowski 2013; Czornik, Klamka, and Niezabitowski 2014). Controllability, observability and reachability of linear standard, positive and fractional electrical circuits (Kaczorek 2011a, 2011b, 2011c). The new stability tests of positive and fractional linear systems have been proposed in (Kaczorek 2011d). Fractional and positive continuous-time systems have been addressed in (Kaczorek 2008) and fractional descriptor standard and positive time-varying systems in (Kaczorek 2015a). Positivity and stability of standard and fractional descriptor time-varying discrete-time have been investigated in (Kaczorek 2015b, 2015c). Positive linear systems consisting of n subsystems with different fractional orders have been analyzed in (Kaczorek 2011e). Positivity and stability of time-varying of discrete-time and continuous-time systems have been considered in (Kaczorek 2015d, 2015e, 2015f). Stability of positive continuous-time linear systems with delays have been investigated in (Kaczorek 2009). Stability and stabilization of positive fractional linear systems by state-feedbacks have been addressed in (Kaczorek 2010).

In this paper positivity and asymptotic stability of fractional time-varying continuous-time linear electrical systems will be addressed.

The paper is organized as follows. In section 2 necessary and sufficient conditions for the positivity and sufficient conditions for the asymptotic stability of fractional time-varying continuous-time linear systems are established. The fractional positive electrical circuits with time-varying parameter are addressed in section 3. Concluding remarks are given in section 4.

The following notation will be used: $\Re$ - the set of real numbers, $\Re^{n\times m}$ - the set of $n\times m$ real matrices, $\Re_+^{n\times m}$ - the set of $n\times m$ matrices with nonnegative entries and $\Re_+^n = \Re_+^{n\times 1}$, M_n - the set of $n\times n$ Metzler matrices (real matrices with nonnegative off-diagonal entries), I_n - the $n\times n$ identity matrix.

2. PRELIMINARIES

Consider the fractional time-varying linear system

$$\frac{d^\alpha x}{dt^\alpha} = A(t)x \tag{1}$$

where $x = x(t) \in \Re^n$ and $A(t) \in \Re^{n\times n}$ with entries a_{ij} being continuous-time functions of time $t \in [0,+\infty)$ and

$$\frac{d^\alpha x(t)}{dt^\alpha} = {}_0D_t^\alpha x(t) = \frac{1}{\Gamma(1-\alpha)}\int_0^t (t-\tau)^{-\alpha}\frac{dx}{d\tau}d\tau, \quad 0<\alpha<1, \tag{2}$$

is the Caputo definition of $\alpha \in \Re$ order derivative of $x(t)$ and

$$\Gamma(\alpha)=\int_0^\infty e^{-t}t^{\alpha-1}dt \quad (3)$$

is the Euler gamma function.

To find the matrix $X = X(t) \in \Re^{n\times n}$ of solutions to the equation (1) the following extension of the Picard method to fractional equations will be used

$$\frac{d^\alpha X_k}{dt^\alpha} = A(t)X_{k-1} \text{ for } k=1,2,... \quad (4)$$

with $X_0(0) = I_n$.

From (2) and (4) we obtain

$$X_k = I_n + \frac{1}{\Gamma(1+\alpha)}\int_0^t (t-\tau)^\alpha A(\tau)X_{k-1}(\tau)d\tau, \quad (5)$$

and $0<\alpha<1$, $k=1,2,...$.

Using (5) for $k=1,2,...$ we obtain

$$X = X(t) = I_n + \frac{1}{\Gamma(1+\alpha)}\int_0^t (t-\tau)^\alpha A(\tau)d\tau + \frac{1}{[\Gamma(1+\alpha)]^2}\int_0^t (t-\tau)^\alpha A(\tau)\int_0^\tau (t-\tau_1)^\alpha A(\tau_1)d\tau_1 d\tau + ... \quad (6)$$

Using (5) in the form

$$X_k = I_n + \int_0^t \overline{A}(\tau)X_{k-1}(\tau)d\tau \quad (7a)$$

where

$$\overline{A}(\tau) = \frac{(t-\tau)^\alpha A(\tau)}{\Gamma(1+\alpha)}, \; 0 \le \tau \le t \quad (7b)$$

the proof can be accomplished in a similar way as in (Gantmacher 1959, Idezak and Kamocki 2011, Kaczorek 2009) since $\overline{A}(\tau)$ is continuous and bounded function of τ. □

Definition 1. The fractional time-varying linear system (1) is called the internally positive fractional system if $x(t)\in\Re_+^n$, $t \ge 0$ for any initial conditions $x_0 \in \Re_+^n$.

Theorem 1. The fractional time-varying linear system (1) is internally positive if and only if

$$A(t) \in M_n \text{ for } t \in [0,+\infty). \quad (8)$$

Proof. From (7b) it follows that $\overline{A}(t)\in M_n$ if and only if the condition (8) is satisfied. In (Kaczorek 2009) it was shown that the time-varying linear system

$$\dot{x}(t) = \overline{A}(t)x \quad (9)$$

is positive if and only if $\overline{A}(t)\in M_n$. □

In a similar way as in (Gantmacher 1959, Kaczorek 2009) the above considerations can be easily extended to the fractional time-varying linear systems

$$\frac{d^\alpha x(t)}{dt^\alpha} = A(t)x(t) + B(t)u(t) \quad (10a)$$

$$y(t) = C(t)x(t) + D(t)u(t) \quad (10b)$$

where $x(t)\in\Re^n$, $u(t)\in\Re^m$, $y(t)\in\Re^p$ are the state, input and output vectors and $A(t)\in\Re^{n\times n}$, $B(t)\in\Re^{n\times m}$, $C(t)\in\Re^{p\times n}$, $D(t)\in\Re^{p\times m}$ are real matrices with entries depending continuously on time $t\in[0,+\infty)$.

Definition 2. The fractional time-varying linear system (10) is called positive if $x(t)\in\Re_+^n$, $y(t)\in\Re_+^p$, $t\in[0,+\infty)$ for any initial conditions $x_0\in\Re_+^n$ and all inputs $u(t)\in\Re_+^m$, $t\in[0,+\infty)$.

Theorem 2. The fractional time-varying linear system (10) is positive if and only if

$$A(t)\in M_n,\; B(t)\in\Re_+^{n\times m},$$
$$C(t)\in\Re_+^{p\times n},\; D(t)\in\Re_+^{p\times m},\; t\in[0,+\infty). \quad (11)$$

3. FRACTIONAL POSITIVE TIME-VARYING LINEAR ELECTRICAL CIRCUITS

It is well-known that for fractional time-varying electrical circuits we have

$$\frac{d^\alpha C(t)u(t)}{dt^\alpha} \ne \frac{d^\alpha C(t)}{dt^\alpha}u(t) + C(t)\frac{d^\alpha u(t)}{dt^\alpha} \quad (12a)$$

and

$$\frac{d^\alpha L(t)i(t)}{dt^\alpha} \ne \frac{d^\alpha L(t)}{dt^\alpha}i(t) + L(t)\frac{d^\alpha i(t)}{dt^\alpha} \quad (12b)$$

where $\alpha\in\Re$ is the fractional order derivative, $C(t)$ is the capacitance, $L(t)$ is the inductance, $u(t)$ is the voltage and $i(t)$ is the current.

To overcome this drawback the following definitions of the capacitance and inductance for fractional time-varying linear electrical circuits are proposed.

Definition 3. The ratio of the current $i(t)$ and α-order derivative of voltage $u(t)$, $\frac{d^\alpha u(t)}{dt^\alpha}$ i.e.

$$C(t) = \frac{i(t)}{\frac{d^\alpha u(t)}{dt^\alpha}} \quad (13)$$

is called the capacitance of fractional electrical capacitor.

Definition 4. The ratio of the voltage $u(t)$ and α-order derivative of current $i(t)$, $\frac{d^\alpha i(t)}{dt^\alpha}$ i.e.

$$L(t) = \frac{u(t)}{\frac{d^\alpha i(t)}{dt^\alpha}} \tag{14}$$

is called the inductance of fractional electrical inductor (coil).

Example 1. Consider the fractional electrical circuit shown in Figure 1 with given resistances $R_1(t)$, $R_2(t)$, $R_3(t)$, inductances $L_1(t)$, $L_2(t)$ and source voltages $e_1(t) \geq 0$, $e_2(t) \geq 0$ for $t \in [0,+\infty)$. $0 < \alpha < 1$

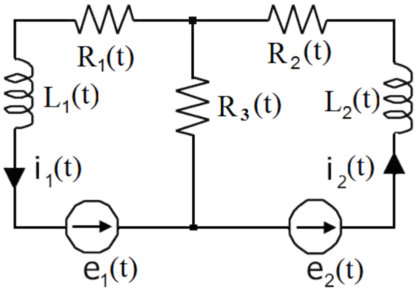


Figure 1: Electrical circuit.

Using the Definition 4 and Kirchhoff's laws we can write the equations

$$\begin{aligned} e_1(t) &= R_3(t)[i_1(t) - i_2(t)] + R_1(t)i_1(t) + L_1(t)\frac{d^\alpha i_1(t)}{dt^\alpha}, \\ e_2(t) &= R_3(t)[i_2(t) - i_1(t)] + L_2(t)\frac{d^\alpha i_2(t)}{dt^\alpha} + R_2(t)i_2(t) \end{aligned} \tag{15}$$

which can be written in the form

$$\frac{d^\alpha}{dt^\alpha}\begin{bmatrix} i_1(t) \\ i_2(t) \end{bmatrix} = A(t)\begin{bmatrix} i_1(t) \\ i_2(t) \end{bmatrix} + B(t)\begin{bmatrix} e(t) \\ e(t) \end{bmatrix} \tag{16a}$$

where

$$A(t) = \begin{bmatrix} -\frac{R_1(t)+R_3(t)}{L_1(t)} & \frac{R_3(t)}{L_1(t)} \\ \frac{R_3(t)}{L_2(t)} & -\frac{R_2(t)+R_3(t)}{L_2(t)} \end{bmatrix}, \quad B(t) = \begin{bmatrix} \frac{1}{L_1(t)} & 0 \\ 0 & \frac{1}{L_2(t)} \end{bmatrix}. \tag{16b}$$

The electrical circuit is positive, since the matrix A is the Metzler matrix and the matrix B has nonnegative entries.

It is easy to show (Kaczorek 2015g) that in general case the linear fractional electrical circuit composed of time-varying resistors, coils and voltage sources is positive for any values of the resistances, inductances and source voltages if and only if the number of coils is less or equal to the number of its linearly independent meshes and the directions of the mesh currents are consistent with the directions of the mesh source voltages.

Example 2. Consider the fractional time-varying electrical circuit shown in Fig. 1 with given nonzero resistances $R_1(t)$, $R_2(t)$, $R_3(t)$, inductance $L(t) > 0$, capacitance $C(t) > 0$ and source voltage $e(t) \geq 0$ for $t \in [0,+\infty)$.

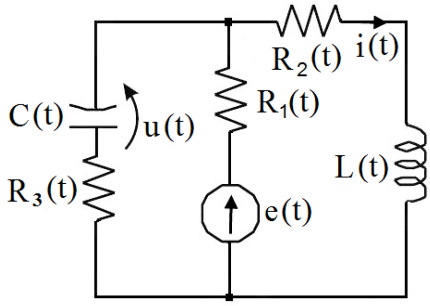


Figure 2: Fractional time-varying electrical circuit

Using the Definitions 3, 4 and Kirchhoff's laws, we can write the equation

$$\begin{aligned} e(t) &= R_1(t)\left[i(t) + C(t)\frac{d^\alpha u(t)}{dt^\alpha}\right] + R_3(t)\left[C(t)\frac{d^\alpha u(t)}{dt^\alpha}\right] \\ &+ u(t), \\ e(t) &= R_1(t)\left[i(t) + C(t)\frac{d^\alpha u(t)}{dt^\alpha}\right] + L(t)\frac{d^\alpha i(t)}{dt^\alpha} \\ &+ R_2(t)i(t). \end{aligned} \tag{17}$$

The equations (17) can be written in the form

$$\frac{d^\alpha}{dt^\alpha}\begin{bmatrix} i(t) \\ u(t) \end{bmatrix} = A(t)\begin{bmatrix} i(t) \\ u(t) \end{bmatrix} + B(t)e(t), \tag{18a}$$

where

$$A(t)=\begin{bmatrix}0 & [R_1(t)+R_3(t)]C(t)\\ L(t) & R_1(t)C(t)\end{bmatrix}^{-1}\begin{bmatrix}-R_1(t) & -1\\ -R_1(t)-R_2(t) & 0\end{bmatrix}$$
$$=\begin{bmatrix}\frac{R_1^2(t)}{[R_1(t)+R_3(t)]L(t)}-\frac{R_1(t)+R_2(t)}{L(t)} & -\frac{R_1(t)}{[R_1(t)+R_3(t)]L(t)}\\ -\frac{R_1(t)}{[R_1(t)+R_3(t)]C(t)} & -\frac{1}{[R_1(t)+R_3(t)]C(t)}\end{bmatrix},$$
$$B(t)=\begin{bmatrix}0 & [R_1(t)+R_3(t)]C(t)\\ L(t) & R_1(t)C(t)\end{bmatrix}^{-1}\begin{bmatrix}1\\ 1\end{bmatrix}$$
$$=\begin{bmatrix}\frac{R_3(t)}{[R_1(t)+R_3(t)]L(t)}\\ \frac{1}{[R_1(t)+R_3(t)]C(t)}\end{bmatrix}. \tag{18b}$$

From (18b) it follows that $A(t)\in M_2$ if and only if $R_1(t)=0$ for $t\in[0,+\infty)$. Therefore, the electrical circuit is a fractional positive time-varying system if and only if $R_1(t)=0$ for $t\in[0,+\infty)$.

Now let us consider electrical circuit shown on Fig. 3 with given positive resistances $R_k(t)$, $k=0,1,\dots,n$, inductances $L_i(t)$, $i=2,4,\dots,n_2$, capacitances $C_j(t)$, $j=1,3,\dots,n_1$ depending on time t and source voltages $e_1(t),e_2(t),\dots,e_n(t)$. We shall show that this electrical circuit is a fractional positive time-varying linear system.

Using the Definitions 3, 4 and Kirchhoff's law we can write the equations

$$e_1(t)=R_k(t)C_k(t)\frac{d^\alpha u_k(t)}{dt^\alpha}+u_k(t) \tag{19a}$$
for $k=1,3,\dots,n_1$,
$$e_1(t)+e_k(t)=L_k(t)\frac{d^\alpha i_k(t)}{dt^\alpha}+R_k(t)i_k(t)+u_k(t) \tag{19b}$$
for $k=2,4,\dots,n_2$,

The equations (19) can be written in the form

$$\frac{d^\alpha}{dt^\alpha}\begin{bmatrix}u(t)\\ i(t)\end{bmatrix}=A(t)\begin{bmatrix}u(t)\\ i(t)\end{bmatrix}+B(t)e(t) \tag{20a}$$

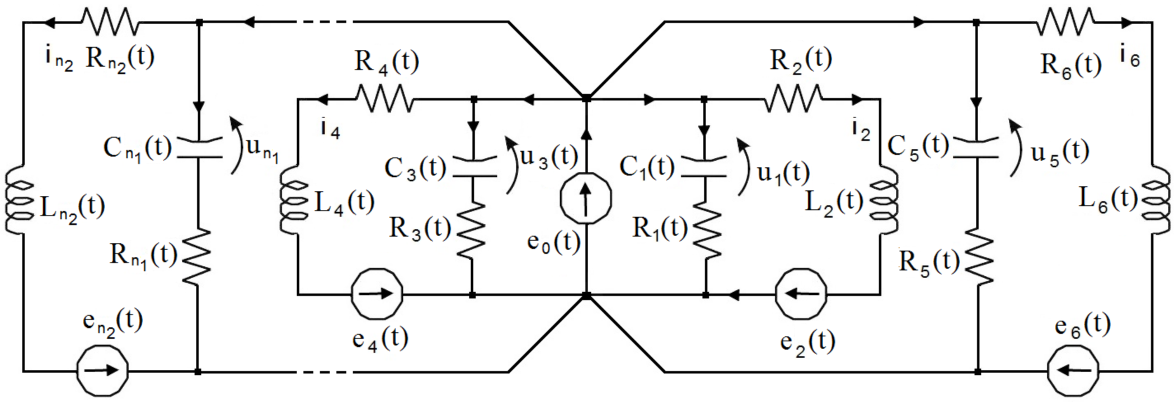


Figure 3: Fractional positive time-varying electrical circuit.

where

$$u(t)=\begin{bmatrix}u_1(t)\\ u_3(t)\\ \vdots\\ u_{n_1}(t)\end{bmatrix},\quad i(t)=\begin{bmatrix}i_2(t)\\ i_4(t)\\ \vdots\\ i_{n_2}(t)\end{bmatrix},\quad e(t)=\begin{bmatrix}e_1(t)\\ e_3(t)\\ \vdots\\ e_n(t)\end{bmatrix}, \tag{20b}$$
$(n=n_1+n_2)$

and

$A(t)=\mathrm{diag}[-a_1(t),-a_3(t),\dots,-a_{n_1}(t),-a_2(t),-a_4(t),\dots,-a_{n_2}(t)]$,

$$a_k(t)=\frac{1}{R_k(t)C_k(t)}\ \text{for}\ k=1,3,\dots,n_1,$$
$$a_k(t)=\frac{R_k(t)}{L_k(t)}\ \text{for}\ k=2,4,\dots,n_2,$$

$$B(t)=\begin{bmatrix}B_1(t)\\ B_2(t)\end{bmatrix},\quad B_1(t)=\begin{bmatrix}\frac{1}{R_1(t)C_1(t)} & 0 & 0 & \dots & 0\\ \frac{1}{R_3(t)C_3(t)} & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \dots & \vdots\\ \frac{1}{R_{n_1}(t)C_{n_1}(t)} & 0 & 0 & \dots & 0\end{bmatrix}, \tag{20c}$$
$$B_2(t)=\begin{bmatrix}\frac{1}{L_2(t)} & \frac{1}{L_2(t)} & 0 & \dots & 0\\ \frac{1}{L_4(t)} & 0 & \frac{1}{L_4(t)} & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \frac{1}{L_{n_2}(t)} & 0 & 0 & \dots & \frac{1}{L_{n_2}(t)}\end{bmatrix}.$$

The electrical circuit is a fractional positive time-varying linear system since all diagonal entries of the matrix $A(t)$ are negative functions of $t \in [0,+\infty)$ and the matrix $B(t)$ has nonnegative entries for $t \in [0,+\infty)$.

4. CONCLUDING REMARKS

The positivity of fractional time-varying continuous-time linear electrical circuits have been addressed. Necessary and sufficient conditions for the positivity of the system and electrical circuits have been established. New definitions of the capacitance and inductance of fractional time-varying electrical circuits have been proposed. It has been shown that there exists a large class of fractional positive electrical circuits with time-varying parameters. The considerations have been illustrated by fractional positive electrical circuits. The consideration can be extended to a class of fractional time-varying nonlinear electrical circuits.

ACKNOWLEDGMENTS

This work was supported by National Science Centre in Poland under work No. 2014/13/B/ST7/03467.

REFERENCES

Czornik A., Klamka J., Niezabitowski M., 2014. On the set of Perron exponents of discrete linear systems, Proc. of World Congres of the 19th International Federation of Automatic Control, Kapsztad, 11740-11742.

Czornik A., Newrat A., Niezabitowski M., 2013. On the Lyapunov exponents of a class of the second order discrete time linear systems with bounded perturbations, Dynamical Systems: an International Journal, vol. 28, no. 4, 473-483.

Czornik A., Newrat A., Niezabitowski M., and Szyda A., 2012. On the Lyapunov and Bohl exponent of time-varying discrete linear systems, 20th Mediterranean Conf. on Control and Automation (MED), Barcelona, 194-197.

Czornik A., Niezabitowski M., 2013a. Lyapunov Exponents for Systems with Unbounded Coefficients, Dynamical Systems: an International Journal, vol. 28, no. 2, 140-153.

Czornik A., Niezabitowski M., 2013b. On the stability of discrete time-varying linear systems, Nonlinear Analysis: Hybrid Systems, vol. 9, 27-41.

Czornik A., Niezabitowski M., 2013c. On the stability of Lyapunov exponents of discrete linear system, Proc. of European Control Conf., Zurich, 2210-2213.

Farina L., Rinaldi S., 2000. Positive Linear Systems; Theory and Applications, J. Wiley, New York.

Gantmacher F. R., 1959. The Theory of Matrices, Chelsea Pub. Comp., London.

Idezak D., Kamocki R., 2011. On the existence and uniquness and formula for solution of R-L fractional Cauchy problem, Fractional Calculus and Applied Analysis, vol. 14, no. 4, 538-553.

Kaczorek T., 2002, Positive 1D and 2D Systems, Springer Verlag, London.

Kaczorek T., 2008. Fractional positive continuous-time linear systems and their reachability, Int. J. Appl. Math. Comput. Sci., vol. 18, no. 2, 223-228.

Kaczorek T., 2009. Stability of positive continuous-time linear systems with delays, Bull. Pol. Acad. Sci. Techn., vol. 57, no. 4, 395-398.

Kaczorek T., 2010. Stability and stabilization of positive fractional linear systems by state-feedbacks, Bull. Pol. Acad. Sci. Techn., vol. 58, no. 4.

Kaczorek T. 2011, Selected Problems of Fractional System Theory, Springer Verlag.

Kaczorek T., 2011a. Controllability and observability of linear electrical circuits, Electrical Review, vol. 87, no. 9a, 248-254.

Kaczorek T., 2011b. Positive electrical circuits and their reachability, Archives of Electrical Engineering, vol. 60, no. 3, 283-301.

Kaczorek T., 2011c. Positivity and reachability of fractional electrical circuits, Acta Mechanica et Automatica, vol. 5, no. 2, 42-51, 2011.

Kaczorek T., 2011d. New stability tests of positive standard and fractional linear systems, Circuits and Systems, no. 2, 261-268.

Kaczorek T., 2011e. Positive linear systems consisting of n subsystems with different fractional orders, IEEE Trans. Circuits and Systems, vol. 58, no. 6, 1203-1210.

Kaczorek T., 2015a. Fractional standard and positive descriptor time-varying discrete-time linear systems, in Progress in Automation, Robotics and Measuring Techniques, Advances in Intelligent Systems and Computing Vol. 350, 101-112.

Kaczorek T., 2015b. Positive descriptor time-varying discrete-time linear systems and their asymptotic stability, Proc. of Conf. TransNav.

Kaczorek T., 2015c. Positivity and stability of fractional descriptor time-varying discrete-time linear systems, Int. J. Appl. Math. Comput. Sci., vol. 26, no. 1, 5-13.

Kaczorek T., 2015d. Positivity and stability of time-varying discrete-time linear systems, Intelligent Information and Database Systems, Lecture Notes in Computer Science, Vol. 9011, 295-303.

Kaczorek T., 2015e. Positive and stable time-varying continuous-time linear systems and electrical circuits, Computer Applications in Electrical Engineering, no. 81, 11-19.

Kaczorek T., 2015f. Normal positive electrical circuits, IET Control Theory and Applications, vol. 9, no. 5, 691-699.

AUTHORS BIOGRAPHY

Tadeusz Kaczorek, born 27.04.1932 in Elzbiecin (Poland), received the MSc., PhD and DSc degrees from Electrical Engineering of Warsaw University of Technology in 1956, 1962 and 1964, respectively. In the period 1968 - 69 he was the dean of Electrical Engineering Faculty and in the period 1970 - 73 he was the prorector of Warsaw University of Technology. Since 1971 he has been professor and since 1974 full professor at Warsaw University of Technology. In 1986 he was elected a corresp. member and in 1996 full member of Polish Academy of Sciences. In the period 1988 - 1991 he was the director of the Research Centre of Polish Academy of Sciences in Rome. In June 1999 he was elected the full member of the Academy of Engineering in Poland. In May 2004 he was elected the honorary member of the Hungarian Academy of Sciences. He was awarded by the title doctor honoris causa by 13 Universities.

His research interests cover the theory of systems and the automatic control systems theory, specially, singular multidimensional systems, positive multidimensional systems and singular positive and positive 1D and 2D systems. He has initiated the research in the field of singular 2D and positive 2D linear systems and in the fractional positive linear and nonlinear systems. He has published 28 books (7 in English) and over 1100 scientific. He has presented more than 100 invited papers on international conferences and world congresses. He has given invited lectures in more than 50 universities in USA, Canada, UK, Germany, Italy, France, Japan, Greece etc. He has been a member of many international committees and programme committees.

He supervised 69 Ph.D. theses. More than 20 of this PhD students became professors in USA, UK and Japan. He is editor-in-chief of Bulletin of the Polish Academy of Sciences, Techn. Sciences and editorial member of about ten international journals.

Analysing collaborative performance and cost allocation for the joint route planning problem

L. Verdonck[(a)], K. Ramaekers[(b)], B. Depaire[(c)], A. Caris[(d)] and G.K. Janssens[(e)]

[(a),(b),(c),(d),(e)] UHasselt, research group Logistics, Agoralaan, 3590 Diepenbeek, Belgium
[(a)] Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussel, Belgium

[(a)]lotte.verdonck@uhasselt.be, [(b)]katrien.ramaekers@uhasselt.be,
[(c)]benoit.depaire@uhasselt.be, [(d)]an.caris@uhasselt.be, [(d)]gerrit.janssens@uhasselt.be

ABSTRACT
Although organisations become increasingly aware of the inevitable character of horizontal collaboration, surveys report failure rates up to 70 percent for starting strategic partnerships. While a growing body of research acknowledges the importance of the partner selection and cost allocation process, no extensive study has been performed on the numerical relationship between specific company traits, applied allocation mechanisms and collaborative performance. This paper investigates the impact of coalition characteristics on attainable collaborative savings and cost allocation values in a joint route planning context. Results indicate that collaborative order sharing provides significant operational benefits and can be based on intuitively appealing cost allocation techniques, which may reduce alliance complexity and enforce the strength of mutual partner relationships.

Keywords: horizontal carrier collaboration, joint route planning, partner selection, cost allocation

1. INTRODUCTION

Horizontal collaboration between transport companies has become an important research area, since severe competition in global markets, rising costs, a growing body of transport legislation and heightened customer expectations have caused profit margins of organisations to shrink. (Cruijssen et al. 2007b). Horizontal logistics cooperation may be defined as collaboration between two or more firms that are active at the same level of the supply chain and perform comparable logistics functions (Cruijssen et al. 2007c). Through partnering with fellow transport organisations, carriers may extend their resource portfolio, reinforce their market position, enhance their service levels and create a more efficient transport planning (Hernández et al. 2011).

Although transport companies become increasingly aware of the inevitable character of collaboration, surveys report failure rates from 50 to 70 percent for starting partnerships (Schmoltzi and Wallenburg 2011). The success of achieving collaborative benefits strongly depends on the degree of fit between cooperation participants (Verstrepen et al. 2009). While a growing body of collaboration research acknowledges the importance of partner characteristics (Cruijssen et al. 2007a, Lozano et al. 2013, Guajardo and Rönnqvist 2015, Guajardo et al. 2016), no extensive study has been performed on the numerical relationship between specific company traits and the performance of the alliances in which these carriers are involved. The first goal of this paper is thus to provide practical recommendations on which partnership structures may provide the highest collaborative benefits.

Although selecting the right partners is crucial for the success of any horizontal alliance, it is not sufficient to guarantee long-term coalition stability. Dividing the collaborative gains in a fair manner constitutes a key issue, since the proposed allocation mechanism should induce partners to behave according to the collaborative goal and may improve cooperation stability. However, a wide range of possible allocation mechanisms have been developed in recent literature, each with its specific benefits, drawbacks and fairness properties. In this context, the second goal of this paper is to perform an extensive comparative analysis examining the applicability and suitability of different cost allocation methods in varying cooperation scenarios.

Based on the literature review on collaborative logistics described in Verdonck et al. (2013), the impact of coalition characteristics on attainable collaborative savings and cost allocation values will be investigated in a joint route planning context. In the majority of carrier alliances customer orders from all participating carriers are combined and collected in a central pool and efficient route schemes are set up for all requests simultaneously using appropriate vehicle routing techniques (Cruijssen et al. 2007a). Since existing studies mainly focus on demonstrating the benefits associated with joint route planning, an empirical analysis of the influence of cooperation structure on partnership performance and the impact of cost allocation mechanisms on coalition stability could provide useful insights for carriers considering collaboration.

The novelty of this paper lies in the application and empirical analysis of an existing routing problem in a practical context with the aim of providing guidelines to practitioners. Using a well-known statistical research method, recommendations can be made to transport organisations considering joint route planning on how they should tackle the partner selection and gain sharing process.

The remainder of this paper is organised as follows. First, current research on joint route planning, partner selection and cost allocation is summarised. Second, the research methodology is described. Third, results of a factorial ANOVA on the impact of coalition characteristics and cost allocation mechanisms are presented and discussed. Finally, conclusions and possible directions for future research are formulated.

2. STATE OF THE ART

2.1. Joint Route Planning

Scientific research on horizontal carrier collaboration can be divided into two main research streams: order sharing and capacity sharing (Verdonck et al. 2013). The majority of carrier cooperation literature focuses on carrier alliances in which customer requests are exchanged between the participating organisations through various techniques. A mechanism which has been generally accepted in a horizontal carrier cooperation context is joint route planning. Creating a joint route plan for all requests simultaneously may lead to reductions in travel distance, empty vehicle movements and number of required trucks (Cruijssen et al. 2007a).

Cruijssen and Salomon (2004) use a simulation study to demonstrate that joint route planning may lead to reductions in transport costs up to 15%. Similarly, Cruijssen et al. (2007a) define a framework based on the VRP with time windows (VRPTW) to determine the synergy value of horizontal carrier cooperation through joint route planning. Case study results show that joint route planning between three frozen food distributors saves about 30% in distance travelled. Nadarajah and Bookbinder (2013) study horizontal carrier collaboration within urban regions. Computational experiments indicate distance savings up to 15% when collaborating at the entrance of the city and additional reductions in kilometres driven up to 15% when carriers are involved in intra-city collaboration. Pérez-Bernabeu et al. (2015) compare a cooperative route planning scenario with various non-cooperative scenarios, differing in geographical customer distribution, in terms of distance-based and environmental costs. Another variant of the traditional VRP used to model the collaborative carrier order sharing problem is the multi-depot pickup and delivery problem (MDPDP), as described in Krajewska et al. (2008) under time windows. The authors test their approach both on artificial instances and real-life data from a German freight forwarder. Dahl and Derigs (2011) examine the order sharing problem in a collaborative network of carriers performing express customer orders. Based on a simulation study using real data from 50 European express carriers, the authors demonstrate that cost reductions up to 13% may be achieved when applying dynamic joint route planning. Contrary to the previous articles considering the entire transport network for collaboration, Bailey et al. (2011) and Juan et al. (2014) focus on order sharing opportunities for the backhaul routes of partnering companies.

The existing research work described above mainly focuses on demonstrating the cost reduction potential of order sharing between logistics service providers. Following this observation, the main contribution of this paper is to provide decision support on the partner selection and cost allocation process within a joint route planning context. The next sections provide relevant details on both collaboration challenges.

2.2. Partner Selection

Selecting the right partners constitutes a crucial phase in the development of a horizontal collaboration. According to Brouthers et al. (1995) cooperating with an unsuitable partner is more damaging to an organisation than not collaborating at all. Carriers also seem to be aware of the crucial importance of partner selection, as indicated in a survey by Cruijssen et al. (2007b).

Van Breedam et al. (2005) distinguish four key factors that should be considered when selecting possible collaboration partners: trust and engagement, operational fit, strategic fit and cultural fit. Trust refers to each company’s conviction that the other partners will refrain from opportunistic behaviour. Engagement reflects the preparedness of each alliance partner to make a contribution to the collaboration, evoking a mutual sense of responsibility towards alliance success (Schmoltzi and Wallenburg 2012). Other focal points are the operational, strategic and cultural fit with a potential partner. Operational fit concerns organisational characteristics on a financial and operational level such as company size, proprietary structure and profitability. In order for strategic fit to be present, the organisational strategies of the partners need to be compatible and mutually strengthen each other. A final key factor in partner selection is cultural fit. Compatibility between organisational cultures is crucial when a stable collaboration is aspired. In line with these four factors, Schmoltzi and Wallenburg (2011) define six dimensions associated with the structure of the cooperation that may impact its performance. First, the contractual scope defines the formality of the cooperation project. Second, the organisational scope refers to the number of companies taking part in the alliance. Third, the functional scope is associated with the activity domains in which organisations join forces. A cooperation might be limited to non-core activities or may involve core business’ operations. Fourth, the geographical scope is related to the markets that are covered by the alliance. In line with this geographical dimension, the service scope defines the products or services offered by the collaboration. Finally, the resource scope refers to the degree of resource overlaps between the cooperation participants. A distinction is made between overlaps in business activities, customer base and company size. Based on the partner selection criteria discussed above, the effect of five measurable coalition characteristics on alliance performance and stability is investigated and statistically analysed in this paper: number of partners, carrier size, geographical coverage, order time windows and order size. In section 3.2 the studied hypotheses are discussed in detail.

2.3. Cost Allocation

Although selecting the right partners is crucial for the success of any horizontal alliance, it is not sufficient to guarantee long-term coalition stability. Dividing the collaborative gains in a fair manner constitutes a key issue.

Verdonck et al. (2016) provide a structured review of allocation mechanisms applied in horizontal logistics collaborations distinguishing between (1) proportional sharing mechanisms, (2) allocation mechanisms using game theory concepts and (3) allocation techniques designed to cope with additional cooperation properties. First, the most commonly used profit or cost division mechanism in practice is the proportional allocation method (Liu et al. 2010). In this case, the collaborative profit is allocated to the cooperating organisations equally, on the basis of, among others, their individual cost level or the volume they have to transport as a consequence of their engagement in the cooperation (Verdonck et al. 2016). Second, a logistics cooperation clearly matches the structure of a cooperative game. Collaborating partners share and consolidate freight and receive or make payments in return. This cooperation process results in an allocation of benefits or costs to each participant that may be considered equivalent to the outcome of a cooperative game. A well-known allocation method based on the foundations of game theory is the Shapley (1953) value. This value allocates to each participant the weighted average of his contributions to all (sub)coalitions, assuming the grand coalition is formed one company at a time. A more complex allocation mechanism supported by game theory is the nucleolus. This profit or cost sharing procedure, developed by Schmeidler (1969), has the distinct property of minimising the maximal excess, which constitutes the difference between the total cost of a coalition and the sum of the costs allocated to its participants. Finally, several authors have developed distinct, more intuitively clear allocation mechanisms that account for certain specific cooperation characteristics, some of them partly based on game theory ideas (Verdonck et al. 2016). Tijs and Driessen (1986) discuss three allocation techniques based on the division of the total collaborative costs in separable and non-separable costs. Frisk et al. (2010) create profit sharing mechanisms with the goal of finding a stable allocation that minimises the largest relative difference in cost savings between any pair of cooperating partners. Özener and Ergun (2008) develop allocation mechanisms ensuring that existing partners do not loose savings when an additional company joins the collaboration, while Hezarkhani et al. (2016) define allocations preserving the competitive positions of cooperation participants.

The overview provided in the previous paragraph demonstrates that a wide range of possible allocation mechanisms exists. As each method has its specific benefits and drawbacks, it remains ambiguous which technique(s) could guarantee stability and sustainability in a joint route planning setting. In order for partners to make an informed decision on the allocation mechanism that suits their collaborative needs, an extensive comparative analysis examining the applicability and suitability of three different cost allocation methods in varying cooperation scenarios is performed in this paper. The three allocation methods selected for their application are the Shapley value, the Alternative Cost Avoided Method (Tijs and Driessen 1986) and the Equal Profit Method (Frisk et al. 2010). Details on their theoretical foundation and mathematical formulas can be found in Appendix A.

3. RESEARCH METHODOLOGY

To investigate the impact of coalition characteristics on collaborative performance and cost allocation values the statistical approach of experimental design is used. Using factorial ANOVA, the value of the performance measure associated with various levels of the independent parameters or factors can be statistically derived. Based on the partner selection criteria discussed in section 2.2, the effect of five measurable coalition characteristics on alliance and cost allocation performance is investigated within a joint route planning context. Section 3.1 briefly describes the joint route planning problem faced by the collaborating carriers. In section 3.2 the research hypotheses are discussed in detail. Since no test instances are available for the specific collaboration problem investigated here, the method used to generate artificial instances is described in section 3.3, together with a presentation of the experimental factors coinciding with the relevant cooperation characteristics.

3.1. Problem Statement

The joint route planning problem of collaborating carriers studied in this paper can be defined as follows. Carriers receive pickup and delivery requests from different types of customers. In a static context, it is assumed that customer demand is known and fixed at the start and no additional requests are acquired during the execution of already determined transport schedules. Each route has to satisfy coupling and precedence constraints, meaning that for each order, the origin must precede the destination and both locations need to be visited by the same vehicle. In addition, hard time windows are associated with each request. In a non-cooperative environment, the routing problem associated with each individual carrier, may be classified as a single depot PDPTW. The objective of the PDPTW is to identify an optimal set of routes for a fleet of vehicles to serve all customers without violating vehicle capacity, time windows, precedence and coupling constraints. The optimality characteristic coincides with an objective function that minimises total customer service time, distance travelled, number of used vehicles or a weighted combination of these goals (Li and Lim 2003, Krajewska

et al. 2008, Parragh et al. 2008, Ropke and Cordeau 2009).
If carriers cooperate horizontally, pooling all their customer orders together to achieve potential savings, additional constraints have to be added to the PDPTW in order to optimally solve the joint route planning problem. The most important modification that needs to be made, is the adoption of a multi-depot perspective. As requests from all carriers are considered simultaneously, vehicles may depart from multiple depots. The joint route planning problem may thus be defined as a multi-depot PDPTW with the general purpose of identifying optimal routes for all customer requests simultaneously. This set of routes minimises total cost, guarantees that all requests are served within their time windows, all vehicles return to their respective depots and vehicle capacities are never exceeded (Krajewska et al. 2008).

3.2. Research Hypotheses

In the first part of the numerical experiment, the effect of specific coalition characteristics on alliance performance is investigated. For this purpose, the following five relationships based on theoretical, qualitative collaboration literature are analysed. First, the influence of the **number of partners** on cooperation performance is examined. Based on statements made by Park and Russo (1996), Griffith et al. (1998) and Lozano et al. (2013) it is investigated whether the number of collaborating partners has a positive impact on coalition performance. Second, in line with the operational fit concept described by Van Breedam et al. (2005), the impact of similarity in **size of the collaborating companies** is studied. Size of a carrier is measured in terms of the amount of customer requests it initially needs to serve before considering the cooperation. In accordance with experimental results of Cruijssen et al. (2007a) and Vanovermeire et al. (2013), it is examined whether coalition performance is higher for cooperations established between equally sized carriers compared to collaborations between organisations differing in size. Third, the effect of resource overlaps between alliance partners is analysed in three ways. The resource scope is first defined as the **degree of overlapping geographical coverage** between cooperating carriers. Here, it is studied whether coalition performance is higher for cooperations established between carriers operating within the same geographical area compared to collaborations between companies active in unrelated customer markets. Next, the effect of similarities and differences in customer base characteristics is investigated. This concept is translated in two partner characteristics. On the one hand, it is studied whether overlap between cooperation participants in terms of **customer order time windows** has a positive effect on coalition performance. On the other hand, it is investigated whether coalition performance is higher for cooperations formed by partners with differing **order sizes** compared to collaborations established between carriers serving orders of similar size. Throughout all these hypotheses the dependent variable, coalition performance (CP), is defined as the difference between the total distance-dependent cost of the coalition after applying joint route planning and the sum of the stand-alone distance-dependent costs of the companies considering operating independently.
The goal of the second part of the numerical experiment is to examine the applicability and suitability of **the Shapley value, the ACAM and the EPM** in varying cooperation scenarios. First, the influence of the five coalition characteristics described above on cooperation stability is examined. Second, it is analysed whether significant differences exist between allocation solutions defined by the three applied mechanisms. Third, the relationship between the cooperation structure and the cost share allocated to the different alliance partners is studied.

3.3. Instance Generation

First, test instances are created for individual carriers differing in terms of the partner characteristics presented in the previous section. Table 1 provides an overview of the characteristics associated with these individual carrier instances together with their implementation details. Regarding the chosen implementation values, experienced practitioners were consulted in order to create realistic partnership structures fitting in a joint route planning setting. Second, the individual carrier instances are combined in a factorial experiment to represent horizontal alliances with varying structures.
Considering the individual carrier instances (Table 1), organisations of three different sizes are created. ’Small’ carriers have to serve between 15 and 25 customer requests, ’medium’ carriers are responsible for 60 to 70 customer orders and ’large’ carriers are assigned 100 to 120 requests. This implementation is in line with the European logistics environment comprised of a significant amount of SMEs (small and medium-sized enterprises). To examine the impact of resource overlaps between alliance partners, within each of the three carrier categories just described, distinct carrier profiles are created. First, the Li and Lim (2003) distinction between LR (random customer locations) and LC (clustered customer locations) instances is used to cope with the geographical coverage associated with individual carriers. Second, a distinction is made between carriers serving customers with broad time windows and carriers performing orders with narrow time windows. The average time window width of customer orders characterised by ‘broad’ time windows is two to three times larger than that of orders with ‘narrow’ time windows. Third, carrier instances may differ in terms of the average size of the orders that need to be served. A ’small’ order takes up 5% to 15% of vehicle capacity, while a ’large’ order occupies 30% to 40% of vehicle space. Transported goods and used vehicles are considered to have homogeneous characteristics among participating transport organisations.

Table 1: Characteristics of Individual Carrier Instances

Characteristic	Levels	Implementation
Carrier size	Small Medium Large	*U*(15, 25) orders *U*(60, 70) orders *U*(100, 120) orders
Geographical coverage	R C	Random Clustered
Order time windows	1 2	Narrow Broad
Order size	Small Large	U(0.05, 0.15) * cap. U(0.30, 0.40) * cap.

The five experimental factors and their associated factor levels are listed in Table 2. Horizontal carrier alliances with different coalition characteristics are generated by combining the individual carrier instances as follows. Regarding the number of partners in a coalition, two-carrier, three-carrier, four-carrier and five-carrier partnerships are considered. Next, due to the stated importance of operational fit between coalition partners (e.g. Van Breedam et al. 2005), a distinction is made between alliances consisting of equally sized organisations and alliances comprised of companies differing in size for each of the studied coalition sizes. As such, 'equal size' coalitions are established either between small carriers, medium-sized carriers or large carriers. In order to get a balanced experimental design, for the 'different size' coalitions a random selection is made of three coalition structures containing a mix of small, medium and large carriers. As a consequence, the experimental design can be considered fractional instead of full since not all factor level combinations are included. Within each of these 24 alliance classes, coalitions are then created between carriers operating in the same geographical area (combination of LR instances) and carriers serving customers in different regions (combination of LC instances). In addition, a distinction is made between coalitions established between carriers who are similar in terms of average order time windows (combination of all 'narrow time windows' or all 'broad time windows' instances, each representing half of the number of instances) and carriers responsible for customers with different time window widths (mix of 'narrow time windows' and 'broad time windows' instances, divided equally within every instance). Finally, both alliance structures with only small or large average order sizes and coalitions servicing a mix of small and large orders are created. For comparison and analysis purposes, three instances are generated for each of the described coalition profiles, leading to a total of 1152 test instances.

Table 2: Experimental Factors and Factor Levels

Factors	Factor levels (number of levels)
Number of partners	Two, three, four, five (4)
Carrier size	Small, medium, large, mix_1, mix_2, mix_3 (6)
Geographical coverage	Random, clustered (2)
Order time windows	Equal, mix (2)
Order size	Small, large, mix_1, mix_2 (4)

4. RESULTS AND DISCUSSION

This section is devoted to the presentation and discussion of the joint route planning outcomes, both in terms of collaborative savings and allocation values. The effects of coalition characteristics on collaborative performance are analysed by factorial ANOVA. In the cost allocation analyses paired *t*-tests are used. All statistical experiments are performed using SPSS for Windows Release 24 and are carried out on a Xeon CPU at 2.8 GHz with 64GB of RAM. Coalition performance of all considered cooperation structures is determined using a metaheuristic framework based on Adaptive Large Neighbourhood Search (Pisinger and Ropke 2007).

4.1. Collaborative Savings Results

The savings level associated with joint route planning ranges from 1.64% to 38.57% over all experiments, with an average savings level of 17.14%. Horizontal collaboration through order sharing can hence produce large operational benefits to carriers. However, because of the wide spread in possible savings and because 1.64% may not be a sufficient gain to compensate for additional overhead of collaboration, a further investigation of the main effects of the five factors on the savings attained by the collaboration is in order.

Table 3 presents the ANOVA based significance values for the main effects of the considered alliance characteristics on coalition performance. For each of the studied characteristics the ω^2 value (Olejnik and Algina 2000) is also reported, indicating their respective effect size. The mean coalition performance (CP) for the studied factor levels are displayed in Tables 4, 5 and 6 in Appendix B. Bonferroni and Games-Howell post hoc *t*-tests were used to define the statistical significance of the different factor levels (Field 2013).

Table 3: Main Effects of Coalition Characteristics

Factor	p	ω^2
Number of partners	0.0000*	0.242
Carrier size	0.0000*	0.281
Geographical coverage	0.0000*	0.052
Order time windows	0.0096*	0.005
Order size	0.0000*	0.039

Note: * Significant at α of 0.01

Table 3 indicates that all of the main effects exhibit a statistical significance of less than 0.01. As such, each of the five studied coalition characteristics has a significant impact on coalition performance. The next paragraph will discuss the experimental factors and the proposed hypotheses (section 3.2) independently.
Reviewing the ω^2 values reveals that the size of the carriers involved in the coalition has the most profound impact on its performance. In accordance with experimental results of Cruijssen et al. (2007a) and Vanovermeire et al. (2013), coalitions with the largest profits are achieved when a lot of orders are combined. The larger the pool of joint orders, the larger the potential to find a more profitable route plan for the collaboration. While large transport organisations thus best seek for partners that are equal in size, small companies best join forces with a significant amount of equal-sized organisations and/or attract a large partner in order to enjoy savings levels associated with large order pools. Next, the hypothesis which states that the number of partners in a collaboration influences its performance in a positive way, can be confirmed in a joint route planning context. Increasing the coalition size from two to five partners leads to a more than tripled profit level. However, companies need to be aware that coalition size cannot be enlarged infinitely. Collaborating with a large number of partners also increases alliance complexity and may dilute the strength of mutual partner relationships. In this context, Lozano et al. (2013) proof that there exists a limit above which the synergy increase generated by adding another company to the collaboration is negligible. Then, results demonstrate that coalitions between partners operating within the same geographical service area gain on average 45% more compared to collaborations between companies active in completely unrelated customer markets. Increased geographical coverage may provide more cooperation opportunities and could thus lead to larger cost reductions. Overlapping customer markets seem to constitute an important aspect of coalition sustainability, as was also stated by Van Breedam et al. (2005), Cruijssen et al. (2007a), Schmoltzi and Wallenburg (2011) and Guajardo and Rönnqvist (2015). In terms of order sizes, transport organisations involved in joint route planning best seek for partners that serve requests differing in size. A company with large orders may experience difficulties combining them in a single trip. As such, small orders can be useful to fill the remaining vehicle capacity. Moreover, organisations with small orders could avoid performing a multitude of routes, possibly with many detours, to deliver all its orders by combining them with larger ones. Following these statements, coalitions formed by partners with differing order sizes may achieve on average 26% more compared to collaborations established between carriers serving orders of similar size. Similar results were found by Vanovermeire et al. (2013) and Palhazi Cuervo et al. (2016) for two-partner shipper coalitions. Finally, differences in order time windows seem to complement each other and increase the number of possible improvement opportunities for the joint route plan. This result is supported by Schmoltzi and Wallenburg (2011) who found that, in practice, the majority of multi-lateral horizontal cooperations between logistics service providers are characterised by complementary customer portfolios of partners. However, the remark needs to be made here that, although the main effect of the time window width is significant, its explaining power is rather limited as shown by its low ω^2 value.

4.2. Cost allocation results

In order to ensure sustainability of the joint route planning project, incurred costs need to be divided in a fair way among the participants. For this reason, the collaborative costs are now allocated to the carriers applying the Shapley value, the ACAM and the EPM.
To identify whether the cost allocations defined for the studied experiments guarantee cooperation **stability**, compliance of the Shapley and ACAM solutions with individual, subgroup and group rationality needs to be verified. A cost allocation satisfying the individual rationality property guarantees that no carrier pays more than his stand-alone cost. Subgroup rationality avoids that players leave the grand coalition to form a subgroup because they could be better off excluding certain partners. Group rationality, also labelled efficiency, ensures that the total cooperative cost is shared as the grand coalition forms. Since core constraints are included in the EPM linear program, feasibility of the EPM solution indicates whether the grand coalition is stable. In case of a non-stable grand coalition, additional allocations are calculated for comparison purposes, namely the 'Stability relaxation EPM' and 'ϵ-EPM'. Regarding the calculation of these cost allocations for non-stable collaborations, two modifications are applied to the EPM in order to find a feasible solution. First, allocation values are calculated while relaxing core constraints that could not be satisfied for the respective cooperative game. Second, EPM is combined with the ϵ-core concept, as suggested by Frisk et al. (2010). Applying the ϵ-core, cooperation participants are penalised with a cost $\epsilon > 0$ for quitting the grand coalition. In this way, stable cost allocations may be calculated for cooperative games with an empty core (Shapley and Shubik 1966).
Analysing cost allocations over all instances reveals that stability of the grand coalition is guaranteed in 73% of the studied experiments. In the remaining 27% the core of the cooperative game is empty. If the grand coalition is stable, then no subgroup of partner companies has the incentive to leave the grand coalition and be better off acting alone. Results demonstrate that in the experimental design stability either holds or not, that is, that this outcome is independent of the allocation technique used. The non-stable coalition instances demonstrate the influence of cooperation structure on the longevity of joint route planning projects. The analysis reveals that increasing the number of coalition participants has a negative impact on its long-term sustainability. While two-carrier cooperations are always

related with stable outcomes, only 45% of the five-carrier cooperations are associated with stability. Although increasing the coalition size from two to five partners leads to a more than tripled profit level, companies need to be aware that collaborating with a large number of partners also increases alliance complexity and may dilute the strength of mutual partner relationships. Regarding the other experimental factors, the influence on coalition stability is not so clear. When cooperations with varying levels of partner size, order size, geographical coverage or order time windows are compared the number of stable versus unstable experiments is divided almost equally.

Investigating the **allocation values** defined by means of the Shapley value, the ACAM and the EPM variations over all instances, the following observations can be made. First, when comparing over the division mechanisms using paired *t*-tests, no significant differences exist in the allocation values. The share of logistics costs allocated to the cooperation participants is thus fairly similar with respect to the used allocation technique. On average, the smallest differences are associated with coalitions of limited size between equal partners. For all two-partner coalitions, Shapley and ACAM even lead to identical cost allocations. Second, examining the cost share allocated to the different cooperation participants reveals that the division of cost savings is related to the collaborative efforts made by the participants, regardless of the used sharing mechanism. As such, organisations that contribute more to the partnership receive a higher share of the collaborative savings. For example, consider a coalition of three partners A, B and C joining their orders. When partner A has to serve significantly more shared orders than partner B and C when executing the joint route plan, partner A is rewarded for this effort with a higher share in the collaborative gains. Third, the original EPM and the EPM with relaxed stability constraints provide the most equally spread cost savings among the partners of the coalition. Although the ϵ-EPM also aims to minimise maximal pair wise differences between allocated savings, increased variation in carrier savings is caused by adding ϵ-core constraints. Finally, the Shapley value benefits small carriers in case of a coalition comprised of participants of different size. On average, collaborative savings of companies with a smaller amount of customer orders are highest when costs are divided by means of the Shapley value.

5. CONCLUSIONS AND FUTURE RESEARCH

Although transport companies become increasingly aware of the inevitable character of horizontal collaboration, surveys report failure rates up to 70 percent for starting strategic partnerships. While a growing body of collaboration research acknowledges the importance of partner characteristics, no extensive study has been performed on the numerical relationship between specific company traits and the performance of the alliances these organisations are involved in. The first contribution of this paper is thus to provide practical recommendations on which partnership structures may provide the highest collaborative benefits by means of analysing the results of an experimental design. Although selecting the right partners is crucial for the success of any horizontal alliance, it is not sufficient to guarantee long-term coalition stability. Dividing the collaborative gains in a fair manner constitutes a key issue. In this context, the second contribution of this paper is to perform a comparative analysis examining the applicability and suitability of three different cost allocation methods in varying cooperation scenarios.

Based on extensive numerical experiments analysing the influence of alliance characteristics on the amount of attainable collaborative savings using factorial ANOVA, the following managerial insights may be formulated. First, results reveal that coalitions with the largest profits are achieved when a lot of orders are combined. The larger the pool of joint orders, the larger the potential to find a more profitable route plan for the collaboration. While large transport organisations best seek for partners that are equal in size, small companies best join forces with a significant amount of equal-sized organisations and/or attract a large partner in order to enjoy savings levels associated with large order pools. Second, considering the positive influence of the number of partners on collaborative performance, the importance of the total number of orders is confirmed. However, companies need to be aware that coalition size cannot be enlarged infinitely. Collaborating with a large number of partners also increases alliance complexity and may dilute the strength of mutual partner relationships. Third, broad geographic coverage and/or overlapping customer markets seem to constitute an important aspect of coalition sustainability. The larger the service region of the coalition, the more possibilities for efficient order sharing there are. Moreover, when the supply areas of the companies overlap each other the average transport distances decrease. Finally, transport organisations involved in joint route planning best seek for partners that serve requests differing in size. In this way, the coalition can take full advantage of unused vehicle capacity.

When participants have to decide on the mechanism of how to share collaborative savings, the following observations can be made. Regardless of the used sharing mechanism, allocation techniques account for differences in partner contributions to the collaborative goal. Participants that make notable efforts to execute the joint route plan are rewarded with a higher share of the collaborative savings. The original EPM and the EPM with relaxed stability constraints may be most useful in collaborations between carriers with similar characteristics as they provide the most equally spread cost savings. In addition, both allocation techniques may also be valuable in the early phases of a growing horizontal cooperation, in which having an initial allocation with similar benefits for all participating organisations may suit communication and negotiation purposes. Small carriers may prefer costs to be allocated by means of the Shapley value. This division mechanism favours companies with a smaller share in customer demand by allocating them a higher percentage of collaborative savings in comparison with the ACAM and the EPM. Next, results show that although increasing the coalition size from two to five partners leads to a more than tripled profit level, increasing the size of the alliance has a negative impact on its long-term sustainability. Companies need to be aware that collaborating with a large number of partners increases alliance complexity and may dilute the strength of mutual partner relationships. Finally, the most striking finding is that no significant differences were observed in the allocation values when comparing over the division mechanisms.

To conclude, the following relevant suggestions for future research can be made. First, when exploring joint route planning, the focus may be expanded from considering cost minimisation exclusively to account for customer service effects. Besides its impact on cost and efficiency levels, cooperation with fellow transport companies may also have an influence on the service that can be provided by each participating carrier. Second, considering the overview of collaboration strategies in Verdonck et al. (2013), a similar impact study of cooperation characteristics and allocation mechanisms could be done in other collaborative logistics environments. Third, another natural avenue of research is to examine the efficacy of other cost allocation techniques in a joint route planning setting. Finally, the consideration of specific factors and factor levels may influence the general validity of the findings. As such, an adapted experimental design with other experimental factors and/or factor levels could be the subject of future research work.

ACKNOWLEDGMENTS

This research is supported by the Research Foundation Flanders (FWO) and the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office (research project COMEX, Combinatorial Optimization: Metaheuristics & Exact Methods). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Centre), funded by the Hercules Foundation and the Flemish Government - department EWI.

APPENDIX

Appendix A

Shapley value

The Shapley value (Shapley 1953) allocates to each participant the weighted average of his contributions to all (sub)coalitions, assuming the grand coalition is formed one company at a time. The Shapley allocation to participant *i* can be mathematically expressed as:

$$y_i = \sum_{S \subset N:\, i \in S} \frac{(|S|-1)!(|N|-|S|)!}{|N|!} [c(S) - c(S \setminus \{i\})] \qquad (1)$$

With |.| denoting the number of participants in the considered (sub)coalition, $c(.)$ the cost of the respective (sub)coalition, *N* the grand coalition and *S* a cooperation of a subset of partners of the grand coalition.

Alternative Cost Avoided Method

The ACAM (Tijs and Driessen 1986) allocation to participant *i* can be mathematically expressed as:

$$y_i = m_i + \frac{c(i) - m_i}{\sum_{j=1}^{n}[c(j) - m_j]} * (c(N) - \sum_{j=1}^{n} m_j) \qquad (2)$$

With m_i, denoting the separable or marginal cost of company *i*, which may be calculated as $c(N) - c(N \setminus i)$, and $j \in N$ representing all other coalition partners.

Equal Profit Method

The EPM (Frisk et al. 2010) guarantees stable allocations that minimise the maximum difference between the cost savings allocated to the cooperating partners. In order to find the EPM allocations to all participants, the following linear pro- gram needs to be solved to optimality:

$$Min\ f \qquad (3)$$

Subject to

$$f \geq \frac{y_i}{c(i)} - \frac{y_j}{c(j)} \qquad \forall\, i, j \in N \qquad (4)$$

$$\sum_{j \in S} y_j \leq c(S) \qquad \forall\, S \subseteq N \qquad (5)$$

$$\sum_{j \in N} y_j = c(N) \qquad (6)$$

The first constraint set (4) measures the pair wise difference between the relative savings of the participants. The objective function (3) minimises the largest difference using variable *f*. Constraint sets (5) and (6) ensure that the allocation is stable and belongs to the core. As such, the cost allocation guarantees that no subcoalition *S* exists in which a set of partners would be better off (5) and that the total collaborative cost is shared as the grand coalition forms (6).

Appendix B

Table 4: Mean Coalition Performance for Factor Levels

Number of partners	Mean CP	Carrier size	Mean CP
2	3892.209	Small	3187.501
3	7232.228	Medium	9015.907
4	10950.386	Large	14797.365
5	13511.741	Mix	8882.020

Tale 5: Mean Coalition Performance for Factor Levels

Geographical coverage	Mean CP	Order time windows	Mean CP
Random	10725.29	Equal	8494.768
Clustered	7390.856	Mix	9621.381

Table 6: Mean Coalition Performance for Factor Levels

Order size	Mean CP
Small	8243.017
Large	6890.533
Mix	10233.673

REFERENCES

Bailey E., Unnikrishnan A., Lin D.Y., 2011. Models for minimizing backhaul costs through freight collaboration. Transportation Research Record 2224(1):51-60.

Brouthers K., Brouthers L., Wilkinson T., 1995. Strategic alliances: Choose your partners. Long Range Planning 28(3):18-25.

Cruijssen F., Bräysy O., Dullaert W., Fleuren H., Salomon M., 2007a. Joint route planning under varying market conditions. International Journal of Physical Distribution and Logistics Management 37(4):287-304.

Cruijssen F., Cools M., Dullaert W., 2007b. Horizontal cooperation in logistics: Opportunities and impediments. Transportation Research Part E: Logistics and Transportation Review 43(2):129-142.

Cruijssen F., Dullaert W., Fleuren H., 2007c. Horizontal cooperation in transport and logistics: A literature review. Transportation Journal 46(3):22-39.

Cruijssen F., Salomon M., 2004. Empirical study: Order sharing between transportation companies may result in cost reductions between 5 to 15 percent. CentER discussion paper. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=607062 [Accessed 5 April 2017].

Dahl S., Derigs U., 2011. Cooperative planning in express carrier networks - An empirical study on the effectiveness of a real-time decision support system. Decision Support Systems 51(3):620-626.

Field A., 2013. Discovering statistics using IBM SPSS Statistics. London: SAGE Publications Ltd.

Frisk M., Göthe-Lundgren M., Jörnsten K., Rönnqvist M., 2010. Cost allocation in collaborative forest transportation. European Journal of Operational Research 205(2):448-458.

Griffith D.A., Hu M.Y., Chen H., 1998. Formation and performance of multi-partner joint ventures: A Sino-foreign illustration. International Marketing Review 15(3):171-187.

Guajardo M., Rönnqvist M., 2015. Operations research models for coalition structure in collaborative logistics. European Journal of Operational Research 240(1):147-159.

Guajardo M., Jörnsten K., Rönnqvist M., 2016. Constructive and blocking power in collaborative transportation. OR Spectrum 38(1):25-50.

Hernández S., Peeta S., Kalafatas G., 2011. A less-than-truckload carrier collaboration planning problem under dynamic capacities. Transportation Research Part E: Logistics and Transportation Review 47(6):933-946.

Hezarkhani B., Slikker M., Van Woensel T., 2016. A competitive solution for cooperative truckload delivery. OR Spectrum 38(1):51-80.

Juan A., Faulin J., Pérez-Bernabeu E., Jozefowiez N., 2014. Horizontal cooperation in vehicle routing problems with backhauling and environmental criteria. Procedia - Social and Behavioral Sciences 111:1133-1141.

Krajewska M., Kopfer H., Laporte G., Ropke S., Zaccour G., 2008. Horizontal cooperation among freight carriers: Request allocation and profit sharing. Journal of the Operational Research Society 59(11):1483-1491.

Li H., Lim A., 2003. A metaheuristic for the pickup and delivery problem with time windows. International Journal on Artificial Intelligence Tools 12(2):173-186.

Liu P., Wu Y., Xu N., 2010. Allocating collaborative profit in less-than-truckload carrier alliance. Journal of Service Science and Management 3(1):143-149.

Lozano S., Moreno P., Adenso-Díaz B., Algaba E., 2013. Cooperative game theory approach to allocating benefits of horizontal cooperation. European Journal of Operational Research 229(2):444-452.

Nadarajah S., Bookbinder J.H., 2013. Less-Than-Truckload carrier collaboration problem: modeling framework and solution approach. Journal of Heuristics 19(6):917-942.

Olejnik S., Algina J., 2000. Measures of effect size for comparative studies: Applications, interpretations and limitations. Contemporary Educational Psychology 25(3):241-286.

Özener O.Ö., Ergun Ö, 2008. Allocating costs in a collaborative transportation procurement network. Transportation Science 42(2):146-165.

Palhazi Cuervo D., Vanovermeire C., Sörensen K., 2016. Determining collaborative profits in coalitions formed by two partners with varying characteristics. Transportation Research Part C: Emerging Technologies 70:171-184.

Park S.H., Russo M.V., 1996. When competition eclipses cooperation: An event history analysis of joint venture failure. Management Science 42(6):875-890.

Parragh S., Doerner K., Hartl R., 2008. A survey on pickup and delivery problems, part II: Transportation between pickup and delivery locations. Journal für Betriebswirtschaft 58(2):81-117.

Pérez-Bernabeu E., Juan A.A., Faulin J., Barrios B.B., 2015. Horizontal cooperation in road transportation: a case illustrating savings in distances and greenhouse gas emissions. International Transactions in Operational Research 22(3):585-606.

Pisinger D., Ropke S., 2007. A general heuristic for vehicle routing problems. Computers & Operations Research 34(8):2403-2435.

Ropke S., Cordeau J.F., 2009. Branch-and-cut-and price for the pickup and delivery problem with time windows. Transportation Science 43(3):267-286.

Schmeidler D., 1969. Nucleolus of a characteristic function game. Siam Journal on Applied Mathematics 17(6):1163-1170.

Schmoltzi C., Wallenburg C.M., 2011. Horizontal cooperations between logistics service providers: Motives, structure, performance. International Journal of Physical Distribution and Logistics Management 41(6):552-575.
Schmoltzi C., Wallenburg C.M., 2012. Operational governance in horizontal cooperations of logistics service providers: Performance effects and the moderating role of cooperation complexity. Journal of Supply Chain Management 48(2):53-74.
Shapley L.S., 1953. Contributions to the Theory of Games. Princeton: Princeton University Press.
Shapley L.S, Shubik M., 1966. Quasi-cores in a monetary economy with nonconvex preferences. Econometrica 34(4):805-827.
Tijs S.H., Driessen T.S.H., 1986. Game theory and cost allocation problems. Management Science 32(8):1015-1028.
Van Breedam A., Krols K., Verstrepen S., 2005. Logistiek samenwerken praktisch bekeken. Flanders Institute for Logistics (VIL). Available from: anet.uantwerpen.be/link/irua/57584 [Accessed 5 April 2017].
Vanovermeire C., Palhazi Cuervo D., Sörensen K., 2013. Estimating collaborative profits under varying partner characteristics and strategies. Antwerp University. Available from: http://antor.uantwerpen.be/members/christine-vanovermeire/ [Accessed 5 April 2017].
Verdonck L., Beullens P., Caris A., Ramaekers K., Janssens G.K., 2016. Analysis of collaborative savings and cost allocation techniques for the cooperative carrier facility location problem. Journal of the Operational Research Society 67(6):853-871.
Verdonck L., Caris A., Ramaekers K., Janssens G.K., 2013. Collaborative logistics from the perspective of road transportation companies. Transport Reviews 33(6):700-719.
Verstrepen S., Cools M., Cruijssen F., Dullaert W., 2009. A dynamic framework for managing horizontal cooperation in logistics. International Journal of Logistics Systems and Management 5(3-4):228-248.

AUTHORS BIOGRAPHY

Lotte Verdonck graduated as Master of Applied Economic Sciences: Business Engineer at Hasselt University, Belgium, in 2011. In May 2017, she obtained her PhD in Business Economics at Hasselt University funded by the Research Foundation Flanders (FWO). Her research mainly considers modelling collaborative logistics from the perspective of logistics service providers.

Katrien Ramaekers is Assistant Professor in Operations Management and Logistics at the Faculty of Business Economics (BEW) of Hasselt University (Belgium). She graduated as master in Business Engineering at the Limburg University Centre, Belgium, in 2002. She obtained her PhD in Applied Economics in 2007 at Hasselt University. In her PhD she developed a simulation optimization framework for inventory management decision support based on incomplete information. Her research interest goes to the application of Operations Research techniques in the field of operations management and logistics, with a strong focus on simulation (optimization). Current research domains are warehouse operations, healthcare logistics and cost allocation in intermodal barge transport.

Benoît Depaire is Associate Professor at the Faculty of Business Economics, Hasselt University. He belongs to the research group Business Informatics within the Quantitative Methods Department. His research interests focus on the application of data mining, statistics and process mining within a business-related context.

An Caris was appointed in 2012 as Assistant Professor at the Faculty of Business Economics (BEW) of Hasselt University. She obtained her PhD in Applied Economics in 2010 and was first a postdoctoral research fellow of the Research Foundation - Flanders (FWO). Her research interest goes to the application of Operations Research (OR) techniques in the field of operations management and logistics. In her PhD thesis she focused on the competitiveness of intermodal transport making use of inland navigation. Currently she studies planning problems in warehouse operations, intermodal rail transport, collaborative logistics and healthcare logistics.

Gerrit K. Janssens received degrees of M.Sc. in Engineering with Economy from the University of Antwerp (RUCA), Belgium, M.Sc. in Computer Science from the University of Ghent (RUG), Belgium, and Ph.D. from the Free University of Brussels (VUB), Belgium. After some years of work at General Motors Continental, Antwerp, he joined the University of Antwerp until the year 2000. Currently he is Professor Emeritus at Hasselt University within the Faculty of Business Economics. His main research interests include the development and application of operations research models in production and distribution logistics.

SUPPORTING ORGANISATIONAL DECISION MAKING IN PRESENCE OF UNCERTAINTY

Vinay Kulkarni[(a)], Souvik Barat[(b)], Tony Clark[(c)], Balbir Barn[(d)]

[(a),(b)] Tata Consultancy Services Research, India
[(c)] Sheffield Hallam University, UK
[(d)] Middlesex University London, UK
[(a)]vinay.vkulkarni@tcs.com, [(b)]souvik.barat@tcs.com, [(c)]T.Clark@shu.ac.uk
[(d)]B.Barn@mdx.ac.uk

ABSTRACT
Effective organizational decision-making often requires deep understanding of various aspects of an organisation such as goals, structure, business-as-usual operational processes etc. The large size of the organisation, its socio-technical characteristics, and fast business dynamics make this a challenging endeavor. Current industry practice relies on human experts thus making organisational decision-making time-, effort- and intellectually-intensive. This paper proposes a simulatable language capable of specifying the relevant aspects of enterprise in a machine-processable manner so as to support simulation-driven decision-making in presence of uncertainty. A possible implementation of the language is outlined. Validation of the proposed approach using a real-life example is discussed.

Keywords: Decision making; organisational decision making; uncertainty.

1. INTRODUCTION

Modern enterprises need to respond to a variety of change drivers in order to stay competitive in a rapidly changing business context. The cost of erroneous decisions is often prohibitively high and there may not be an opportunity for subsequent diversion (Daft, R., 2012). Minimizing such undesired consequences calls for a-priori judicious evaluation of the available courses of action as regards their influence on the desired objective. The decision-makers are thus expected to understand, analyze and correlate existing information about various aspects of enterprise such as its goals, its structure, business-as-usual operational processes, change drivers and their influences on overall organisation *etc*. Scale and complexity, inherent socio-technical features (McDermott et al, 2013), and multiple stakeholders with possibly conflicting goals all contribute to the complexity of organisational decision-making. Moreover, increasing demands for agility and certainty make this endeavor even more challenging.

A modern enterprise is a large and complex system and the sheer volume of information makes manual analysis ineffective as well as inefficient. Moreover, with the required information pieces typically strewn across multiple sources such as documents, spreadsheets, pictures, logs etc., these pieces need to be stitched together to form a coherent, consistent and integrated view.

Modern enterprises operate in increasingly dynamic environment that must be kept up-to-date at an increasingly rapid rate. The current inability to stitch together an integrated, complete, and timely view makes manual analysis further untenable. The inability to address all these factors to the desired level of sophistication is the principal contributing reason for the current state of organisational decision-making[1].

A pragmatic approach to organisational decision-making therefore seems to hinge on the availability of: (i) the information required for decision-making in a machine-processable form, (ii) suitable machinery for effective processing of this information, and (iii) a method to enable repetitive use of this machinery at the hands of knowledgeable users. Moreover, the form as well as the machinery needs to be capable of respectively representing and processing the inherent uncertainty.

A variety of Enterprise Modeling (EM) languages exist that provide information-capture and analysis support across a wide spectrum of sophistication. Majority of these languages can be traced to Zachman framework (Zachman, J., 1987) advocating that capture of the *why*, *what*, *how*, *who*, *when* and *where* aspects leads to the necessary and sufficient information for addressing a given problem. Thus it can be argued that complete specification of enterprise is possible using Zachman framework, however, there exists no support for automated analysis as the information is captured typically in the form of texts and pictures. It can be observed of the existing EM languages that: the languages capable of specifying all the relevant aspects of enterprise for organisational decision-making lack support for automated analysis (Zachman 1987, Krogstie 2008, Jonkers et al 2004), and the languages capable of automated analysis can cater to specifying only a subset of the aspects required for decision-making (Yu et al 2006, Meadows and Wright 2008, White 2004).

Co-simulation using a relevant set of EM languages can be a pragmatic solution. For instance, as shown in Fig 1, i* (Yu et al, 2006) to specify the why aspect, Stock-n-flow (Meadows and Wright, 2008) to specify the what aspect, and BPMN (White, S., 2004) to specify the how aspect can be used collectively to come up with the

[1]http://www.valueteam.biz/why-72-percent-of-all-business-transformation-projects-fail

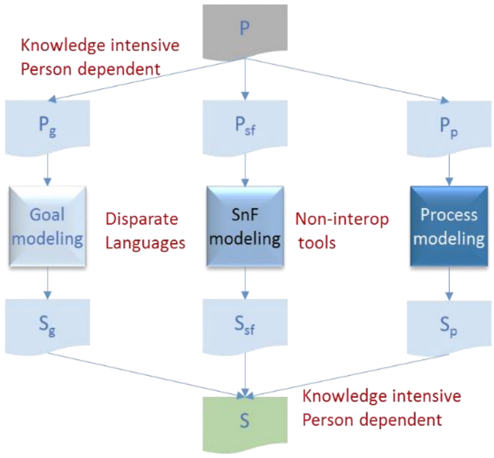


Fig 1. Co-simulation based approach for decision-making

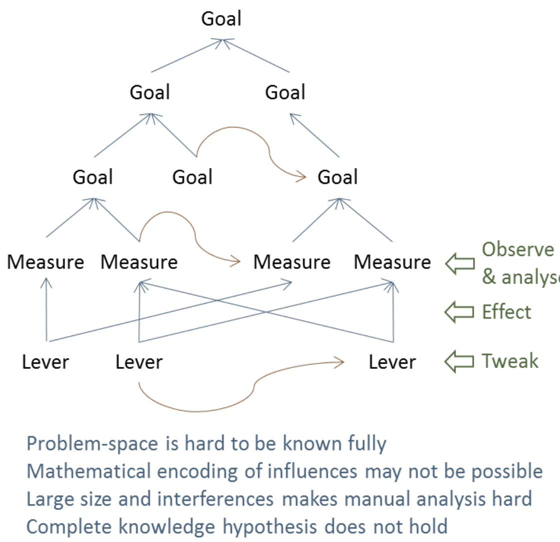


Fig 2. Desired approach to decision-making

necessary and sufficient specification which is amenable for analysis albeit in parts. However, human expertise is still needed to stitch together the results of part analyses into a consistent whole. This is an intellectually intensive activity further exacerbated due to paradigmatically diverse nature of the three languages and non-interoperable tools[2,3,4].

Moreover, increasingly fast business dynamics makes it very difficult to have the complete information required for decision-making available and/or with full certainty. This puts new demands on enterprise specification as well as processing that are not well supported by existing EM languages and associated processing machinery. This paper proposes an approach to support organisational decision-making in presence of uncertainty. Contributions of this paper are twofold. First, it proposes a language capable of specifying the necessary and sufficient aspects of enterprise for organisational decision-making. Second, it describes how the language and its simulation engine can support uncertainty.

This paper is organized as follows: Section 2 describes requirements of a desired approach. Section 3 presents available support for organisational decision-making in the form of EM languages and tools. The proposed language is presented in Section 4 along with a discussion on support for uncertainty. Section 5 presents validation of the proposed approach using a representative case-study from real life. An evaluation with respect to the current state of art and practice is also provided. Section 6 provides a summary and a brief outline of future work.

[2]http://www.cs.toronto.edu/km/ome/
[3]http://www.iseesystems.com/Softwares/Business/ithinkSoftware.aspx
[4]http://www.bizagi.com/en/products/bpm-suite/modeler

2. DESIRED APPROACH

An approach based on the principles of "separation of concerns" and "divide and conquer" seems required for organisational decision-making. It should be possible to decompose the overall goal into sub-goals, sub-sub-goals *etc*., to the desired level of granularity. It should be possible to identify a set of variables (*i.e.*, Measures) that need to be observed in order to determine whether the finest-level goal is met – a goal once met is a measure. It should be possible to identify a set of variables (*i.e.*, Levers) that influence a given Measure and be able to specify the influence in a machine-processable manner. It should be possible to make explicit the dependencies between levers, between measures and between goals.

The goal-measure-lever graph structure of Fig 2 captures the understanding of problem domain in a manner that is amenable to automation. Decision-making is then a bottom-up walk of this graph structure provided it is possible: (i) to compute values for the measures based on the values of levers, (ii) to evaluate whether a goal is met based on the values of measures, and (iii) to honour lever-to-lever, measure-to-measure and goal-to-goal dependencies in the bottom-up walk terminating with evaluation of the overall goal.

Therefore, organisational decision-making can be viewed as a human-guided exploration of the design space as shown by the decision loop of Fig 3. The ability to specify the influence of a lever on a set of measures is the key. Based principally on the Zachman framework (Zachman, J., 1987) it can be argued that an organisation can be understood well by knowing *what* an organisation is, *why* it is so, and *how* it operates involving the key stakeholders (the *who*) thereby constituting the necessary and sufficient information for decision-making. Two primary requirements emerge: (i) the ability to capture the *why*, *what*, *how* and *who* aspects in a formal manner and (ii) the ability to perform what-if analyses of the formal specification.

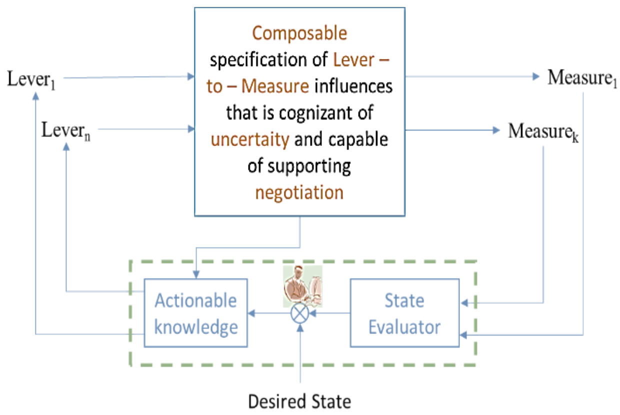


Fig 3. The decision loop

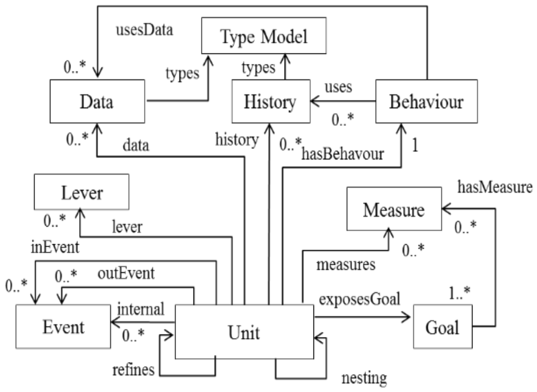


Fig 4. Conceptual Model

In addition, the specification for decision-making should capture the key characteristics of organisation such as: *reactive* (*i.e.*, an organisation reacts to the events taking place in its environment), *adaptable* (*i.e.*, an organisation responds to changes taking place in its environment by transforming itself), *modular* (*i.e.*, an organisation as a whole is a sum of its parts each can be viewed as an organisation), *autonomous* (*i.e.*, an organisation is capable of determining its own course of action so as to achieve its stated goals), *intentional* (*i.e.*, an organisation works towards achieving its stated objectives), and *uncertain* (*i.e.*, an organisation exhibits probabilistic behaviour).

The socio-technical nature of enterprise and the inability to have complete understanding of the problem space means it is difficult to specify lever-to-measure influence in pure mathematical terms. In other words, a closed-form solution to the decision-making problem is unattainable. That leaves an open-form solution using, say, simulation as the only pragmatic recourse. However, simulation is known to deliver only in situations where mechanistic world view holds (Barjis, J., 2008). Modern enterprises, on the other hand, are socio-technical systems that rely on automation systems as well as human agents for their operation. Also, modern enterprise is a system of systems (Ackoff, R., 1971) whose behaviour is too complex to be fully known a-priori. Instead, the overall system behaviour emerges from the interactions between its various agents whose behaviour is simple enough to be fully known a-priori and hence specifiable. The specification language and its simulation engine should cater to these additional requirements as well.

3. STATE OF ART AND PRACTICE

Table 1 summarizes an evaluation of existing EM languages as regards their adequacy to support organisational decision-making in presence of uncertainty. The evaluation criterion comprises of ability to support: (i) specification of aspect views, unit views, socio-technical characteristics, and multiple perspectives, and (ii) processing of these views to support data-driven analysis. As can be seen, the languages capable of specifying all relevant aspects are

Table 1 Current state of EM languages

Modelling Language	To support aspect views				To support unit views		To support socio-technical characteristics					To support data-driven analysis		
	Why	What	How	Who	Modularity	Composability	Reactive	Autonomous	Intentional	Adaptability	Uncertainty	Machine Processability	Quantitative Analysis	Qualitative analysis
Zachman	S	S	S	S	S	N	N	N	S	N	N	N	N	N
Archimate	S	S	S	S	S	I	S	I	S	N	N	N	N	N
EEML	I	S	S	S	N	N	N	N	S	N	N	N	N	N
i*	S	N	N	I	S	S	N	N	S	N	N	S_{Why}	S_{Why}	S_{Why}
SnF	N	S	I	I	I	N	S	S	N	N	N	S_{What}	N	S_{What}
BPMN	N	I	I	S	S_{How}	S_{How}	S	I	N	N	N	S_{How}	S_{How}	S_{How}
S=Supported		Sx =Supported for Aspect X					I = Inadequate Support					N=No Support		

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

found wanting in terms of analysis capabilities e.g., Archimate (Jonkers et al, 2004), EEML (Krogstie, J., 2008), and UEML (Vernadat, F., 2002). On the other hand, the languages capable of sophisticated analysis can cater to specification of only a subset of relevant aspects. For instance, i* (Yu et al, 2006) can specify the why aspect only, BPMN (White, S., 2004) can specify the how aspect only, and Stock-n-flow (Meadows and Wright, 2008) can specify the What aspect only. Therefore, a language capable of supporting all the desired characteristics of organisation specification needs to be designed and the relevant analysis machinery needs to be developed.

4. ENTERPRISE SPECIFICATION LANGUAGE

4.1. Conceptual Model

Looking outside-in, an organisation raises and responds to a set of events as it goes about achieving its stated goals. It consists of several autonomous units, organised into dynamically changing hierarchical groups, operating concurrently, and managing goals that affect their behaviour. We describe structure and behaviour of an organisation using a small set of concepts and their relationships as depicted in Fig 4.

Organisation is a Unit that comprises a set of Units and strives to accomplish its stated Goal. It does so by responding to Events taking place in its environment (InEvents), processing them, and by interacting with other external Units in terms of Events raised/responded (OutEvents). A Unit may choose not to expose all events to the external world (InternalEvents). A declarative specification of event processing logic constitutes the behaviour of a Unit. Thus, looking outside-in, a Unit is a Goal-directed agent that receives events (InEvents), processes them, and raises events (OutEvents) to be processed by other Units. Also, Unit is a parameterized entity whose structure and behaviour can be altered through Levers. A Unit interacts with other Units in a-priori well-defined Role-playing manner. TypeModel provides a type system for structural as well as behavioural aspects of a Unit.

Unit, Event, Data, History and unit nesting together specify the *what* aspect, Goal specifies the *why* aspect, Behaviour specifies the *how* aspect, and Unit, as an individual stakeholder, specifies the *who* aspect of an organisation. Event helps to capture the *reactive* nature of Unit. Also, Unit is *adaptable* as it can construct and reconstruct its structure; *modular* as it encapsulates the structure and behaviour of an organisation; *intentional* as it has its own goals; and *compositional* as it can be an assembly of Units.

The Unit abstraction draws from a set of existing concepts. Modularization and reflective unit hierarchy are taken from fractal component models (Bruneton et al, 2006). Goal-directed autonomous behaviour can be traced to agent behaviour (Bonabeau, E., 2002). Defining states in terms of a type model is borrowed

```
type ::= Var                                   type variable
       | Act { export dec* Mes* }              actor type
       | (type*) → type                        λ-type
       | tt                                    term type
       | Int | Bool | Str                      constant type
       | Void                                  undefined
       | [ type ]                              lists
       | Fun(Name*) type                       parametric type
       | ∀(Name*) type                         polymorphic type
       | rec Name . type                       recursive type

tt   ::= Name(type*)                           term type

exp  ::= var                                   variables
       | num | bool | str                      constants
       | self                                  active actor
       | null                                  undefined
       | new name[type*] (exp*)                create actor
       | become name[type*] (exp*)             change behaviour
       | exp op exp                            binary exp
       | not exp                               negation
       | λ(dec*)::type exp                     λ-abstraction
       | let bind* in exp                      local bindings
       | letrec bind* in exp                   local recursion
       | case exp* arm*                        pattern matching
       | for pat in exp { exp }                looping
       | { exp* }                              block
       | if exp then exp else exp              conditional
       | [ exp* ]                              list
       | []                                    empty list
       | exp(exp*)                             application
       | Name(exp*)                            term
       | exp ← exp                             message
       | name := exp                           update
       | exp . name                            name reference
       | probably(exp)::type exp exp           uncertainty
       | exp[type*]                            type limitation

bind ::= dec = exp                             value binding
       | name(pat*)::type=exp when exp         λ-binding
       | type Name[Name*] = type               type declaration
       | data Name[Name*] = tt*                algebraic type
       | act name(dec*)::type {                behaviour def
           export name*                        interface
           bind*                               locals
           → exp                               initial action
           arm*                                behaviour
         }

dec  ::= name[Name*] :: type                   declaration

arm  ::= pat* → exp when exp                   guarded exp

pat  ::= dec                                   binding
       | dec = pattern                         naming
       | num | bool | str                      const pattern
       | pat : pat                             cons pair
       | [ pat* ]                              list
       | [][type]                              empty list
       | Name[type*](pat*)                     term pattern
```

Fig 5. ESL syntax

from UML[5]. An event driven architecture (Michelson, B., 2006) supports flexible interactions between components, and the concept of intentional modelling (Yu et al, 2006) is adopted to enable specification of component goals.

4.2. Implementation

We have provided an implementation of the conceptual model in the form of Enterprise Simulation Language (ESL). ESL is an extension of an existing event-driven lan guage LEAP (Clark and Barn, 2013) with concepts borrowed from actor model of computation (Agha, G., 1985), multi-agent systems (Van Harmelen et al, 2008), goals (Yu et al, 2006) and linear temporal logic (Pnueli, A., 1977). ESL and its associated development and run-time environment is a language that has been designed to support our thesis that simulation and emergent behaviour can be used to support organisational

[5] http://www.omg.org/spec/UML/

decision-making. ESL together with a supporting toolset is currently in development[6].

The syntax of ESL is shown in Fig 5. It is statically typed and includes parametric polymorphism, algebraic types and recursive types. An ESL program is a collection of mutually recursive bindings. Behaviour types Act {...} are the equivalent of component interfaces and behaviours act {...} are equivalent to component definitions. A behaviour b is instantiated to produce an actor using new b in the same way that class definitions are instantiated in Java. Once created, an actor starts executing a new thread of control that handles messages that are sent asynchronously between actors. Pattern matching is used in arms that occur in case-expressions and message handling rules. Uncertainty is supported by probably(p) x y that evaluates x in p% of cases, otherwise it evaluates y. Functions differ from actors because they are invoked synchronously.

A minimal ESL application defines a single behaviour called main, for example:

```
1 type Main = Act{ Time (Int) };
2 act main :: Main {
3 Time (100) ! stopAll ();
4 Time (n:: Int) ! {}
5 }
```

An ESL application is driven by time messages. The listing defines a behaviour type (line 1) for any actor that can process a message of the form Time(n) where n is an integer. In this case, the main behaviour defines two message handling rules. When an actor processes a message it tries each of the rules in turn and fires the first rule that matches. The rule on line 3 matches at time 100 and calls the system function stopAll() which will halt the application. Otherwise, nothing happens (line 4).

4.3. Supporting Uncertainty

4.3.1. Decision space

Decision space can be broadly classified into four categories namely Known Knowns (KK), Known Unknowns (KU), Unknown Knowns (UK), and Unknown Unknowns (UU) as shown in Fig 6.

In the context of decision making, KK space denotes full certainty about what are the possible next states as well as availability of information to decide which of them to be the next state. KU space denotes full certainty about what are the possible next states but uncertainty as regards determining the next state – instead a probability distribution is considered available. UK space denotes partial information i.e. only a subset of possible next states are known and there could well be uncertainty as regards which would be the next state. UU space denotes lack of information about

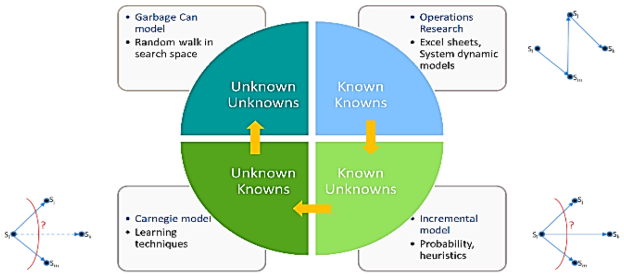


Fig 6. Decision spaces and kinds of uncertainty

the possible next states as well as uncertainty about which would be the next state.

This paper addresses decision making for KU space only.

4.3.2. Specification of uncertainty

Organisation is a set of interacting Units. A Unit interacts with other Units by sending messages.

Unit is a 5-tuple < I, S, B, St, Q > where,

I denotes immutable identity of the Unit,

S denotes its structure in terms of Unit and Unit relations shown in Fig 4,

B denotes behavior of the Unit in terms of event-handling i.e., a set of < e_i, eh_i > tuples where 'e' corresponds to an event (or a message) and 'eh' corresponds to the corresponding event handling logic

St denotes state of the Unit in terms of its Data and History i.e., a set of < var_i, val_i > tuples where 'var' corresponds to an attribute (of the Unit) and 'val' corresponds to its value

Q denotes a FIFO queue wherein messages received by the Unit are arranged in temporal order

The notion of Uncertainy impacts Unit specification in the following manner,

Unit.State takes the form $var_i = val_i$ @ p_i | val_j @ (1 - p_i) where 'p' is the probability. This can be extended to 'n' values with the constraint that sum of the probabilities equals 1.

Unit.Behaviour: < e_i, eh_i > association takes the form < e_i, eh_i @ p_i | eh_i @ (1 - p_i) where 'p' is the probability. This can be extended to 'n' event-handlers with the constraint that sum of the probabilities equals 1.

Unit.Queue gets impacted with "Send m_i to U_i" taking the form "Send m_i to U_i @ p_i | U_j @ p_j where 'p' is the probability. This can be extended to 'n' values with the constraint that sum of the probabilities equals 1.

On similar lines, Becomes takes the form "Becomes U_i @ p_i | U_j @ (1 - p_i) where 'p' is the probability. This can be extended to 'n' values with the constraint that sum of the probabilities equals 1.

Unit.Identity has no impact

Unit.History has no impact

[6] The current version of ESL is available at https://github.com/TonyClark/ESL

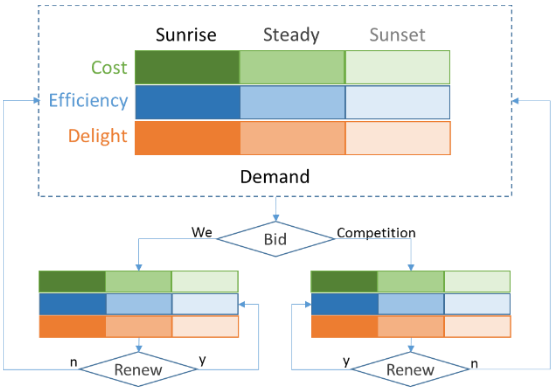


Fig. 7. Business Process Outsourcing

The above mentioned concept of Uncertainty is supported in ESL by the construct probably(p) x y.

4.3.3. Simulation

Language outlined in the previous section helps specify behavior of a system i.e. the input-output transfer function of Fig 3. Given such specification, an initial setup, and a sequence of events, it is possible to observe the impact of perturbation of *levers* onto the *measures* using simulation. In the interest of space, we do not describe simulation of the language outlined in the previous section, instead, we focus on the impact of uncertainty on simulation.
As discussed earlier, introduction of uncertainty impacts handling of an event by a Unit, sending of a message to a Unit, and assigning of a value to a variable of a Unit. Each of these, in fully certain situation, have exactly one course of action available. Presence of uncertainty introduces multiplicity of actions along with probability for each. In other words, a single value is replaced by a probability distribution.
Let's assume that uncertainty manifests at 'n' places in system specification. As a result, there are 'n' probability distributions – one for each place. Identifying a value from the corresponding probability distribution for every single place where uncertainty has manifest leads to fully certain specification that can be simulated without any change (to the simulator) whatsoever. However, the resultant output can be thought of as a true representation of the system behavior as a whole only if the 'n' probability distributions are fully covered. This identifies two additional demands on simulation: *how many simulation runs are required to fully cover the 'n' distribution functions*? *which value to pick from the probability distribution for a given simulation run*? These questions can be readily answered by making use of established results from probability and statistics (Dellino and Meloni, 2015).
Let's assume that the 'n' probability distributions necessitate 'k' simulation runs. This leads to 'k' sets of output variables each describing just a snapshot of the analysis. Thus emerges the question: *how to present the 'k' output datasets to the user so that meaningful inference can be drawn*? This question too can be readily answered by making use of established results from probability and statistics. For instance, the 'k' datasets can be collapsed into a single dataset describing value for an output variable, its confidence level, and confidence interval.
Thus, impact of introduction of uncertainty in system specification can be fully and effectively addressed by making use of established results from probability and statistics while leaving the simulation engine largely untouched.

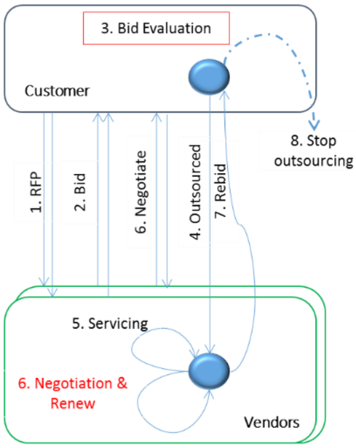


Fig. 8. Interactions and transitions

5. VALIDATION

5.1. Illustrative Example

Consider business process outsourcing (BPO) space. Customers outsource business processes for a variety of reasons such as reducing cost (C), increasing efficiency (E), bringing about a major transformation (D) *etc*. The outsourced processes can be classified into three buckets based on maturity of BPO for the specific process and the vertical. For instance, Transcript Entry process of Healthcare vertical was one of the first to take to BPO and has derived almost all potential benefits accruable from outsourcing (i.e. Sunset or SS). On the other hand, IT Infrastructure Management process being a late adopter of BPO has a large unrealized potential to be tapped (i.e. Sunrise or SR). And there are processes such as Help Desk, Account Opening, Monthly Alerts etc., that fall somewhere in between the two extremes as regards benefits accrued from BPO (i.e. Steady or ST). Thus, BPO demand space can be viewed as a 3 x 3 matrix of Fig 7. A Customer invites bids from the vendors for a specific BPO project or takes help of an external agent to identify a vendor. Typically, factors such as *Quadrant* (i.e. ranking as per independent agency such as analysts), *FTE Count Range* (i.e. min-max count of full time employees to be deployed on the outsourced process), *Billing Rate Range* (i.e. min-max range for per hour rate of full time employee), *Market Influence* (i.e. perception of the market as regards delivery certainty with acceptable quality) etc., influence who wins the bid. Other soft issues such as familiarity with the processes being

	Quadrant	Billing Rate ($/Hr)		FTE Quotation	Negotiation Levers		Market Influence	Delivery Excellence
		Min	Max	Deviation from Standard	FTE Deviation (%)	Billing Rate deviation (%)		[Probability of Excellent,Probability of Good, Probability of Normal, Probability of Below Normal]
Efficiency	Leader	8	12	8	2	5	Excellent	[60, 40,10, 0]
Effectiveness	Visionary	16	24	6	0	5	Good	[50, 40,10, 0]
Delight	Contender	120	140	4	5	0	Normal	[20, 60,20, 0]

Fig 9.a Characteristics of ‘We’ Unit

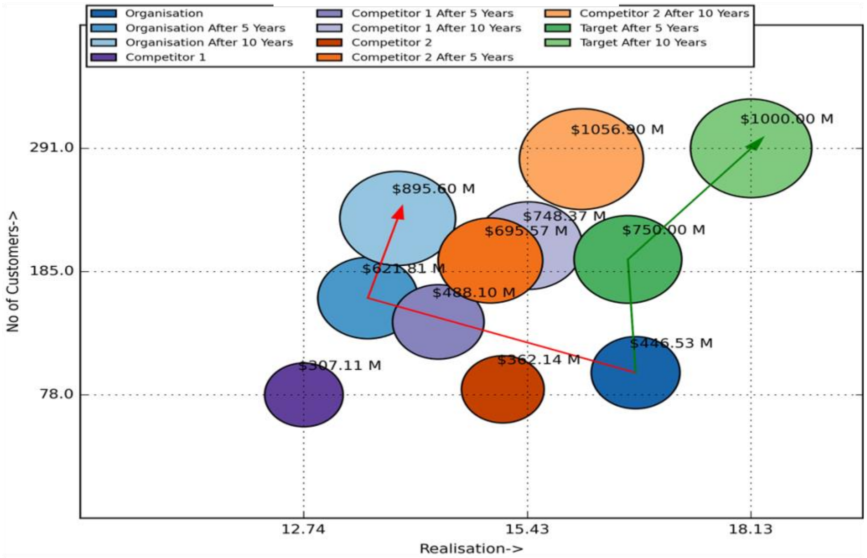


Fig 9.b Comparative Progression

outsourced, rapport with the vendor *etc*., also play a part in selection of the vendor. It is common observation that BPO contracts come up for renewal after few years. Customer may renew the contract with the existing vendor on modified terms (typically advantageous to the customer) or may opt for rebidding. Factors influencing the renewal decision are reduction offered in *FTE Count*, *Billing Rate*, number and degree of escalations, perception the external agent has as regards ability to meet the project requirements *etc*. Contracts that fail to get renewed become candidates for later bidding. Fig 7 shows a high level schematic of BPO space with events of interest therein detailed out in the form of a state transition diagram of Fig 8.

The demand space exhibits temporal dynamism. For instance, new processes emerge as candidates for outsourcing and some of the existing processes no longer need to be outsourced as, say, technology advance eliminates the need for human intervention in the process thus making it straight-through. Thus BPO space can be viewed as an event-driven system where events have a certain frequency and are stochastic in nature. The frequency and stochastic characteristics typically vary from process to process. While operating in this uncertain space, a BPO vendor needs to make decisions of the following kind: *Will continuation with the current strategy keep me viable ‘n’ years hence? What alternative strategies are available? How effective will a given strategy be? By when a given strategy will start showing positive impact? Will I be growing at the expense of competition or vice versa?* And so on.

Answers to the above questions are essentially linked to the evaluation of portfolio basket i.e., 3 x 3 matrix of Fig 7, of the organisation in terms of revenue accruable and expense. Therefore, ability to predict portfolio basket of the organisation and its competitors after a given time period becomes critical to support data-driven decision making.

5.2. Simulation setup and results

In the interest of space, we only address the question: *Will continuation with the current strategy keep me viable ‘n’ years hence with respect to the competition*?

We model Demand as a Unit comprising of set of SR, ST and SS Units each comprising of set of C, E and D Units. Increase (or decrease) in SR demand is modelled as increase (or decrease) in SR.C, SR.E and SR.D Units at a pre-defined frequency and probability – and on similar lines for ST and SS processes.

We and Competition are modelled as Units each comprising of a set of SR, ST and SS Units such that each member of the set comprises of a set of C, E and D Units. Fig 9.a shows sample characteristics of ‘We’ Unit. Values of its attributes such as *Quadrant*, *Billing Rate*, *FTE count*, *Market Influence* and *Delivery Excellence* contribute towards its ability to win a bid. As can be seen from Fig 9.a, ‘We’ Unit is best equipped to win BPO contracts aimed at cost reduction. Value of *Delivery Excellence* attribute is a probability distribution. For instance, ‘We’ Unit is confident of delivering ‘Excellent’ quality on 60% of ‘C’ kind of BPO projects won. The values for ‘Good’, ‘Normal’ and ‘Below Normal’ quality for this kind of BPO projects are 30%, 10% and 0% respectively. The Unit is

	Quadrant	Billing Rate ($/Hr)		FTE Quotation	Negotiation Levers after term completion		Market Influence	Delivery Exc
		Min	Max	Less FTE (in %) Quotation in Bid	FTE Deviation (%)	Billing Rate deviation (%)		[Probability of Excellent, Proba Probability of Normal, Probab
Efficiency	Leader	12	14	10	5	1	Excellent	[90, 10,0, 0]
Effectiveness	Visionary	25	30	10	5	1	Good	[80, 10,10, 0]
Delight	Contender	110	120	10	5	1	Excellent	[70, 30,0, 0]

Fig 10.a Modified characteristics of 'We' Unit

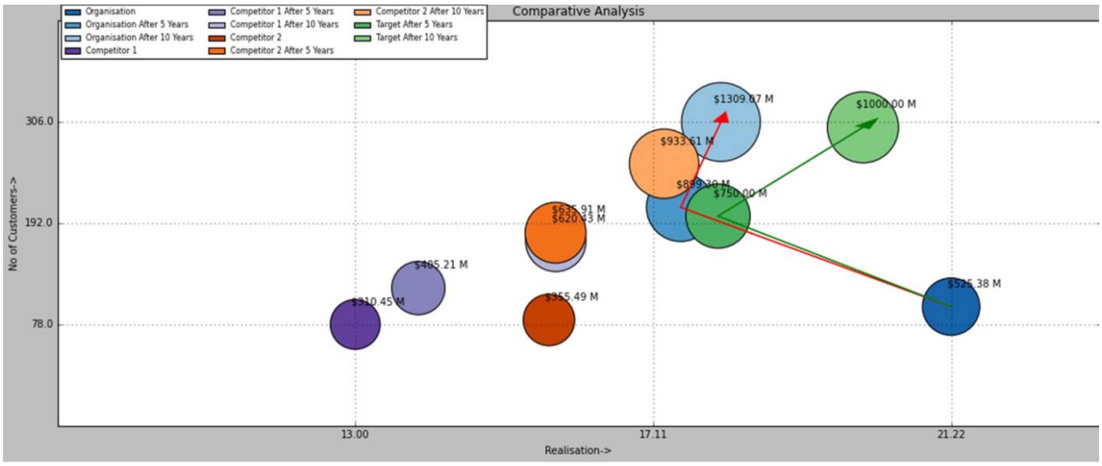


Fig 10.b Comparative Progression

equipped with two negotiation levers namely, what's the percent reduction possible in number of full time employees billed against the outsourced process, and what's the percent reduction possible in per hour billing rate for full time employee. The variables *Billing Rate* and *FTE Count* are both ranges and a value is picked at random from the specified range.

The 'Competition' Unit is modelled on the same lines as 'We' Unit. We can model multiple 'Competition' Units each having different characteristics.

Simulator raises RFP events at random. Each RFP event is characterized by the kind of process being outsourced (i.e. SR or ST or SS), the objective for outsourcing (i.e. C or E or D), size of the process in terms of FTE count, and the desired billing rate. All vendors respond to the RFP event by picking suitable values for their characteristics at random. Bid evaluation function is a weighted aggregate of the various elements of RFP response and a random valu e to capture effect of inherent uncertainty. The vendor with the lowest bid wins the project which gets executed as defined by the characteristics of the particular vendor. In other words, we acknowledge that it won't be possible to know everything about the domain and we can't be fully certain of the domain knowledge acquired so far.

The decision to renew existing contract is modelled on similar lines but with a different set of characteristic attributes influencing the decision. Here too, we take cognizance of incomplete and uncertain knowledge about the problem domain by introducing a random variable in the evaluation function.

This completes the simulation setup and the system is allowed to run for 10 years. Results of the simulation run are shown in Fig 9.b. As can be seen, our current revenue is 446.54 MUSD from 90 customers with a realization of nearly 17 USD per hour per person. Corresponding numbers for competitor#1 and competitor#2 respectively are < 307.11, 78, 12.74 > and < 362.14, 80, 15 >. In short, at present we are doing much better than competition.

We are set a goal to deliver < 750, 200, 17 > after 5 years and < 1000, 290, 18 > after 10 years. As can be seen, by continuing to operate the same way we will be delivering < 621.81, 160, 13.5 > after 5 years and < 895.6, 215, 14 > after 10 years thus missing both the targets by a considerable margin. More importantly, competitor#2 will be overtaking us after 5 years and both the competitors will be significantly ahead of us after 10 years.

Clearly, we cannot afford to continue with our current way of operation. Further detailed analysis, involving model elements not described in this paper for want of space, shows that much of our current revenue is from sunset processes outsourced for cost reasons. Over time this market is going to shrink considerably with demand for steady as well as sunrise processes (for objectives other than pure cost reduction) increasing significantly. We need to bring about a change in our characteristics so as to be able to win more bids in this demand situation. Fig 10.a shows the modified characteristics of 'We' Unit leading to the improved performance as shown in Fig 10.b. We are able to beat both revenue and customer targets while failing to meet the realization target narrowly.

5.3. Evaluation

For the kind of decision making problem illustrated in this paper, industry practice relies extensively on excel sheets. Such an approach typically represents the influence of lever onto measures in terms of static algebraic equations. However, value of a lever and

influence of a lever onto a set of measures can vary over time. This behaviour cannot be captured using excel sheet. There is no support for encoding stochastic behaviour either.
Systems dynamics models are also used for this kind of decision-making wherein the system is specified in terms of stocks, flows of stocks, and equations over system variables that control the flows. Value of a stock or a flow or a variable can be a discrete number or a range or a distribution. The quantitative nature of systems dynamics models and sophisticated simulation support enables decision making through what-if scenario playing. It is possible for a stock or an individual variable to have a value that is a probability distribution, however, structure of the stock-n-flow model must remain unchanged. Thus, systems dynamics modeling provides only a partial support for specifying and processing the inherent uncertainty within a system. Moreover, it is best suited for an aggregated and generalized view of a system where individual details get eliminated through averaging, and sequences of events are grouped as continuous flows. This generalized approach and ignorance of individual characteristics that significantly influence the system over time often leads to a model that is somewhat removed from reality. Though not designed to specify specialized behaviour, it can be done using systems dynamics modeling. But this is an effort-intensive endeavour, and more importantly leads to model size explosion. For example, modeling of 4 competitors each having special characteristics leads to roughly 4 times increase in the size of systems dynamics model.
The proposed language enables modelling of a system as a set of units each listening/responding/raising events of interest and interacting with other units by sending messages. A unit encapsulates state (i.e. a set of attributes), trace (i.e. events it has responded to and raised till now) and behavior (i.e. encoding of lever-to-measure influence). Thus, the language subsumes systems dynamics model. As the language supports 'time' concept, value of a variable and relationships between variables can change with respect to time. Consider the example of determining the impact of track record on winnability of organisation where the value of track-record variable changes over time thus affecting winnability. Since a process is an individual actor, simulator can determine the impact of successful contract completion, renewal with/without negotiation etc., for that specific process – systems dynamics model falls short here. A trace of events serves as a memory that can be queried to establish more complex relationships between levers. For example, successful completion of contract leads to improved track record as well as better rapport with the customer thus improving winnability of future outsourcing bids everything else remaining the same. Thus, the language provides primitives for creating models that closely mimic reality.
Thus, it can be said that the proposed language subsumes Excel sheets as well as systems dynamics model. Excel sheets provide no support for specifying or processing uncertainty. Systems dynamics model provides only partial support for uncertainty as it is possible for variables to have value as a probability distribution (as opposed to a discrete value or a range) but the flows remain fully deterministic. The proposed language provides for the flows to be probabilistic too. Therefore, proposed approach presents a pragmatic solution for supporting decision-making in presence of uncertainty using simulation. The solution addresses uncertainty by using established results from probability and statistics while leaving the simulator implementation largely unchanged. The proposed approach is validated using a representative example from real world.

6. SUMMARY AND FUTURE WORK

Effective decision-making is a challenge all modern enterprises face. It requires deep understanding of aspects such as organisational goals, structure, operational processes etc. Large size, siloed structure, and increasingly fast business dynamics means the available information is either incomplete or uncertain or both. Inability to handle this inherent uncertainty is a present lacuna in current industry practice of organisational decision-making. We began by outlining an approach based on the principles of "separation of concerns" and "divide and conquer" that enables creation of goal-measure-lever graph to capture understanding of the problem domain at a higher level of abstraction. We outlined a language for specifying the graph structure in a manner that is amenable to what-if scenario playing. Decision-making thus is a bottom-up walk of this graph structure i.e., human-guided simulation-aided exploration of solution space. We then discussed where and how uncertainty affects the graph structure and proposed extensions to the language so as to be able to externalize the uncertainty at specification level. We showed how well-established results from statistics and probability can be used to implement the manifest uncertainty without changing the core simulation engine of the language. We presented a validation of the approach using a representative example from real life.
The approach has been illustrated with a substantive example from Business Process Outsourcing domain. We have shown the example can be modeled and simulated leading to the ability to influence the strategically selected measures. However, we recognise that the current implementation of ESL is not sufficiently high-level for direct adoption by decision-makers. Our immediate next step is to develop higher level abstractions to support the core ESL concepts in a business-facing manner. In doing so, we will adopt language processing and model transformation technology to enable support for defining domain specific languages geared for specific problems. We note that decision-making is more a satisfaction problem rather than an optimisation problem. Consequently, we will draw upon game theory and

computational economics to consider extending our proposed solution to impart this characteristic.

REFERENCES

Ackoff, R., 1971. Towards a system of systems concepts. Management science, 17(11), pp.661-671.

Agha, G., 1985. Actors: A model of concurrent computation in distributed systems. No. AI-TR-844. Massachusetts Institute Of Technology Cambridge Artificial Intelligence Lab.

Barjis, J., 2008. Enterprise, organization, modeling, simulation: putting pieces together. Proceeding of EOMAS.

Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences, 99(suppl 3), pp.7280-7287.

Bruneton, E., Coupaye, T., Leclercq, M., Quéma, V. and Stefani, J.B., 2006. The fractal component model and its support in java. Software: Practice and Experience, 36(11-12), pp.1257-1284.

Clark, T. and Barn, B., 2013. Goal Driven Architecture Development using LEAP. Enterprise Modelling and Information Systems Architectures 8(1): 40-61 (2013).

Daft, R., 2012. Organization theory and design. Nelson Education.

Dellino, Gabriella, and Carlo Meloni. Uncertainty management in simulation-optimization of complex systems. Springer, 2015. Evans W.A., 1994. Approaches to intelligent information retrieval. Information Processing and Management, 7 (2), 147–168.

Jonkers, H., Lankhorst, M., Van Buuren, R., Hoppenbrouwers, S., Bonsangue, M. and Van Der Torre, L., 2004. Concepts for modeling enterprise architectures. International Journal of Cooperative Information Systems, 13(03), pp.257-287.

Krogstie, J., 2008. Using EEML for Combined Goal and Process Oriented Modeling: A Case Study,[w:] T. Halpin, J. Krogstie, E. Proper. In Proceedings of EMMSAD'08. Thirteenth International Workshop on Exploring Modeling Methods for Systems Analysis and Design.

McDermott, T., Rouse, W., Goodman, S. and Loper, M., 2013. Multi-level modeling of complex socio-technical systems. Procedia Computer Science, 16, pp.1132-1141.

Meadows, D. and Wright, D., 2008. Thinking in systems: A primer. Chelsea green publishing.

Michelson, B., 2006. Event-driven architecture overview. Patricia Seybold Group, 2.

Pnueli, A., 1977, October. The temporal logic of programs. In Foundations of Computer Science, 1977., 18th Annual Symposium on (pp. 46-57). IEEE.

Van Harmelen, F., Lifschitz, V. and Porter, B. eds., 2008. Handbook of knowledge representation (Vol. 1). Elsevier.

Vernadat, F., 2002. UEML: towards a unified enterprise modelling language. International Journal of Production Research, 40(17), pp.4309-4321.

White, S, 2004. Introduction to BPMN. IBM Cooperation, 2(0).

Yu, E., Strohmaier, M. and Deng, X., 2006, October. Exploring intentional modeling and analysis for enterprise architecture. In Enterprise Distributed Object Computing Conference Workshops, 2006. EDOCW'06. 10th IEEE International (pp. 32-32). IEEE.

Zachman, J.A., 1987. A framework for information systems architecture. IBM systems journal, 26(3), pp.276-292.

AUTHORS BIOGRAPHY

Vinay Kulkarni is Chief Scientist at Tata Consultancy Services Research (TCSR). His research interests include model-driven software engineering, self-adaptive systems, and enterprise modeling. His work in model-driven software engineering has led to a toolset that has been used to deliver several large business-critical systems over the past 20 years. Much of this work has found a way into OMG standards. Vinay also serves as Visiting Professor at Middlesex University London.

Souvik Barat is a Senior Scientist at Tata Consultancy Services Research (TCSR). His research interests include model-driven software engineering, and enterprise modeling.

Tony Clark is Professor of Software Engineering at Sheffield Hallam University in the UK. His academic research on meta-modelling led to the development of a tool called XModeler that has been used in a number of commercial applications including the development of tool support for a new Enterprise Architecture modeling language.

Balbir Barn is Professor of Software Engineering at Middlesex University London with extensive experience in industrial Software Engineering including the design and implementation of the IEF.

INFLUENCE OF MANUFACTURING PROCESS IN MECHANICAL BEHAVIOR OF INJECTED PLASTIC COMPONENTS.

Manuel Muniesa[(a)], Isabel Clavería[(b)], Carlos Javierre[(c)], Daniel Elduque[(d)], Ángel Fernández[(e)]

[(a), (b), (c), (d), (e)]Universidad de Zaragoza, Spain

[(a)]mmuniesa@unizar.es , [(b)]iclaver@unizar.es , [(c)]sabicjl@unizar.es
[(d)]danielduque88@gmail.com , [(e)]afernan@unizar.es

ABSTRACT

Mechanical behavior of automotive injection components is usually analyzed without regard the conditions of manufacturing process. Some consequences due to plastic injection process must be taken into account, such as fiber orientation, residual stress or weld line locations. These factors are specially relevant for reinforced materials. However, traditionally CAE tools are limited to analyze stress, strain or displacement results from a isotropic material model.

This paper shows an integrated analysis with CAE tools, including results of manufacturing process in mechanical calculations. Two types of reinforced thermoplastic material are analyzed: PP and PA6. Glass fiber is used for the reinforcement of both materials and fiber orientation tensor due to injection process is taken into account. Also, influence of fiber material percentage is analyzed. Results obtained using integrated methodology show important differences in comparison with traditionally method. Is specially relevant the difference obtained in displacement results.

Keywords: part design, plastic injection process, fiber orientation, mechanical behavior.

1. INTRODUCTION

The development of components made with reinforced thermoplastic materials produced by injection process is on the increase nowadays, and therefore the accuracy of design and the correct calculations involved in making these products has become very important. By using CAE tools we can predict the behavior of these components when in use (Lin et al. 2004). These simulation analyses can save a lot of time and money when used in the design phases of said components. Up to now it has been habitual to make simulations in separate phases of production depending on the types of analysis behavior used. On the one hand analyzing the fabrication process of the components and on the other hand seeing how they behave under conditions of use. However, in the case that concerns us in relation to reinforced materials, these different phases are strongly related. Manufacturing process determines fiber orientation on the matrix of reinforced material (Parveeen et al. 2014), and therefore strongly conditions its later response to the tests which it will undergo (Adam and Assaker 2014).

This article presents a work methodology for the integrated simulation of manufacturing process of reinforced materials and their mechanical responses as well as analyzing the influence of the percentage of fiber in the material and the type of polymer matrix (Doghri and Tinel 2004). For this study a component of a turbo compressor for a vehicle has been used (Ruiz et al. 2008). The said component is subjected to internal pressure and it made with thermoplastic material and reinforced with fiberglass (Bicart et al. 2011). The design of this component is determined, mechanically speaking, by the restrictions of the injection process used in its fabrication, the material used and the strength required (Fernández et al. 2013, Demirer and Deniz 2012). The consequences that the injection process has on the component are very well known, such as the appearance of shrinkage and warpage, residual stress and fiber orientation once it is made (Fernández et al. 2014, Yu et al. 2014, Ozcelik and Sonat 2009). These results should be kept in mind when making a mechanical-structural analysis of the component by increasing the load or applying greater security coefficients.

2. METHODOLOGY

This article presents a work method that objectively integrates the reactions over the component by the injection process as well as the conditions of the load when working with composite materials in the mechanical analysis. More specifically, the fiber orientation tensor of composite material is integrated into the mechanical-structural analysis.

In most cases, only structural analysis with commercial software are made. Start point for a lineal static analysis is a mathematical material model with elastic modulus and Poisson´s ratio coefficients.

For the integrated methodology first step is to create a structural mesh (for example with software Abaqus©) of the part. In this case a component of a turbo compressor. Also, a rheological analysis of manufacturing process is made by Moldflow© software. A different mesh with more elements and nodes is used for calculations along part thickness. Injection time, number of plastic entrance for injection process, material viscosity, process temperature and pressure are different parameters to be considered.

A result obtained of this analysis is the fiber orientation tensor. A file with .xml extension contain all the information about principal directions and elastic modulus in each element of the rheological mesh.

Digimat© software coupled fiber orientation tensor file, structural mesh and a mathematical material model based on experimental analysis. The output of this software is a new structural mesh file with two additional files (.mat, .dof) wich are ready to use in a structural analysis with anisotropic influence by fiber orientation added.

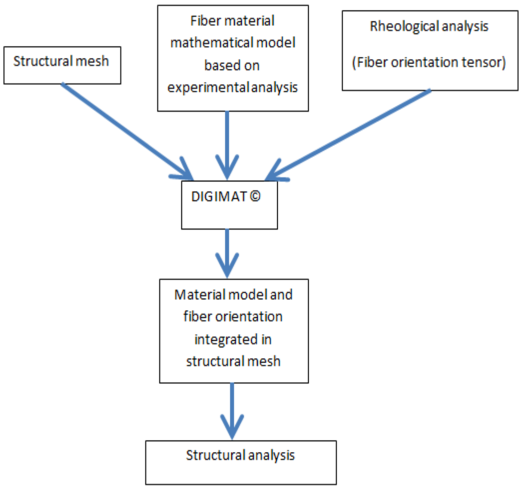


Figure 1: Scheme of integrated methodology applied.

In order to see the influence that the additional procedural and mechanical calculations make, we will analyze different cases with different material compositions. To do this, we will apply the methodology already developed for mechanical analysis of said vehicle component made with: polypropylene 20% fiberglass (Hostacom©G2U02), polypropylene 50% fiberglass (Celstran©PP-GF50-03), polyamide 6 20% fiberglass (Durethan©BKV20FND) and polyamide 6 50% fiberglass (Durethan©BKV50H20).

Material Type	PP+20%GF	PP+50%GF	PA+20%GF	PA+50%GF
Trade Name	Hostacom ©G2U02	Celstran©P P-GF50-03	Durethan© BKV20FND	Durethan ©BKV50H 20
Density	1,05 g/cm^3	1,33 g/cm^3	1,35 g/cm^3	1,57 g/cm^3
Young Modulus	4800 MPa	10300 MPa	6100 MPa	16000MPa
Injection Temp.	220°C	260°C	270°C	290°C
Mold Temp.	40°C	50°C	80°C	80°C

Table 1: Material properties

Moldflow©, Digimat© and Abaqus© CAE software have been used for analysis. In first step, process manufacturing is analyzed with Moldflow© software. Injection parameters like injection time, melt temperature, mold temperature, packing and cooling time are scheduled. Results of injection pressure, bulk temperature, weld line location, air traps or fiber orientation tensor are obtained. Digimat© software is used to couple process analysis with mechanical analysis through a material numerical model including fiber orientation and mechanical properties. Finally, mechanical analysis is made by Abaqus© software using this new material input data.

3. RESULTS

A comparison between integrated methodology and single mechanical analysis has been made. Results from manufacturing process simulation are showed in Figure 1 and Table 1.

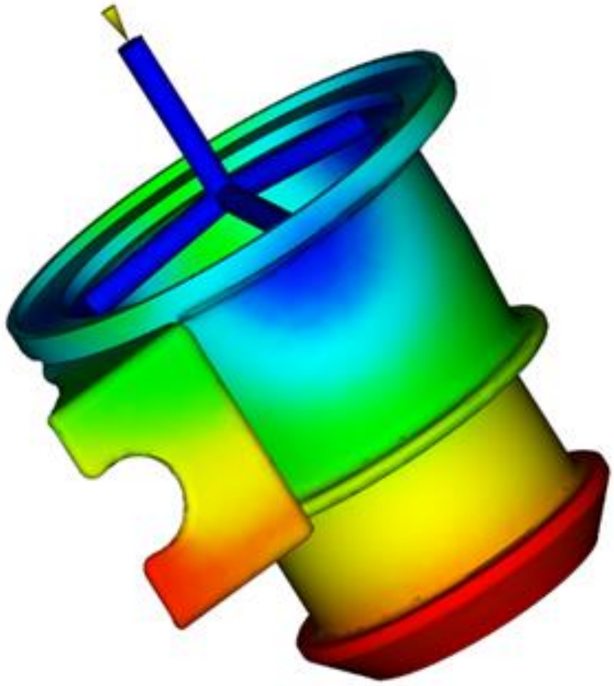


Figure 2:Melt front advance

Material	Max. Injection Pressure	Flow Front Temperature Difference
PP + 20%GF	8 MPa	3°C
PP + 50%GF	14 MPa	4°C
PA6 + 20%GF	16 MPa	3°C
PA6 + 50%GF	22 MPa	6°C

Table 2: Injection results

Higher values of maximum injection pressure is given for polyamide 6 with 50 % glass fiber reinforcement due to viscosity of polymer. Results of fiber orientation tensor are showed in Table 2. This result reveals the importance influence than percentage of fiber reinforcement has in the structural morphology of reinforced material.

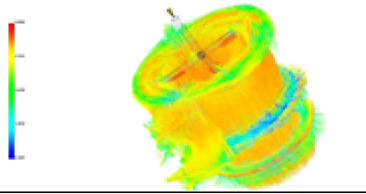
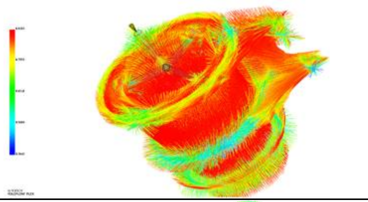
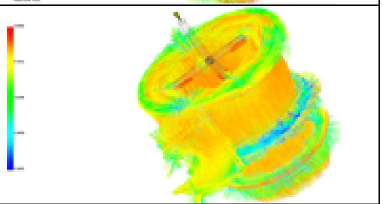
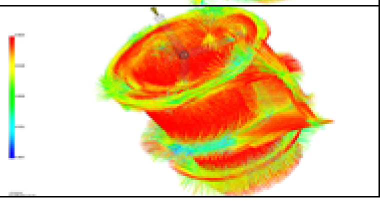
Material	Fiber orientation
PP + 20%GF	
PP + 50%GF	
PA6 + 20%GF	
PA6 + 50%GF	

Table 3: Fiber orientation tensor results

Abaqus© analysis show stress, strain and displacement results under boundary conditions applied, like load or fixation points, Figure 2. In this case, stress results with material model input calculated previously by Digimat© software are showed in Figure 3.

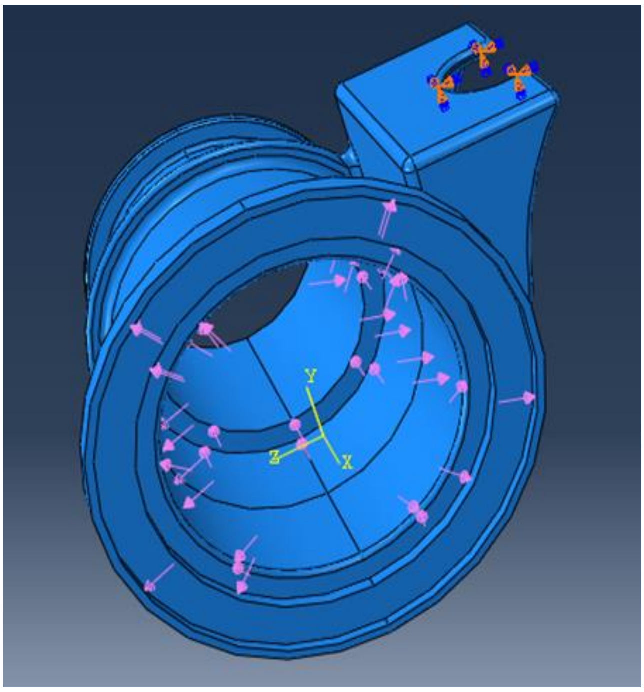


Figure 3: Boundary conditions applied

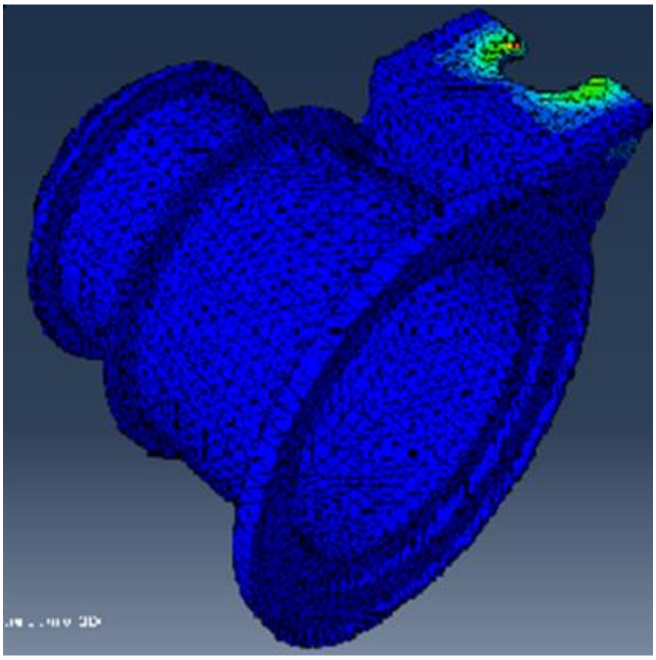


Figure 4: Von Mises stress results for integrated analysis

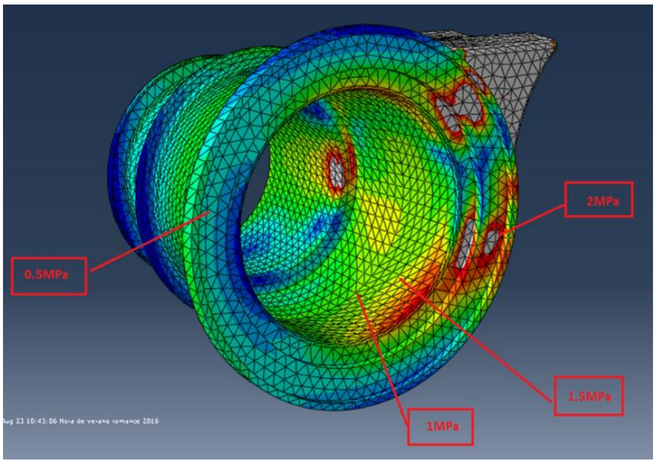


Figure 5: Von Mises stress detail results for integrated analysis in the inner zone of the part.

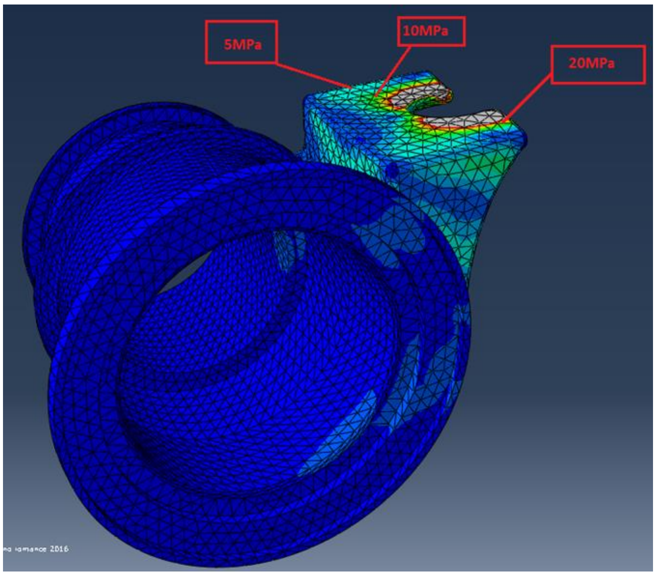


Figure 6: Von Mises stress detail results for integrated analysis in the fixing zone of the part.

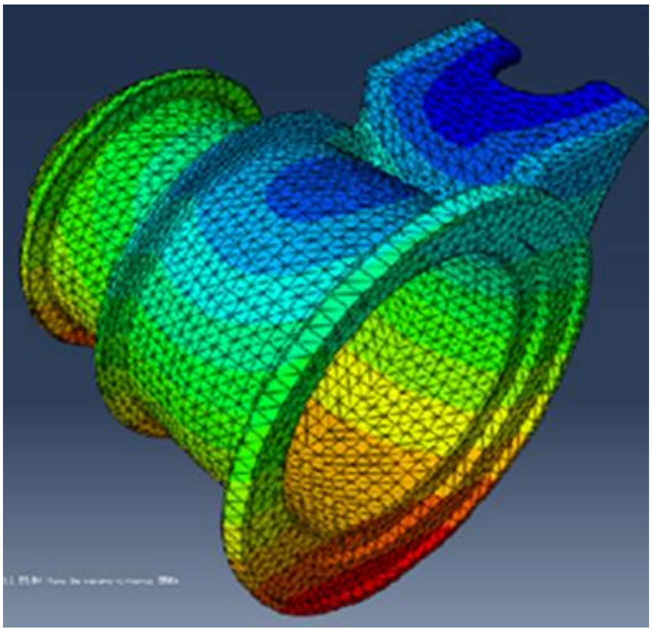


Figure 7: Displacement results for integrated analysis.

One will observe that the resulting displacements nearly double when taking into consideration the anisotropic behavior of the material using the fiber orientation tensor, calculated in the process simulation step. Results with integrated methodology and with traditionally isotropic material assumption are shown in Table 3.

Material	Displacements with integrated methodology	Displacements with isotropic material assumption	Difference (%)
PP + 20%GF	2.73 mm	1.65 mm	64.7 %
PP + 50%GF	1.43 mm	0.62 mm	128.8 %
PA6 + 20%GF	1.51 mm	1.14 mm	32.1 %
PA6 + 50%GF	0.82 mm	0.45 mm	82.8 %

Table 3: Displacements comparation

Displacement result difference is much more noticeable when working with materials with 50% fiber than ones with only 20%. Additionally, the displacements are seen to increase in greater measure when working with polypropylene. If both of these variants occur at the same time the displacement is more than double that which is obtained without applying the integrated methodology.

4.-CONCLUSIONS

A study of the variation of results in mechanical simulations of an automotive part, implementing all of the information gathered in the manufacturing process rather than only the results that do not take into account such information, has been realized. Additionally the influence of the percentage of fiber and the material used has been analyzed.

The results obtained using a composite material with a polypropylene matrix and another with a polyamide matrix, both made of fiberglass, have been compared. A greater influence has been seen in the results of the polypropylene examples, probably due to the different relation between matrix and reinforcement, which induces a better orientation of the fiber on this type of matrix than on the polyamide one.

The percentages of fiber compared, in the polyamide as well as in the polypropylene components, have been 20% and 50%. Upon adding a higher percentage of fiberglass to the material the resistance is increased, as can be seen in the values determined in Young´s model and Poisson´s coefficient. That is why the values of fiber displacement is higher in components with only 20% of said material. However, upon applying the methodology, the values increase much more for the 50% examples that for those of 20%. This contrast is

attributed to the fact that the fiber is better oriented in the case of the 50% samples, but the way of orientation with regard to load applied, make it less efficiency.
To conclude, the mechanical-structural analysis highly depends upon the fabrication process, because that is where the real displacements that the product undergoes are taken into account. On the contrary, if the manufacturing process of the item is not considered, we run the risk of have failed designs if the values of stress, strain or displacements differ much between one case and another.

REFERENCES

Adam, L., Assaker, R., 2014. Integrated nonlinear multi-scale material modelling of fiber reinforced plastics with DIGIMAT RP®: application to short and continuous fiber composites. 11th World Congress On Computational Mechanics; 5th European Conference On Computational Mechanics; 6th European Conference On Computational Fluid Dynamics II – IV, 2322-2333.

Bicart, J., Robert, G., Moulinjeune, O., 2011. Multiscale Modeling of the Orthotropic Behaviour of PA6-6 overmoulded Composites using MMI (TM) Approach. 14th International Conference On Material Forming Esaform, 2011 Proceedings 1353, 726-731.

Demirer, A., Deniz, S., 2012, Investigation of the effects of weld lines on the mechanical properties and energy consumption for injection moulded thermoplastics. Energy Education Science and Technology. Part A-Energy Science and Research 29, 1055-1062

Doghri, I., Tinel, L., 2004. Micromechanical modeling and computation of elasto-plastic materials reinforced with distributed-orientation fibers. International Journal of Plasticity 21, 1919-1940.

Fernández, A., Muniesa, M., González, J., 2013. Characterisation and Processing of Reinforced PA 6 with Halloysite Nanotubes (HNT) for Injection Molding. Journal of Mechanical Engineering 59, 183-192.

Fernández, A., Muniesa, M., Javierre, C., 2014. In-line rheological testing of thermoplastics and a monitored device for an injection moulding machine: Application to raw and recycled polypropylene. Polymer Testing 33, 107-115.

Lin, B.J., Jin, X.S., Zheng, R., Costa, F.S., Fan, Z.L., 2004, 3D fiber orientation simulation for plastic injection molding. Materials Processing and Design Modeling Simulation and Applications, Parts 1 & 2. 712. 282-287

Ozcelik, B., Sonat, I., 2009, Warpage and structural analysis of thin shell plastic in the plastic injection molding Materials & Design 30, 367-375

Parveeen, B., Caton-Rose,P., Costa,F., Jin, X., Hine, P., 2014, Study of injection moulded long glass fibre reinforced polypropylene and the effect on the fibre length and orientation distribution. Proceedings of PPS-29: The 29th International Conference of the Polymer. Conference Papers, 432-435, Jul 15-19,2013, Nuremberg, Germany

Ruiz, I., Jimenez, E., Blanco, J., Pérez, M.,2008. Design and simulation of production of injection pieces in automobile industry. Second UKSIM European Symposium on Computer Modeling and Simulation, 335-340.

Yu, Y., Y., Zhang, B., Tang, Z., Qi, G., 2014. Stress transfer analysis of unidirectional composites with randomly distributed fibers using finite element method. Composites Part B 69, 278-285.

USE OF GAMES TO STUDENT TRAINING IN THE CIVIL ENGINEERING UNDERGRADUATION COURSE: A BIBLIOMETRIC ANALYSIS

FREIRE, A.[(a)], ELY, D.[(b)], SANTANA, L.[(c)], JUNGLES, A.[(d)], DALMAU, M.[(e)]

[(a)] Instituto Federal do Piauí
[(b)] Centro Federal de Educação Tecnológica de Minas Gerais
© Instituto Federal da Bahia
[(d)] Universidade Federal de Santa Catarina

[(a)]ailton.freire@ifpi.edu, [(b)]danielaely@civil.cefetmg.br,[(c)]santana.leiliane@gmail.com
[(d)]ajungle@ceped-ufsc.com, [(e)]marcos.dalmau@ufsc.br

ABSTRACT
This article is the result of a bibliographic portfolio selection process and bibliometric analysis made from a survey of information, knowledge and experiences from use of games to student training in civil engineering undergraduation courses available at periodicals in relevant databases to Teaching and Engineering area. The structure of this bibliographic review is based on the ProKnow-C (Knowledge Development Process – Constructivist) method (ENSSLIN, L et al., 2010) developed by the laboratory of Milticriteria Methodologies in Decision Support (Laboratório de Metodologias Multicritério em Apoio à Decisão - LabMCDA) from the Universidade Federal de Santa Catarina (UFSC). As a result of this analysis the keywords most used; the geographical location of the conducted researches highlighting the authors, universities and countries that publish the most; the journal with the greatest number of publications; the annual evolution of the scientific research related to this theme, furthermore also the identification relevant contributions of these researchers to student formation from the use of Serious games were pointed out.

Keywords: engineering education, simulation, gaming simulation, Case studies, ProKnow-C

1. INTRODUCTION

With each passing year, the search for use of games in undergratuation teaching to support the teaching and learn process increases (Sherif and Mekkawi 2010; Khenissi et. al. 2015).
The intention of using games in student training is to combine the serious aspect (learning, instruction, etc.) with the aspect of digital games (Sorense and Meyer 2007).
Learning from experimentation and simulation, together with the application of theoretical classes' results in a solid foundation, for students who need knowledge for decision making in real scenarios (Latorre and Jiménez 2012).
The main objective of this article is to present a bibliographic portfolio containing the relevant articles for the theme of use of games for student training in the engineering course, to serve as a guide to the construction of learning tools using games.
This set of articles raised will point to the most used keywords; The geographical location of the research carried out, highlighting the authors, universities and countries that publish the most; the journal with the greatest number of publications; the annual evolution of the scientific research related to this theme, furthermore also the identification relevant contributions of these researchers to the student formation from the use of Serious games were pointed out.

2. LITERATURE REVIEW

In this section the main concepts about the bibliometric studies and games in school and professional training will be presented.

2.1. Bibliometric studies

The bibliometric methods have been a widely used tool in many areas of science to measure scientific progress, to estimate the contribution of countries, institutions and researchers to world scientific production, and to identify new paths to be followed by knowledge (Noronha 2000).
The term Statistical Bibliography, now known as Bibliometrics, was first used in 1922 by researcher E. Wyndham Hulme to define a stage of knowledge responsible for counting and cataloging scientific documents.
In 1944, this term reappears in a work by Gosnell on the obsolescence of world scientific literature.
In an article entitled Statistical Bibliography in the Health Sciences written by L. Miles Raisig and published in the year 1962, the term Statistical Bibliography reappears, in this work which deals with analysis of citations in the area of health sciences for a period between the year 1931 and 1957.
Only in the year 1969, the researcher Alan Pritchard coined the term bibliometry as being the study that quantifies the processes of written communication (Prichard 1969).

Similar definition to the current, which says bibliometrics is a field of knowledge that uses mathematical and statistical technique, to quantify, describe and prognosticate the process of written communication (Pao 1989; Diodato 2012).

2.2. Games in education

Engineering is a profession that modifies the resources energy, materials, and information to transform into goods used by mankind, and the general objective of engineering education is "to prepare students to practice engineering and, in particular, to deal with the forces and materials of nature" (Feisel and Rosa 2005, p. 121).

The military were the first to realize the great advantage of using simulation and games in teaching - learning process (Michael and Chen 2006). In addition, this technique is also widely used in health and business training.

Games and simulations have been present in society for the past two centuries. It has been perceived that the use of these techniques can go beyond simple fun, and can add value to technical, scientific and intellectual training (Magee 2006).

The difference between serious games and games of entertainment is summarized in Table 1.

Table 1: Differences between Serious games and Entertainment games

	Serious games	**Entertainment games**
Task vs. rich experience	Problem solving in focus	Rich experiences preferred
Focus	Important elements of learning,	To have fun
Simulations	Assumptions necessary for workable simulations	Simplified simulation processes
Communication	Should reflect natural (i.e., non-perfect) communication	Communication is often perfect

Serious games usually refer to "games used for training, advertising, simulation, or education that are designed to run on personal computers or video game consoles" (Susi, Johannesson and Backlund 2007, p.3).

The use of serious games in teaching and learning process allows the student to experience real-world situations that are difficult to perform due to safety, cost or time to perform (Squire and Jenkins 2003; Corti 2006).

In this sense, Serius games cover all aspects of education - education, training and information (Michael and Chen 2006).

Serious games have been applied in several areas of knowledge for student formation, among these works we have: Herbsman 1986, AbouRizk and Sawhney 1994, Schumann et. al. 1997, Scott et. al. 2004, Philpot et. al. 2005, Rovner 2006, Freire et. al. 2016.

Serious game are also used in other domains (e.g. Logistics, Industry, Healthcare, Defense, etc.), how can it be found in: Raybourn 2007, Arnab et. al. 2013, Chittaro and Sioni 2015, Longo et. al. 2015, El-Beheiry et. al. 2017.

3. SELECTION OF THE BIBLIOMETRIC PORTFOLIO

In order to reach the result of selection and quantitative analysis of a bibliographic portfolio about the use of games in student training in the engineering courses, the ProKnow-C (KnowledgeDevelopmentProcess – Constructivist) method was used. It was developed by the laboratory of multicriteria Methodologies in Decision Support - LabMCDA (Multicriteria Methodologies in Decision Support) from Federal University of Santa Catarina (UFSC). This method has four main steps, as illustrated by figure 1:

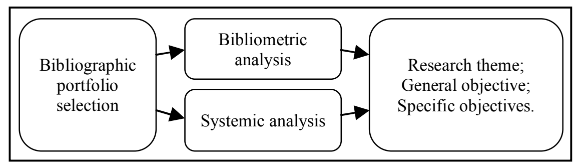


Figure 1: Process steps of ProKnow-C
Source: adaptaded from Ensslin et. al. 2010.

In general terms, the steps that were followed in the application of this method to reach the proposed goal in this work were to the selection of the bibliographic portfolio and the bibliometric analysis, made from the following steps: (1) Definition of keywords; (2) Selection of databases containing scientifically qualified journals on the researched topic; (3) Selection of articles related to the researched topic; (4) Creation of bibliographic portfolio relevant to the researched topic.

The first step was to define keywords related to the research theme; these should be in the title, abstract or keywords of the articles, when the database search was performed.

When these keywords were selected, a verification was performed in order to check if these actually discriminated scientific articles focus on the research topic, Defined the keywords (Table 2), the next step was to select the databases

Table 2: Keywords

Keywords
Engineering education
Serious games
Gaming simulation
Active learning
Innovative pedagogy

The selection of databases for constructing the gross articles bank should have a proven scientific relevance with periodicals related to the Engineering area. From this criterion, the selected databases were: ACM Digital Library; Elsevier (Science Direct); IEEE Xplore; Scopus and Springer.

With the databases defined, searches were performed using the keywords in each of these bases, thus generating the database of gross articles bank (Table 3).

The articles selected in these databases should be between January 2005 and December 2015.

Table 3: Data base

Data base	Papers
ACM Digital Library	936
Elsevier (Science Direct)	1142
IEEE Xplore	421
Scopus	831
Springer	495
Total	**3.825**

From this gross articles bank began the refining stage where repeated articles were separated when they appeared in different bases, this time database went from 3.825 articles to 1.426 articles.
This was followed by reading the article title for items that do not have grating to the subject of research were removed. These removed articles may have been incorporated since the keyword research was also done in the abstracts, which could contain the keyword without having a direct relationship with the searched subject. At this time database was reduced to 223 articles, representing approximately 15.64% of the total gross database after the withdrawal of the repeated articles.
Following, the moment of verification of scientific relevance of the article was performed, this relevance was made through the Google Scholar search tool, raising the amount of citations that each article has. ProKnow-C method suggests as a cut-off point for the permanence of the article in the portfolio, that it is in the representatively of 85% of the citations, here we selected articles with up to seventeen citations. This filtering moment resulted in 41 articles with confirmed scientific relevance and 82 articles with scientific relevance still to be confirmed.
In the next step, for articles with confirmed scientific relevance, the abstract was read and to confirm alignment with the research objective, thus discarding 30 articles remaining 11 articles.
For items with no proven scientific relevance an analysis was carried out in two stages: (1) reading the abstract of articles with up to two years of publication, as these articles usually have few or no quotations; (2) for other articles, the reading of abstracts which had authors in the group of articles with scientific relevance confirmed. These two steps resulted in 05 articles in relation to the objective of this work. The current bibliographic portfolio then contained 16 articles.
The next moment was the filtering regarding availability of the full articles for reading, the ones which could be accesses directly in search databases or journals in which this articles were published and likely to be accessed from an institutional library. Here the database lost 03 articles, getting a number of 13.
After a complete reading of the available texts, alignment with the objective of this work was carried out and it was verified that 02 could be removed from the bibliographic portfolio, resulting in 11 articles which integrate the bibliographic portfolio referring to the theme of this work. These articles are presented in table 4 highlighting the main contribution to the objective of this article:

Table 4: Bibliographic portfolio obtained

Author	Quotations	Contribuition
Ebner and Holzinger, 2007.	487	To know to what extent the serious games have power to contribute to the student's learning. Use of questionnaire as a tool to evaluate the game and as a feedback tool for the implementation of new versions.
Kebritchi and Hirumi, 2008.	300	To identify teaching strategies and pedagogical bases used to design educational games.
AbouRizk and Hague, 2009.	144	It presents a model for use in computer that simulates all the stages of the process of a construction, from the stage of design until the construction. Plans, analyzes and controls all stages of construction.
Whitton, 2007.	97	It identifies the motivations of the students to play and learn from serius games. Evaluation carried out from the application of questionnaires. Some serious games can provide pedagogical benefits beyond learning.
Mayer et al., 2014	86	It brings a review of the literature on the evaluation of serius games highlighting the following questions: what are the requirements and principles of the design; the extent to which these games contribute to learning; what factors contribute to or determine this learning; to what extent this learning can be transferred to the real world.
Deshpande and Huang, 2008.	72	It presents a review of the state of the art in the area of games for engineering education. Proper application maximizes the transfer of knowledge to professional practice.
Wall and Ahmed, 2008	65	It shows that simulation games can play a very effective role in providing lifelong learning opportunities geared to the construction industry.
Marfisi-Schottman, George and Tarpin-Bernard, 2010.	61	It details the process of designing a serius game and enumerating the various actors who collaborate in the process, such as: project manager, cognitive specialist, area experts, storyboard writer, artistic director, pedagogic specialist, programmers.
Kosmadoudi et. al., 2013	49	It reviews techniques and game mechanisms that can be developed from CAD systems in engineering, in particular to maintain cognitive engagement. It highlights CAD user graphical user interfaces and how they can be improved.
Philpot et. al. 2005.	47	Serius games developed for mechanical engineering, but with perfect application to civil engineering. Use of questionnaire as an instrument to evaluate the moment of learning.
Juang, Hung	17	The result of the simulation shows

and Kang, 2011.		the advantages of developing a construction simulation based on a game mechanism. It emphasizes that the quality of the simulation graphics should provide a high degree of reality and there should be interactivity between user and computer in the performance of the game.

The formatting of the chart above from the number of citations, obtained on scholar.google.com in December 2015, is based on the premise that authors cite work that is important for development of their research so the most cited documents will have a greater influence on an area of knowledge than the less quoted ones (Culnan1987, Tahai1999).

4. BIBLIOMETRIC ANALYSIS

In order to achieve at the presented result, a bibliometric analysis was performed in the 123 featured articles, a quantity of these with confirmed scientific relevance and others with relevance to be confirmed.
The analysis was conducted through the searching of the most relevant keywords; the authors, universities and countries COM DESTAQUE in the theme area of this work; the most relevant articles. With the portfolio presented in chart 2 it was possible to verify the evolution of the research in the area to see the current contribution to construction of games for pedagogical purposes.

4.1. Keywords

In the screening phase of the scientific articles, which resulted in the 123 articles with scientific relevance confirmed and yet to be confirmed, an analysis of key words was carried out with greater frequency and of these the first five, presented in the table 5.

Table 5: Most mentioned keywords

Keyword	% ofthesample
Game-based learning	21,4
Simulations	15,6
Educational computer games	9,5
Serious Games	8,3

These key words represent approximately 54.8% of the references, the surprise was due to the word serious games only in the fourth position among the most cited key words.

4.2. Geographic location

For the geographical location of the scientific articles, Europe and North America are highlighted with more than 75% of the world production in this theme, as seen in the graph shown in figure 2 below.

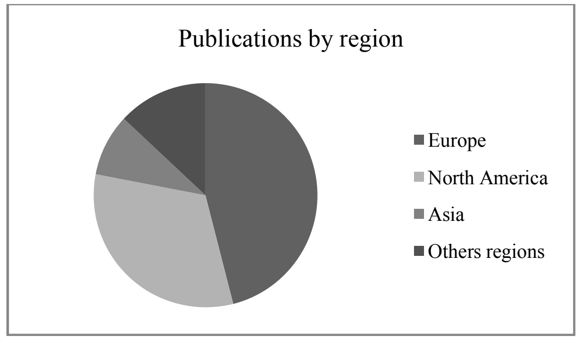


Figure 2: Geographical location of the articles raised.

During the construction of the portfolio, in the identification of articles, some universities appeared repeatedly among those that produced pedagogical tools for engineering teaching, among them we have: University of Alberta, University of Central Florida and University of Calabria. It is believed that this fact occurred due to the proximity of the terms "serious games" and "simulation" in the construction of these pedagogical tools.
Among the authors, the distribution in the portfolio did not highlight any, but each of these authors had a prominent importance in the "contributions" column presented in chart 2.

4.3. Featured journals

Among the journals that stood out (Figure 3); the Computers & Education e British Journal of educational Technology, with a contribution of approximately 54% of the references presented, among the 123 articles with scientific prominence.

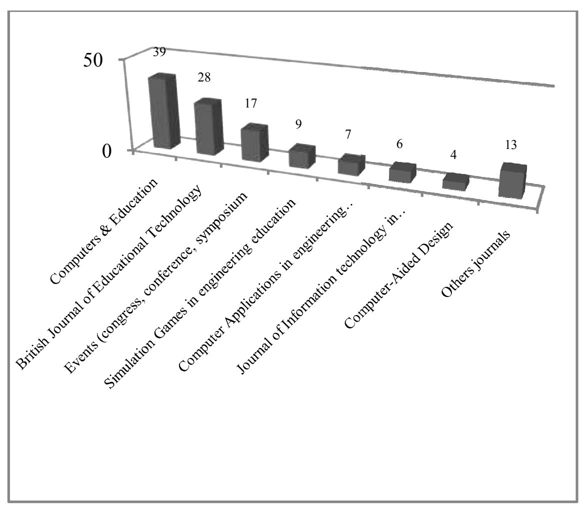


Figure 3: Featured scientific journals

The articles published in events (congress, conference, symposium) also deserve prominence, since they contributed with approximately 7% of the articles.
Among the events (congress, conference, symposium) that include articles in this bibliometric analysis, we highlight: European Conference on Games Based Learning, Winter Simulation Conference e International

Multidisciplinary Modeling & Simulation Multiconference.

It is also worth highlighting the journals: Simulation Games in engineering education, Computer Applications in Engineering Education, Journal of Information Technology in Construction e Computer-Aided Design.

4.4. Evolution of research

Over the years analyzed it is possible to perceive a passage from simple application of the games, almost always adapted games, for application of games built for specific purposes of the subject to be taught.

In the analysis of the bibliographic portfolio one can arrive at the construction of the flow chart below, from contributions identified in the table 4.

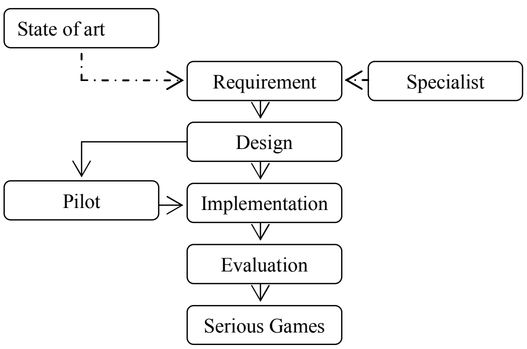


Figure 4: Construction of the pedagogical tool.

The flowchart presented, which summarizes the logical sequence of the portfolio is in line with the one presented by Preece, Rogers and Sharp 2011, which starts from the establishment of requirements arriving at the presentation of a framework for evaluation.

Literature review comes in to provide the state of the art and make contributions when defining the requirements.

The expert consultation to help define the requirements from some method of consultation, such as the Delphi Delphi method[1].

The next steps are the construction of the pedagogical tool (Serious Games), following the construction of the design, passing through the pilot/implementation and arriving at the evaluation process, in this review, based on the application of questionnaires.

5. FINAL CONSIDERATIONS

It was possible to verify in this bibliographic review that the application of serious games to the formation of the engineer is something promising and viable, however, there is still a long way to be covered for the inclusion of serious games as a pedagogical tool.

In the reading of the articles it was possible to perceive, with regard to construction or choice of games, a need already verified by Squire et. Al. 2005, where students are more involved with games that have interactivity similar to their entertainment games.

[1] O método Delphi foi desenvolvido pela *Rand Corporation* para estudar o impacto da tecnologia sobre a guerra (Dalkey and Helmer 1962).

REFERENCES

AbouRizk, S. M., Hague, S., 2009. An overview of the COSYE environment for construction simulation, In Proceedings of the 2009 Winter Simulation Conference (WSC), pp. 2624 – 2634. Dec. 13-16, Austin (Austin – Texas, USA).

AbouRizk, S. M., Sawhney, A., 1994. Simulation and Gaming in Construction Engineering Education, ASEE/C2E2/C2EI Conference, Edmonton (Edmonton – Alberta, CANADÁ).

Arnab, S., Brown, K., Clarke, S., Dunwell, I., Lim, T., Suttie, S. L., Louchart, S., Hendrix, M., Freitas, S. 2013.The development approach of a pedagogically-driven serious game to support Relationship and Sex Education (RSE) within a classroom setting. Computers & Education, 69 (November), 15 – 30.

Chittaro, L., Sioni, R., 2015. Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack. Computers in Human Behavior, 50, 508 – 519.

Corti, K., 2006. Games-based Learning; a serious business application. PIXELearning Limited. Available from:www.pixelearning.com/docs/games_basedlearning_pixelearning.pdf [accessed 15 July 2005].

Culnan, M. J., 1987. Mapping the intellectual structure of MIS, 1980 – 1985: a co-citation analysis. Management of Information System Quartely, 11 (3), 341 – 353.

Dalkey, N. C.; Helmer, O., 1963.An experimental application of the Delphi method to the use of experts. Management Science, 9 (3), 458 – 467.

Deshpande, A. A., Huang, S. H., 2011.Simulation games in engineering education: A state-of-the-art review. Computer Applications in Engineering Education, 19 (3), 399 - 410.

Diodato, V. P., 2012. Dictionary of bibliometrics.Taylor & Francis Group. New York, New York state.204 p.

Ebner, M., Holzinger, A., 2007.Successful implementation of user-centered game based learning in higher education: An example from civil engineering. Computers & Education, 49 (3), 873 – 890.

El-Beheiry, M., McCreery, G., Schlachta, C. M., 2017. A serious game skills competition increases voluntary usage and proficiency of a virtual reality laparoscopic simulator during first-year surgical residents' simulation curriculum. Surgical Endoscopy and Other Interventional Techniques, 31 (4), 1643-1650.

Ensslin, L., Ensslin, S. R., Lacerda, R. T. O., Tasca, J. E., 2010.ProKnow-C, Knowledge Development Process – Construtivist. Technical file with registration patent pending with INPI, Brazil.

Feisel, L. D., Rosa, A. J., 2005. The Role of the Laboratory in Undergraduate Engineering

Education. Journal of Engineering Education, 1, 121 – 131.
Freire, A., Cavalheiro, C., Ely, D. M., Jungles, A. E., 2016. The competencies acquisition with simulation application in the course 'construction planning and controlling'. International Journal of Simulation and Process Modelling, 11 (6), 443 – 452.
Herbsman, Z., 1986. Project management training using microcomputers, Journal of Management in Engineering, 2 (3), 165 - 176.
Juang, J. R., Hung, W. H., Kang, S. C., 2011. Using game engines for physics-based simulations - A forklift.Journal of Information Technology in Construction, 16, 3 - 22.
Kebritchi, M.; Hirumi, A., 2008.Examining the pedagogical foundations of modern educational computer games. Computers & Education, 51 (4), 1729 – 1743.
Khenissi, M. A., Essalmi, F., Jemni, M., 2015. Comparison Between Serious Games and Learning Version of Existing Games. Procedia - Social and Behavioral Sciences, 191, 487– 494.
Kosmadoudi, Z., Lim, T., Ritchie, J., Louchart, S., Liu, Y., Sung, R., 2013.Engineering design using game-enhanced CAD: The potential to augment the user experience with game elements. Computer-Aided Design, 45 (3), 777 - 795.
Latorre, J.I., Jiménez, E., 2012. Simulation for Education in Business decision-Making. Magazine of the Society for Modeling & Simulation International (SCS M&S Magazine), 2, 59–65.
Longo, F., Chiurco, A., Musmanno, R., Nicoletti, L., 2015. Operative and procedural cooperative training in marine ports. Journal of Computational Science, 10 (September 01), 97-107.
Magee, M., 2006. State of the Field Review: Simulation in Education. Internal report. Alberta Online Learning Consortium. Calgary AB.
Marfisi-Schottman, I.,;Sébastien, G.,Tarpin-Bernad, F. 2010. Tools and methods for efficiently designing serious games.In: 4th European Conference on Games-Based Learning: ECGBL 2009. Oct 21-22, Copenhagen (Copenhgen, DENMARK).
Mayer, I., Bekebrede, G., Harteveld, C.,Warmelink, H., Zhou, Q., Ruijven, T., Lo, J., Kortmann, R., Wenzler, I., 2014.The research and evaluation of serious games: Toward a comprehensive methodology. British Journal of Educational Technology, 45 (3), 502 – 527.
Michael, D., Chen, S., 2006. Serious games: Games that educate, train, and information. Boston, MA.: Thomson Course Technology.
Noronha, D. P., Ferreira, S. M. S. P., 2000. Revisão da literature. In: B. S. Campello, B. V. Cendón, J. M. Kremer, Fontes de informação para Pesquisadores e Profissionais, Federal Universityof Minas Gerais, 2000, 197 – 198.
Pao, M. L., 1989. Concepts of information retrieval. Englewood.Colorado: Libraries Unlimited, Inc., 1989. 285 p.
Philpot, T. A., Hall, R. H., Hubing, N., Flori, R., 2005.Using games to teach statics calculation procedures: Application and assessment. Computer Applications in Engineering Education, 13 (3), 222 - 232.
Preece, J.; Rogers, Y.; Sharp, H., 2011. Interaction Design: Beyond Human-Computer Interaction. Chichester: John Wiley & Sons Ltd.
Pritchard, A., 1969. Statistical bibliography or bibliometrics?. Journal of documentation, 25, (4), 348 – 349.
Raybourn, E. M., 2007. Applying simulation experience design methods to creating serious game-based adaptive training systems. Interacting with Computers, 19 (2), 206-214.
Rovner, S. L., 2006. Video game aims to engage students. Chemical & Engineering News, 84 (15), 76-77.
Schumann, P. L., Anderson, P. H., Scott, T. W., 1997. Using computer based simulation exercises to teach business ethics. Teaching business ethics, 1 (2), 163 – 181.
Scott, D., Mawdesley, M., Al-Jibouri, S., 2004.The use and efficacy of a simulation model for teaching project control in construction. In First International Conference on world of construction project management, pp. 563 - 568, Toronto (Toronto, CANADÁ).
Sherif, A., Mekkawi, H., 2010.Excavation Game: Computer-Aided-Learning Toolfor Teaching Construction Engineering Decision Making. Journal of Professional Issues in Engineering Education and Practice, 136 (4), 188-196.
Sorense, B. H., Meyer, B., 2007.Serious games in language learning and teaching-a theoretical perspective. In Proceedings of the 2007 Digital Games research Association Conference, pp. 559-566.Sep 24-28, Tokyo (Tokyo, JAPAN).
Squire, K. & Jenkins, H., 2003. Harnessing the power of games in education. Insight, 3(1), 5- 33.
Squire, K., Giovanetto, L., Devane, B., 2005. From users to designers: Building a self-organizing game-based learning environment. TechTrends: Linking Research &Practice to Improve Learning, 49(5), 33-44.
Susi, T., Johannesson, M., Backlund, P., 2007.Serious Games: An Overview. IKI Technical Reports, Skövde: Institutionen för kommunikation och information , 28 p.
Tahai, A., Meyer, M. J., 1999. A reveled preference study of management journals' direct influences. Strategic Management Journal, 20 (3), 279 – 296.
Walla, J., Ahmed, V., 2008.Use of a simulation game in delivering blended lifelong learning in the construction industry – Opportunities and Challenges. Computers & Education, 50 (4), 1383 - 1393.

Whitton, N., 2007. Motivation and computer game based learning. In Proceedings ascilite Singapore 2007, pp. 1063 – 1067.Dec 2 – 5, Singapore (Singapore, REPUBLIC OF SINGAPORE).

AUTHORS BIOGRAPHY

Ailton Soares Freire – Professor at *Instituto de Educação, Ciência e Tecnologia do Piauí* in Brazil, Master in Civil Construction by *Universidade Federal de São Carlos* in Brazil and doctor in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil.

Daniela Matchulat Ely – Professor at *Centro Federal de Educação Tecnologica de Minas Gerais* in Brazil, Master in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil and doctor in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil.

Leiliane Santana Souza – Professor at *Instituto de Educação, Ciência e Tecnologia da Bahia* in Brazil, Master in Civil Engineering by *Universidade Estadual de Goias* in Brazil and doctoral student in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil.

Antônio Edésio Jungles – Associate professor at *Universidade Federal de Santa Catarina*, with doctorade in Production Engineering by *Universidade Federal de Santa Catarina* and internship (sandwich) at University of Waterloo in Canada. General coordinator at CEPED/UFSC (University Center for Studies and Research on Disasters), and GestCon (Construction Management Group) in Brazil.

Marcos Baptista Lopez Dalmau – Associate professor at *Universidade Federal de Santa Catarina*, with doctorade in Production Engineering by *Universidade Federal de Santa Catarina*. Technical director at CEPED/UFSC (University Center for Studies and Research on Disasters) in Brazil.

A METHOD FOR OBTAINING THE CREDIBILITY OF A SIMULATION MODEL

Ke Fang[a], Ming Yang[b], Kaibin Zhao[c]

[a],[b],[c]Control & Simulation Center, School of Astronautics, Harbin Institute of Technology, Harbin, China

[a]hitsim@163.com, [b]myang@hit.edu.cn, [c]kaibin.zhao.HIT@hotmail.com

ABSTRACT
Model validation is one of the key problems of simulation systems V&V (Verification & Validation). How to obtain the credibility of a simulation model is still the core issue prior to its application. Based on the introduction of model credibility, a method of workflow for obtaining the credibility of a simulation model is brought forward, the essentials including factor space establishment, similarity analysis, result transformation and defect tracing are explained, then a case study which provides a walk through application of the method is presented. The method corrects and clarifies some misunderstanding in model validation, and provides a practical way to obtain the model credibility.

Keywords: model validation, V&V, credibility, factor space, similarity analysis, result transformation

1. INTRODUCTION

A simulation model uses mathematical modeling and coding techniques to build a run-able program to simulate the object in the real world. There is a quantitative similarity between the simulation model and its origin in application domain named "credibility", to measure at what level the model behaves as the origin. Research and practice have been being conducted to compute the credibility of a simulation model, which is the goal of model V&V (Verification & Validation) (Sargent 2015). However it is not an easy job.

Many recent studies focus on the problem of model validation, and propose constructive walkthroughs to conduct it. Roy addressed the uncertainty quantification problem in model validation, and proposed a comprehensive framework for model V&V (Roy 2011). You summarized several statistics similarity analysis methods and proposed a decision process in quantitative model validation (You 2012). Eek established a dependency graph of a simulator at Saab Aeronautics, and developed a quantified table for credibility evaluation (Eek 2015). Kutluay focused on vehicle dynamics simulation models, and gave a literature survey on validation approaches (Kutluay 2014). Song used Colored Petri Nets to rebuild the model of train-to-train distance measurement system, and established a multi-layer factor space to reveal the model validation aspects (Song 2017).

Since every simulation model has different modeling strategies, it is hard to propose a straight mathematical expression of credibility in general. Researchers and engineers often use decision making method to synthesize the final result, after they obtain the influence level of all attributes related to the credibility. See the expression below.

$$F = \{<N, V>; <L, A>\}$$

$$C = f(v_1, v_2, ..., v_k) \qquad (1)$$

In Eq.(1), F is a factor space (Zhou, Sun, and Li 2016) created by the model decomposition, where $N = (n_1, n_2, ..., n_k\}$ is the node set, which contains all the credibility influencing factors to the model; $V = \{v_1, v_2, ..., v_k\}$ is the value set, which contains all partial credibility of each factor and maps to N; $L = \{l_1, l_2, ..., l_k\}$ is the link set, which contains all relationship between each two factors; and A is the attribute set, which contains all attributes on each link such as weight etc. and maps to L. C is the credibility, and it is a synthesized result of $v_1 \sim v_k$ handled by f.

Obviously, how to establish factor space F, how to get partial credibility v_k, and how to select a synthesis function f are the key jobs. Especially, f will influence the final result significantly. An weighted average algorithm and variations are frequently used as f. However, the relations $l_1 \sim l_k$ between factors are not bound to be linear, and the synthesis way of partial credibility to the final one should be determined by the computational process of the model but not a general expression.

There is a misunderstanding that factor space F is for providing a pathway to synthesize the final credibility, and it should be built through structural decomposition. This is not true. Model credibility is mainly affected by its computational process but not the structure, so the usage of the factor space in model V&V has to be corrected.

The main idea of this paper is to bring forward a practical method for obtaining the credibility of a simulation model, which may change the way of using factor space and similarity analysis etc. to get model credibility. A case study of a flying vehicle's mass center motion model is also presented to describe a walk through application of the method.

2. THE WORKFLOW OF THE METHOD

The fact is, the direct outputs of the simulation model have to be pointed out before acquiring the credibility, and if the credibility traceability is required, the intermediate outputs of the simulation model also have to be demonstrated. This actually forms the factor space of the model V&V.

By going through the factor space, each node of model output can be analyzed of the similarity between model and its origin by simulation and observed data, and then the result is transformed into the partial credibility.

If the credibility of direct outputs is acceptable, the work is done. If not, a defect tracing should be conducted to analyze lower level nodes of intermediate outputs. When all the analysis is done a final conclusion is drawn to show the grand credibility and nodes of defect.

The figure below shows the workflow of the method for obtaining the credibility of a simulation model, where sold line represents execution flow, and dot line represents data flow.

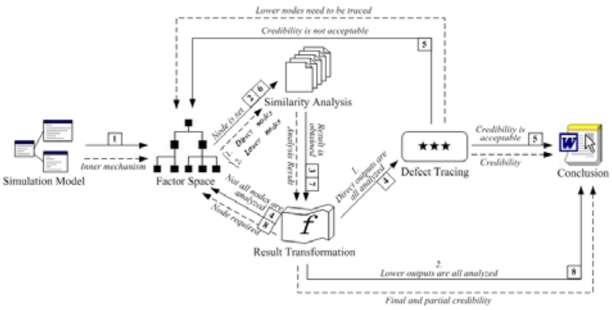


Figure 1: The Workflow for Obtaining Model Credibility

The workflow should be conducted as the following procedures:

1. Establish the factor space for evaluating the model credibility by reproducing the model computational process, to point out all factors which influence the credibility.
2. Use time domain analysis such as TIC (Theil Inequality Coefficient) (Andrei, Gary and Tryphon 2014), GRA (Gray Relational Analysis) (Wei and Li 1997) etc., frequency domain analysis such as Welch's periodogram analysis (Jiang and Mahadevan 2011), maximum entropy spectrum analysis (Mullins, Ling, Mahadevan and Sun 2015) etc., and statistics analysis such as parameter estimation, hypothesis testing etc. to perform the similarity analysis on higher nodes of direct outputs.
3. Use appropriate transformation formula to transform the similarity analysis result to credibility description.
4. If direct outputs are all analyzed, go to next step. If not, go back to Step 2 and continue with other direct outputs.
5. If the credibility is acceptable, draw the conclusion and the work is done. If not, perform defect tracing and go back to the factor space.
6. Similar to Step 2 but the nodes analyzed are lower nodes of intermediate outputs.
7. Similar to Step 3 but the nodes handled are lower nodes of intermediate outputs.
8. If traced lower nodes are all analyzed, gather partial and grand credibility to locate the defect nodes and draw the conclusion. If not, go back to Step 6 to continue with other nodes.

3. THE ESSENTIALS OF THE METHOD

3.1. Factor Space Establishment

The conventional way of building a factor space is the tree view, which is suitable to express the decomposition of a simulation model. However, a tree view is unable to reveal the computational process of the model, which is important to the grand credibility.

Actually, the tree view method only uses partial credibility on leaf nodes and applies weighted average algorithm etc. to synthesize the grand credibility. It does not need similarity analysis among branch nodes, which is problematic. The computing of credibility is not bound to be linear, and the direct outputs similarity should not be synthesized through intermediate outputs similarity. To resolve this, MADN (Multi-Attribute Decision Network) (Fang, Ma and Yang 2012) is used here.

See the example of a flying vehicle's motion model below. When the branches reach direct outputs, the factor space cannot be further built by structural decomposition but has to reproduce the model's computational process.

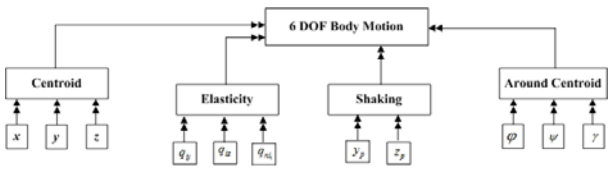


Figure 2: Part of A 6-DOF Body Motion Model's Factor Space

Take the vehicle's position x, y, z as an example. Further expand the factor space as below. The meaning of variables in Figure 3 is presented in Table 1 in the following Section 4.1. The process is conducted by revealing the computation process of the model but not the structural components. All the expanded nodes are credibility influencing factors n_k in Eq. (1), and the links between them are factors relationship l_k .

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

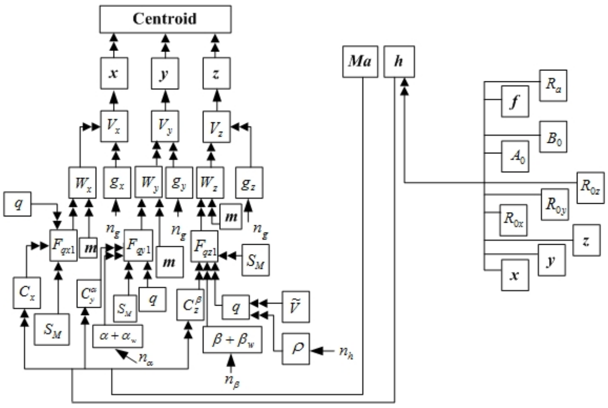


Figure 3: Expanded Factor Space of A 6-DOF Body Motion Model

A factor space establishment should be conducted as the following procedures:

1. Decompose the simulation model to the nodes which cannot be divided due to the model structure.
2. Continue to expand the factor space by computational process of the model. Position the outputs at upper level, and the inputs at lower level.
3. Terminate the expansion when the nodes which represent the model's initial inputs are revealed.
4. Use compose, derive, surpass, substitute, contradict, inherit etc. types of link (Fang, Yang and Zhang 2011) to connect the nodes in the factor space.
5. Determine the parameters for the factor space, such as weights for composition links, thresholds for surpassing links, conditions for uncertain links etc.

3.2. Similarity Analysis

A model's credibility has to be analyzed by comparing the simulation outputs with the real ones. In factor space, it has to be performed similarity analysis on the nodes, such as the vehicle's position x, y, z in Figure 2. According to the outputs feature, various analysis methods are correspondingly used.

If the output is time irrelevant, such as miss distance, hit probability etc., statistics analysis like *u* testing is often used. If the output is a series of sequential data which is gradually changing along with the elapsing time, time domain analysis like GRA is often used. If the output is changing frequently, frequency domain analysis like maximum entropy spectrum analysis is often used.

For example, the formula below shows the grey relational degree of a single output which comes from the simulation and the real world. (Wei and Li 1997)

$$\gamma = \frac{1}{n}\sum_{k=1}^{n}\frac{\min\limits_{k=1\sim n}|s_k - o_k| + \rho\max\limits_{k=1\sim n}|s_k - o_k|}{|s_k - o_k| + \rho\max\limits_{k=1\sim n}|s_k - o_k|} \quad (2)$$

where s_k is the simulation data series, and o_k is the observed data series. ρ is the resolution coefficient. $\min\limits_{k=1\sim n}|s_k - o_k|$ selects the minimum value in the series result set $|s_k - o_k|$, $k = 1 \sim n$, which is a single value, and vice versa for $\max\limits_{k=1\sim n}|s_k - o_k|$. The conclusion γ is named as grey relational degree, which indicates the similarity of the two series of data. The higher γ comes, the higher similarity is obtained.

Other methods have specific computational ways of acquiring the similarity between simulation and observed data respectively. Some have a Boolean result of "acceptable" or "unacceptable" like hypothesis testing and spectrum analysis. Some have a quantitative result like TIC. However, these similarity analysis results have to be transformed into credibility description.

3.3. Result Transformation

The result formation is various due to the similarity analysis method. It has different ranges, monotonies etc. In order to judge and synthesize the similarity analysis results, they have to be transformed into a unanimous credibility description, which has the range of [0, 1] and is monotonically increasing with the credibility.

The transformation should be realized by the nature of similarity analysis method itself. It has to adapt to the physical meaning of the method. For example, the *T* testing possesses a Boolean analysis result, which bases on an original and alternative hypothesis below:

$$H_0 : \mu_1 = \mu_2, H_2 : \mu_1 \neq \mu_2 \quad (3)$$

There is a β which represents the pseudo probability when the original hypothesis is accepted. Thus there comes a conclusion that when the alternative hypothesis is accepted, the transformed analysis result into credibility should be 0, and when the original hypothesis is accepted, the result should be 1-β.

If the significance level is α, according to the *T* distribution and the probability density function, 1-β can be calculated as below: (Zhang 2011)

$$1\text{-}\beta = 2 - T[T^{-1}(1-\frac{\alpha}{2}) - \frac{\delta}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}] - T[T^{-1}(1-\frac{\alpha}{2}) + \frac{\delta}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}]$$

$$S_w = \sqrt{\frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}} \quad (4)$$

where δ is the maximum tolerance of deviation, $\mu_1 = \mu_2 + \delta$; n and m are sample numbers of

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

simulation data and observed data; s_1 and s_2 are the variances of two samples.
Other methods have their own transformation formula, but meet the range of [0, 1] and the increased monotony. Some is simple, like 1-ρ for TIC. Some is relatively complex, like the transformation of F testing result for maximum entropy spectrum analysis (Zhang 2011).

3.4. Defect Tracing

If a node of higher level results in unacceptable credibility, we want to know what induces that "bad" credibility. In this case, a further and deeper validation is required, which should be conducted as the following procedures:

1. Perform similarity analysis on nodes in the factor space downwards, and locate those unacceptable nodes at upper level.
2. Perform further similarity analysis till the leaf nodes are reached.
3. On analysis route, those unacceptable nodes are defect points of the model, and the unacceptable leaf nodes are the origins of the defect.
4. If all sibling nodes of the upper unacceptable node are analyzed but no more defect points are found, that means more strict acceptability criteria (Oldrich and Andreas 2014) is required.

4. CASE STUDY

Use a flying vehicle's mass center motion model as an example, to present a walkthrough of the method. The model computes the position of the flying vehicle via aerodynamics coefficients and geophysics constants.

4.1. Factor Space Establishment

According to the computational process of the model, place final outputs at upper levels, intermediate outputs and inputs at lower levels, and build the factor space of model validation. Figure 3 shows the factor space. The table below shows the meaning of the variables:

Table 1: The Meaning of Variables

Variable	Meaning	Variable	Meaning
x,y,z	Position	ρ	Air density
V_x,V_y,V_z	Velocity	$\alpha+\alpha_w$	Velocity transition
W_x,W_y,W_z	Acceleration	$\beta+\beta_w$	Velocity transition
g_x,g_y,g_z	Acceleration of gravity	M_a	Mach
$F_{qx_1},F_{qy_1},F_{qz_1}$	Aerodynamic force	h	Height
m	Mass	f	Wind velocity
$C_x,C_y^{\alpha},C_z^{\beta}$	Aerodynamic coefficient	R_a	Equator radius
q	Dynamic pressure	A_0	Launch direction
$\widetilde{V}$	Relative flow velocity	B_0	Launch spot latitude
S_M	Sectional area	R_{0x},R_{0y} R_{0z}	Launch spot core radius

4.2. Similarity Analysis

First perform similarity analysis on higher nodes, that is, the flying vehicle's position x, y, z . The tables below show the simulated and observed data of the trajectory.

Table 2: Simulation & Observed Data of Position x

time(s)	simulation data *x_sim*(m)	observed data *x_ref*(m)
0.005	6022942.548	6022942.549
0.010	6023002.710	6023002.713
0.015	6023062.829	6023062.834
……	……	……
4.515	6064014.043	6069515.665
4.520	6064019.502	6069531.364
4.525	6064024.717	6069546.836
……	……	……
6.135	6055118.998	6058912.748
6.140	6055118.998	6058802.077
6.145	6055118.998	6058691.265

Table 3: Simulation & Observed Data of Position y

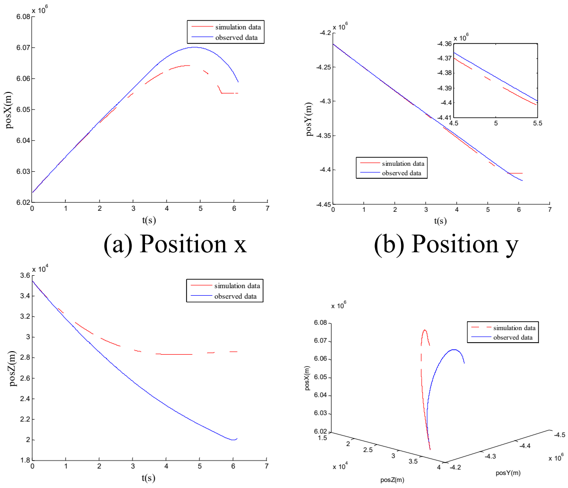
time(s)	simulation data *y_sim*(m)	observed data *y_ref*(m)
0.005	-4216050.798	-4216050.798
0.010	-4216224.172	-4216224.171
0.015	-4216397.545	-4216397.543
……	……	……
4.515	-4370194.423	-4366433.045
4.520	-4370366.670	-4366598.879
4.525	-4370538.900	-4366764.726
……	……	……
6.135	-4404860.781	-4415681.560
6.140	-4404860.781	-4415755.711
6.145	-4404860.781	-4415828.882

Table 4: Simulation & Observed Data of Position z

time(s)	simulation data *z_sim*(m)	observed data *z_ref*(m)
0.005	35471.429	35471.429
0.010	35452.188	35452.187
0.015	35432.944	35432.943
……	……	……
4.515	28272.912	22350.570
4.520	28273.351	22341.140
4.525	28273.801	22331.720

……	……	……
6.135	28528.336	20086.733
6.140	28528.336	20098.080
6.145	28528.336	20109.940

According to the simulation and observed data, draw the flying vehicle's trajectory as below:



(a) Position x (b) Position y

(c) Position z (d) 3D trajectory

Figure 4: Trajectory of the Flying Vehicle

Use GRA method of Eq.(2) to perform the similarity analysis of trajectory. Set $\rho = 0.5$, and we get $\gamma_x = 0.727,\ \gamma_y = 0.789,\ \gamma_z = 0.623$.

4.3. Result Transformation

When Eq. (2) meets $\gamma \geq \gamma_{th}$, it can be considered as the simulation and observed data have acceptable similarity. Set C_{th} as the acceptability criteria, the formula below can be used to transformed the GRA result into credibility:

$$C(\gamma) = \begin{cases} \frac{1-C_{th}}{1-\gamma_{th}}(\gamma - \gamma_{th}) + C_{th}, & \gamma \in [\gamma_{th}, 1] \\ \frac{C_{th}\gamma}{\gamma_{th}}, & \gamma \in [0, \gamma_{th}] \end{cases} \quad (5)$$

Select $\gamma_{th} = 0.8, C_{th} = 0.7$, the figure below shows the transformation:

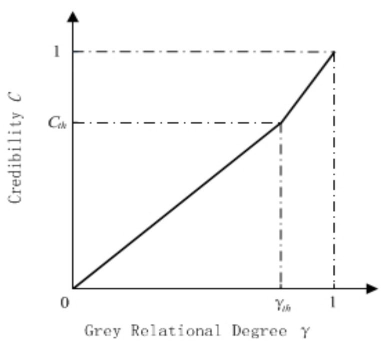


Figure 5: Transformation from γ to C

Use Eq.(5) to perform a transformation with the results we get in Chapter 4.2, the credibility of *x*, *y*, *z* can be derived as $C_x = 0.636,\ C_y = 0.691,\ C_z = 0.545$. It indicates the credibility is not acceptable, and the model validation needs a defect tracing.

4.4. Defect Tracing

Make further validation downwards along the factor space routes shown in Figure 3. Here to explain the process in general, we only analyze node *h* and its sub-networks due to the limited content of the paper.
Based on the principle of tracing downwards, the node of height *h* is analyzed first. Its simulation and observed data are shown as the following table.

Table 5: Simulation and Observed Data of Height *h*

time(s)	simulation data *h_sim(m)*	observed data *h_ref(m)*
0.005	10000.0778	10000.078
0.010	9999.489	9999.490
0.015	9998.888	9998.890
……	……	……
4.505	6340.700	8424.085
4.510	6324.899	8412.098
4.515	6309.027	8400.041
……	……	……
6.135	21.894	110.314
6.140	21.894	68.455
6.145	21.894	26.649

According to the simulation and observed data, draw the figure of height *h* as below:

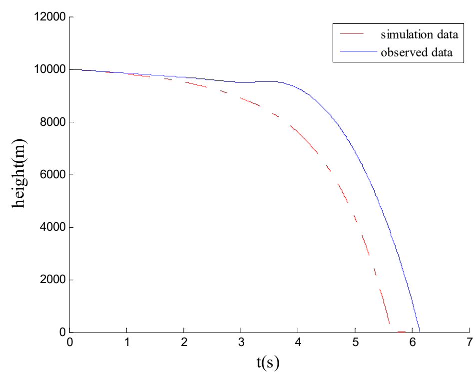


Figure 6: Height of the Flying Vehicle

Use GRA method of Eq.(2) to perform the similarity analysis of height *h*. Set $\rho = 0.5$, and we get $\gamma_h = 0.7161$. Use Eq.(5) to perform a result transformation, and we get $C_h = 0.6266$, which shows it is unacceptable, and needs further defect tracing.
The sibling nodes of height *h* include equator radius R_a , launch direction A_0 , launch spot latitude B_0 and

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

launch spot core radius vector R_{0x}, R_{0y}, R_{0z}, which are equal to each other respectively between simulation and observed data. Meanwhile, the initial values of the iteration variable x, y, z are also equal to each other respectively between simulation and observed data. Then we focus on the last sibling node wind velocity *f*. The table below shows the simulation and observed data of *f*.

Table 6: Simulation and Observed Data of *f*

Time(s)	simulation data *f_sim*(m/s)	observed data *f_ref*(m/s)
0.005	69.230	-69.230
0.010	69.227	-69.227
0.015	69.223	-69.223
……	……	……
4.505	47.560	-59.899
4.510	47.467	-59.828
4.515	47.373	-59.757
……	……	……
5.515	16.608	-36.353
5.520	16.373	-36.186
5.525	16.137	-36.018

According to the simulation and observed data, draw the figure of wind velocity *f* as below:

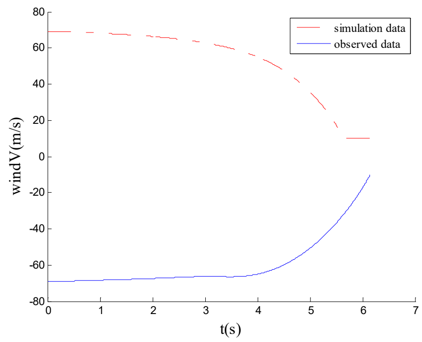


Figure 7: Wind Velocity of the Model

Use GRA method of Eq.(2) to perform the similarity analysis of wind velocity *f*. Set $\rho = 0.5$, and we get $\gamma_f = 0.5150$. Use Eq.(5) to perform a result transformation, and we get $C_f = 0.4506$, which shows the node of wind velocity *f* has unacceptable credibility.

4.5. Result Analysis

First, since x, y, z, the direct outputs of the model get unacceptable credibility $C_x = 0.6363 < 0.7$, $C_y = 0.6906 < 0.7$, and $C_z = 0.5447 < 0.7$, the model failed to pass the validation, which indicates the model is lack of credibility and has to be revised prior to its application.
Second, when analyze further along the factor space routes, the vehicle's height h gets unacceptable credibility $C_h = 0.6266 < 0.7$, and wind velocity f gets unacceptable credibility $C_f = 0.4506 < 0.7$ at leaf level. Other constants and variables are proved to be credible beneath the node of vehicle's height *h*. (The parts other than height *h* are omitted in the paper)
According to the procedures of defect tracing, it can be concluded that wind velocity f is the main reason which causes the direct outputs x, y, z has unacceptable credibility. In another word, as long as the sub-model of wind velocity is corrected, the model of mass center motion will have a good chance to be credible.

5. CONCLUSION

There is some misunderstanding in model validation nowadays, which hinders us to get an objective credibility of the model. Factor space has to be built by structural and computational decomposition combined, especially with the latter one on emphasis of obtaining objective results. Meanwhile, similarity analysis result has to be transformed into credibility description by the nature of similarity analysis method itself. Finally, the computational decomposition is mainly used to fulfill the potential defect tracing of the model, but not to make credibility synthesis.
Model validation is one of the key problems of simulation systems V&V. The method explained here provides a practical way to obtain the credibility of a simulation model, and corrects and clarifies some misunderstanding in model validation. Since the method is insensitive to the model which is being validated, it can be applied to all kinds of simulation models, no matter the model is built by continuous-time nature or discrete event. Further study should be conduced on how to determine an accurate acceptability criteria for the nodes in factor space according to the computational process of the model, and how to solve the problem of validate the model with iteration operation inside etc. Meanwhile, we should notice it is hard to get observed data on some intermediate outputs of the model, such as aerodynamic force etc. On these nodes subjective evaluation is often used to substitute the objective similarity analysis.

ACKNOWLEDGMENTS

The paper is supported by the National Natural Science Foundation of China (Item No. 61374164).

REFERENCES

Sargent R. G., 2015. Model verification and validation. London: Springer.

Roy C. J., Oberkampf W. L, 2011. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Computer Methods in Applied Mechanics & Engineering, 200(25–28):2131-2144.

You L., Mahadevan S., 2013. Quantitative model validation techniques: New insights. Reliability Engineering & System Safety, 111(2):217-231.

Eek M., Kharrazi S., Gavel H., et. al., 2015. Study of industrially applied methods for verification, validation and uncertainty quantification of

simulator models. International Journal of Modeling Simulation & Scientific Computing, 6(02): 1-29.
Kutluay E., Winner H., 2014. Validation of vehicle dynamics simulation models - A review. Vehicle System Dynamics, 52(2): 186-200.
Song H., Liu J., Schnieder E., 2017. Validation, verification and evaluation of a Train to Train Distance Measurement System by means of Colored Petri Nets. Reliability Engineering & System Safety, 164:10-23.
Zhou L, Sun K, Li H., 2016. Multifactorial decision making based on type-2 fuzzy sets and factor space approach. Journal of Intelligent & Fuzzy Systems, 30(4): 2257-2266.
Andrei D., Gary J. B., Tryphon T. G., 2014. Validating aircraft models in the gap metric. Journal of Aircraft, 51(6): 1665-1672
Wei H. L., Li Z., 1997. Grey relational analysis and its application to the validation of computer simulation models for missile system. System Engineering and Electronics, (2): 55-61
Jiang X. M., Mahadevan S., 2011. Wavelet spectrum analysis approach to model validation of dynamic systems. Mechanical Systems and Signal Processing, 25(2): 575-590
Mullins J., Ling Y., Mahadevan S., Sun L., 2015. A strachan separation of aleatory and epistemic uncertainty in probabilistic model validation. Reliability Engineering & System Safety, 147: 49-59
Fang K., Ma P., Yang M, 2012. The MAD network for virtual protocol systems credibility evaluation. CIMS, 18(5): 1054-1060
Fang K., Yang M., Zhang Z, 2011. The MAD network for credibility evaluation of computer simulation. Proceedings of The 13th IEEE Joint International Computer Science and Information Technology Conference, pp.636-642., Chongqing, China.
Zhang J. Y., 2011. Validation methods and assistant tools based on coherence of data. Thesis (Master). Harbin Institute of Technology.
Oldrich P., Andreas B., 2014. A new approach to define criteria for rail vehicle model validation. Vehicle System Dynamics, 52(1): 125-141

AUTHORS BIOGRAPHY

Ke Fang is an Associate Professor of the Control & Simulation Center, School of Astronautics at Harbin Institute of Technology, China. He holds a Ph.D. in control science and engineering from Harbin Institute of Technology, and was a visiting scholar at Arizona State University from year 2014 to 2015 in AZ, USA. His research interests include complex simulation systems, model validation and VV&A. His e-mail address is hitsim@163.com.

Ming Yang is a Professor of the Control & Simulation Center, School of Astronautics at Harbin Institute of Technology, China. He holds a Ph.D. in control science and engineering from Harbin Institute of Technology. He involved in major simulation conference and committees in China. His research interests include distributed simulation, hardware-in-loop simulation and VV&A etc. His e-mail address is myang@hit.edu.cn.

Kaibin Zhao received the M.S. degree in control science and engineering from Harbin Institute of Technology, China, in 2013. He is currently pursuing the Ph.D. degree in control science and engineering at HIT. His current research interests is the model validation for complex system. His e-mail address is kaibin.zhao.HIT@hotmail.com.

QUANTUM COGNITION MODELS OF ETHICAL DECISION-MAKING IN HUMAN BEHAVIOR SIMULATION

Levent Yilmaz

Department of Computer Science and Software Engineering *and*
Department of Industrial and Systems Engineering
Samuel Ginn College of Engineering
Auburn University

yilmaz@auburn.edu

ABSTRACT
Ethical decision-making is a unique aspect of human behavior. When confronted with situations that require careful deliberation over multitude of options that have ethical implications, human behavior tends to resolve dilemmas by resorting to a range of heuristics and principles that view the situation from different perspectives. While constraint and utility-driven decision-making strategies are relevant, the incompatibility among these perspectives can invalidate the underlying premises of models of probabilistic utility-based decisions that rely on classic Kolmogorov axioms. In this paper, it is posited that quantum cognition models can provide an alternative and credible representation of human behavior modeling in simulations that involve ethical decision-making.

Keywords: ethical decision-making, quantum cognition, machine ethics, computational ethics

1. INTRODUCTION

Ethical dilemmas often arise due to conflicts among different facets observed in a situation (Anderson and Anderson, 2011; Wallach and Allen, 2009). The consequences, obligations, duties, and principles present incompatibilities that cannot be simultaneously evaluated to make judgments based on joint assessments. For instance, consider humanitarian response and recovery operations, which require situation awareness and allocation of resources. Is the objective to increase the "number of lives" or "the number of years of life" saved. Also, to what extend the decision-making capacity need to consider the necessity to save the lives of the first responders? Acceptable decisions in such situations rely on the priorities distributed over the principles adhered to (Arkoudas and Bello, 2005) and the consequences expected (Gips, 1995; Gert, 1998). The resolution of conflicts among such competing decisions rely on strategies, which facilitate bringing the uncertainty and ambiguity into a conclusion (Yilmaz et al., 2016; Yilmaz et al., 2017).

The resolution of such conflicts is complicated when the dilemma involves multiple decisions that are framed from the perspective of different stakeholders. In the presence of multiple competing perspectives, the order in which each perspective is evaluated can influence the outcome due to potential interference between judgements. In an ethical dilemma, the decision-maker may start the evaluation from the perspective of an organization that prioritize specific rules and principles, which may be incompatible with the consequentialist view that is prioritized from an individual's perspective. Yet another perspective can be based on the propositions that involve cost-benefit analysis. In the presence of multiple perspectives, judgments can be incompatible in that the set of features necessary to encode and evaluate one perspective may not be shared by the set of attributes used to evaluate another (Busemeyer and Bruza, 2014). Therefore, no common set of features are available to jointly evaluate information in different perspectives.

Quantum causal reasoning and inference models (Bruza et al., 2015) have been put forward to account for a variety of such cognitive phenomena, including interference effects. Quantum models of cognition (Busemeyer and Bruza, 2014) postulate that the state of a cognitive system is undetermined and stays in an indefinite superposition state, reflecting the conflict and uncertainty that is akin to the state of ambiguity observed in ethical dilemmas. Beliefs stay in a superimposed state until a decision must be reached. Making a decision is then the process of transforming a thought wave into a particle in a quantum model. Furthermore, judgements with respect to a perspective create a context that influences the construction of an opinion. As such, quantum models do not require adherence to the principle of unicity. That is, because incompatible perspectives cannot be evaluated on the same basis, they require constructing distinct, but possibly related sample spaces, without complete joint assessments.

In the rest of this paper, the plausibility of using quantum cognition models in modeling ethical dilemmas is examined to bring resolution to cognitive conflicts in accordance with the principles of quantum causal reasoning and inference. We demonstrate the application of quantum causal reasoning using a practical example that involves an ethical dilemma viewed from multiple perspectives. Following the illustration of the strategy, we evaluate the utility of the strategy, discuss its limits,

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

and conclude by discussing potential areas of further inquiry.

2. BACKGROUND

The field of machine ethics has generated wealth of knowledge, including methods, tools, and strategies for engineering artificial systems that can exhibit the characteristics of moral behavior (Wallach and Allen, 2009; Herman, 2014). As a corollary, the use of computational methods is facilitating both the identification of gaps and the advancement of normative theories of ethical decision-making. Machine ethics has emerged as a field to explore the nature, issues, and computational approaches for ascribing ethical values to agents. Recently, AI-assisted ethics (Etzioni and Etzioni, 2016) is promoted as a perspective to better understand the ethical norms and principles that individuals adhere to so as to provide effective guidance for human-agent interaction.

2.1. Computational Ethics

The most commonly used AI methods in computational ethics are planning, deductive reasoning with deontic logic, analogical reasoning, constraint satisfaction, decision-theory, and (social) learning. In (Kurland, 1995), the role of theory of planned behavior and reasoned action in predicting ethical intentions towards others is demonstrated in experiments with human subjects. Besides planning and goal-directed behavior, deductive logic is proposed (Saptawijaya and Pereira, 2012) to infer judgments and their ethical implications. In (Arkoudas and Bello, 2005), authors demonstrate the use of deontic logic in formalizing ethical codes and discuss the utility of mechanized formal proofs for instilling trust in autonomous systems. Analogy-based reasoning has also emerged as an effective method to address practical problems. For instance, the Truth-Teller system (McLaren, 2006) implements a computational model of casuistic reasoning, by which a decision is made via comparing the given problem to paradigmatic, real, or hypothetical cases.

Constraint-based methods such as the Ethical Governor model (Arkin, 2009) are proposed promote moral and ethical behavior of autonomous systems by using explicit rules that avoid harmful consequences. Recently, a similar strategy is promoted in (Greene et al., 2016) to manage collective decision-making by filtering behavior per explicitly defined constraints. To address decision-making under uncertainty, Dennis et al. (2013) provide a strategy for generating plausible strategies and preferences in unforeseen situations.

In the absence of pre-defined constraints and preferences, the application of machine learning facilitated discerning rules and principles that are often left implicit in discussions involving moral philosophy and psychology. Among the applications of machine learning to ethics are the MedEthEx (Anderson et al., 2006) and the GenEth (Anderson and Anderson, 2014) systems, both of which use Inductive Logic to discern rules that resolve ethical dilemmas that emerge due to conflicting obligations and duties. Consequentialist theories of ethics are also among the common models for which solutions have been applied in the form of agents that are driven by utility-based cognitive architectures. In consequentialist theories of ethics, actions are evaluated by their consequences; that is, the objective is often to optimize the well-being of the largest group of individuals and achieve the most desirable situation possible among many options (Gips, 1995).

2.2. Machine Ethics

In relation to using ethics for agents, Moor (2006) proposed four types agency: (1) ethical impact agent, (2) implicit ethical agent, (3) explicit ethical agent, and (4) fully ethical agent. Ethical impact agents are indirectly ethical agents, because they are not endowed with models of ethics. Rather, the presence of a computing technology indirectly brings an effect that facilitates the emergence of a situation that can be viewed as morally desirable outcome. On the other hand, if the ethical behavior is intentional, an agent can be either an implicitly or explicitly ethical agent. Implicitly ethical agents are entities, whose actions are constrained to avoid unethical outcomes. Constraints are defined to inhibit undesirable behavior or allow only legal actions permissible by the rules of engagement and general laws of the domain of interest (Arkin, 2009). However, explicitly ethical agents embody knowledge-based models of ethical decision-making that allow representing ethical categories and performing analysis in each situation to select actions recommended by the model. Such models are guided by theories of ethics, including consequentialist, deontological, and virtue-based theories. Therefore, whereas explicit ethical agents are knowledge-based entities, implicit ethical agents can be considered as rule-based entities. Building on knowledge-based ethical agents, fully ethical agents bring a level of expertise, which enables agents to justify their moral reasoning and learn from experience (e.g., failures) to improve their own models of ethics.

3. QUANTUM INFERENCE MODELS

Both classic and quantum probability theories are concerned with the problem of assigning probabilities to events such as the decisions in an ethical dilemma. Quantum approach to reasoning (Trueblood and Busemeyer, 2012) and decision-making is predicated on three major observations: (1) human judgments are constructed from both the current context and the question; hence, the state of a cognitive system is undetermined and the resolution is facilitated by measurements/questions; (2) the ambiguity with respect to multiple beliefs is akin to a thought wave over multiple cognitive states that are superimposed until a final decision is made; (3) the changes in a context driven by a judgement affect later judgements due to incompatibility between events disturbing each other to generate uncertainty.

The quantum inference model has been used to account for order effects in various application domains, including medical diagnostic tasks and jury decision-making problems. In explaining the order effects, the notion of incompatibility is used to model different aspects of a problem along with their interactions. Specifically, the concept of incompatibility recognizes that the set of features used to evaluate one aspect may not be shared by the set used to model another so that no common set of features can be used to jointly evaluate information from different aspects. This is particularly the case when knowledge about all combinations of information from different sets is not known by the decision-maker. Hence, different perspectives need to be adopted to evaluate different pieces of information.

3.1. Events

Events in quantum theory are modeled as subspaces, which are characterized in terms of basis vectors. Each basis vector corresponds to an elementary outcome. Figure 1 illustrates a hypothetical vector space with three events *X*, *Y*, and *Z*. The basis vectors are orthogonal to each other, indicating that the events that they represent are mutually exclusive.

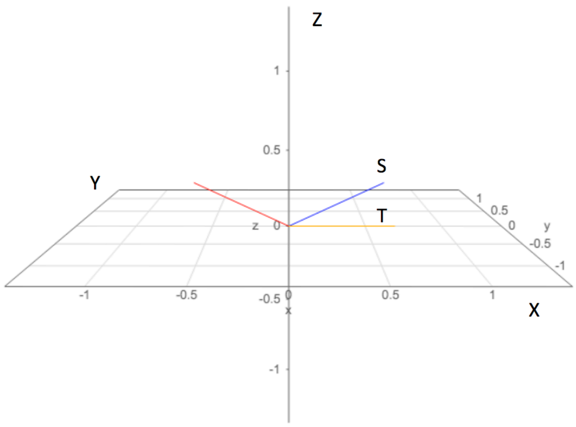


Figure 1: Vector Space

Because events are defined as subspaces, the logic of subspaces are used to combine events so that conjunction of events are defined in terms of the intersection of subspaces. That is, the span of the rays formed by $|X\rangle$ and $|Y\rangle$

3.2. System State

In quantum theory, the system state is defined by a unit-length vector in the *N*-dimensional vector space. The vector, symbolized as $|S\rangle$, is used to map events into probabilities. The state in Figure 1 refers to the line segment labeled as *S*, which is located at 0.696 units along the *x* axis, 0.696 units along the *y* axis, and 0.175 units up the *z* axis. Updating the state based on observations involves projection of the state vector onto the subspace of the observed event.
More specifically, the state is projected onto the subspace of the event, and projected vector is divided by the length of the projection to make sure that the revised state has length one. Also, given the original state vector, *S*, the probability of an event associated with a basis vector is equal to the squared length of the projection. The state $|S\rangle$, shown in Figure 1 is defined with respect to basis vectors $|X\rangle$, $|Y\rangle$, $|Z\rangle$ as follows:

$$|S\rangle = (-06963).|X\rangle + (0.6963).|Y\rangle + (0.1741).|Z\rangle$$

The coordinates of the state vector are defined in terms of the amplitudes of individual basis vectors:

$$|S\rangle \rightarrow \begin{bmatrix} -0.693 \\ 0.693 \\ 0.1741 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \alpha$$

The symbol α denotes the 3×1 (single column) matrix, which is comprised of coordinates of the abstract state vector $|S\rangle$ when defined in terms of the $|X\rangle$, $|Y\rangle$, $|Z\rangle$ basis. These coordinates can also be derived from the canonical coordinate system.

$$|X\rangle \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, |Y\rangle \rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, |Z\rangle \rightarrow \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

The inner product between $|X\rangle$ and $|S\rangle$ computes the transition amplitude from the state $|S\rangle$ to state $|X\rangle$, and it is denoted by the matrix formula.

$$\langle X|S\rangle = [1 \quad 0 \quad 0]. \begin{bmatrix} -0.693 \\ 0.693 \\ 0.1741 \end{bmatrix}$$

$$= 1.(-0693) + 0.(0.693) + 0.(0.1741) = -0.693$$

Next, we use the concepts defined above to model the scenario presented in section 3 to illustrate the process of quantum inference model of causal reasoning to support ethical decision.

4. MULTI-PERSPECTIVE ETHICS CONFLICTS

Ethical dilemmas often involve consideration of decisions from the perspectives of multiple stakeholders. Next, we introduce a motivating scenario, which is adopted from Ethics Case Studies, published by APS Physics (2017). The scenario will be used to demonstrate the basic tenets of the proposed approach.

4.1. Motivating Scenario

Consider a graduate student who is in the final stages of completing her dissertation project and has applied for several tenure-track assistant professor positions. She is invited for interview at her undergraduate alma mater, which is a prestigious research university at which she has connections and would like to work. At the end of the seminar, during the question-and-answer session, the chair of the department asks for specific and detailed information about her research. However, her research group is currently preparing for a patent application that relates to the requested information. Moreover, her research group has agreed to hold information until the paper that is currently being prepared is submitted for publication. Also, her advisor is scheduled to give a presentation at a major international conference to release the details of the technique used in her research.

So, she kindly answers the department chair by indicating that she and her colleagues are in the process of developing a paper and submitting a patent application, and that she would be glad to share the information when an early print of the paper becomes available.

After the seminar, during the private interview session, the search committee presses harder to learn more about the details of her research, highlighting that the department is seeking team players and that a viable candidate for the position should share such information.

In this scenario, there are multiple stakeholders: the graduate student applying for the position, her graduate advisor, the colleagues in her research group, the chair, the university where she is interviewing, her graduate university, and the research field. Each one of the stakeholders have different interests. The graduate student is conflicted for the decision made by her research group to delay the release of the information and to keep it confidential reduces her chances of getting hired by the department she is interviewing with. Her options include (1) giving the information and not telling her colleagues, (2) giving the information and telling her colleagues, (3) contacting her supervisor for permission, (4) convincing the chair out of his urgency to acquire the information.

From the perspective of the graduate applicant, demonstrating that she is a person of her word has the highest importance, but she is also conflicted and uncertain about which judgement is appropriate. The preferences of different stakeholders constitute distinct judgements, which may not be assessed simultaneously or jointly due to the incompatibilities among preferences. That is, for instance, a clear valuation over joint preferences of events from the perspective of her team and the department chair may not be readily available or even be meaningful. Therefore, before reaching a conclusion, she can view the problem from the perspectives of her research group, including her advisor, and the chair of the department to which she is applying. The interference and order effects should be taken into consideration in the reasoning process.

4.2. Quantum State Model

The motivating scenario involves a graduate student applying for a tenure-track position. She is expected to decide whether to give the requested information. There are three perspectives. The personal perspective of the graduate student, and the viewpoints of her research team and the department chair that will make the hiring decision. There are two judgments, giving the information or not, represented by events G versus $\overline{G}$ The evidence or observation that influences the judgement is whether it is permissible to release the information. The permissibility of releasing the information is represented by the event P, and its complement $\overline{P}$. In the context of classical probability or classical Bayesian model, the decision process can be modeled with at least 8 events, which consider combinations of giving the information, its complement, as well as the permissibility (and its complement) from the perspectives of the team and the department chair. Yet, these perspectives are incompatible, and the order in which each perspective is considered affects the final moral judgment.

To define a quantum inference model for this problem, we need to define a vector space, which is comprised of orthogonal and mutually exclusive events. One approach is to define eight basis vectors, one for each one of the combinations of three features: Giving the information (G), permissible from the perspective of the research team (P_t), and permissible from the perspective of the chair (P_c). Including the complements of these features, there are eight distinct joint events. Instead, the joint event space can be reduced to a four-dimensional space based on four combinations of two hypotheses (i.e., judgments of giving information or not) and two types of effects (permissible or not).

These four dimensions can be framed from three perspectives: applicant's perspective, team's perspective, and chair's perspective. The applicant's perspective represents the student's state of belief before any facts are considered from other perspectives. Considering her own perspective, the applicant generates the first judgment. If the research team's view is considered next, then the team's perspective denotes the state of belief when the applicant views the dilemma from the lens of the team, resulting in the second judgment. If the chair's arguments are considered next, then the department chair's perspective defines the applicant's state of belief when viewed from the chair's point of view, generating the third and final judgment.

To determine the probabilities associated with the judgments at each step of the process as different perspectives are considered, we need to clarify the relationships between the applicant basis, team basis, and chair basis. The same belief state $|S\rangle$ can be defined from the perspective of these three basis configurations as follows:

$$
\begin{aligned}
|S\rangle \quad &= a_{GP}.|A_{GP}\rangle + a_{G\overline{P}}.|A_{G\overline{P}}\rangle + a_{\overline{G}P}.|A_{\overline{G}P}\rangle \\
&\qquad + a_{\overline{GP}}.|A_{\overline{GP}}\rangle \\
&= t_{GP}.|T_{GP}\rangle + t_{G\overline{P}}.|T_{G\overline{P}}\rangle + t_{\overline{G}P}.|T_{\overline{G}P}\rangle \\
&\qquad + t_{\overline{GP}}.|T_{\overline{GP}}\rangle \\
&= c_{GP}.|T_{GP}\rangle + c_{G\overline{P}}.|T_{G\overline{P}}\rangle + c_{\overline{G}P}.|T_{\overline{G}P}\rangle \\
&\qquad + c_{\overline{GP}}.|T_{\overline{GP}}\rangle
\end{aligned}
$$

For each basis, the index GP stands for the pattern "Giving the information and Permissible", $G\overline{P}$ for "Giving and not Permissible", $\overline{G}P$ for not "Giving and Permissible", and $\overline{GP}$ for not "Giving and not Permissible." Each basis can be represented by 3×1 coordinate matrix defining the amplitudes for each basis vector.

$$
a = \begin{bmatrix} a_{GP} \\ a_{G\overline{P}} \\ a_{\overline{G}P} \\ a_{\overline{GP}} \end{bmatrix}, t = \begin{bmatrix} t_{GP} \\ t_{G\overline{P}} \\ t_{\overline{G}P} \\ t_{\overline{GP}} \end{bmatrix}, c = \begin{bmatrix} c_{GP} \\ c_{G\overline{P}} \\ c_{\overline{G}P} \\ c_{\overline{GP}} \end{bmatrix}
$$

Transformation from one basis to another requires mapping the amplitude distribution of one perspective to the other so that events can be evaluated from a different viewpoint. Figure 2 illustrates a mapping from the perspective of the applicant to the perspective of the research team.

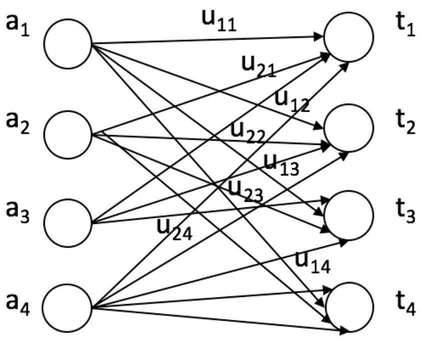


Figure 2: Transformation of the Amplitudes

The connection weights u_{ij} are used to map input amplitude distribution $\boldsymbol{a}$ across $a_1$to a_4 into the output distribution $\boldsymbol{t}$ across $t_1$to t_4. The mapping uses a linear combination strategy $t_j = \sum_i u_{ij}\, a_j$. This mapping can be characterized in terms of the following matrix operation:

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} & u_{31} & u_{41} \\ u_{12} & u_{22} & u_{32} & u_{42} \\ u_{13} & u_{23} & u_{33} & u_{43} \\ u_{14} & u_{24} & u_{34} & u_{44} \end{bmatrix} . \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

This can be abbreviated as

$$t = U.a$$

where U is a unitary matrix that preserves probability amplitudes in such a way that $t^T.t = 1$ just as $a^T.a = 1$. That is, sum of probabilities across events in each basis equal to 1. Therefore, a unitary matrix U maps one set of coordinates defined by the first basis to another set of coordinates defined by the second basis. Because we need the basis vectors to be of unit length, pairwise orthogonal, a unitary matrix need to preserve the lengths and inner products. Therefore, U needs to satisfy the property $U^T.U = I = U.U^T$. Any unitary matrix can be expressed in terms of matrix exponential of a Hamiltonian matrix, $U = e^{-iH}$, where $H^T = H$ is an$N \times N$ Hermitian matrix. The elements of an Hermitian matrix has elements $h_{jk} = h_{jk}^*$.

Next, we assume the presence of unitary transformation matrices among the perspectives. One unitary matrix U_{at} is needed to map the initial personal perspective of the applicant to the viewpoint of her research team. By using the unitary matrix, the coordinates of the belief state vector of the applicant is mapped on to the coordinates $t = U_{ta}.a$ of what the applicant thinks the team's perspective is. Similarly, the mapping from the applicant's perspective to the department chair's viewpoint requires the unitary matrix U_{ca}.

4.3. Applicant's Perspective

The initial belief state is defined in terms of the applicant's perspective and its orthogonal basis vectors. The initial state is akin to prior probability in a Bayesian network.

$$|S\rangle = a_{GP}.|A_{GP}\rangle + a_{G\overline{P}}.|A_{G\overline{P}}\rangle + a_{\overline{G}P}.|A_{\overline{G}P}\rangle + a_{\overline{G}\overline{P}}.|A_{\overline{G}\overline{P}}\rangle$$

The coordinates of the initial state are based on the probability amplitudes, which can be used to determine the probabilities of individual events. Probabilities of events are equal to square of the amplitudes defined in the coordinate matrix.

$$a = \begin{bmatrix} a_{GP} \\ a_{G\overline{P}} \\ a_{\overline{G}P} \\ a_{\overline{G}\overline{P}} \end{bmatrix} = \begin{bmatrix} \sqrt{0.643} \\ \sqrt{0.643} \\ \sqrt{0.643} \\ \sqrt{0.643} \end{bmatrix}$$

Given this coordinate matrix of probability amplitudes for the applicant perspective, before the evaluation of other perspectives, the probability of giving the information can be derived by projecting the state vector onto the subspace spanned by the give-information basis vectors: $a_{GP}\,.\,|A_{GP}\rangle + a_{G\overline{P}}\,.|A_{G\overline{P}}\rangle$. Using the respective amplitudes, the probability of the decision (D) of giving the information equals $D(G) = |a_{GP}|^2 + |a_{G\overline{P}}|^2$.

4.4. Research Team's Perspective

To view the ethical dilemma from the perspective of the research team, the applicant needs to rotate her perspective and align it with the team's viewpoint. This can be done by either rotating the state vector or the basis. In the previous section, we defined the team's viewpoint from a distinct set of orthogonal basis vectors to interpret the applicant's initial state $|S\rangle$ from a different viewpoint.

$$|S\rangle = t_{GP}.|T_{GP}\rangle + t_{G\overline{P}}.|T_{G\overline{P}}\rangle + t_{\overline{G}P}.|T_{\overline{G}P}\rangle + t_{\overline{G}\overline{P}}.|T_{\overline{G}\overline{P}}\rangle$$

The amplitudes for the team's basis can be derived from the applicant's basis via an appropriate unitary matrix that meaningfully relates probability amplitudes across the source and target coordinate matrices.

$$t = \begin{bmatrix} t_{GP} \\ t_{G\overline{P}} \\ t_{\overline{G}P} \\ t_{\overline{G}\overline{P}} \end{bmatrix} = U_{ta}.a$$

From the perspective of the team, the probabilities associated with each event are the squared magnitudes of the coordinate matrix. These probabilities are

$$\begin{bmatrix} |t_{GP}|^2 \\ |t_{G\overline{P}}|^2 \\ |t_{\overline{G}P}|^2 \\ |t_{\overline{G}\overline{P}}|^2 \end{bmatrix}$$

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Suppose that the team responds with "No Permission". The state vector $|S\rangle$ is then projected onto the subspace, which is spanned by the set of vectors $\{|T_{G\overline{P}}\rangle, |T_{\overline{GP}}\rangle\}$. This is followed by normalization to produce the following unit-length state vector.

$$|S_{\overline{P}}\rangle = \frac{0.|T_{GP}\rangle + t_{G\overline{P}}.|T_{G\overline{P}}\rangle + 0.|T_{\overline{G}P}\rangle + t_{\overline{GP}}.|T_{\overline{GP}}\rangle}{\sqrt{|t_{G\overline{P}}|^2 + |t_{\overline{GP}}|^2}}$$

The probability amplitudes of the projected and normalized vector, which is predicated on the research team's permission of the release of the information, is as follows:

$$t_{\overline{P}} = \frac{1}{\sqrt{|t_{G\overline{P}}|^2 + |t_{\overline{GP}}|^2}} \begin{bmatrix} 0 \\ t_{G\overline{P}} \\ 0 \\ t_{\overline{GP}} \end{bmatrix}$$

Following the selection of the "No permission" event from the perspective of the research team, the probability of releasing information, *D(G)*, is derived by projecting the state $|S_{\overline{P}}\rangle$ onto the subspace spanned by $\{|T_{GP}\rangle, |T_{G\overline{P}}\rangle\}$ and then squaring its length.

$$D(G|S_{\overline{P}}) = ||\ |T_{GP}\langle T_{GP}|S_{\overline{P}}\rangle\rangle + |T_{G\overline{P}}\langle T_{G\overline{P}}|S_{\overline{P}}\rangle\rangle\ ||$$
$$= ||\frac{t_{G\overline{P}}}{\sqrt{|t_{GP}|^2 + |t_{G\overline{P}}|^2}}\ |T_{G\overline{P}}\rangle||$$
$$= \frac{|t_{G\overline{P}}|^2}{|t_{GP}|^2 + |t_{G\overline{P}}|^2}$$

4.5. Department Chair's Perspective

After evaluation of the research team's perspective, suppose the applicant considers the events of giving the information and permissibility of releasing the information, along with their complements, from the perspective of the department chair. The revised belief state, which was derived based on the evaluation of the team's perspective, is now viewed from the perspective of the department chair.

$$|S_{\overline{P}}\rangle = c_{GP}.|C_{GP}\rangle + c_{G\overline{P}}.|C_{G\overline{P}}\rangle + c_{\overline{G}P}.|C_{\overline{G}P}\rangle + c_{\overline{GP}}.|C_{\overline{GP}}\rangle$$

The belief state, as viewed from the chair's perspective, is characterized by the probability amplitudes, which can be defined by the following 4×1 matrix.

$$c_p = \begin{bmatrix} c_{GP} \\ c_{G\overline{P}} \\ c_{\overline{G}P} \\ c_{\overline{GP}} \end{bmatrix} = U_{tc}.t_{\overline{P}}$$

The probability amplitudes are derived by using U_{ct}, which maps the calculated probability amplitudes of the team's perspective onto the department chair's perspective. The probabilities for each event are then defined as the squares magnitudes of the coordinates of c_p.

$$\begin{bmatrix} |c_{GP}|^2 \\ |c_{G\overline{P}}|^2 \\ |c_{\overline{G}P}|^2 \\ |c_{\overline{GP}}|^2 \end{bmatrix}$$

The unitary matrix U_{tc} is expected to be designed to reflect the fact that the department chair is more likely to choose the "Permission" event than not. That is $c_{GP} + c_{\overline{G}P}$ should be significantly larger than $c_{G\overline{P}} + c_{\overline{GP}}$. Following the research team's decision, if the applicant considers the department chair to choose to request the information, the belief state $|S_{\overline{P}}\rangle$ is projected onto the subspace spanned by $\{|C_{GP}\rangle, |C_{\overline{GP}}\rangle\}$. The projection produces the following belief state

$$|S_{P\overline{P}}\rangle = \frac{c_{GP}.|C_{GP}\rangle + 0.|C_{G\overline{P}}\rangle + c_{\overline{G}P}.|C_{\overline{G}P}\rangle + 0.|C_{\overline{GP}}\rangle}{\sqrt{|c_{GP}|^2 + |c_{\overline{G}P}|^2}}$$

The probability amplitudes with respect to the department chair's perspective as viewed by the applicant is predicated on department chair's preference (i.e., permission), which follows the research team's negative response. The amplitudes is represented by the following coordinate matrix.

$$c_{P\overline{P}} = \frac{1}{\sqrt{|c_{GP}|^2 + |c_{\overline{G}P}|^2}} \begin{bmatrix} c_{GP} \\ 0 \\ c_{\overline{G}P} \\ 0 \end{bmatrix}$$

Following the provision of the choices from the perspective of both the research team and the department chair, the judged probability of releasing the information is derived from the squared magnitude of the projection of the belief state onto the subspace spanned by $\{|C_{GP}\rangle, |C_{G\overline{P}}\rangle\}$. This probability equals

$$(G|T_{\overline{P}}, C) = ||\ |C_{GP}\langle C_{GP}|S_{P\overline{P}}\rangle\rangle + |C_{G\overline{P}}\langle C_{G\overline{P}}|S_{P\overline{P}}\rangle\rangle\ ||$$
$$= ||\frac{c_{GP}}{\sqrt{|c_{GP}|^2 + |c_{G\overline{P}}|^2}}\ |C_{GP}\rangle||$$
$$= \frac{|c_{GP}|^2}{|c_{GP}|^2 + |t_{G\overline{P}}|^2}$$

Given the probabilities of distinct events derived from the resultant state vector, a final judgment is made in accordance with the probability amplitudes of the computed coordinate matrix.

5. EVALUATION AND DISCUSSION

The quantum probability model of the inference process shown in the previous section illustrates how an ethical dilemma can be viewed from the perspectives of different stakeholders. By evaluating different perspectives in the preference order imposed by the decision-maker alternative conclusions can be drawn. The value of the quantum approach and the use of the vector-based projections of the belief states are consistent with the notion of ambiguity and uncertainty involved in ethical conflicts.

At any point in the process, the decision-maker does not have a definite answer; rather, by asking questions and making evaluations, the uncertainty is reduced or projected onto relevant dimensions. Also, the order in which the perspective considerations are made influence the outcome. This is akin to the notion of giving priorities to principles, consequences, or obligations, as well as the perspectives of the stakeholders in a context-sensitive manner. For instance, in one context, the applicant may choose to evaluate the situation from the perspective of the department chair, followed by an assessment from the point of view of the research team. Depending on the transition amplitudes or the rotation structure between the basis vectors of the subspaces, which represent distinct perspectives, a choice made in one perspective may either positively or negatively interfere with one or more events in the other perspective.

5.1. Learning Unitary Transformation Matrices

One of the critical challenges in devising a quantum approach to with such an interference mechanism is the need for defining unitary transition matrices between pairs of coordinate systems that represent different perspectives. Learning the connection weights between basis vectors using a gradient descent optimization approach is one plausible approach for deriving the relations between perspectives.

5.2. Hybrid Modeling of Quantum Cognition Models

The quantum inference model presented in this paper utilizes the Hilbert space to model belief state of an individual in terms of vector projections. The revision of belief states is driven by manipulation via unitary matrices of the amplitudes of the basis vectors in a coordinate system. The explicit representation of symbolic propositions and the revision of their degree of expectancy facilitate clear communication of the cognitive state of an agent. These symbolic propositions are not only succinct, but are also accessible to the decision-makers. However, the symbolic representation lacks a holistic view of the emergence of the probability amplitudes associated with the basis vectors. A sub-symbolic or implicit cognition layer can provide a unified mechanism to support the implementation of quantum cognition model.

The symbolic versus sub-symbolic representation dichotomy in cognitive systems is well-recognized. At the symbolic level, situations and objects are represented by atomic symbols that define labels and categories, which can be modified by syntactic manipulation. On the other hand, such categories and types emerge from a complex pattern of distributed and interactive sub-symbols that characterize the formation of the symbolic types and concepts. At the symbolic level, entities are atomic, discrete, and content-defined, whereas in the sub-symbolic level, there is a distributed and complex adaptive mechanism that characterize the emergence of categories and types at the symbolic level. The coupling of the explicit symbolic level and the implicit sub-symbolic level aims to mitigate the symbol grounding problem and variable-binding problem of the sub-symbolic phenomena.

Metaphorically, the implicit cognition layer is consistent with the view of making decision as the process of transforming a thought wave into a particle in a quantum model (Busemeyer and Bruza, 2014). Because even classical dynamical systems such as connectionist models could exhibit quantum-like properties (Blutner and Graben, 2014), the basic idea is to find a general and abstract formulation of quantum models that align with and validate against the structures of quantum probabilities and vectors in a Hilbert space. The wave propagation can be mimicked via coherence maximization in the constraint network in a way similar to simultaneous firing of dendrites in the connectionist model of the Pattern Recognition Theory of Mind (Kurzweil, 2012). The activations of nodes in the network is analogous to acceptability of the respective propositions.

This strategy is akin to viewing state as a unit-length vector in an N-dimensional vector space. Each node receives input from every other node that it relates to. The inputs can then be moderated by the weights of the link from which the input arrives. These weights can be construed as the elements of a unitary matrix that connect the probability amplitudes of propositions in one perspective to the probability amplitudes of the propositions in another perspective. The activation value of a unit is updated as a function of the weighted sum of the inputs it receives. The process continues until the activation values of all units settle. The activation values at the equilibrium state need to be projected and mapped onto probability amplitudes so that the probability of events sum up to 1.

In quantum cognition models, each coordinate basis is considered in sequence, resulting in order and interference effects that are consistent with observed human behavior. The provision of an implicit connectionist and sub-symbolic processing layer can help demonstrate a process by which beliefs as viewed from different perspectives can be adjusted in both directions simultaneously without following a strict sequence in evaluation. That is, evaluation from the view of multiple perspectives can interfere with each other. In the most general sense, the equilibrium can be construed as an attractor state in a complex adaptive system. The attractor state emerges at the end of a deliberation process by which we reflect on and revise our beliefs.

6. CONCLUSIONS

Ethical decision-making is emerging as a critical challenge as autonomous systems and software agents continue to immerse in our daily lives. Such systems face ethical dilemmas and conflicting situations when their course of actions may have both desirable and undesirable consequences when viewed from different perspectives. Moreover, the conflicts between obligations, duties, and consequences require viewing situations from different perspectives to reach a decision that considers multiple points of view. The use of quantum probabilistic and causal inference models is promoted as a plausible strategy to frame the perspectives in terms of distinct coordinate systems, which are related to each other in terms of vector orientations. Furthermore, the capacity of quantum cognition to specify judgements in terms of indefinite states facilitates capturing psychological experience of conflict, ambiguity, confusion, and uncertainty. By using a simple and hypothetical example, we illustrated how quantum probabilistic reasoning can resolve conflict by reducing an indefinite state to a certainty and decision via projection over subspaces defined by multitude of basis vectors.

REFERENCES

Anderson, M. and Anderson, S.L., 2014. GenEth: A General Ethical Dilemma Analyzer. In *AAAI* (pp. 253-261).

Anderson M. and Anderson, S.L. eds., 2011. Machine ethics. Cambridge University Press.

Anderson M., Anderson S.L. and Armen C., 2006. An approach to computing ethics. *IEEE Intelligent Systems*, *21*(4), pp.56-63.

APS Physics, 2017. Ethics Case Studies. https://www.aps.org/programs/education/ethics/

Arkin R., 2009. Governing lethal behavior in autonomous robots. CRC Press.

Arkoudas S. B. and P. Bello, 2005. Toward ethical robots via mechanized deontic logic. In AAAI Fall Symposium on Machine Ethics.

Blutner R. and beim Graben, P., 2016. Quantum cognition and bounded rationality. Synthese, 193(10), pp.3239-3291.

Bruza P.D., Wang, Z. and Busemeyer, J.R., 2015. Quantum cognition: a new theoretical approach to psychology. Trends in cognitive sciences, 19(7), pp.383-393.

Busemeyer J.R. and Bruza P.D., 2012. Quantum models of cognition and decision. Cambridge University Press.

Gert B., 1998. Morality: Its nature and justification. Oxford University Press on Demand.

Greene, J., Rossi, F., Tasioulas, J., Venable, K.B. and Williams, B., 2016, February. Embedding ethical principles in collective decision support systems. In AAAI (pp. 4147-4151).

Dennis L., Fisher, M., Slavkovik, M., and Webster, M., 2013. Ethical choice in unforeseen circumstances. In Conference Towards Autonomous Robotic Systems (pp. 433-445). Springer Berlin Heidelberg.

Etzioni A. and Etzioni, O., 2016. AI assisted ethics. Ethics and Information Technology, 18(2), pp.149-156.

Gips J., 1995. Towards the ethical robot. Android epistemology, pages 243–252, 1995.

Herman M., 2014. Moral heuristics and biases. Journal of Cognition and Neuro-ethics, 73(1), pp.127-142.

Kurland N.B., 1995. Ethical intentions and the theories of reasoned action and planned Behavior1. Journal of applied social psychology, 25(4), pp. 297-313.

Kurzweil R., 2012. How to create a mind: The secret of human thought revealed. Penguin.

McLaren B.M., 2006. Computational models of ethical reasoning: Challenges, initial steps, and future directions. IEEE intelligent systems, *21*(4), pp.29-37.

Moor J.H., 2006. The nature, importance, and difficulty of machine ethics. *IEEE intelligent systems*, *21*(4), pp.18-21.

Saptawijaya A. and Pereira L.M., 2012, March. Moral reasoning under uncertainty. In International Conference on Logic for Programming Artificial Intelligence and Reasoning (pp. 212-227). Springer Berlin Heidelberg.

Trueblood, J.S. and Busemeyer, J.R., 2012. A quantum probability model of causal reasoning.

Wallach W. and Allen C., 2009. Moral machines: Teaching robots right from wrong. Oxford University Press.

Yilmaz L., Franco-Watkins, A. and Kroecker, T.S., 2016. Coherence-driven reflective equilibrium model of ethical decision-making. In Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2016 IEEE International Multi-Disciplinary Conference on (pp. 42-48). IEEE.

Yilmaz, L., Franco-Watkins, A. and Kroecker, T.S., 2017. Computational models of ethical decision-making: A coherence-driven reflective Equilibrium Model. Cognitive Systems Research. http://doi.org/10.1016/j.cogsys.2017.02.005

AUTHORS BIOGRAPHY

Levent Yilmaz is Professor of Computer Science and Software Engineering at Auburn University with a courtesy appointment in Industrial and Systems Engineering. He holds M.S. and Ph.D. degrees in Computer Science from Virginia Tech. His research interests are in agent-directed simulation, cognitive autonomous systems, and model-driven science and engineering for complex adaptive systems. He is the former Editor-in-Chief of Simulation: Transactions of the Society for Modeling and Simulation International and the founding organizer and general chair of the Agent--Directed Simulation Conference series. He published over 200 contributions in journals, book chapters, and conference proceedings and edited or co-authored several books on Modeling and Simulation. He received the Distinguished Service and Distinguished Professional Achievement Awards from the Society for Modeling and Simulation International, as well as the inaugural Auburn University External Consulting Award. He also served as the member of the Board of Directors and Vice President for Publications of the SCS. His email address is yilmaz@auburn.edu.

VALIDATION OF DESIGN SOLUTIONS AND MATERIALS FOR A HIGH PERFORMANCE BEARING CAGE

Aleida Lostalé[(a)], Isabel Clavería[(a)], Ángel Fernández[(a)], Manuel Muniesa[(a)], Carlos Javierre[(a)], Daniel Elduque[(a)], Sergio Santodomingo[(a)]

[(a)]University of Zaragoza I+

[(a)]isabel.claveria@unizar.es,

ABSTRACT
Bearings are one of the most used components in mechanical engineering. One element in the bearing assembly having a great influence in its performance is the cage. Cages work mainly under friction and inertia loads, so the selection of a proper material with low density and good mechanical and friction properties is essential. This paper analyzes the injection molding feasibility of different polymeric matrix to be used in composite materials for cage design. Alternative geometrical design of the cage optimized for lubrication purposes will be also analyzed in order to make sure its manufacturing feasibility. Results conclude that polyamide, mechanically improved with glass fiber, is the best option for polymeric matrix regarding injection molding process. On the other hand new cage geometry requires some modifications in injection molding parameters to meet process requirements.

Keywords: bearing, cage, mechanical design, injection molding.

1. INTRODUCTION

Bearings are the most used components in mechanical engineering to allow relative rotational motion between two concentric elements. Conical bearings are one of the most widespread bearings because of their stiffness, load capacity, rotation speed and equilibrium between radial and axial load. One of the components of the conical bearing assembly is the cage, which have a great influence on the suitability of the bearing for a specific purpose. Cages on bearings are required for keeping the rolling elements separated, for achieving a uniform distribution of the rolling elements, for guiding rolling elements to avoid sliding, and for retaining the rolling elements during assembly.

Cages usually have a passive role during bearing performance, and this study pretends them to become an active role thanks to the usage of new materials and new geometries.

Cages usually work under loads derived from friction and inertia (Kohar and Hreck 2016), as well as deformation due to pressure. Besides, the usage of lubricants and greases can affect chemically the cage material reducing its properties. That is why a proper selection of matrix materials for the cage and additives are strongly relevant to make sure the functionality of the bearing and improve its performance (Gunst Zabel Valle 2010).

New trends in cage design are focused on solid cages that are the most used for conical bearings and the most suitable for innovation. These cages combine in a very profitable way flexibility and resistance (Fang, Pugh, and Themudo 2007). If polymeric materials are used, the advantages of a serial manufacturing process like injection molding process can be combined with the good friction properties of some polymers. Injection molding process also allows to achieve an excellent surface quality to get in touch with the rollers. Another advantage of using polymers is their low density, minimizing the effect of inertia loads on the cage, which make them optimum for high speed and long term performances.

The most important limitations that solid cages have are two: high temperatures reached due to low conductivity of polymers under continuous work conditions, and ageing caused by some kind of lubricants on the polymer matrix.

New trends for high performance bearings are focused on new materials (Foster, Rosado, Brown et al. 2002), with low friction and high mechanical properties and high working temperature, new methods of lubrication for high temperatures and higher limits of speed-pressure (Ettefaghi 2013; Liu 2011; Huang 2011), new surface treatments for the rolling roads (Ciarsolo Fernández Ruiz de Gopegui Zubizarreta Abad Mariscal caretti Jiménez sánchez-López, 2014; Manier Ziegele Barriga, Goikoetxea Woydt, 2010) and new design of the components.

The goal of this research is to optimize cage geometry and material in order to obtain a more effective design for high performance purposes. Different cage geometries and materials will be analyzed from the point of view of mechanical and rheological behavior to meet the optimum design according to mechanical and manufacturing process requirements.

Regarding to geometry, the goal of this paper is validate new shape of the cage from the point of view of injection molding process, establishing strong and weak points regarding the original geometry. On the other hand, regarding to material, another aim of the paper is to analyze the mechanical and rheological behavior of different potential materials for the polymeric matrix of the cage.

2. METHODOLOGY

Two different geometrical models are going to be analyzed varying some geometrical features that could be relevant to get a better tribological performance as well as different materials. Figure 1 shows a basic geometry that is now being used in some applications. From this geometry, a modification in the size of rolling elements cells will be carried out to create an alternative model that could be tested from tribological point of view. Anyway, a rheological and mechanical analysis is before required to guarantee the feasibility of the component.

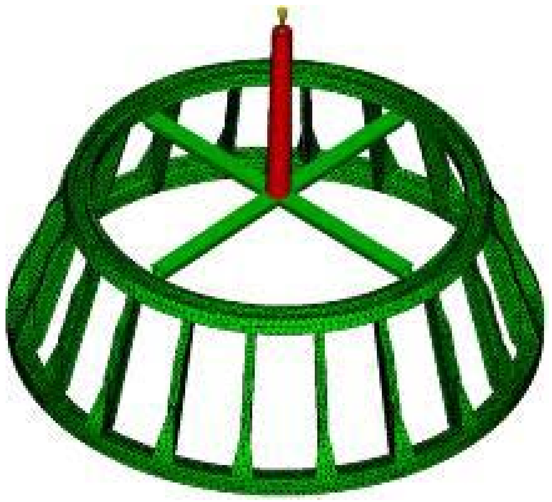


Figure 1: Basic model to analyze

Regarding to materials, different polymeric matrix have been analyzed. A first group of technical polymers including Polyamide 66 (PA66), Polyamide 46 (PA46), Polyacetal (POM) and Polytetrafluoethylene (PTFE) will be tested. On the other hand a set of high purpose polymers including Polieteretercetone (PEEK) and Polyeterimide (PEI) will be also tested in order to validate mechanical and rheological behavior.

A matrix of tests has been designed to compare rheological and mechanical behavior according to Table 1.

Table 1: Study cases

	Material	Geometry
M1	PA66	Basic
M2	PA46	Basic
M3	POM	Basic
M4	PEEK	Basic
M5	PTFE	Basic
M6	PEI	Basic
G1	PA66	Alternative
G2	PEEK	Alternative

For each of the study cases a rheological analysis evaluating their manufacturing feasibility by injection molding process will be carried out by means of software MOLDFLOW PLASTICS INSIGHT®. The methodology follows a flow+warp analysis including filling, packing and cooling stages of injection molding process as well a warpage calculation. Results related to both process feasibility such as melt front advancement, injection pressure and shear stress, and related to mechanical behavior such as deflection of the part will be discussed. These results are essential to determine the suitability of the material or design for bearing performance because the rolling elements must fit perfectly between the cage spaces to minimize friction between components.

3. RESULTS

3.1. Rheological analysis

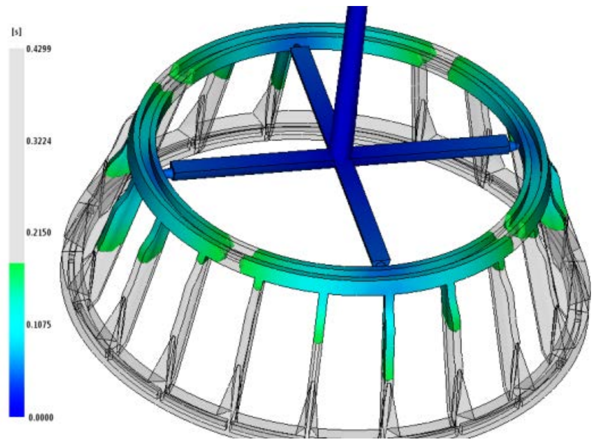


Figure 2: Melt front advancement at the beginning of injection time

Figure 2 shows melt front advancement result for M1 study case. It provides information about the proper flow pattern along part during injection time. This pattern is required to be both as uniform as possible and finishing filling at the same time all around the part contour. This fact can be seen on figure 3 for a total injection time of 0.5 seconds. It can be observed that flow near part entrance points goes further and this area of the part completes its filling slightly before. That is the reason why short filling times are considered in order not to let the first material freeze while the part completes the filling. Melt front advancement depends strongly on the injection point. As the feeding system remains the same for all the study cases this result is similar for all the materials and geometries tested.

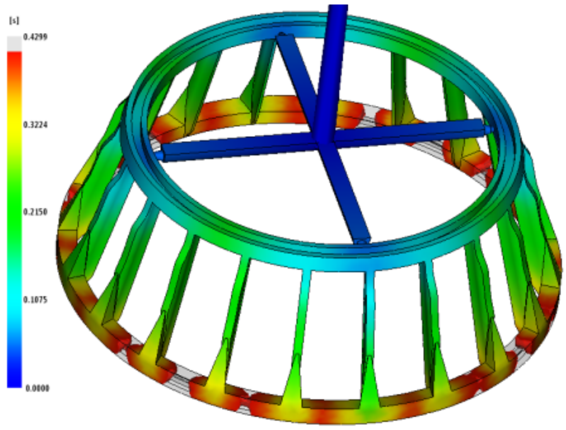


Figure 3: Melt front advancement at the end of injection time

One of the most important results to validate the feasibility of an injected part is maximum injection pressure reached at the end of filling stage when the whole cavity of the mold part is 99% completely filled. This pressure value has to do with part thickness, part flow length, material viscosity and flow rate. For this kind of geometry with short flow length values under 60 MPa are recommended. For the base cases analyzed M1 a maximum value of 25 MPa is obtained in figure 4. Pressure distribution at the end of filling is another parameter related to injection pressure to be taken into account. It should homogenous around the outer contour of the part, which indicates that the part has completed its filling at the same practically at the same time as described in Figure 4.

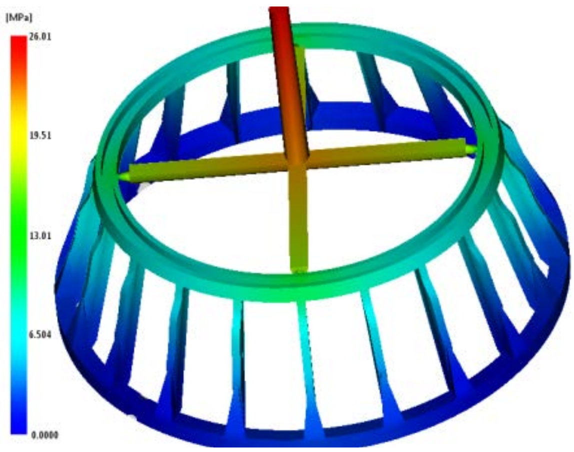
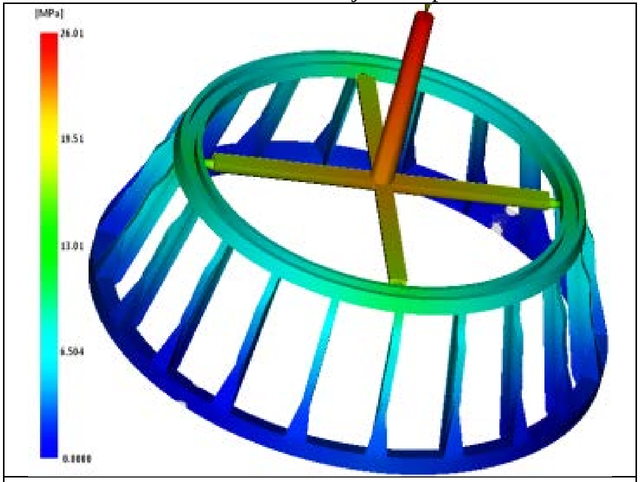
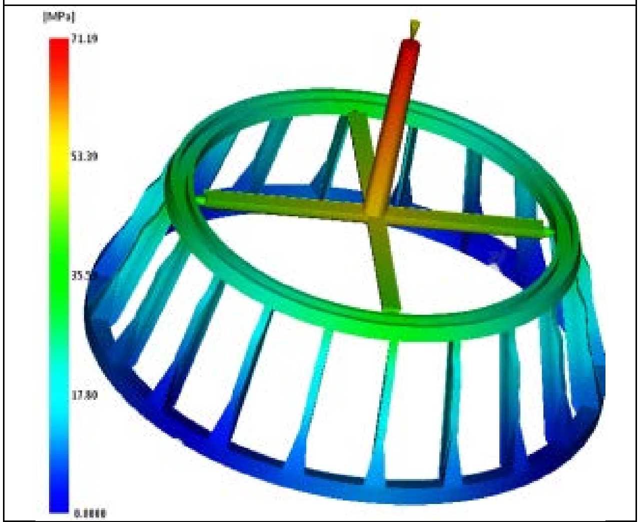
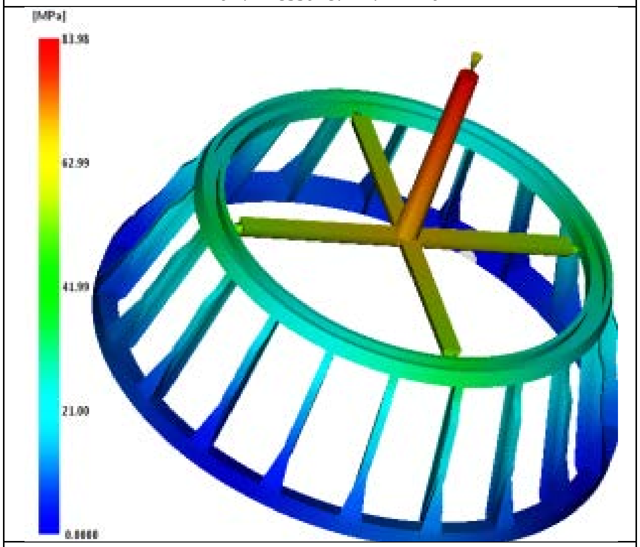
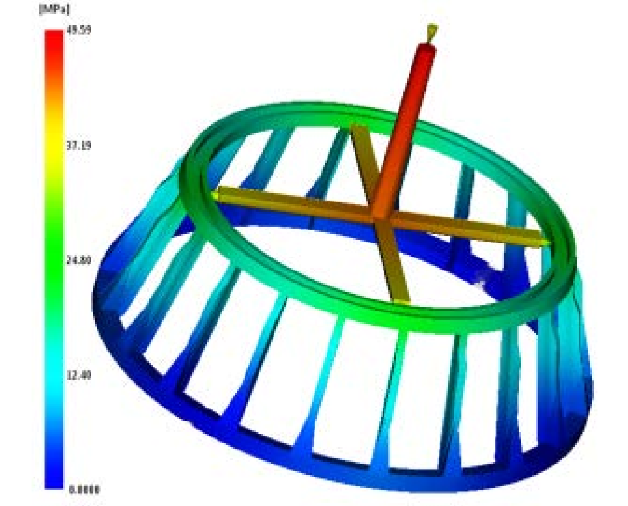


Figure 4: Maximum injection pressure and distribution for study case M1

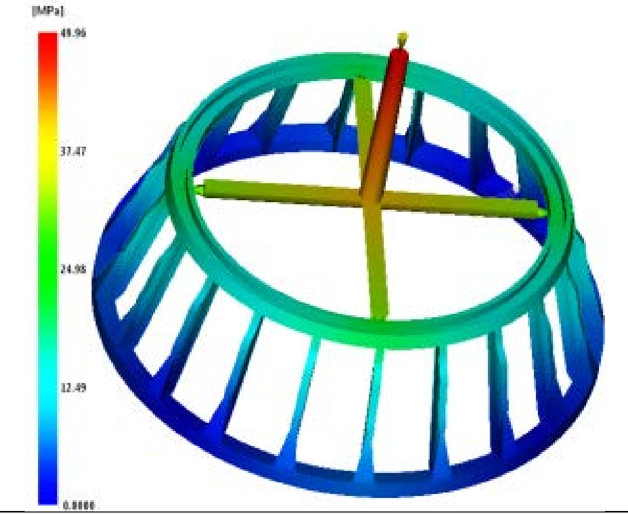
Table 2 shows pressure results of different material cases M1 to M6 analyzed. It can be observed that PEI, PTFE and POM require higher injection pressures above 60 MPa even over 100 MPa for PEI. It is mainly due to high viscosity values of theses material for the shear rates reached during injection molding. PEEK and PA46 reach medium values around 50 MPa that would be an adequate limit pressure for this kind of geometry, and finally PA66 reaches low pressure values under 30 MPa due to mainly its lower viscosity. Maximum injection pressures for alternative geometry are around 10% lower due to the higher thickness of the rolling elements cells walls that makes easier the flow.

Table 2: Maximum injection pressure

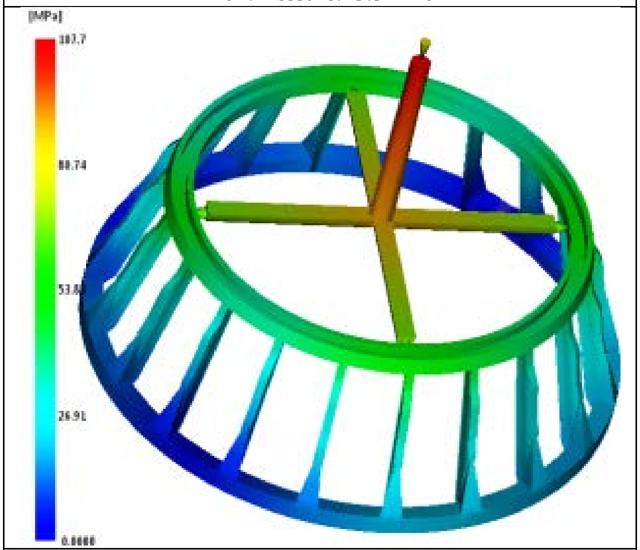

M1-(PA66) Max. Pressure: 26.8 MPa

M3-(POM) Max. Pressure: 71.1 MPa

M5-(PTFE) Max. Pressure: 83.9 MPa



M2-(PA46)
Max.Pressure: 49.5 MPa



M4-(PEEK)
Max.Pressure:49.9 MPa



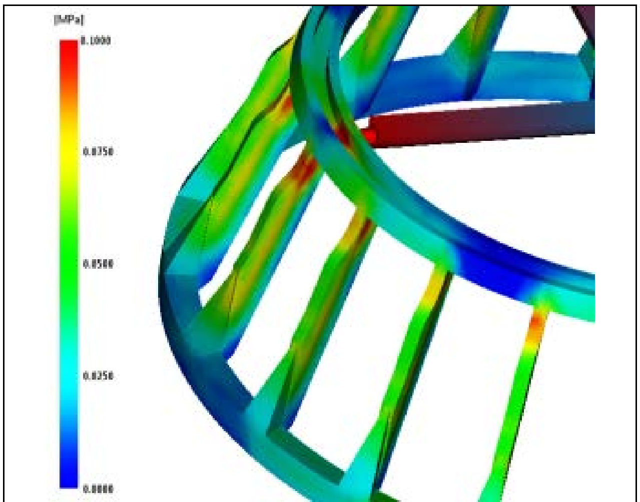
M6-(PEI)
Max. Pressure: 107.7 MPa

G2 (PEEK)
Max. Pressure: 37.9 MPa

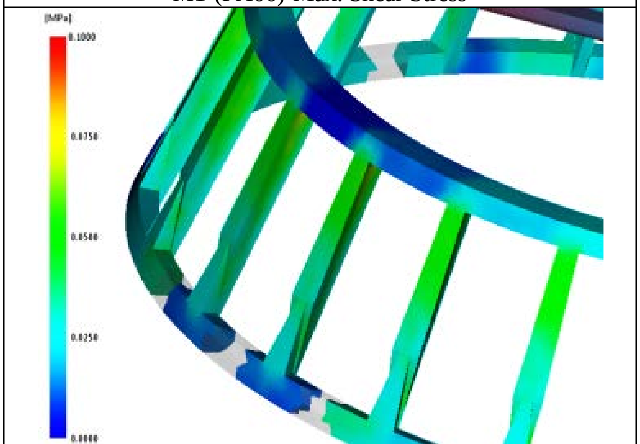
G1 (PA66)
Max. Pressure: 23.2 MPa

Shear stress is another result to be taken into account for this kind of components subjected during its working life to mechanical loads. Shear stress provides information about the residual stress that remains into the injected part due to manufacturing injection process. This residual stress will be added to the stresses generated by the external loads applied to the component. It is advisable to reduce to a minimum this stress, and, on the other hand to be confident about not to overpass maximum allowable shear stress values for different materials. Shear stress generated has to do mainly with maximum injection pressures reached and part thickness.

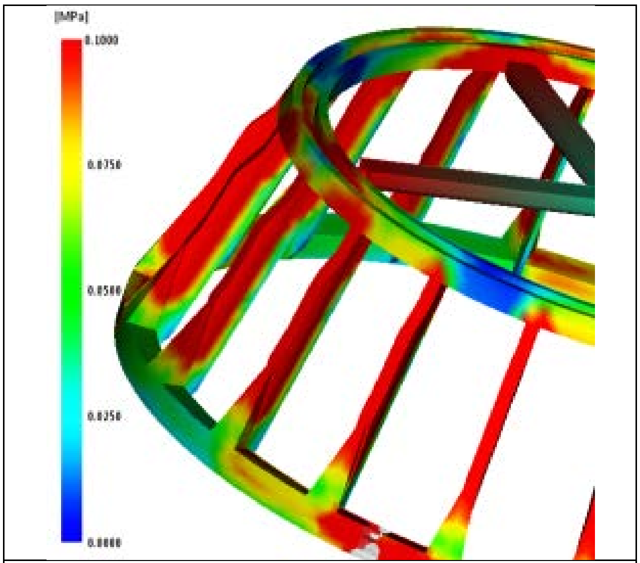
Figure 5 shows shear stress results for base geometry (M1, M4) and alternative geometry (G1, G2). All the results have been scaled to 0.1 MPa, therefore red areas indicate higher shear stress values. Study cases with base geometry (M1, M4) shows red areas with top stress values. These areas correspond to a change in the cross section thickness of the part, where it is more likely to appear shear stress problems. For alternative geometry (G1, G2), these areas have been softened with transition thickness areas to avoid abrupt changes in cross section of the part, which leads to lower shear stress values.



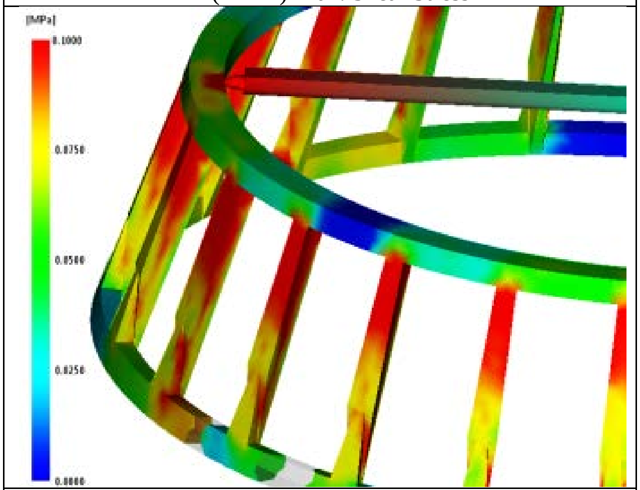
M1-(PA66)-Max. Shear Stress



G1-(PA66)-Max. Shear Stress



M4-(PEEK)-Max. Shear Stress



G2-(PEEK)-Max. Shear Stress

Figure 5: Shear stress for base geometry (M1, M4) and alternative geometry (G1, G2)

3.2. Mechanical analysis

Main purpose of the bearing cages is housing the rolling elements. To achieve an optimum performance of the bearing friction between rolling elements and the cage must be reduced at minimum. The way how injection process contributes to meet this requirement is finding the material, geometry and process conditions to obtain a post-molding deflection as low as possible to avoid undesirable additional friction between components.

Figure 6 shows qualitative results for cage deflection due to injection manufacturing process. From the graphical result can be observed that part tends to close radial and tangentially by reducing both outer and inner diameters and leaving less space for the rolling elements. Because of the same filling pattern obtained for all the study cases the deflection mechanism is the same for all the materials and geometries. In spite of it, different mechanical and shrinkage properties of the different materials maximum deflection values change according to values shown in table 3.

Higher deflections values are reached for PEEK and POM near 1 mm. Lower values are achieved for PTFE and PEI and medium values are obtained for PA66 and PA46. According to functionality tests, deflection values under 0.5 millimeters are allowed for the cage performance, so POM and PEEK would be discarded by this reason. Both PA66 and PA46 reach values next to the limit, however these values could be reduced by optimizing process parameters, especially at cooling stage.

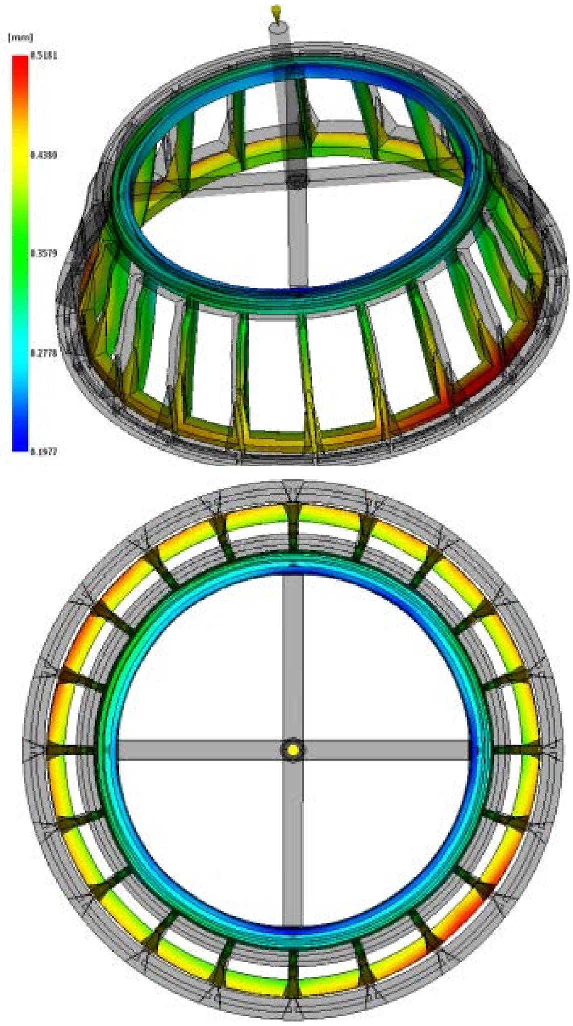


Figure 6: Qualitative post molding deflection for the analyzed study cases

For all different materials the deflection results are shown in Table 3:

Table 3: Results of maximum deflection

CASE	Max. Deflection (mm)
M1-(PA66)	0.5
M2-(PA46)	0.43
M3-(POM)	0.78
M4-(PEEK)	0.97
M5-(PTFE)	0.32
M6-(PEI)	0.35

4. CONCLUSIONS

A FEM analysis has been carried out to determine the most suitable material for a polymeric matrix in a composite focused to a bearing cage component from the point of view of process feasibility and induced mechanics induced by the injection process.

Melt front advancement is adequate for both proposed geometries with an uniform flow and an homogenous end of filling along the part contour. Maximum injection pressure is more adequate for PA46 and PEEK offering values under 60 MPa, and specially PA66 with values under 30 MPa. Regarding to mechanical analysis, better results are achieved by PTFE and PEI, followed by PA66 and PA46 that are also under performance specifications. It can be conclude that polyamide 66, mechanically improved with glass fiber, is the best option for polymeric matrix regarding injection molding process.

On the other hand, a new cage geometry has been tested. The change of geometry, especially cross section thickness, has a direct impact on maximum injection pressure and shear stress induced into the part. Results show that a new geometry with a slightly higher thickness at rolling element cells walls provides lower injection pressures and lower shear stress.

ACKNOWLEDGMENTS

This research has been carried out with the aid of project INNOVA-A1-008/15 granted by Aragon Regional Government in Spain.

REFERENCES

Ciarsolo I., Fernández X., Ruiz de Gopegui U., Zubizarreta C., Abad MD., Mariscal A., Caretti I., Jiménez I., Sánchez-López JC., 2014. Tribological comparison of different C-based coatings in lubricated and unlubricated conditions. Surface & Coatings Technology 257, 278–285.

Ettefaghi, 2013. Preparation and thermal properties of oil-based nanofluid from multi-walled carbon nanotubes and engine oil as nano-lubricant. International. Communications. in Heat and Mass Transfer 46, 142–147.

Fang N., Pugh D., Themudo R. 2007. On stress concentration factors of rolling elements bearing cages. Tribology Transactions 50(4), 445-452.

Forster NH, Rosado L., Brown JR., et al. 2002. The development of carbon-carbon composite cages for rolling element bearings. Tribology Trransactions 45 (1), 127-131.

Gunst U., Zabel W., Valle N. et al, 2010. Investigation of laminated fabric cages used in rolling bearings by ToF-SIMS. Tribology International 43 (5-6), 1005-1011.

Hwang, (2011). Effect of the size and morphology of particles dispersed in nano-oil on friction performance between rotating discs. Journal of Mechanical Science and Technology 25 (11), 2853–2857.

Kohar R., Hrcek S., 2016. Dynamic analysis of Rolling bearings with Elastic Cage. Modern Methods of Construction Design, Lecture notes in Mechanical Engineering 249-254.

Liu, 2011. Lubrication Effect of the Paraffin Oil Filled with Functionalized Multiwalled Carbon Nanotubes for Bismaleimide Resin. Tribology Letters 42, 59–65.

Manier CA, Ziegele H., Barriga J., Goikoetxea J., Woydt M, 2010. Zirconium-based coatings in highly stressed rolling contacts as alternative solution to DLC and ta-C coatings. Wear 269, 770–781.

SIMULATION MODEL OF PREFABRICATED HOUSE FROM MANUFACTURING TO ON-SITE INSTALLATION

Mohammed Sadiq Altaf[(a)], Mohamed Al-Hussein[(b)]

[(a)]Post-Doctoral Fellow, Civil and Environmental Engineering, University of Alberta
[(c)]Professor, PhD Student, Civil and Environmental Engineering, University of Alberta

[(a)]msaltaf@ualberta.ca, [(b)] malhussein@ualberta.ca

ABSTRACT

Prefabricated home building has been becoming a popular construction technique among home builders due to the superior quality, fast production, and design flexibility. Prefabricated home manufacturing is unique from traditional manufacturing industry as every house is somewhat unique due to the customization. This makes it difficult to predict the cycle time of the full production and installation process without a sophisticated model of the entire process. This paper developed a complete simulation model of the prefabricated residential home production process from manufacturing to on-site installation. The simulation model is then utilized for the purpose of calculating required amount of resources such as transportation trucks, cranes and direct labor. Based on the simulation results, different types of house can be scheduled in a way that will provide maximum utilization of different resources. A prefabricated home production facility located in Edmonton, Canada has been used as a case study to develop the simulation model.

Keywords: Prefabricated home production, simulation model

1. INTRODUCTION

Prefabricated building process starts with the virtual creation of the entire house in 3D building information modelling (BIM) software. From the 3D model, all the structural component of the house is verified and then all the materials are ordered and machine files are generated. A typical residential home has mainly three components – wall, floor and roof. All three of these components are manufactured in different areas based on the 3D drawing and then transported to the site for installation. However, the productivity of these three areas is not same and it is important to balance the production of wall, floor and roof components in order to have a smooth flow. After production, wall, floor and roof components/panels are transported and then installed on site using crane. The installation process follows the following sequence – main floor panels, main floor wall panels, 2nd floor panels, 2nd floor wall panels and finally roof panels. Back framing and roof shingling work follows after the completion of the panel installation.

The complete cycle time of a house production depends on the square footage of the house, number of wall panels, floor panels, roof pitch, roof type (gable/hip), number of trailer required, number of windows and doors, distance from manufacturing plant and so on. These different factors contribute towards the productivity of different components, travel time and site installation time. For this reason, it is difficult to predict the total duration of a project and necessary resources required for the job without a detailed model. In order to address this issue, a simulation model has been developed in this study starting from plant production, transportation to on-site installation. ACQBUILT Inc., an Edmonton based prefabricated home builder was partnered with University of Alberta to develop the simulation model of their home production process.

The simulation model has been utilized as a planning tool in various sectors of the construction and manufacturing industry. Its application can improve the understanding of a complex system and can be a useful decision support tool. Many researchers have used simulation models in construction and production planning in order to schedule activities, perform “what-if” analysis, allocate resource, and implement layout optimization. Halpin (1977) has introduced CYCLONE, a simulation environment that created the foundation for the progress of construction simulation. AbouRizk and Hajjar (1998) have proposed a framework for the application of simulation in construction, specifically focusing on construction practitioners. They have presented the concept of special-purpose simulation (SPS), which is a computer-based environment specially built for experts in the area, the advantage of this environment being that the user does not need to have knowledge of simulation. AbouRizk and Mohamed (2000) have introduced Simphony.NET, an integrated environment to model construction activities. This simulation software supports both DES and continuous simulation. It can provide different model outputs, such as standard statistical averages, resource utilization, standard deviation, minima and maxima, and charts such as histograms, cumulative density functions (CDFs), and time graphs. Al-bataineh et al. (2013) have presented a case study in which a simulation model for a tunneling project in Edmonton, Canada was developed in Simphony.NET as a decision support system for the project management team. Alvanchi et al. (2012) have developed a DES model of the steel girder bridge fabrication process for the purpose of providing a solution for the complex process of planning off-site girder bridge construction. Lui et al. (2015) have introduced a Special Purpose Simulation (SPS) template for the panelized construction process and linked the simulation model with building information modelling (BIM). Altaf et al. (2015a) has used the SPS template to build a simulation model of the prefabricated wall panel production process. Lu et al. (2008) have developed an automated resource-constrained critical path analysis using DES and particle swarm optimization (PSO). Based on their study, simulation modelling enables engineers to precisely examine different approaches in order to complete the project. Performing this type of analysis in advance yields reduced costs, improved quality, and improved productivity (AbouRizk 2010). Altaf et al. (2015b) has used simulation and RFID to develop on-line simulation model to get real time feedback from the simulation based on actual production performance. Simulation model have also been in used in prefabricated panelized home production process to resource allocation and optimization of the production schedule. (Altaf et al. 2014a and Altaf et al. 2014b). In the following sections of this paper, the productivity data collection system of wall, floor and roof production area has been presented, followed by simulation input modelling. Then simulation logic and the modeling environment are presented followed by the results and analysis.

2. SIMULATION MODEL

Figure 1 shows the development process of the simulation model. At first, productivity data has been collected from wall, floor and roof production areas as well as transportation time. Based on the collected data, simulation input modeling has been performed to develop mathematical model to estimate production time based on different parameters of the job such as number of wall panel, square feet of floor, roof pitch and type, distance from site, and so on. Then a discrete-event simulation model has been developed using symphony.NET general purpose modeling template. Different parameters of the job are stored in the database which is read by the simulation model to calculate production time, transportation time and number of required trailer per area. The model output will provide total cycle time to complete all the jobs, resource utilization, waiting and bottleneck of the system.

2.1. Process Mapping of the Production Process

The panelized production process starts with drafting the entire house in Building information modeling (BIM) software. Figure 2 shows a wood framed house model drawn in SEMA Software. This model contains all the detail information of the house such as, stud location, nailing pattern, sheathing location, mechanical/HVAC location, crane lift points and so on.

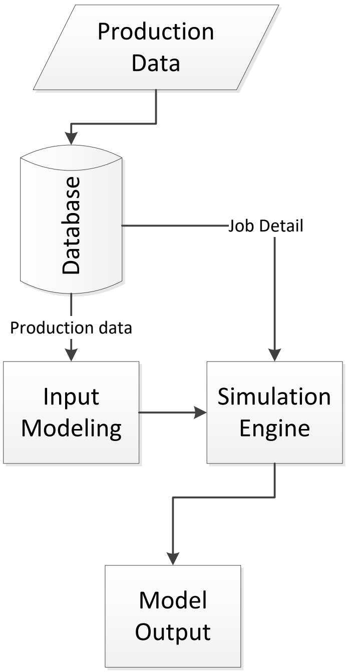


Figure 1: Simulation Model Development Process

After modeling the entire house, machine files are generated from the model, which is read by the CNC machines to pre-cut all the pieces and then build the wall and floor panels in the CNC workstation. Roof panel production is not automated; all the panels are manually built inside the plant in smaller sections and then shipped to the site for installation. After installing the wall, floor and roof panels, back-framing process (installation of the blocking) starts. Figure 3 shows the panelized home building process from manufacturing to on-site installation.

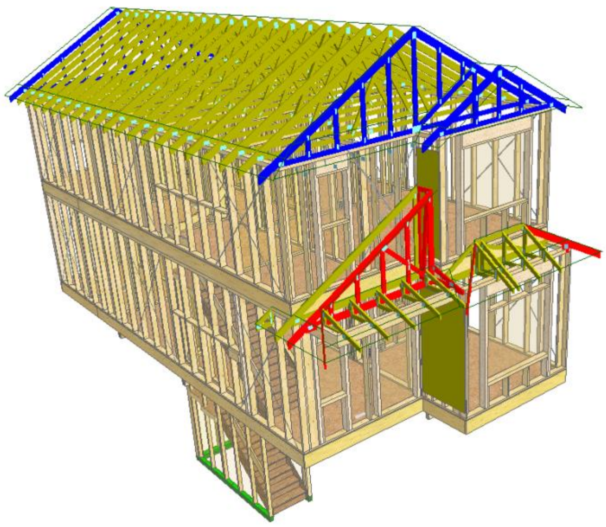


Figure 2: BIM Model drawn in SEMA Software

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

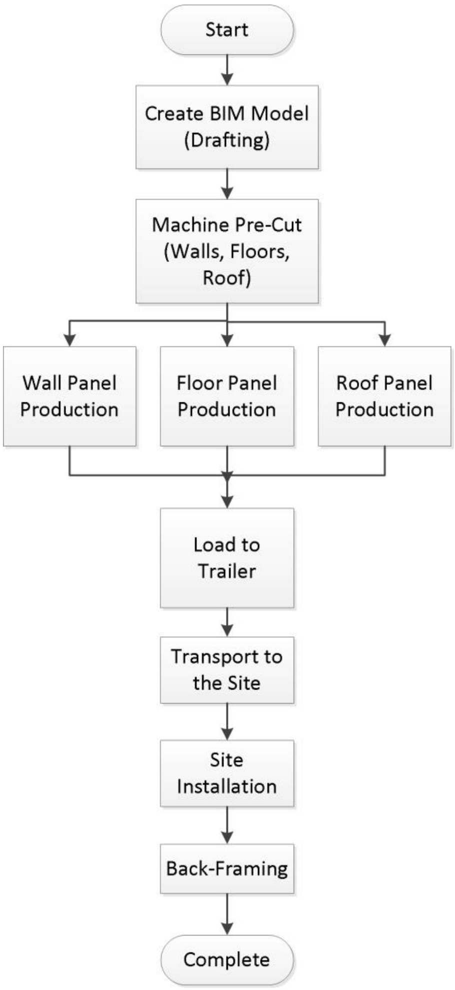


Figure 3: Panelized production process

2.2. Data Collection

Production data has been collected from wall, floor and roof area. The productivity of wall panel production area has been collected using RFID system which automatically collects panel processing time. The floor and roof panel processing time has been collected manually. This data is then used to develop production rate per square feet for each production area. All the transportation vehicles are equipped with GPS tracking device which allows collecting travel time from and to the installation site. Wall productivity is function of wall square footage, floor and productivity is function of floor square footage. The transportation cost is function of number of trailer and travel distance. The field installation time is function of number of panel and square footage of the house.

2.3. Modeling Environment

The simulation model is built in Simphony.NET, a simulation modeling environment, developed by University of Alberta. Simphony's general purpose modeling template has been used to develop the simulation model. The simulation model is linked to a database where all the productivity information as well as job information is stored. After creating the model entity, which represents a job, it read the database to collect number of trailer required for that job for each production area. Then it generates three model entities to represent wall, floor and roof production area. Based on the required number of trailer for each area, duplicate entities are created. Once each entity (now representing one trailer load) is loaded from the factory, transported to the site and lifted with the crane, all the entities marge together into one entity to represent completion of site installation. After that back-framing process starts and then completes the entire job. Figure 4 shows the developed model in Simphony.NET.

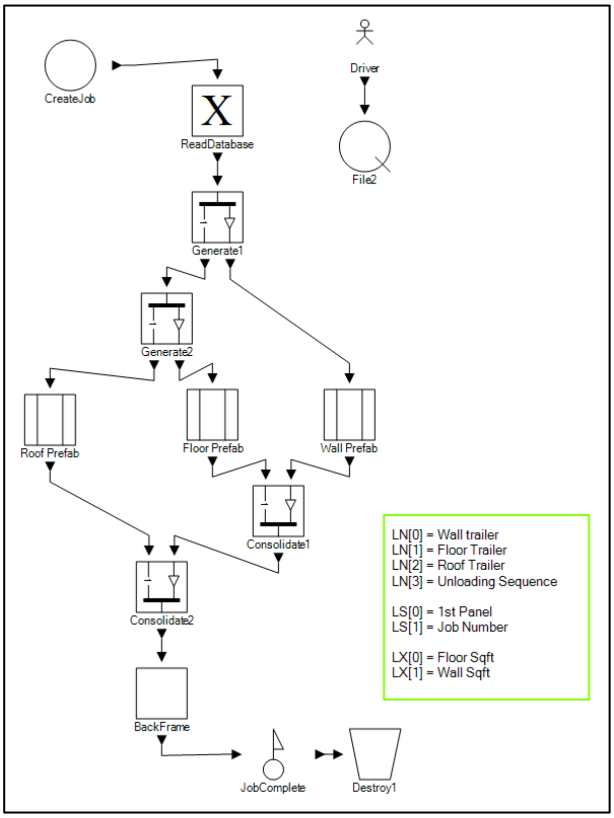


Figure 4: Simulation Model in Simphony.NET

Simphony's general purpose template contains several generic modeling elements which are used to create the model of the panelized production process. These are – *create*, *execute*, *generate*, *task*, *consolidate*, *counter*, *destroy*, *capture*, *resource*, *branch*, *set attribute*, *composite*, and *release* element. *Create* element starts the simulation process by creating entity that passes through different modeling elements. At first the entity goes through the *execute* element which contains the structure query language (SQL) to read the database to get job details. Then using *generate* element multiple entities are created to represent wall, floor and roof panels. In order to keep the model simple, *composite* elements are used to represent wall, floor and roof production process. Each *composite* element contains all the modeling elements required to complete the process. *WallPrefab* composite element is shown in detail in Figure 5 as an example.

At first the entity goes through a *task* element which contains the equation to calculate the panel processing

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

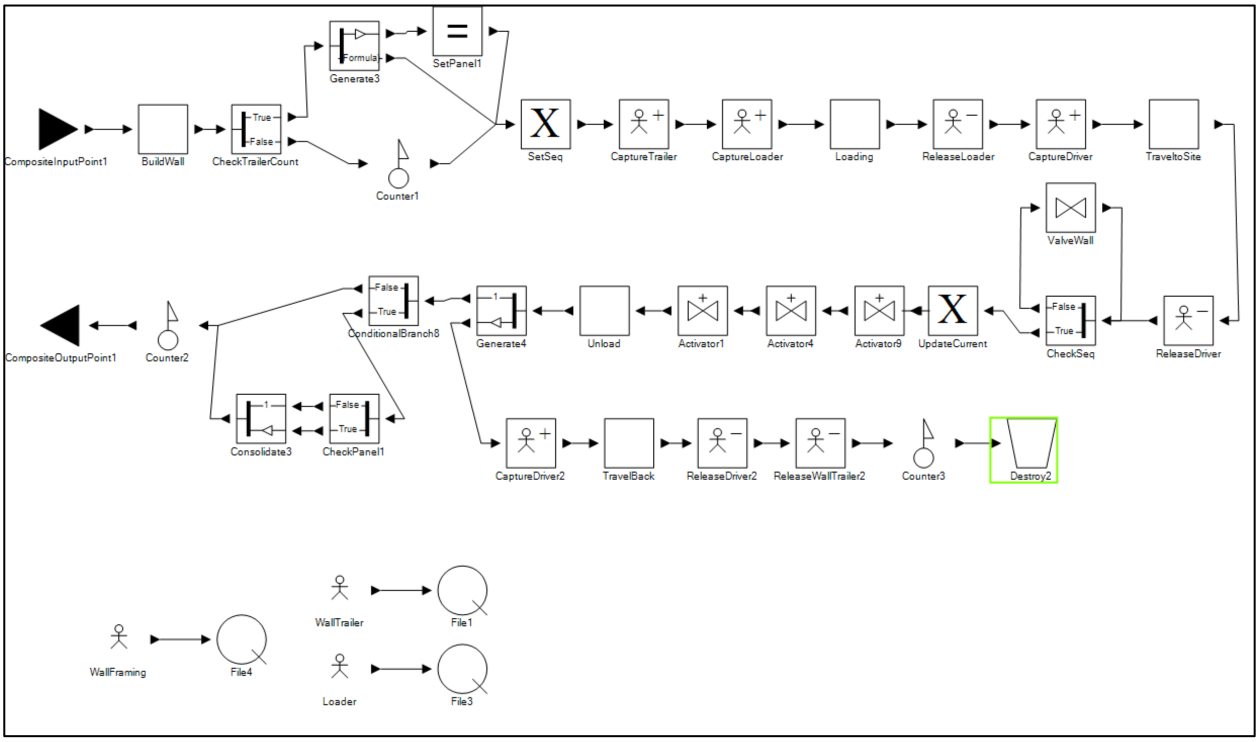


Figure 5: Wall Prefab *composite* element detail in Simphony model

time. The processing time equation is developed based on historical data collected. The model *entity* holds the required square footage information to calculate the processing time. Then another *generate* element is used to generate entities based on the number of trailer required to transport all the panels. This information is also read from the database. Trailer has been set as a *resource* in the model. Based on the availability, the model *entity* captures a trailer and start loading panels into the trailer. This process is simulated using another *task* element which has the loading time equation. In order to complete the loading process, a loader (resource) is required. Once the loading process is completed, the loader is released and any following job can capture the loader. After loading, the entity captures the truck driver (resource) and goes to another *task* element to complete the travel time. After that the drive resource is released and then the entity goes to a *branch* element to check if the current trailer is in order. This is required to simulate the unloading sequence at site. If 2nd floor panel trailer arrives at the site before main floor wall panel trailer, the 2nd floor trailer has to wait until the main floor wall trailer is unloaded and installed. Until the current sequence arrives, the trailer (model entity) will be blocked by a *valve*. All entities will be consolidated into one entity after the completion of the unloading process. Then the entity will go through the final *task* element which represents the back-framing operation. After that, a *counter* is used to collect the model data and the simulation process is terminated.

3. RESULTS AND DISCUSSION

The simulation model of the manufactured home from prefabrication to on site installation has been tested for different house models (attached, detached garage, duplex) and the results are compared with actual performance to validate the model. The simulation model provides the production manager and scheduler the opportunity to plan and schedule the jobs in a most optimize manner. The number of trailer and driver can be adjusted based on the demand which helps to minimize the operation cost. The output from the simulation model is the total time to complete multiple house models, utilization of truck driver and trailer, waiting time for trailer and driver. The simulation model was run for 2 attached garage house and 1 detached garage house model. The square footage of each house and required number of trailers for each work area are summarized in Table 1.

Table 1: House detail for Simulation Modeling

Type	SQFT		No. of Trailer		
	Floor	Wall	Wall	Floor	Roof
Attached Model #1	1656	4263	3	2	2
Attached Model #2	2104	4907	2	2	2
Detached Model #1	1611	3426	1	2	2

Total time required to complete all three houses will be 26 hours from plant production to in site installation and back framing. Numbers of resources inputted in the model are – 7 wall trailers and 10 flat trailers for floors and roof; and 3 truck drivers (Wall trailer is different

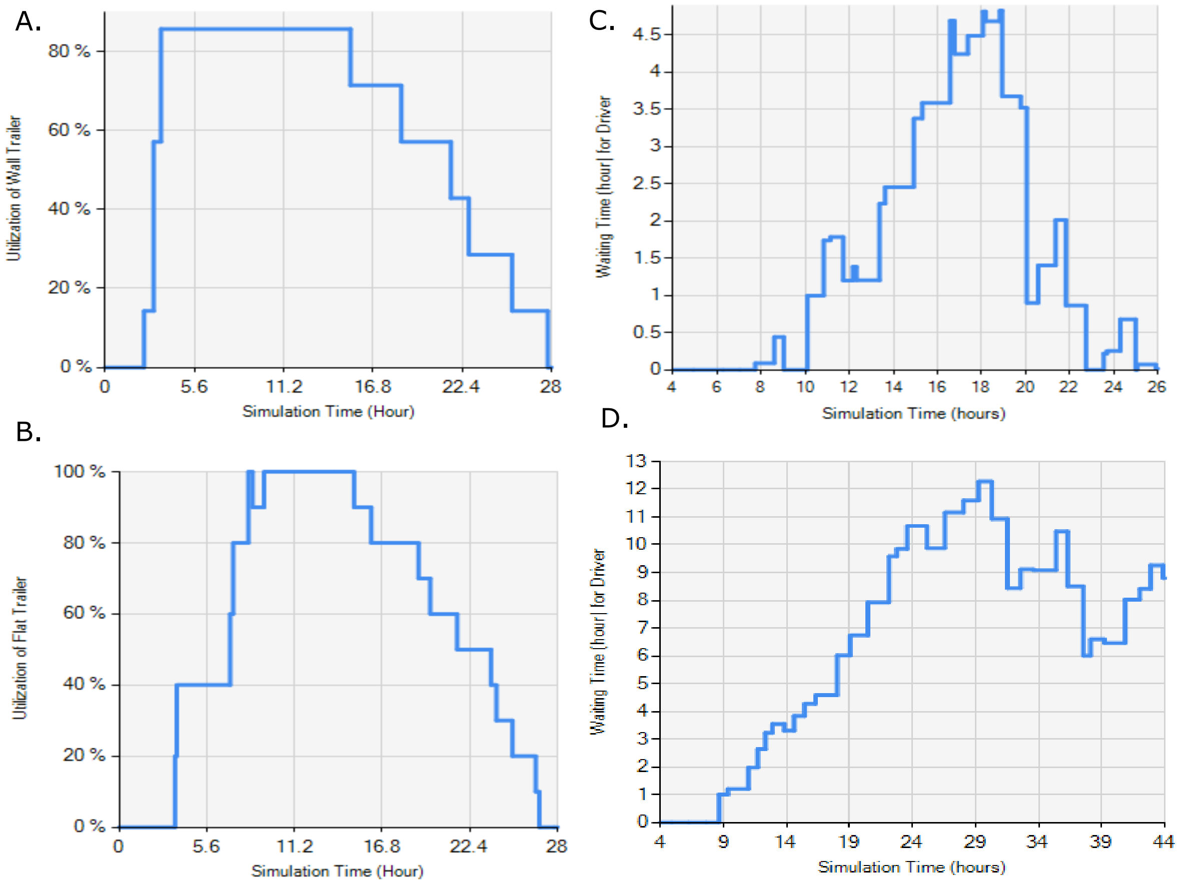


Figure 6: Utilization and waiting time chart from the simulation model – A. Utilization of wall trailer; B. Utilization of Flat trailer; C. Waiting time for driver when 2 driver available; D. Waiting time for driver when 1 driver is available

than floor and roof trailer). Figure 6A and 6B shows the utilization of the wall and flat trailers. The average utilization of the wall and flat trailers are 58% and 56% respectively. If the number of available wall and flat trailers are reduced to 4 and 6 the total time to complete all three houses is also 26 hours. However, the utilization of the wall and flat trailers are 74% and 70%. Similar analysis is done with the number of driver. The result shows that the optimum number of driver is 2 to produce these three homes. If the number of driver is reduced to one, the completion time increases to 37 hours. As shown in the Figure 6C and Figure 6D, the mean waiting time increases from 1.5 hour to 6.5 hour if the number of driver is reduced from 2 to 1. Because of the complex nature of the production process, there are different factors that contribute to the total production time. Even if some panels are waiting for the wall trailer does not always mean that it will delay the entire process as that trailer may need to wait in the site due to the unloading sequence. Thus, it becomes challenging to calculate the required number of resources for a group of house models. A simulation model of the entire process can be a useful tool to determining this critical information for the production planner and scheduler prior to the production date.

4. CONCLUSION

Off-site construction technique has become a popular construction method due to the associated improved quality, reduced waste and environmental impact. Among different off-site construction method, panelized system can accommodate various type of design. However, unlike the traditional manufacturing industry (e.g. car manufacturing), panelized production does not end at the plant; however, a significant amount of work needs to happen on site. Due to this fact, logistics plays an important role in the success of panelized home production. Co-ordination between plant and site activities, optimized use of trucks and trailers, scheduling of different type of jobs based on plant and site capacity – all these factors contributes to the success of the prefabricated home manufacturing process.

In this paper, a simulation model is developed from the prefabricated panel production process to on site installation using simphony.Net modeling environment. This simulation model can act as a decision support tool for the production coordinator by providing utilization of trucks, waiting time, total production time to produce a batch of houses in advance. Without the help of such model, it is challenging to optimize the entire process and obtain maximum plant and site capacity. In future, more elaborate simulation model can be developed for this purpose. This is study, a simplified simulation model is developed where the entire home production is divided into three main areas. All the sub processes in these three areas (wall, floor and roof) are ignored and the average productivity is used to estimate the production time. In future studies, all the individual workstation can be simulated which can be a part of the entire simulation mode.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC), as well as the assistance from ACQBUILT Inc. in Edmonton, Canada.

REFERENCES

Abourizk, S., 2010. Role of Simulation in Construction Engineering and Management. Journal of Construction Engineering and Management, (10), 1140–1153.

AbouRizk, S.M. & Hajjar, D., 1998. A framework for applying simulation in construction. Canadian Journal of Civil Engineering, 25(3), 604–617.

AbouRizk, S.M. & Mohamed, Y., 2000. Simphony- An integrated environment for construction simulation. In Proceedings of the 2000 Winter Simulation Conference.

Al-bataineh, M. et al., 2013. Using Simulation to Plan Tunnel Construction. Journal of Construction Engineering and Management, 139(5), 564–571.

Altaf, M., Al- Hussein, M., & Yu, H., 2014a. Wood-frame wall panel sequencing based on discrete-event simulation and particle swarm optimization. The 31st International Symposium on Automation and Robotics in Construction and Mining. Sydney, Australia.

Altaf, M., Liu, H., Al-Hussein, M., & AbouRizk, S.M., 2014b. Resource Planning Based On Discrete-Event Simulation For Panelized Construction. 14th International Conference on Construction Applications of Virtual Reality in Construction (ConVR2014), Sharjah, UAE, Nov. 16-18.

Altaf, M., Liu, H., Zhang, Y., Al-Hussein, M., and Bouferguene, A. 2015a. Discrete-event simulation modelling of prefabricated wall production line. Proceedings, 27th European Modeling & Simulation Symposium, Bergeggi, Italy, Sep. 21-23, pp. 234-239.

Altaf, M., Liu, H., Yu, H., and Al-Hussein, M. 2015b. Online simulation modeling of prefabricated wall panel production using RFID system. Proceedings, Winter Simulation Conference, Huntington Beach, CA, USA, Dec. 6-9, pp. 3379-3390.

Alvanchi, A. et al., 2012. Off-Site Construction Planning Using Discrete Event Simulation. Jounrnal of Architectural Enginnering, (6), 114–122.

Halpin, D.W., 1977. CYCLONE: method for modeling of job site processes. ASCE Journal of the Construction Division, 103(3), 489–499.

Liu, H., Altaf, M. S., Lei, Z., Lu, M., and Al-Hussein, M., 2015. Automated production planning in panelized construction enabled by integrating discrete-event simulation and BIM. Proceedings, International Construction Specialty Conference, Vancouver, BC, Canada, Jun. 7-10.

AUTHORS BIOGRAPHY

MOHAMMED SADIQ ALTAF is a Post-doctoral fellow in Civil and Environmental Engineering Department at University of Alberta. He completed his PhD in Construction Engineering Management. Currently, Mr. Altaf is appointed at ACQBUILT Inc., where he focuses on construction simulation, modular and panelized construction, production planning, and schedule optimization. His email address is msaltaf@ualberta.ca.

MOHAMED AL-HUSSEIN is a professor and Industrial Research Chair (IRC) in the Industrialization of Building Construction at the University of Alberta, and a highly sought researcher and consultant in the areas of lean manufacturing, construction process optimization, CO_2 emission quantification, and building information modelling (BIM), with the development of modular and offsite construction technologies and practices forming the hub of his research.. His email address is malhussein@ualberta.ca.

A HYBRID SIMULATION-BASED OPTIMIZATION APPROACH FOR SCHEDULING DINAMIC BLOCK ASSEMBLY IN SHIPBUILDING

Natalia P. Basán[(a)], Victoria G. Achkar[(b)], Carlos A. Méndez[(c)], Alejandro García-del-Valle[(d)]

[(a),(b),(c)] INTEC (UNL – CONICET), Güemes 3450, 3000 Santa Fe, Argentina
[(d)] University of A Coruña and UMI Navantia-UDC, C/ Mendizábal s/n, 15403 Ferrol, Spain

[(a)]nbasan@intec.unl.edu.ar, [(b)]vachkar@intec.unl.edu.ar, [(c)]cmendez@intec.unl.edu.ar, [(d)]alejandro.garcia.delvalle@udc.es

ABSTRACT
Shipbuilding is a complex and long-term process which requires the coordination of a large amount of resources and several types of manufacturing process planning. To achieve scheduling optimization of block production line and therefore, greater competitiveness in shipbuilding market, an efficient medium-term and short-term operations strategies are used. This work focuses on a real-world case study involving a large amount of blocks. Hence, a novel simulation-based optimization approach is developed for the efficient scheduling of production and assembly operations in a system of multi-stage multi-product production of a shipyard. The goal is to generate good schedules with modest computational effort minimizing the makespan (the total processing, waiting, and assembly time) while satisfying a large set of hard constraints. Mathematical model results are generated by using data from a real shipyard and improved through the simulation. Several examples are solved and reported to illustrate the capabilities of the approach proposed.

Keywords: discrete-event simulation, scheduling, shipbuilding, continuous time-slot, MILP model

1. INTRODUCTION

The management of workshops in a shipyard requires spatial and resource restrictions. Each ship has a high degree of customization and there are few units having a similar design. Shipbuilders have considered the assembly and production process as a bottleneck since every panel for every ship has to be processed through a set of workshops and unexpected uncertainties. Hence, the shipbuilding process involves a largest amount of work and many decisions.

Lee et al. (2009) point out that shipyards use a conveyer line production system with a combination of human and automatic equipment resources at the panel assembly workshop promoting the efficiency of conventional management and predictive planning. Nonetheless, this management has not allowed satisfactory results in terms of productivity. Therefore, in the last decades, a modular approach has been considered applying Lean principles and standardizing processes to improve the productivity of the shipyard shop (Cebral-Fernandez et al. 2016). A modular design has been developed allowing the pre-fabrication of steel blocks or structures, which are then assembled in the block erection process. According to this approach, the common unit of production for most steps of the process is a block or sub-block.

Furthermore, Cho et al. (1998) highlight that the block assembly process takes more than half of the total shipbuilding processes. Hence, it is essential to have a useful block assembly process planning system which allows building plans of maximum efficiency requiring minimum man-hours. Following this purpose, the main objective of the present work is to find an optimal or near optimal solution for the scheduling problem on a shipbuilding production and assembly process. This process involves multi-stage production system with variable processing times for each module and strong storage restrictions.

To achieve the aim, simulation and optimization tools have been evaluated to generate an optimal schedule. On the one hand, simulation technology has been used in shipbuilding problems in several works involving different aspects such as the spatial scheduling problem (Liu et al. 2011); data-based simulation model generation (Back et al. 2016); dynamic effects for design step (Park et al. 2016); and the assembly scheduling problem using heuristics (Lee et al. 2009, Chen et al. 2013).

On the other hand, mathematical optimization models and algorithms were applied to solve scheduling problems. Seo et al. (2007) and Kim et al. (2002) model the problem of the block assembly planning as a constraint satisfaction problem where the precedence relations between operations are considered constraints. To optimize the block spatial scheduling, Shang et al. (2013) present an allocation algorithm and mathematical model. Different methods and algorithms have been proposed recently to solve the scheduling problem in shipbuilding from different approaches, but they do not ensure an optimal solution of the scheduling problem. A research made by Xiong et al. (2015) consider a hybrid assembly-differentiation flowshop scheduling problem and introduce a mixed integer programming (MIP) model to present some properties of the optimal solution. This approach could be useful

to the shipbuilding issue due to it could also be considered an assembly flowshop scheduling problem.
In previous work, a mixed integer linear mathematical formulation (MILP) has been considered to efficiently solve the scheduling problem up to ten blocks (Basán et al. 2017). The mathematical models developed present an increased computational complexity associated to the big number of blocks and sub-blocks and the possible combinations which this type of real-world scheduling problem involves. Hence, a new hybrid simulation and optimization approach is proposed to solve the scheduling problem aiming at minimizing the makespan of blocks and sub-blocks in the yard. Several MILP formulations are used and their results are combined to become an input into a discrete-event simulation model. The MILP model is based on continuous time-slot concept. GAMS® is the software chosen for mathematical modeling and Simio® software is the one selected for discrete-event simulation.
This paper is organized as follows. In Section 2, the block assembly process with all stages is described. Then, Section 3 contains the proposed solution methodology. The mathematical model developed with assumptions and nomenclature is presented in Section 4, and the simulation model in Section 5. Next, in Section 6, computational results obtained of the hybrid simulation-based optimization approach are shown. Finally, conclusions are given in Section 7.

2. THE BLOCK ASSEMBLY PROCESS

The manufacturing process of shipbuilding is based on blocks production and assembly. Hence, this process begins with block division. A block is a basic component in the ship construction which varies in size, type, and consists of one or several sub-blocks assembled, depending on type of ship. A sub-block is composed of steel plates in accordance with the design drawing of the ship.
Therefore, the block division in shipbuilding process depends on the ship design. A ship is usually divided into many blocks of specific size and these, in turn, can be divided into same sub-blocks. Both blocks and sub-blocks are considered types of basic intermediate products in the modular design and construction. Figure 1 illustrates an example of the method of division into blocks and sub-blocks.

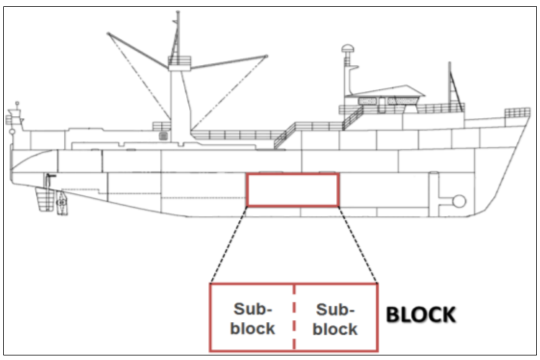


Figure 1: Method of division into blocks - Modular construction

Following is the steps involved in the block assembly process. As shown in the Figure 2, in the early stages of the shipbuilding process, sub-blocks are processed and assembled in specific workshops to form large blocks. Then, in the following stages, the blocks are assembled in a dock to form the hull of the ship.

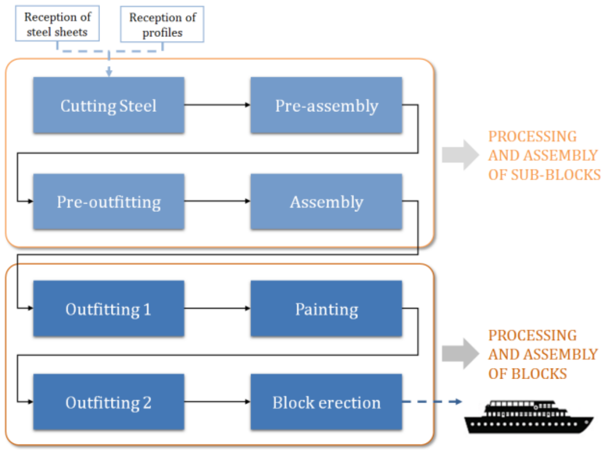


Figure 2: Shipbuilding process

The first stage of the block assembly process is *Cutting Steel*. This stage begins with reception of steel sheets and profiles. Then, these elements are cut into small parts and assembled to form small units according to the requirements of the sub-blocks designs. As result of welding and cutting processes panels and webs are obtained that will constitute the structural components of the blocks.
In the following step, *Pre-assembly*, welding operations are used to assembled the small structural elements and, therefore, formed the sub-blocks. Then, in the *Pre-outfitting* stage, different components as pipes, brackets and auxiliary elements, are installed in each sub-block. Hence, completed sub-blocks are obtained from this outfitting process.
The next step of the shipbuilding process is the assembly to sub-blocks (*Assembly*). According to the specifications of each block, the sub-blocks are positions together to carry out welding operations and form the blocks. Then, outfitting process is performed to install pipes, and electrical and lighting lines inside blocks (*Outfitting 1* stage).
Once the sub-blocks have been assembled and equipped, are blasted and painted in the painting booths. In this *Painting* stage the protection and design requirements of blocks are taken into account.
After painting process, all equipment that could be deteriorated in this process, such as wires and electronic components, is installed at the *Outfitting 2* stage. Therefore, a second outfitting process of blocks is performed before are moved to the dock.
The last stage is *Block erection*, where the prefabricated blocks are positioned in the dry dock to build the ship. The erection process consists of assembling the blocks, one after another, by a pre-established sequence, respecting the specifications of the ship. In accordance with given sequence, if a block arrives earlier, it must to

wait until its precedent is completed. There are different processing times in this step depending on the type of block being assembled (base block, lateral block or superior block). Figure 3 shows these different erection times between the lateral and superior blocks.

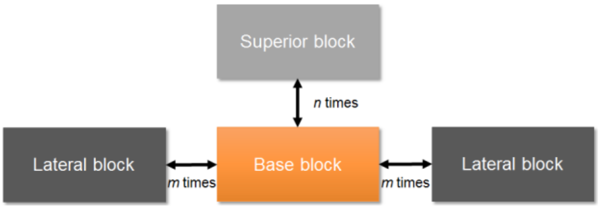


Figure 3: Block erection processing times

A possible workshop configuration of a shipyard is presented in Table 1. Note that capacity and processed product are described in each stage detailed above. This configuration belongs to the case study addressed in this work and is based on a real world problem. The data shown below refer to a simplified model due to confidentiality reasons. Hence, the real configuration of the block assembly process is not explicitly mentioned.

Table 1: Stages

Stage	Capacity	Entity to process
Cutting Steel	2	Sub-blocks
Pre-Assembly	6	Sub-blocks
Pre-Outfitting	3	Sub-blocks
Sub-blocks Assembly	6	Sub-blocks
Blocks Outfitting 1	3	Blocks
Painting	2	Blocks
Blocks Outfitting 2	3	Blocks
Erection	1	Blocks in a defined order

3. PROPOSED SOLUTION METHODOLOGY

This work introduces an algorithm to generate a complete schedule of shipbuilding process incorporating a mixed-integer linear programming continuous time-slot formulation and a discrete-event simulation model. The MILP model developed uses sequencing variables for the processing and assembling tasks keeping the complexity at a manageable level.
Moreover, an efficient discrete-event simulation framework is developed to represent the assembly operations in a system of multi-stage production of ships in a shipyard. The major advantage of this computer-aided methodology is the possibility of reproducing highly complex manufacturing process in an abstract and integrated form, visualizing the dynamic behavior of its constitutive elements over time (Banks et al. 2005). Figure 4 presents a brief description of how the GAMS® model and Simio® model are associated to obtain a final result.

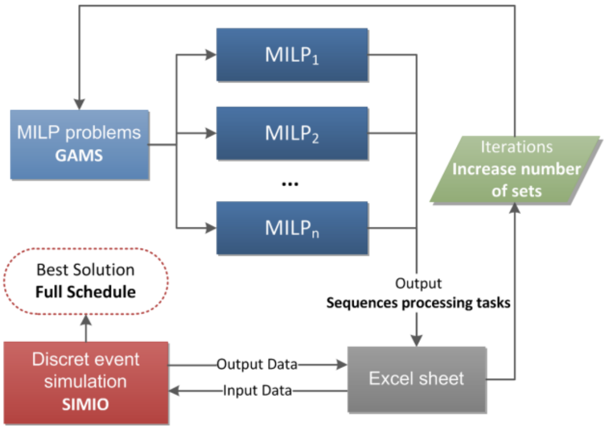


Figure 4: Steps of the solution methodology

Note that mathematical models find difficulties to converge into a solution in this kind of problems including a large amount of blocks and sub-blocks. The reason is that possible combinations increase with every additional block. Hence, several iterations are carried out to allow the solution of this large-scale problem in reasonable time. Each iteration represents a constrained version of the global model given in the following section. As a result, the MILP model generated achieves an effective solution for the whole shipyard scheduling problem. It also becomes useful for making and testing alternative decisions to enhance the current process performance. Castro et al. (2011), uses a similar hybrid simulation optimization approach to address a similar scheduling problem.
On the other hand, predefined job-sequences obtained from the set of optimization-based formulations combined are used to create an input to the simulation model. In other words, these sequences are generated by MILP model and written in Excel as input data to the Simio® model.
The simulation model structure allows easily, by using tables, change arrival sequences of sub-blocks to the system according to the output of mathematical models defined. Multiple runs are performed and written in Excel to generate statistical data as output data.
Therefore several scenarios are defined changing the quantity of blocks to enter as an input into the mathematical model, simulate the chained sequence results and compare computational effort and solutions. The aim is to analyze the impact that industrially size problems has on solutions obtained from the interaction of a mathematical and simulation models.

4. MATHEMATICAL MODEL

In order to develop an efficient mathematical formulation all production and assembly operations and the described characteristics of the shipbuilding problem are taken into account. The block assembly process requires coordination of many different resources. Hence, it is a complicated and long-term process.
Following is mathematical model based on the continuous time-slot batches concept to optimize the

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

processing sequence of the blocks at each stage and, therefore, minimizing the total processing time. The shipyard is considered a multi-stage and multi-product plant. The nomenclature used in the model and all constraints involved, including the objective function, are detailed in this section.

Considerer the shipbuilding system in Figure 2, where a set J of sub-blocks must be processed and assembled to form then a set I of blocks. These blocks must also be processed and then assembled in the dock. A total of S stages are considered with K_s identical parallel units. Let K denote the number of units (or workshop) in the shipyard.

Note that there are two types of products in the shipyard: sub-blocks $(j = 1,2, \dots, m)$ and blocks $(i = 1, 2, \dots, n)$, assuming that each block is made up of two known sub-blocks. Although each product has its own requirements, follows the same sequence $1, 2, \dots, c$ of processing stages, where $s \in S$.

Moreover, each workshop has capacity to process one block (or sub-block) at a time and, likewise, more than one workshop cannot process a single block in each stage. In addition, each workshop servers as intermediate storage if processing finished and the next step is not yet available.

Processing times of each block are known a priori ($TP_{j,s}$ and $TPb_{i,s}$), and transfer times between the units are considered negligible.

The assembly sequence on slipway (*Erection* stage) is known a priori. In addition, the output order of finished blocks of *Outfitting 2* stage is the same order in which they will be assembled in the last stage of shipbuilding process. Hence, the MILP model proposed determines the production schedule until the *Outfitting 2* stage.

Sets

I	set of blocks (index i, $i = 1,2, \dots, n$)
J	set of sub-blocks (index j, $j = 1,2, \dots, m$)
S	set of stages (index s,$s = 1,2, \dots, c$)
K	set of machines (index k, $k = 1,2, \dots, q$)
P	set of slots (index p, $p = 1,2, \dots, m$)
J_i	set of sub-blocks of each block i
K_s	set of parallel machines in stage s
I_s	set of blocks that can be processed in stage s
J_s	set of sub-blocks that can be processed in stage s

Parameters

$TP_{j,s}$	processing time of sub-block i at stage s
$TPb_{i,s}$	processing time of block i at stage s
mc_s	parallel units in stage s
M	big constant in big-M constraints

Continuous variables

$Ti_{j,s}$	initial processing time of sub-block j in stage s
$Tf_{j,s}$	final processing time of sub-block j in stage s
$Tbi_{i,s}$	initial processing time of block i in stage s
$Tfb_{i,s}$	final processing time of block i in stage s
$TSi_{p,k}$	initial processing time of slot p in machine k
$TSf_{p,k}$	final processing time of slot p in machine k
mk	makespan

Binary variables

$x_{j,p,k,s}$ 1, indicates whether sub-block j is processed in position p of machine k of stage s

$y_{i,p,k,s}$ 1, indicates whether block i is processed in position p of machine k of stage s

Constraints

The block assembly system in a shipyard involves several types of constraints such as resource, allocation, sequencing, and timing constraints. Therefore, all these must be taken into account in the mathematical model to determine the optimal production scheduling.

Firstly, the objective function of the MILP model is defined in the equation (1): makespan minimization.

$$\min \quad mk \tag{1}$$

Due to the early stages of the shipyard process sub-blocks and the last ones process blocks, two binary variables were defined to formulate allocation constraints: $x_{j,p,k,s}$ and $y_{i,p,k,s}$. The first of these is used to determine which unit processes each sub-block in the first three stages. And the last of these defines which unit is used to process and assembly each block of ship. Hence, equations (2)-(7) use these binary variables to introduce the allocation constraints. Equations (2) and (3) are constraints assigning sub-blocks and blocks to units of each stage of shipbuilding, where each product (sub-block and block) must only be processed in one workshop of each stage. And equations (4) and (5) are constraints assigning slots to sub-blocks or blocks in each unit of step s, i.e. these equations assign only one sub-block (or block) in each slot of each workshop.

$$\sum_{p=1}^{N} \sum_{k=1,k\in K_s}^{mc_s} x_{j,p,k,s} = 1 \quad \forall j \in J, s \in S, s \leq 3 \tag{2}$$

$$\sum_{p=1}^{N} \sum_{k=1,k\in K_s}^{mc_s} y_{i,p,k,s} = 1 \quad \forall i \in I, s \in S, s > 3 \tag{3}$$

$$\sum_{j=1}^{M} x_{j,p,k,s} \leq 1 \quad \forall p \in P, k \in K_s, s \in S, s \leq 3 \tag{4}$$

$$\sum_{i=1}^{N} y_{i,p,k,s} \leq 1 \quad \forall p \in P, k \in K_s, s \in S, s > 3 \tag{5}$$

$$\sum_{j=1}^{M} x_{j,(p+1),k,s} \leq \sum_{j'=1}^{M} x_{j',p,k,s} \qquad \forall p \in P, k \in K_s, s \in S, s \leq 3, j \neq j' \tag{6}$$

$$\sum_{i=1}^{N} y_{i,(p+1),k,s} \leq \sum_{i'=1}^{N} y_{i',p,k,s}$$
$$\forall p \in P, k \in K_s, s \in S, s > 3, i \neq i' \quad (7)$$

Equations (8)-(12) introduces the sequencing constraints, where equations (8)-(10) state the processing order of products at each stage identifying those that process sub-blocks, blocks or both products. Moreover, slots must also be sequenced in each unit (eq. 11).

$$Tf_{j,s} \leq Ti_{j,(s+1)} \quad \forall j \in J, s \in S, s < 3 \quad (8)$$

$$Tbf_{i,s} \leq Tbi_{i,(s+1)} \quad \forall i \in I, s \in S, s > 3 \quad (9)$$

$$Tf_{j,s} \leq Tbi_{i,(s+1)} \quad \forall i \in I, j \in J_i, s \in S, s = 3 \quad (10)$$

$$TSf_{p,k} \leq TSi_{(p+1),k} \quad \forall k \in K, p \in P \quad (11)$$

In the last stage of shipbuilding process (*Block erection*) a predefined block assembly sequence must be satisfied according to specification of the Figure 3. Therefore, Equation (12) is introduced to fulfill this given sequence.

$$Tbf_{i,s} \leq Tbf_{i+1,s} \quad \forall i \in I, s \in S, s = |S|, i < |I| \quad (12)$$

The duration of product i or j in stage s must be equal to initial processing time plus the processing time at that stage (eq. 13-14). Similarly, the sum of start processing time of slot p in step s and processing time of product assigned to that slot must be equal to final processing time of the slot (eq. 15-16).

$$Tf_{j,s} = Ti_{j,s} + \sum_{p}^{N} \sum_{k \in K_s} x_{j,p,k,s} * TP_{j,s}$$
$$\forall j \in J, s \in S, s \leq 3 \quad (13)$$

$$Tbf_{i,s} = Tbi_{i,s} + \sum_{p}^{N} \sum_{k \in K_s} y_{i,p,k,s} * TPb_{i,s}$$
$$\forall i \in I, s \in S, s > 3 \quad (14)$$

$$TSf_{p,k} = TSi_{p,k} + \sum_{j}^{M} x_{j,p,k,s} * TP_{j,s}$$
$$\forall p \in P, k \in K_s, s \in S, s \leq 3 \quad (15)$$

$$TSf_{p,k} = TSi_{p,k} + \sum_{i}^{N} y_{i,p,k,s} * TPb_{i,s}$$
$$\forall p \in P, k \in K_s, s \in S, s > 3 \quad (16)$$

The constant M is used in equations (17)-(20) to model the relationship between slots and blocks (or sub-blocks, as appropriate). If a sub-block (or block) is processed in position p of the of machine k of stage s (i.e. $x_{j,p,k,s} = 1$ or $y_{i,p,k,s} = 1$) then the start time of the slot p must match with the start processing of the sub-block (or block).

$$-M(1 - x_{j,p,k,s}) \leq Ti_{j,s} - TSi_{p,k}$$
$$\forall i, j \in J, p \in P, k \in K_s, s \in S, s \leq 3 \quad (17)$$

$$M(1 - x_{j,p,k,s}) \geq Ti_{j,s} - TSi_{p,k}$$
$$\forall i, j \in J, p \in P, k \in K_s, s \in S, s \leq 3 \quad (18)$$

$$-M(1 - y_{i,p,k,s}) \leq Tbi_{i,s} - TSi_{p,k}$$
$$\forall i, i \in I, p \in P, k \in K_s, s \in S, s > 3 \quad (19)$$

$$M(1 - y_{i,p,k,s}) \geq Tbi_{i,s} - TSi_{p,k}$$
$$\forall i, i \in I, p \in P, k \in K_s, s \in S, s > 3 \quad (20)$$

Due to the makespan represents the total processing and assembling time required for the construction of a ship and the shipyard has a sequential processing, it could be calculated considering the longest final processing time of last stage of shipbuilding process.

$$mk \geq Tfb_{i,s} \quad \forall i \in I, s \in S, s > 3 \quad (21)$$

5. DISCRETE-EVENT SIMULATION MODEL

Simulation technology is a type of shipbuilding product lifecycle management solution used to support production planning or decision-making (Back et al. 2016). Banks et al. (2005) point out that a simulation model can be used to investigate a wide variety of "what if" questions about the real-world system. Potential changes to the system can first be simulated, in order to predict their impact on system performance. Thus, from the simulation, data are collected as if a real system were being observed. This simulation-generated data is used to estimate the performance variables of the system. In this work, the simulation model is used to combine the results of the mathematical models and obtain the expected makespan including stochastic variables. The groups of blocks are separately optimized to minimize their local makespan. The outputs of the mathematical models are optimal sequences of groups of blocks. By introducing these sets of sequences into the simulation model we can chain them to obtain a real time in the global production process. Therefore, it includes de *Erection* stage, where blocks are assembled following a defined order. This stage will also affect the global makespan.

In conclusion, simulating the outputs of the mathematical model allows adding more detail, obtaining more reliable results. Hussein et al. (2009) add that there are some cases where the results of a simulation are a confirmation of expectations, but the true benefit is the discovery of the unexpected situation or circumstance. The simulation could be useful to find if the *Erection* process affects optimal sequences and if different scenarios really make an impact on global makespan.

The chosen simulation framework for this study is Simio® software. Figure 5 presents a global view of the model where most modules represent processing stages.

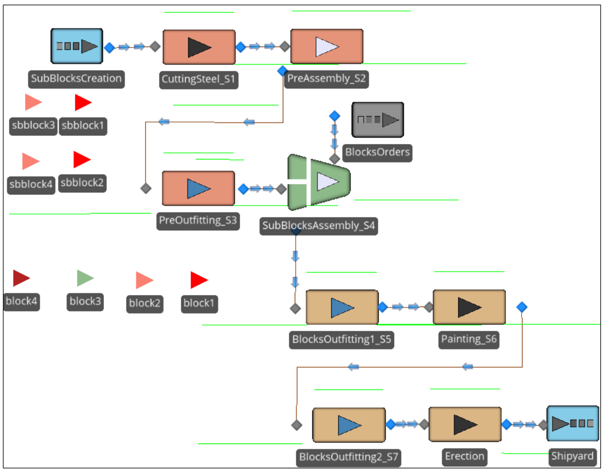


Figure 5: Simulation Model in Simio® 2D

Storage restrictions are modeled using internal logic processes. These processes are sequences of steps with logical actions like assigning values to variables, using conditions to make decisions, waiting until an evet occurs, reserving or unreserving resources, etc. This tool allows to include more customization on system.
Figure 6 presents internal logic processes associated to the Painting stage. These processes principally reserve and unreserve resources to avoid being occupied when an entity is waiting to enter the next stage. They also write on worksheet output values to posterior analysis. The stages that do not accomplish this resctriction are Sub-blocks Assembly and Erectionones.

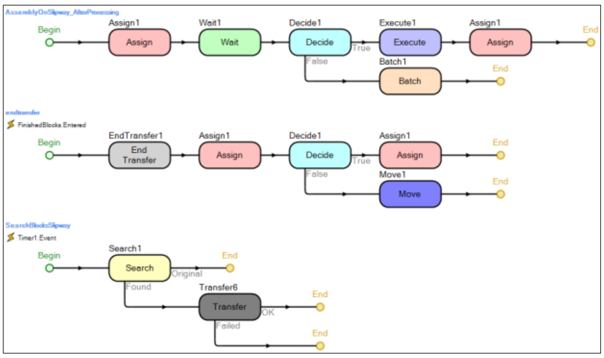
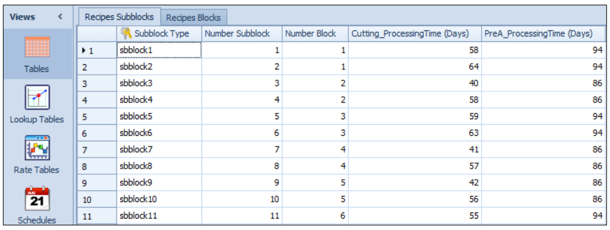


Figure 6: Internal logic processes for Painting stage

Processing times vary depending on the block or sub-block and the stage. In the simulation model, tables are defined for these two types of entities determining processing times on each stage. Simio allows making data tables to hold model data and then they can be referenced by individual entities. The data can be any of the property types provided by software including expressions, object references, class types, etc. Hence, each entity in the model can reference a specific row of data in the table containing, for example, processing times. Data import and export can be used for both Data Tables and Sequence Tables. Figure 7 is an example of the table for sub-blocks. However, several stages present a stochastic behavior with probability distributions, principally normal and discrete ones. Distribution probabilities with their parameter values can be directly entered into processing time properties in module characteristics.


Views
Tables
Lookup Tables
Rate Tables
Schedules
Recipes Subblocks | Recipes Blocks

	Subblock Type	Number Subblock	Number Block	Cutting_ProcessingTime (Days)	PreA_ProcessingTime (Days)
1	sbblock1	1	1	58	94
2	sbblock2	2	1	64	94
3	sbblock3	3	2	40	86
4	sbblock4	4	2	58	86
5	sbblock5	5	3	59	94
6	sbblock6	6	3	63	94
7	sbblock7	7	4	41	86
8	sbblock8	8	4	57	86
9	sbblock9	9	5	42	86
10	sbblock10	10	5	56	86
11	sbblock11	11	6	55	94

Figure 7: Sub-block recipes and processing times table

Once the simulation model is finished, verification is carried out. Verification is concerned with determining if the conceptual model with its specifications and assumptions were correctly traduced in computerized representation (Law 2007). To verify the simulation model there is an iterative comparison between outputs of the GAMS® model and the Simio® model. The Gantt chart obtained as a result from the mathematical model is analyzed in different points of the timeline. Each point is also analyzed in the simulation model looking if the same activities are being performed, initiating or finalizing in all stages. We obtained satisfactory conclusions.
Therefore, the model must be validated. Validation is concerned with determining how closely the simulation model represents the real system (Law 2007). To attain this aim, several comparisons are made with information given from the shipyard, related to stages characteristic such as capacity, inventory policies, processing times. All aspects were discussed with experienced staff and historical information and necessary adjustments were made to achieve the desired values.

6. SCENARIOS AND RESULTS

Once mathematical and simulation models are developed, verified and validated, experimentation is performed. The first scenario considered is the one that has a set of 1 block (2 sub-blocks) for each MILP problem. Hence, there is no global optimization and blocks are processed in order from 1 to *N*. Next, the following scenario consists in adding 2 or 3 blocks to each group (following the conventional order) and running all MILPs generated. Following this logic, the sets of groups continue increasing and the number of MILP problems decreasing up to finding a set that cannot be solved in a reasonable computer time. Computational efforts are compared against solutions obtained. Table 2 summarizes results obtained from experimentation. It contains (i) the number of scenario, (ii) the quantity of sub-blocks per group; (iii) the quantity of MILP models to run on each scenario; (iv) the total computational time in days (*CT*); (v) the makespan as a sum of MILP model outputs (MILP MK); and finally, (vi) the expected makespan obtained from simulation runs (EMK). The last column involves stochastic processing times and 10 replications per scenario.

Table 2: Results

Scenario	Group size	MILP models	CT	MILP MK	EMK
1	-	-	-	-	3427.9
2	6	11	0.02	5561	3391.4
3	8	8	0.11	4802	3385.4
4	10	7	0.66	4642	3405.6
5	14	5	1.21	4194	3412.8
6	16	4	8.17	3940	3413.0
7	22	3	75.44	3855	3426.4

MILP MK is calculated only to have a reference value. As you can observe in Table 2, this makespan appears to be improving as the group size increases. A possible cause is that there are not determined initial conditions on each MILP model considering the previous set of sub-blocks processed previously (all resources are idle) and processing times used are deterministic.
The best scenarios according to the expected makespan are the second and third ones, which involve optimization in small groups. The worst results includes the conventional order, and those scenarios having bigger size of sets. Nonetheless, variations in makespan values are small, and this results could vary due to the stochastic nature of the problem. Figure 8 presents both makespan results on the left axis values and the computational time involved on the right axis.

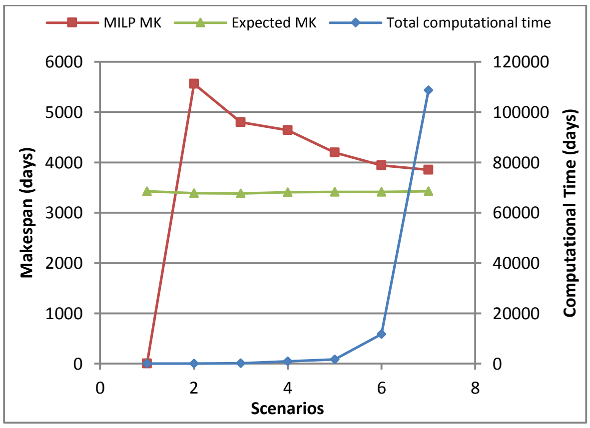


Figure 8: Makespan vs Computational Time

On the one hand, when simulating chained sequences resulting from MILPS per scenario, including stochastic processing times and block-erection last stage, the makespan obtained is more reliable than the simple sum of MILP model outputs. Expected makespans have a maximum variation of days of 2 %. Thus, scenarios do not provide siginificant differences between them. Nonetheless, the computational time has an important increasment rate on each scenario. In conclusion, when increasing group sizes, computational times strongly increases but solutions do not proportionally improve.
Figure 9 shows resulting boxplots of each scenario (1 to 7). No significant differences can be observed. Therefore, an ANOVA test is performed to identify if there is a significative difference between scenarios considering individual makespans from all replications. The result was a p-value of 0.21, discarting a possible best solution. However, in Figure 9, despite of the similarity between all boxplots, it is possible to identify that, from second to seventh scenario, results tend to get worse, and the worst is the first one. Scenario 2 and Scenario 3 seem to have better results.

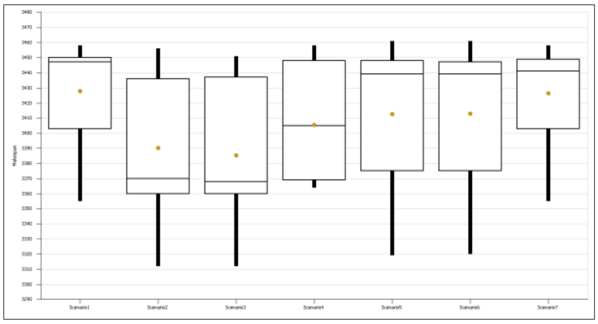


Figure 9: Output variable boxplot per scenario

CONCLUSIONS

A hybrid simulation-optimization approach was developed to solve the scheduling problem of a complex block assembly process of a naval industry. This real-world case study involves a considerable number of blocks and sub-blocks, requiring important efforts for generating the production plan. A combination between a set of MILP models and a discrete-event simulation model was performed to obtain the best processing sequence to improve the system efficiency. MILP models contribute to find optimal sequences for groups of blocks, and simulation provides a more reliable solution taking MILP outputs adding stochasticity and the block-erection process. Different scenarios were proposed and computational experiences were measured, comparing computational time vs solution improvement. Results demonstrate that computational time strongly increases without providing much better solutions. Hence, the major advantage of this tool is that it could find near-optimal solutions without falling into extremely long and unreasonable computational times.

ACKNOWLEDGMENTS

This paper was partially founded by CONICET under Grant PIP 112 20150100641 and from ANPCYT under Grant PICT-2014-2392.

REFERENCES

Back M.G., Lee D.K., Shin J.G., and Woo J.H., 2016. A study for production simulation model generation system based on data model at a shipyard. International Journal of Naval Architecture and Ocean Engineering, 8(5), 496-510.

Banks J., Carson J.S., Nelson B.L., and Nicol D.M., 2005. Discrete-Event System Simulation. 4th ed. New Jersey: Pearson Prentice Hall.

Basán N.P., Achkar V.G., García-del-Valle A., and Méndez C.A., 2017. An effective continuous-time formulation for scheduling optimization in a shipbuilding assembly process. Proceedings of 46° JAIIO. Córdoba (Argetina).

Castro P.M., Aguirre A.M., Zeballos L.J., and Méndez C.A., 2011. Hybrid Mathematical Programming Discrete-Event Simulation. Industrial and Engineering Chemistry Research, 50 (18), 10665–10680.

Chen N., Wang Z., and Wu J., 2013. Simulation-based research on adjustment technology of ship block production plan. Proceedings of International Conference on Information Technology and Applications, 259-262. 16 November, Chengdu (China).

Cho K.K., Oh J.S., Ryu K.R., and Choi H.R., 1998. An Integrated Process Planning and Scheduling System for Block Assembly in Shipbuilding. Annals of the CIRP 1998, Vol. 47/1/1998, 419-422.

Cebral Fernandez M., Crespo-Pereira D., Garcia-del-Valle A., and Rouco-Couzo M., 2016. Improving Planning and Resource Utilization of a Shipbuilding Process Based on Simulation. Proceedings of the European Modeling and Simulation Symposium, 197-204. 26-28 September, Larnaca; Cyprus.

Hussein W.B., Kecker F. and Mitzscherling M., Becker T., 2009. Computer Modelling and Simulation of Bakeries. Production Planning. International Journal of Food Engineering, 5(2), 8.

Kim H., Kang J., and Park S., 2002. Scheduling of Shipyard Block Assembly Process using constraint satisfaction problem. Asia Pacific Management Review, 7(1), 119-138.

Law A., 2007. Simulation Modeling and Analysis. New York: McGraw-Hill.

Lee K., Shin J.G., and Ryu C., 2009. Development of simulation-based production execution system in a shipyard: A case study for a panel block assembly shop. Production Planning and Control, 20(8), 750-768.

Liu Z., Chua D.K.H., and Keng H.W., 2011. A Simulation Model for Spatial Scheduling of Dynamic Block Assembly in Shipbuilding. Journal of Engineering, Project, and Production Management, 1(1), 3-12.

Park K.P., Ham S.H., and Lee C.Y., 2016. Application and validation of production planning simulation in shipbuilding. Ocean Engineering, 114, 154-167.

Seo Y., Sheen D., and Kim T., 2007. Block assembly planning in shipbuilding using case-based reasoning. Expert Systems with Applications, 32(1), 245-253.

Shang Z., Gu J., Ding W., and Duodu E.A., 2013. Spatial scheduling optimization algorithm for block assembly in shipbuilding. Mechanical Information Research Center, 2017, 21.

Xiong F., Xing K., and Wang F., 2015. Scheduling a hybrid assembly-differentiation flowshop to minimize total flow time. European Journal of Operational Research, 240(2), 338-354.

AUTHORS BIOGRAPHY

Natalia P. Basán is an Industrial Engineer and PhD student at National Scientific and Technical Research Council (CONICET). She is professor of Simulation at Universidad Nacional del Litoral (UNL). Her research interests include hybrid optimization & simulation tools for production planning and scheduling of automated production systems. Her email address is nbasan@intec.unl.edu.ar.

Victoria G. Achkar is an Industrial Engineer and PhD student at National Scientific and Technical Research Council (CONICET) and professor of Simulation at Universidad Nacional del Litoral (UNL). Her research includes hybrid simulation & optimization tools for logistic management of complex production and distribution processes of industrial interest. Her email address is vachkar@intec.unl.edu.ar.

Dr. Carlos A. Méndez is a Titular Professor of Industrial Engineering at Universidad Nacional del Litoral (UNL) in Argentina as well as a Senior Researcher of the National Scientific and Technical Research Council (CONICET) in the area of Process Systems Engineering. He has published over 200 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. His group's web page is: http://servicios.intec.santafe-conicet.gob.ar/grupos/capse/.

Alejandro Garcia-del-Valle is a Doctor Industrial Engineer and Professor of Department of Economic Analysis and Business Administration at University of A Coruña in Spain. His scientific publications and researches include simulation and optimization of industrial processes, and logistic and transport. He has numerous journal articles, book chapters, and conference papers in these areas. His email address is alejandro.garcia.delvalle@udc.es.

EDUCATIONAL ANIMATION AS A COMPLEMENTARY TOOL TO CONVEY KNOWLEDGE TO CIVIL CONSTRUCTION WORKERS

Santana, L. [(a)], Freire, A.[(b)], Dagostin, M.[(c)], Jungles, A.[(d)]

[(a)] Instituto Federal da Bahia
[(b)] Instituto Federal do Piauí
[(c)] Universidade Federal de Santa Catarina
[(d)] Universidade Federal de Santa Catarina

[(a)] santana.leiliane@gmail.com, [(b)] ailton.freire@ifpi.edu, [(c)] marilia-dagostin@hotmail.com,
[(d)]ajungle@ceped-ufsc.com

ABSTRACT
The use of resources which enable the modeling of information has been growing in the media and also in the educational context. The objective of this article is to adapt a model for the narrative construction focused on the development in the educational animation field, facing the reality in civil construction as an aid in the process of transferring knowledge for the construction workers. The literature review was performed as a research strategy and after the analysis a reference model was chosen in order to elaborate the educational animation. Then, the adaptation of the model to civil construction took place. The Lean Construction Primer was selected from the literature to exemplify the developed proposal. The analysis of correlated themes confirms that the contemporary languages used for education motivate, involve and assist in the memorization of the learner. It is expected to contribute with the development of educational animation in the civil construction directed to the construction workers as well as encouraging discussions on similar issues.

Keywords: Educational animation, narrative, civil construction.

1. INTRODUCTION

The sectorial development of civil construction is pointed out as one of the reasons which influenced the search for training courses for workers in the area. However, what is perceived is that there is often a gap between the knowledge and the transmission of the craft, which hampers the learning process.
Chiavenato (2010) alerts to the need of professionals with better qualification. It states that unemployment rates are lower where educational levels are higher.
Brandenburg and Byrom (2006) when analyzing construction, state that companies which invest in planning and human resource management strategies have achieved high levels of performance, including high productivity, cost efficiency and company effectiveness overall.
Despite the efforts to overcome disability in vocational training, those are still considered insufficient when analyzing the gap in the quality of the available worker, in relation to what would be considered ideal for a greater sectorial development of the civil construction (ABRAMAT, 2007).
The discussions raised reveal a convergence of opinions when considering the transfer of knowledge to the training in civil construction relevant.
In this context it is important to stick to the new forms of knowledge presentation, since they can prove themselves beneficial in the teaching-learning process.
Therefore, adapting a narrative construction model for the development of educational animation to the reality of civil construction, (as an aid in the process of transmitting knowledge to the construction workers,) is presented as one of these new forms of knowledge presentation.

2. THE TRANSFER OF KNOWLEDGE

Flor et al. (2009) They affirm that, in the last decades, knowledge was recognized as something to be constructed by the learner himself in his relations with the social environment, and no longer as volume of information passed on by the teacher. In this context, the use of media or multimedia is increasingly growing with the aim of interacting with the user, based on information in diverse presentations.
Portugal (2014), Affirms to be fundamental to look for new models, methods and approaches capable of including the contemporary languages in teaching. Among them, we can highlight multimedia, hypertext, audio, video, animation, among others. Such approaches can provide significant information and create pleasant experiences in the teaching-learning process, according to the author.
This research is classified as qualitative and has an exploratory character. The article intends, with the literature review, to choose a reference model for the construction of educational animation and based on the analysis of this model propose an adaptation for the civil construction, directed to the construction workers. It is important to emphasize that the proposal will be the first phase of the conception of educational animation; the moment the narrative is built. This phase is responsible for defining the content needed to make the decisions which will guide the rest of the process.
It is also expected that the discussions raised may encourage the use of educational animation in the

context of civil construction, focusing on the target audience, the construction worker. The literature review was the strategy used to achieve the goal.
To emphasize the use of the adapted model (the proposal of this study) a material available was chosen in social networks for the transfer of specific concepts of civil construction for the workers of the construction.
In this way, the Primer of Lean Construction, (educational instrument available to assist the training of the worker in construction,) will be used to put the proposed objective into practice.

2. CONTEMPORARY APPROACHES IN EDUCATION

The term "transference" has been discussed in different contexts, including education, psychology, and administration. Below, we present concepts which seek to define the theme addressed and its relation with the civil construction sector.
Broad and Newstrom (1992) define the concept of transference in training as the effective and continuous application of knowledge and skills acquired in training for the workplace, whether these inside or outside this environment.
Vygotsky (2001) believes that mental content results from the interaction between the subject and the environment. So the preference for methods that create situations which reproduce the work environment is growing.
Krugüer and Heineck (1997) conceptualize training as an improvement of skills and techniques, in view of the action of performing a task. The authors define education as the transfer of knowledge and ideas, in a perspective focused on life.
Segundo Freire (1996) and the progressive conception, teaching does not take place in one way, since teaching and learning are one and the same. Thus, those who teach also learn.
In this way, knowledge transfer occurs both in training and in education. Therefore, there is the need to know/use tools which can aid in the teaching-learning process.
This research proposes the adaptation of a reference model for the development of educational animation focused on the construction sector, with the aim of transferring knowledge to the workforce.
Alves (2016) admits that when directed towards education, animation needs to add some extra elements related to the learning process, so it is necessary not only to design an animation or a story, but also to consider the context, the learning processes involved, the public and the learner's learning.
The following are some concepts brought to increase understanding of the subject matter.

2.1. Animation

Animation can be described as the art of capturing a series of individual and continuous positions that when played in rapid succession convey the illusion of motion. It can be used as educational material, contributing to the educational process. (Patmore 2003).
Animations can empower, facilitate and engage the learner in learning situations (Ainsworth 2008).
Vygotsky (2001) states that cognition originates in motivation, but according to the author it is not born spontaneously, it requires stimulation.
In this context, Gondim et. Al. (2011) state that the use of animation stimulates cognitive processes, such as perception, memory, language, thought and others, as well as producing a playful environment for the development of learning.
The authors state that animation allows the modeling of real events that temporarily evolve into abstract concepts. They draw attention to the interaction between user and system, achieved through the use of animation, since it is culturally perceived as a little formal language.
Huhnt et. al. (2010) believe that the basic mission of animation is to transmit knowledge, so the learner can use the teaching where the application was planned.
Xiangyu et.al. (2013) reveal that animation is one of the media options used to produce Augmented Reality (RA). Although it is not the focus of this study, it is important to emphasize that the discussions about the use of Virtual Reality (RV) In built environments have been growing in researches in the field of architecture, engineering, construction and other related to assist in matters such as planning, design, security and training of risk operations, among others.
It is important to mention characteristics which differ from the commercial animation of the animation developed with the educational approach. The following item conceptualizes the animation in this last approach.

2.2. Educational Animation

Lowe et. Al. (2008) affirm that animation consists of forms or elements which change in relation to time, in a sequential and continuous way.
Barbosa Júnior (2005) describes animation as art which depends on technology, and thus relies on the resources of visual syntax, as well as drawing and painting, but which aggregates audiovisual elements. Therefore, it becomes "a multimedia art" which has the ability to tell stories and convey information.
Alves (2012) clarifies that the production of commercial animations counts on specialized professionals for each aspect of the production (narrative, script, illustration, animation), however, this does not always happen with the teams of development of educational animations.
The author alerts that in many cases in the design of educational animations, decisions are focused on the presentation of content, based on the subjective knowledge and experience of the author responsible for the project (animator, teacher or animation team and editing). In this way, the process becomes variable and depends on the repertoires of the developers, and these are not always able to do so.
Barbosa Júnior (2005) briefly points out four phases for the generation of commercial animations: 1) design and

development; 2) pre-production; 3) production and 4) post-production.
The first and second phases of the process are characterized by the development of the story to be told. Thus, story, requirements, goals, and graphic choices are defined at that time. The third and fourth are responsible for executing the planned decisions, the production.
In the same way as in the chosen reference model, after reviewing the literature, this article focuses on the first phase of the animation production process. This phase is responsible for the decisions which will guide the rest of the process.
In this context, a survey of the existing knowledge about the narrative construction, the script and a restructuring of the design process is relevant, in order to guide it to meet the expectations of the learners.

3. ELEMENTS OF THE SCRIPT

Clark and Lyons (2011) believe that learning occurs when new content is processed by working memory and then integrated into the learner's long-term memory.
Alves (2016) emphasizes the importance of information organization and also the use of narrative structures for learning. It completes the argument by stating that in learning the use of narrative can aid in memorization and empathy.
The literature review points out as relevant the use of the model developed in Alves' research (2016) as a reference for the development of educational animation for the construction workers, proposed in this article.
The reference model chose to study more language-related authors, cinematographic production and the making of scripts, considering them closer to the object of study, the animations.
Table 1, below, presents the selected elements after comparison with the authors suggested in the literature, according to Alves (2016). They are: Gancho (2002), Field (2001), Comparato (1995), Maciel (2003), Seger (2007), Mckee (2006) and Vogler (2006 The elements highlighted by the authors are: theme, subject, message, space, time, action, structure, character, scene, cohesion and narrator.

Table 1: Selection of script elements

Elements	Definition	Authors
Theme	Idea in which the story is written around	Gancho(2002); Maciel (2003); Seger (2007); Mckee (2006)
Subject	Fullfillment of the theme	Gancho(2002); Field (2001); Comparato (1995)
Message	Thought or conclusion to be drawn from the story	Gancho(2002);
Space	Determining the place and environment of the story	Gancho(2002)
Time	Period of the event	Gancho(2002); Comparato (1995)
Action	Structure of the events and ongoings which will construct the scenes	Field (2001); Comparato (1995)
Structure	Set of shapes or elements which compose the script	Gancho(2002); Field (2001); Maciel (2003); Seger (2007); Mckee (2006)
Character	Beings active or not, present in the story, not necessarily human.	Gancho(2002); Field (2001); Comparato (1995) Maciel (2003); Seger (2007); Mckee (2006)
Scene	An event of the story which presents the action with the conflict	Field (2001); Maciel (2003); Seger (2007); Mckee (2006)
Cohesion	Form of writing and elements which make the narrative attractive	Seger (2007)
Narrator	Character or being that tells the story	Gancho(2002)

From the description of the elements listed, the reference model presents a structure for the narrative construction in educational animations, as represented in Figure 1, below.

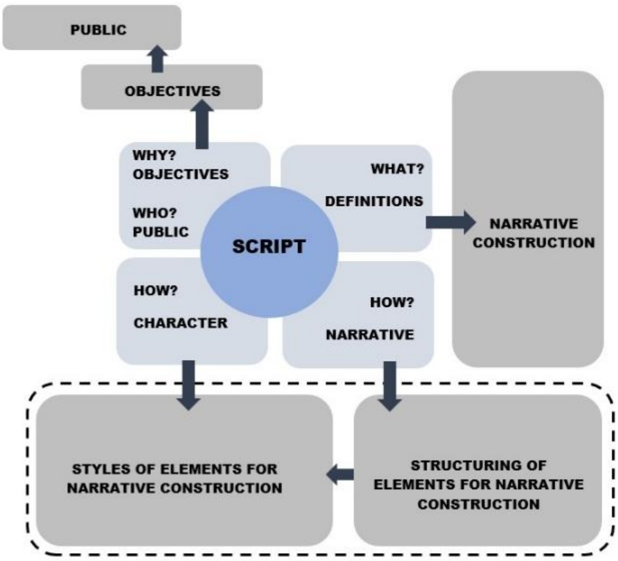


Figure 1: Structure for the construction of narrative in educational animations
Source: adapted from Alves 2016.

4. STRUCTURE FOR THE NARRATIVE CONSTRUCTION IN EDUCATIONAL ANIMATIONS

From the study of the structure for the narrative construction in educational animations, developed by the reference model addressed, an adaptation was elaborated to aid the transfer of the knowledge for the civil construction workers.
The Lean Construction Primer was chosen to exemplify the adapted model.
The items below present the elements defined by the reference model and that will be used to produce the proposed model. It is important to point out that the

elements were reorganized by modules in order to facilitate visualization at the time of structuring the representation. In this way, the content is presented in three modules, according to its purpose: narrative construction, structuring of the elements for the narrative construction and styles of the elements for the narrative construction.

4.1. Narrative conStruction

The narrative construction according to the reference model contains six elements for the definition of the presented module. They are: theme, subject, message, space, time and action. Table 2, below, presents the elements for the construction of narrative in educational animation.

Table 2: Narrative construction - reference model

Theme	Subject	Message
What is the story about?	How does the theme appear and develop?	What is the conclusion to be draw from the story?
Component elements: - Premise; - Governing Idea and - Problem.	Component elements: - Research on the subject (personal memories, imagination about the content and facts that occurred); - Fact / event; and-. - General and detailed view of the subject.	Component elements: - Story - Objectives; - Their personal interest and - Outcome.
Space	**Time**	**Action**
Where? (Place / setting of the story)	What is the time of the event?	Structure of the events and events which will form the scenes
Component elements: - Place / location and - Environment (socioeconomic, moral and psychological).	Component elements: - Period - Duration - how long? What is the time used for ?; - Chronological / psychological time; - Linear Time; and Nonlinear time.	Component elements: - Physical action (character movement); - Emotional action (internal to the character); and When and how the facts will happen.

4.2. Structuring of elements for narrative construction

The structure of the elements for the narrative construction, according to the reference model, contains three elements for the definition of the presented module. They are: structure, character and scenes.
Table 3, below, presents the structuring of the elements for the narrative construction in educational animation.

Table 3: Structuring of the elements for the narrative construction - reference model

Structure	
Definition: Set of shapes or elements which compose the script, selecting events from the life story of the characters composed in a strategic sequence to stimulate specific emotions.	Component elements: - Plot - Acts; - Conflict; - Turning point; - Evolution of the story; - Climax; - Secondary plots; and - Others (unfolding of the items mentioned above)
Character	
Definition: Beings active or not, present in the story, not necessarily human.	Component elements: - Role (who is the character? Protagonist / supporting ... others); - Function in the plot; - Type of character; - Needs; - Characteristics; - Level of veracity; - Motivations of the character; - Evolution; - Applied archetypal structure; and - Conflicts.
Scenes	
Definition: An event of the story which presents the action with the conflict, in more or less continuous time.	Component elements: - Objects - specific unit, an action in a given context. - Events: action and reaction; - Composition of speeches - dialogue and point of view; - Texts and subtexts contained in the action; - Visual composition - movement and actions in time; - Size: beginning and end of an event; - Types of camera shots and angles; - Rubrics: everything that is not dialogue: scenarios and environments; -Indexes: clues to the audience, insinuations; - Sequencing of scenes; and - Dramatic scenes: activities, dialogue and language.

4.3. Styles of elements for narrative construction

The styles of the elements for the narrative construction according to the reference model, contains two elements for the definition of the presented module. They are: cohesion and narrator.
Table 4, below, presents the style of the elements for the narrative construction in educational animation.

Table 4: Style of elements for narrative construction - reference model

Cohesion	
Definition: Form of writing and elements which make the narrative attractive.	Component elements: - Anticipation of the outcome: tips that are coming to an achievement; - Recurring themes: image, rhythm or sound which are repeated throughout the film;

	- Repetition: attributes to confirm a characteristic; - Contrasts: opposites increase the dramaticity; - Unity: vision of the whole and the parts; - Elements to success: appeal, creativity, structure of the script; - Connection: universal appeal, trends, needs and reasons; and - Clarity: Theme and connections to the theme.
Narrator	
Definition: Character or being that tells the story	Component elements: - Type: Third person (omniscience / omnipresence) First person (witness and protagonist).

After the study of the structure for the narrative construction in educational animation, developed by the reference model, it was possible to elaborate a model to help the transfer of knowledge to construction workers.
The following item presents the proposal of this research, taking into account aspects relevant to the adequacy of the reference model used for the civil construction sector.

5. PROPOSAL - EDUCATIONAL ANIMATION ADAPTED TO CIVIL CONSTRUCTION

The literature review allowed choosing the model which after the study of the elements of the narrative, proposed a structure for the narrative construction in educational animations. The reference model used was obtained from Alves (2016). The choice was made for clarity, pertinence and proximity to the theme of this article.
After the brief explanation of the elements of the narrative, expressed in the modules presented previously, some points will be considered in order to conduct the production process in educational animation.

5.1. Relevant considerations to educational animation production

This research chose to focus on the first phase of the animation production process, since it is in the initial texts that the content is conceived and the decisions which guide the rest of the process are made.
It is necessary to consider the interdisciplinary work for the construction of the animation, thus, a team composed of professionals from different areas can be part of the creation process.
The adaptation of the reference model it was possible after consideration of the aforementioned considerations, including the analysis of the context.
The context covered is the construction sector and the target public is the construction worker.
After analyzing the data disclosed by Annual Social Information Report – RAIS/TEM (2014) it was possible to affirm that the age range of the construction worker is concentrated in the range of 25 to 39 years and that this number corresponds to a total of 51.78% of the total workers.

The analysis also reveals that in Brazil, the construction worker is predominantly male and has monthly incomes in the range of up to two minimum wages. The data indicate that the turnover rate is high, since approximately 50% remain for up to two years in the same company.
The brief discussion about the context and the target audience was intended to characterize the universe considered for the development of the one proposed by the research. Basically, it served as support for the necessary premises for the accomplishment of educational animation directed to the civil construction.
The following item presents the proposal adapted to the civil construction, based on the reference model.

5.2. Adaptation of the reference model for educational animation in civil construction

The Civil Construction Primer was developed during Santana (2010) dissertation research and aimed to present and encourage construction workers to perform good practices related to lean construction in the work environment; the construction site. The material was presented in A5 format, printed and colored. Simple texts with easy assimilation. The practices highlighted are usually taken from the daily routine of the worker to facilitate understanding. Figure 2, below, illustrates the Civil Construction Primer.

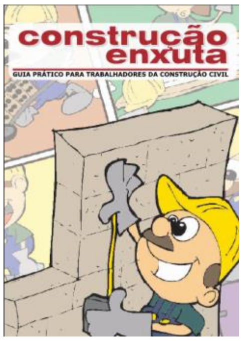


Figure 2: Lean Construction Primer
Source: Santana 2010.

The material studied was not designed for the development of educational animation. However, it can be used to exemplify the proposal developed in this article, adaptation of educational animation for construction workers.
It is important to emphasize that the proposal covers the first stage of the animation which provides the guiding definitions for the implementation of the product. The next phases can be achieved with the help of professionals of the area, who will transform the information (static images and texts) in dynamic graphic, with movement.
Table 5, below, describes the module for the narrative construction, according to the reference model. It contains six elements for the definition of the module presented. They are: theme, subject, message, space, time and action.

Table 5: Narrative construction – proposal

Theme	Subject	Message
What is the story about?	How does the theme appear and develop?	What is the conclusion to be drawn from the story?
The story relates the day to day construction worker and his daily actions at the construction site.	At each scene, it is possible to focus on the action of the worker who consciously performs it (whether in the execution of a task, in the use of a work tool, in the exchange of experience with a colleague, among others).	It is possible to carry out actions directed to the principles of lean construction in the day to day work.
Space	**Time**	**Action**
Where? (Place / environment of the story)	What is the time of the event?	Structure of the events and events that will form the scenes
The scenario is a common day on a construction site.	The story takes place in the present day. Corresponds to a work day, which usually starts at 7am and ends at 5pm at the construction site.	The physical action will be expressed by the correct execution of each service, considering the studies of the ergonomics for the accomplishment of the same ones. Emotional action describes a positive, constantly motivated character to learn.

The module directed to the narrative construction, presents an overview of the theme, including the space and the time where the scenes will take place, in this case, the construction site.

Table 6, below, describes the module for structuring the elements for the narrative construction, according to the reference model. It contains three elements for the displayed module definition. They are: structure, character and scenes.

Structuring the elements for the narrative construction - proposal

Structure	
Definition: Set of shapes or elements which compose the script, selecting events from the life story of the characters composed in a strategic sequence to stimulate specific emotions.	
Plot	How to use good practices to solve the issues arising from the day-to-day construction site, using the basics of lean construction. The worker is aware of his actions and seeks to carry out good practices in the work environment.
Acts	The questioning towards the current action and the one considered satisfactory for a given procedure.
Conflict	What can I do to improve what I already do? What are the consecrated practices I should follow?
Turning Point	New knowledge to carry out good practices.
Evolution of the story	The worker knows and puts into practice what he has learned.
Climax	The worker feels accomplished by practicing the action / thought which made his work even better.
Secondary scenarios	The worker transmits to his colleagues the new learning.
Others (unfolding of the items mentioned above)	The worker is recognized for his performance in the service performed or attitude which has contributed to the improvement of processes in the workplace.
Character	
Definition: Beings active or not, present in the story, not necessarily human..	
Role	Miguelito - Worker (Protagonist) Master of works (Supporting role) Coworker (supporting role)
Function in the plot	The main character, the construction worker is active in the plot.
Type of character	The main character is the caricature of a middle-aged workman.
Needs	The character wants to know the best way to perform the services which are assigned to him, as well as to know the most appropriate tools to execute them.
Characteristics	
Level of veracity	The scenes depict real situations of the day-to-day construction site, however do not stick to the representation of the complementary scenario, but only what is necessary for the transmission of the main message of the story
Motivations of the character	The character is motivated by his daily learning because he understands that he is the first client of the process. He also knows that good practices generate continuous improvement.
Evolution	There was no significant evolution in the actions of the main character
Applied archetypal structure	Middle-aged man with habits coming from a traditional education
Conflicts	How to maintain healthy relationships in the workplace. How to find ways to grow in the work environment.
Scenes	
Definition: An event of the story which presents the action with the conflict, in more or less continuous time.	
Objects	The tools and the specific scenario of each service at the construction site.
Events	What generates each action carried out in practice, day by day, at the construction site
Composition of speeches	The lines almost do not exist, the expressions depict the message that the character wants to convey.
Texts and subtexts contained in the action;	The texts portray the actions considered good practices to be conducted at the construction site.
Visual composition	The scenario is simple. It was chosen to portray only the necessary, to convey the main message of the scene.
Size	The events are fast, have a short duration of time, as they portray the daily life and emphasize what is already done and leads to continuous improvements.

Types of camera shots and angles	To be defined with the responsible team.
Rubrics	To be defined with the responsible team.
Indexes	Use already usual terms of the working environment (slang or audience's own staff)
Sequencing of scenes	The scenes were divided according to the principle of lean construction which was presented.
Dramatic scenes	Some scenes were exaggerated to attract the attention of the target audience.

The module directed to the structuring of the elements of the narrative discussed the structure, the characters and the scenes. It is important to emphasize that the reference model presents elements that, in some cases, may be repetitive according to the level of complexity to describe the story that it is to be told.

Table 7, below, describes the module for styles of elements for the narrative construction, according to the reference model. It contains two elements for the displayed module definition. They are: cohesion and narrator.

Table 7: Style of the elements for the narrative construction - proposal

Cohesion	
Definition: Form of writing and elements which make the narrative attractive.	
Anticipation of the outcome	It was not present in this work
Recurring themes	The construction site scenario (sidings, tools, building materials, signage, EPIs ... among others)
Repetition	Repetition of the main character using the uniform and tools needed for each service
Contrasts	Some drawings were rendered out of the usual pattern, in order to draw attention to the main message of the scene.
Unity	The repetitions already cited, in the scenarios, promoted the unity of the parties.
Elements to success	The casual approach and short, simple language have come together to convey the main message of the story.
Connection	The verisimilitude between story told and reality, bring the target audience closer to the developed product, generating empathy.
Clarity	Short phrases, images focused on the representation of necessary objects were the combination chosen to attribute clarity to scenes
Narrator	
Definition: Character or being that tells the story	
Type	Selective omniscient narrator (third person): narrates the facts always with the concern to report opinions, thoughts and impressions of one or more characters.

The module directed to the style of the elements for the narrative construction presents the form of writing and elements that make the narrative attractive. It also presents the narrator, character or being that tells the story. In the case presented, the main character is the construction worker.

The following item shows the application of the proposal in the context of civil construction after the use of narrative structure adapted for educational animation, as presented in the previous item.

5.3. Application of the proposal: Lean Construction Primer

The base of the reference module helped to organize the elements by purpose, which allowed to visualize the whole process, which in turn can facilitate the structuring of the representation of an educational animation.

The Lean Construction Primer did not emerge from a narrative structure for the development of educational animation. However, the use of the model proposed by this study eased the understanding of the narrative exposed, making more transparent the objectives which could be intrinsic only to the author/developer's desire for this material. Thus, he was able to reduce his subjective manner of the material.

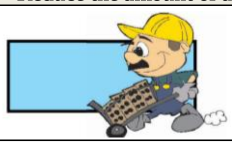
It is worth mentioning that, in addition to the principles of lean construction in practice, it was also prioritized to review and / or encourage the correct use of work tools, materials available in the work and concepts related to cleaning and organization of work environment (5S Program).

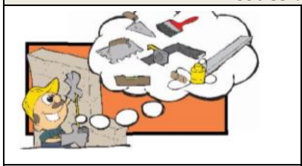
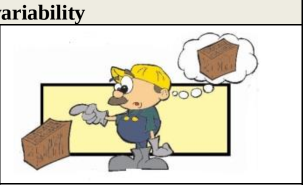
The search for practical examples that could be inserted in the day to day of the worker related to the principles of lean construction was one of the desires for the elaboration of the chosen material. It was due to exemplify the proposed model adapted for the civil construction.

Briefly, the structure of the narrative prepared for the Civil Construction Primer, if the animation production stages are completed, may contribute to: 1) Dynamize, and make the learning process more playful; 2) Review the content required for the specified services (procedures, equipment, tools and materials); 3) Streamline the process of assimilating the steps in each service; 4) Ease the understanding of the processes from the realistic exposition of the facts brought by the animation; 5) Make the overall process more transparent by exposing the relevant information; 6) Value the workspace and interpersonal relationships; 7) Encourage the use of IPE, safety equipment; and 8) Reinforce the policy of good practices and the participation of decisions in the company, suggesting something that adds value to the process, among others.

Table 8, below, illustrates some principles of lean construction and good practices related to the day-to-day construction site.

Table 8: Lean construction principles and practices tips for the construction site

Reduce the amount of activities which do not add value	
	
Suggest to your superior the use of some tool or equipment which will speed up the work.	Transport only the material necessary for the execution.
Increase the value of the product/service from internal/external customer considerations	

	
Clarify any doubts of a customer-modified project.	Do not delay the schedule of activities since the client expects to receive your property within the stipulated time.
Introduce continuous process improvements	
	
Participate in the training offered by the company.	Help to improve the company, give suggestions.
Reduce variability	
	
Use the tools you need to get quality service.	Do not use materials outside the standard.

6. FINAL REMARKS

Training in civil construction is still an issue in constant discussion. According to data from RAIS/TEM (2014), the level of education in this sector is considered inferior when compared to the productive sector.

The search for contemporary languages for learning is recurrent in several areas. The use of narrative in education can help the process to entertain, motivate and facilitate learning. The review of correlated literature states that its importance may be associated with cognitive, motivational and memory-appealing processes.

The reference model was useful for the development of the structure directed to the construction of narrative for educational animation, directed to the construction workers.

The organization of the elements in modules allowed the visualization of the whole process, which facilitated structure the representation. Thus, three modules are generated, according to their purpose: narrative construction, structuring of the elements for narrative construction and styles of the elements for narrative construction.

The Lean Construction Primer, selected from the literature to exemplify the proposal of this research, used the proposed model as a parameter to develop the educational animation aimed at construction workers. The set of recommendations of the reference module were sufficient to assist in the narrative construction.

It was noticed that the knowledge of the elements of the narrative, brought by the reference model, allowed a more structured approach to the Lean Construction Primer. Thus, the elements which make up its structure could be visualized and analyzed from a deeper insight into the development of narrative for the educational purpose.

In this way, knowledge about the narrative elements is valid, since they can improve the construction of educational animations, when used consistently and clearly.

It is important to point out that the systematization of some elements and narrative concepts in the proposed scheme can provide clarity in the processes and ease the application of these elements, even though by non-specialized teams of authors/developers or by the teachers themselves.

Educational animation, a contemporary auxiliary language to the traditional teaching exposed in this article, is not intended to remedy the shortcomings of the civil construction sector, given the need for training, but only to complement existing methods.

It is expected that this research may encourage the discussion and the use of educational animation in the civil construction sector, since such a resource may help teaching, while motivating and involving the learner.

Thus, as the reference model was useful to guide the elaboration of the model proposed by the research, it is expected that the results of this study may also help other authors/developers of varied areas in the process of construction of educational animation.

REFERENCES

Ainsworth, S. 2008. How do animations influence learning? School of Psycology and Learning Sciences Reserch Institute, UK: University of Nottingham.

Alves, M. M. 2012. Design de animações educacionais: recomendações de conteúdo, apresentação gráfica e motivação para aprendizagem. Dissertação de Mestrado. PPGDesign, Ctba: UFPR.

Alves. M. M., Battaiola A. L. & Cezarotto M. A. 2016. Representação gráfica para a inserção de elementos da narrativa na animação educacional. Revista Brasileira de Design da Informação. Brazilian Journal of Information. Design São Paulo.

Associação Brasileira da Indústria de Materiais de Construção (Abramat). Capacitação e Certificação Profissional na Construção Civil e Mecanismos de Mobilização da Demanda. Escola Politécnica da Universidade de São Paulo- Departamento de Engenharia de Construção de São Paulo. São Paulo, 2007.

Barbosa Júnior, A. L. 2005. Arte da Animação: técnicas e estética através da História. 2. Ed. Senac: SP

Brandenburg, S. G. e Byrom, K. 2006. Strategic management of human resource in construction. Journal of Management in Engineering, ASCE. p. 89-96. April.

Broad, M. L e Newstrom, J. W. 1992. Transfer of training: Action-Packed strategies to ensure high payoff from training investiments. Reading, MA: Addison-Wesley.

Chiavenato, I. 2010. Gestão de Pessoas. 7. ed. CAMPUS – RJ.
Clark, R.; Lyons, C. 2011. Graphics for learning. S. Francisco: John Wiley and Sons.
Flor, Carla S.; Beirão Filho, J. A.; Vanzin, Tarcisio. 2009. A hipermídia como ferramenta de aprendizagem na sociedade do conhecimento. In: 4º Conahpa- Congresso Nacional de Ambientes Hipermídia para Aprendizagem, Florianópolis.
Freire, P. (1996). Pedagogia da autonomia: saberes necessários à prática educativa. São Paulo: Paz e Terra (Coleção Leitura).
Goodwin , K. 2009. Designing for the digital age: how to create human-centered products and services. Indianapolis: Wiley Publishing.
Huhnt, W., Richter, S., Wallner, S., Habashi, T., Krämer, T. 2010. Data management for animation of construction processes. *Advanced Engineering Informatics*, *24*(4), 404-416.
Krüger, J. A.; Heineck, L. F. M. 1997. Superando as fronteiras do treinamento tradicional através da disseminação dos conceitos de qualidade de vida no trabalho – um estudo de caso no assentamento de cerâmica. In: XVII Encontro Nacional de Engenharia de Produção- ENEGEP. Gramado- RS.
Lowe, R. e Schnotz, W .2008. Learning with animation: research implications for design. New York: Cambridge University Press.
Relação Anual de Informações Sociais- Ministério do Trabalho - Rais/ Tem. 2014. Grau de instrução na construção civil e no setor produtivo.
Patmore, C. 2003. The complete animation course. London, Thames & Hudson.
Portugal, Cristina. 2014. Design and the Contemporary Languages in Education. Blucher Design Proceedings, v. 1, n. 2, p. 339-350.
Santana, L.; Brandstetter, M. C. G, de O.; Amaral, T. G. do. 2010. Construção Enxuta: Guia Prático para os Trabalhadores da Construção Civil. FUNAPE – UFG. Goiânia.
Vygotsky, L., 2008. Pensamento e linguagem. São Paulo, SP: Martins Fontes, 4ª Edição, 224 p.
Xiangyu Wang, Mi Jeong Kim, Peter E.D. Love, Shih-Chung Kang (2013). Augmented Reality in built environment: Classification and implications for future research. Automation in Construction.

AUTHORS BIOGRAPHY

Leiliane Santana Souza – Graduation in Architecture and Urbanism by *Universidade Estadal de Goias*, Professor at *Instituto de Educação, Ciência e Tecnologia da Bahia* in Brazil, Master in Civil Engineering by *Universidade Federal de Goias* in Brazil and doctoral student in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil.

Ailton Soares Freire – Professor at *Instituto de Educação, Ciência e Tecnologia do Piauí* in Brazil, Master in Civil Construction by *Universidade Federal de São Carlos* in Brazil and doctor in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil.

Marília Martins Dagostin – Bachelor's student at *Universidade Federal de Santa Catarina*, with a education degree in English by *Universidade Federal de Santa Catarina* in Brazil. Translator and schollarship at GestCon (Construction Management Group) in Brazil.

Antônio Edésio Jungles – Associate professor at *Universidade Federal de Santa Catarina*, with doctorade in Production Engineering by *Universidade Federal de Santa Catarina* and internship (sandwich) at University of Waterloo in Canada. General coordinator at CEPED/UFSC (University Center for Studies and Research on Disasters), and GestCon (Construction Management Group) in Brazil.

EFFICIENT PRODUCT REPRESENTATIONS FOR AUTOMOTIVE LOGISTICS

Daniel Fruhner[a], Konrad Pawlikowski[b], Katja Klingebiel[c], Michael Toth[d]

[a]IDiAL Institute, Dortmund University of Applied Sciences and Arts
[b],[d]Department of Business and Management, Bochum University of Applied Sciences
[c]Faculty of Business Studies, Dortmund University of Applied Sciences and Arts

[a]daniel.fruhner@fh-dortmund.de, [b]konrad.pawlikowski@hs-bochum.de, [c]katja.klingebiel@fh-dortmund.de,
[d]michael.toth@hs-bochum.de

ABSTRACT

E-mobility and the increasing introduction of intelligent assistance systems that are based on embedded systems have led to a radical change in complexity of parts and variants in the automotive industry. To guarantee the availability of components and minimize obsolescence risks, dependencies between electronical and non-electronical components and the compatibility between hardware and software components have to be transparently documented in product representations. This is especially relevant for logistics, which acts as a cross-divisional function between technology development, procurement, production, sales and after-sales. This contribution presents a systematical analysis of the requirements on product representations in series production followed by a review of product representation approaches in scientific literature, focused on the automotive industry. Based on these findings, proven and innovative concepts for product representations are classified and rated against the requirements. Especially the advantages of semantic networks and graph structures seem to be promising.

Keywords: automotive product representation, automotive product structure, e-mobility, embedded systems

1. INTRODUCTION

Since mass production was introduced by Henry Ford in the early 19th century, the automotive industry has changed considerably (Holweg and Pil 2004). Nowadays, original equipment manufacturers (OEMs) offer their customers a huge variety of models - which can be individualised by several hundred options - to compete in international markets (Dörmer 2013). These options include design elements (i.e. colors), functional components (i.e. gear system) and recently more and more assistance systems (i.e. navigation and driver assistance systems). Besides, the OEMs constantly update their product range with increasing frequency (Schuberthan and Potrafke 2007).

Especially technological trends such as e-mobility and the increasing integration of intelligent assistance systems (based on embedded systems) have led to a radical increase in complexity of parts and variants (Kampker et al. 2016, Krumm et al. 2014). This digitization of the car has established new interrelations and dependencies among the car components. Here, the continuous compatibility of the various electronical and non-electronical components needs to be ensured. Prominent examples like the incident of the recall of the Takata airbags in 2015 (Sharon O'Malley 2016) show how easily the OEM's reputation is compromised by failures on supplier side. This challenge applies in particular to logistics: logistics has to guarantee material availability and quality under high demand uncertainty and acts as a cross-divisional function between technology developments, procurement, production, sales and after-sales.

In order to guarantee the availability of components and minimize obsolescence risks in parallel, it is essential for all logistics processes that the product representation depicts all dependencies between parts, components and car features and provides transparent holistic information for all involved departments. Especially, dependencies of components and the compatibility between hardware and software components have to be considered. Nevertheless, the current form of the product representation applied in logistics does not adequately document the new technical interrelations of components. Therefore, new forms of product representations or the enrichment of given product representations are required.

This contribution presents a systematical analysis of the requirements on product representations in series production followed by a respective rating of product representation approaches in scientific literature with focus on the automotive industry and the related field of mechanical engineering. Based on these findings, innovative concepts applicable for automotive logistics shall be identified.

The paper is structured as follows: all relevant terms and concepts will be defined in section 2, followed by a detailed discussion of challenges in the automotive industry. This leads to a deduction of requirements on product representations from different angles, but with focus on automotive logistics. Here, relevant characteristics are mapped and integrated in a holistic requirements catalogue. Subsequently, a systematic review of scientific literature on product representations in series production is given in section 3. The methodology pursued in this step is explained beforehand. In section 4, the state of research is classified and promising approaches and concepts are rated against the defined requirements. The contribution concludes with a summary of key insights and an outlook on further research.

2. CHALLENGES FOR THE AUTOMOTIVE INDUSTRY AND REQUIREMENTS ON PRODUCT REPRESENTATION

Before challenges for an efficient product representation for logistics in the field of automotive industry may be identified, a general understanding of the product car, its complexity and the logistics processes is necessary.

Nowadays, automotive customers have to deal with the rapid change of variants and options (Ebel and Hofer 2014). After the customer has chosen a car series and model, the model is typically further individualised by so-called options and option packages. These – sometimes – several hundred options include for example exterior, interior and security equipment, but also assistance systems like navigation systems or driving assistance systems (for example parking aid) (eVchain 2014). The high potential for individual configuration is an essential marketing factor for premium OEMs, but contributes significantly to the complexity of the product car.

A typical car consists of about 3000 to 6000 material items. If different variants and their parts are considered, it results in about 15000 to 20000 items per car (Klug 2010). This is a challenge in itself. But even worse, customers tend to expect that their vehicle orders can be recustomised, i.e. changed even shortly before actual production and that the produced car is rapidly delivered on the formerly planned date (Alford et al. 2000, Krog and Statkevich 2008).

Nowadays, among the items of a car are many simple parts, but automotive suppliers develop more and more complex modules (Trojan 2007). Moreover the proportion of the electric and electronic components increases within the car (eVchain 2014). This shift in the competence of the car manufacturer has been identified and analysed since many years. OEMs focus more and more just on the assembly of supplied parts and modules, the product marketing, the coordination of suppliers, and the distribution of the end product (Meissner 2009). In this context - as mentioned before - logistics plays a significant role as cross-divisional function between technology development, procurement, production, sales and after-sales.

The effective logistics management of the automotive supply chain requires that resource and component requirements resulting from anticipated or realised market demands are synchronised with resource capacities and restrictions of the production and supply chain. Therefore, the logistics planer needs a holistic set of information. Since relevant data is typically kept in a highly fragmented information landscape (Bockholt 2012, Meyr 2004, Stäblein 2008), this data has to be integrated into a transparent and efficient form of product representation.

This product representation needs to bundle all information needed to capture the customer's anticipated or realised demand (e.g. model volumes, option quotas and dependencies among these). It has to allow to determine the required resources capacities and material items needed to satisfy the customer's or market demand.

The compatibility of car models and options for a respective car series is described by a highly complex set of technical rules, while the relationship between the fully-configured car and the corresponding material items is described by the bill of material (BOM) (Pawlikowski et al. 2016). But in particular, technological trends like e-mobility and the increase of embedded systems – which may be subsumed under the term “digitization of the car” – have changed the requirements on this product representation applied by logistics.

E-mobility causes a major challenge in the compatibility of the electronic components and the connection of the energy consuming components to the energy source(s) in the vehicle. Compared to vehicles with a combustion engine, electric vehicles differ in various components (e.g. the battery control system); the share of electronic components is much higher. Since this represents an important bottleneck in e-mobility, a holistic information base with all dependencies has to be considered to ensure the compatibility.

Vehicle functions are enriched by safety-, comfort-, environment friendly-, and drive technology functions. Of course, these functions have effects in the area of driver assistance, but also on chassis engineering, drive technology and electronics as well as body technology (Ebel and Hofer 2014). Due to the increasing number of electronic components, in particular the large number of embedded systems, new forms of dependencies and compatibility questions arise between non-electronical and electronic components, e.g. their software versions. An embedded system is an information processing system that is embedded into enclosing products (Marwedel 2011). For example, the navigation system communicates with driver’s mobile phone and simultaneously provides input for the driver assistance systems. All of these components are subject of continuous development cycles on hardware as well as on software side. Therefore, compatibility has always to be ensured.

Nevertheless, innovation life cycles in the electrical industry (semiconductors, control units, embedded systems) are significantly shorter than vehicle life cycles and corresponding technical component life cycles (Grimm 2003). While new versions of a car series are launched every three to eight years, the innovation cycles of the electrical industry are significantly shorter. The incentive for the OEM to upgrade electronical functions, components or parts (including new partners such as Apple or Google) in series production is rising. The effect is a continuous change in car models during their life cycle and electronical innovations in components must be constantly adapted to given structures. Consumer electronics (e.g. smartphones) innovation cycles are about one year. That is significantly shorter than the life cycle of vehicles (Kampker et al. 2016, Krumm et al. 2014).

Since new characteristics of electronical components like the compatibility of software versions are not (comprehensively) mapped in product representations, the complex dependencies between electronic and other components are not transparent for logistics. This poses a major challenge and in result, these processes are not always under full control (Nagel 2011). Sometimes the customer becomes an involuntary beta tester – as in the case of the Toyota recall of hybrid electronics software (Edmunds.com 2017). The related unplanned costs and the considerable loss of reputation has to be avoided.

Due to the tremendous influence, the serviceability of parts and modules must already be logistically secured in the early development phase. This requires, in addition to the problem of technical feasibility, that new technologies are to be tested for their compatibility with logistics series processes (Weinzierl 2006).

To guarantee an early insight into possible bottlenecks, the transparent access to relevant data is an important criteria. Lack of transparency is not necessarily a consequence of non-existent data. Rather, it is due to the fact that the data to be considered is often extremely comprehensive and at the same time distributed over different software systems, which often do not have interfaces with one another. On top of this, the type of data processing - from the perspective of variant management - is often inadequate (Kesper 2012).

These effects carry on into the after-sales and also spare-part business. After a few years of use, many electronical systems of a vehicle (e.g. navigation systems) are outdated and obsolete. These cars can only be sold with large discounts to second-hand customers. A way to counteract the massive loss in value is to update the components, e.g. by integration of Apple CarPlay in an end-of-life vehicle (Moynihan 2014). In order to handle the respective logistics, a complete insight into the vehicle structure is necessary to ensure interoperability with existing components, control devices, interfaces and connections to energy sources and energy consumers. The availability of components on hardware and software side needs to be guaranteed. But also obsolescence risks of spare part inventories have to be minimised.

Hence, it is necessary to provide transparency on the multiple new dependencies of innovative electronical vehicle components. But today, the challenges and opportunities introduced by embedded systems and e-mobility are not considered sufficiently in product structures, which are an integral part of the product representation, for logistics. The characteristics of these components have to be identified and integrated within the logistics-relevant product representation.

Nevertheless, the logistics strategy is strongly connected to the cost-efficient variant diversity (see for example Lechner et al. 2011), e.g. by preferring an early or late differentiation in the physical processes (variants of an electronic component are delivered to the assembly site or differentiation is postponed to a late configuration at the assembly site).

Thus, the new characteristics of electronic components are relevant for development as well as sales, planning, logistics, production, distribution and after-sales. The objective of the product representation for logistics is the efficient management of dependencies and interrelationships in automotive planning and order management processes, part procurement, production and part distribution processes (Romberg and Haas 2005). The ideal product representation of the future shall support all processes and stakeholders in every phase of the product life cycle. Only this allows to integrate new modules, components or software components safely and quickly into an existing vehicle structure and into the logistics process.

Nevertheless, it is necessary prerequisite to assure consistency and avoid redundancy in and between all data entities when integrating data into one information model. As it is easily understood, an integrated information base could reduce the complexity and increase transparency of the different processes immensely. The advantage of this integration is the faster and easier access to relevant data and its innermost dependencies, as well as the reduction of redundancies (Pawlikowski et al. 2016).

The resulting requirements on product representation based on the described challenges for the automotive industry can be summarised as follows:

- Besides already documented dependencies in BOMs and technical rules, the **integration of new dependencies** like the compatibility of software versions is required.
- It is necessary to **integrate cross-functional information** to support every phase of the product life cycle, i.e. all processes from development over series production to after-sales.
- As innovation cycles for new technology products accelerate, the continuous compatibility of parts and modules needs to be ensured because electronic systems and their software are developed faster as car types. In this context, **modularity** of the product representation allows to replace components more easily.
- The complexity of the product car is further increasing. The **management of comprehensive and complex information** is necessary.
- **Transparent data structures** are necessary to identify and manage possible bottlenecks to avoid intern expenditure and a deterioration of the delivery service to the customer.

Before approaches and concepts for product representations are evaluated against these requirements in section 4, the following section 3 gives a systematic overview over the state of the art.

3. STATE OF THE ART OF PRODUCT REPRESENTATIONS

Methodologically a content analysis has been pursued to provide a reproducible literature overview of the state of the art in product representations in the field of the automotive industry. Conducted in the English-speaking area, the digital databases ScienceDirect and Google Scholar have been searched with defined search terms and a period from 2006 to 2017. The core terms of the search were “product representation” and “automotive”. The term “product representation” was also modified by the synonyms “product architecture”, “product structure”, “product data”, “product graph” and “product tree”. “Embedded systems” as well as “e-mobility” have not been considered as search terms in the context of product representation because the authors did not want to limit the results more than necessary. Moreover, relevant findings from the related field of mechanical engineering also have been considered.
The 107 resulting scientific publications have been preselected based on the abstracts, keywords and titles. After a thorough analysis, 27 papers from the last decade have been identified as relevant and a limited number of types of product structures could be derived. Furthermore, eight older publications and also books and dissertations found while analysing the results have been included.

3.1. Literature Review

In general, product representations are product knowledge decomposed into its elementary components from a technical view (Deng et al. 2012). These components can be either a physical or a non-physical artefact (service and software components) (Kissel 2014). The next larger units are modules. A definition of modules is given by Klug (2010) as an assembly of several components or assembly units. The modules, which may comprise a variety of functions (Rapp 1999), shall generally be easier replaceable than each part of the module separately (e.g. door, seat, cockpit, power pack, roof). Modules are used within a so-called modularization to subdivide a system. Modularization may occur differently within the phases of the product life cycle such as development, procurement, production, distribution, utilization and disposal (Blees 2011, Gu and Sosale 1999, Krause et al. 2014).
According to Schuh (2014) a product structure as an elemental part of the product representation is typically a structured formation of the product and its components. Generally, structure levels are introduced to represent assemblies, which bundle components in the product structure. Product structures support the multiple use of assemblies and parts. Another important objective of product structures are the reduction of production information and the support of the information flow.

Figure 1 illustrates fundamental product structures. Modular systems are designed from a certain number of building blocks (basic body and attachment). The definition of modules has been given before. In contrast to modules, series comprise components of the same design but of different size. Packages combine components to realize a variety of functions.

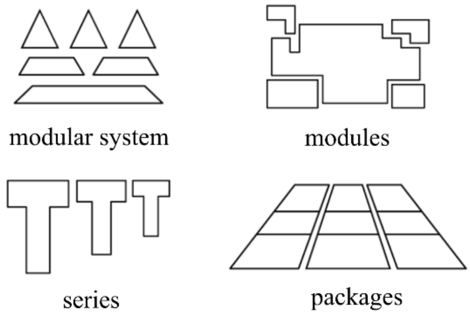


Figure 1: Fundamental product structures [based on (Schuh 1988)]

The most widespread form of product structure is the bill of material (BOM). Within a BOM, the components, i.e. parts and modules that constitute the product in the context of an assembly, subassembly or model are listed (Lee et al. 2012). BOMs are considered as an integral part of the product representation in the automotive industry. In BOMs, the information on components (e.g. compressor, cable, etc.) that are necessarily installed in a product to implement a function (e.g. climate control system) are documented (Wagenitz 2007). Brière-Côté et al. (2010) give a more detailed specification of BOMs: BOMs are described relationally as a list of subassemblies, components, parts, and raw materials which is applied to construct higher-level assemblies. To build a finished product, the BOM lets deduce the type and quantities of each material item used.
In the automotive industry, typically, the relevant data is complex and distributed over several systems in relational data structures. This not only holds true for BOM data, but also for other product information like model descriptions or technical rules. In particular, these different data fragments are not integrated in a common information base (Bockholt 2012).
In literature another common product structure is the tree structure. Literature differentiates between variant trees and feature trees, which differ in their representation and the integrated information (Kesper 2012). While variant trees represent the variety of semi-finished products arising during the assembly process, the feature tree illustrates the variety resulting from the combination of characteristics and their properties (Kesper 2012, Schuh 2014).
Variant trees form the basis for the reduction of variants by means of assembly sequence optimization or product structure optimization. The variant tree is often used to visualize component and product diversity that arises in assembly processes (Kesper 2012, Schuh 1988). Schuh (1988) identifies variant trees as an important instrument

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

to design and evaluate product variants, where different components are symbolised by different boxes (see Figure 2) (Kesper 2012). According to Schuh (2014), variant trees are typically constructed in defined steps. The product characteristics and their properties are captured in a first step. Afterwards, constraints on combinations of properties as well as the prohibitions of combinations are defined and variants are generated. The assembly sequence is determined after the integration of part information and allocation of part usage. The variant tree may be depicted graphically in a last step (Kesper 2012).

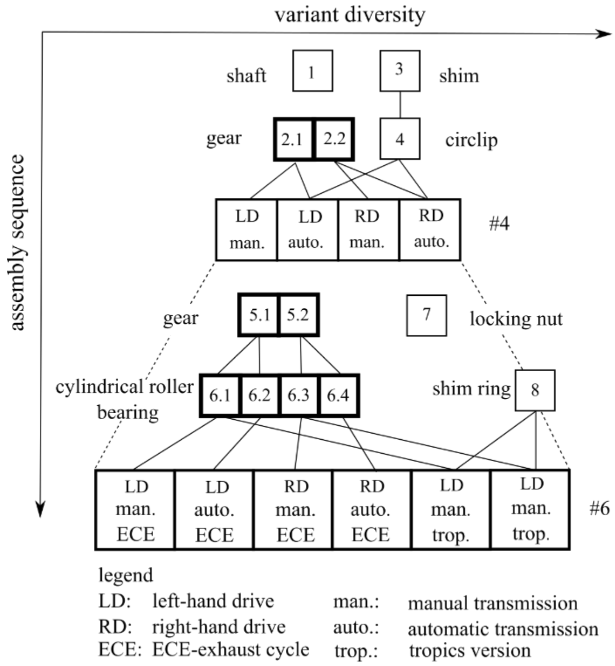


Figure 2: Variant tree [based on (Kesper 2012, Schuh 1988)]

The widely spread feature tree is often incorrectly also designated as a variant tree. It is an instrument to graphically depict variants or spectra with a focus on their characteristics and properties. The feature tree usually starts with a "root" node and then branched from left to right (see Figure 3).

A feature is presented by a vertical level. One variant is depicted by a branch of the tree, where the extent and shape of the feature tree depends on the order of features. A different order changes the total number of the feature expressions to be displayed (Kesper 2012). The visualization of the diversity, resulting from the combinations of characteristics and properties, is also depicted by this kind of tree. Nevertheless, to facilitate the interpretation of the representation, it is recommendable to list the categories of specifications in order of importance (Zagel 2006).

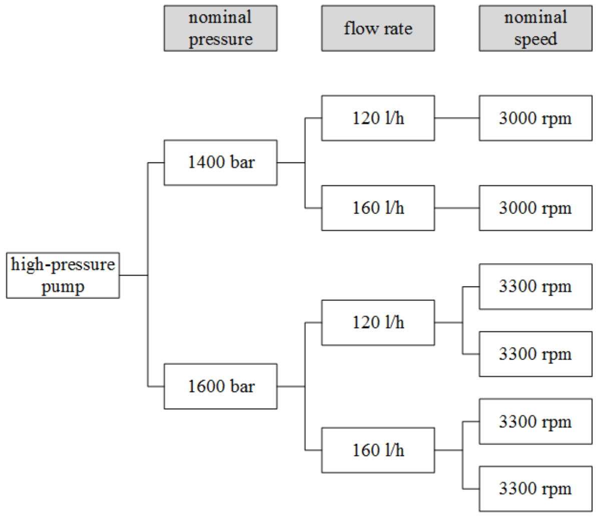


Figure 3: Feature tree [based on (Kesper 2012)]

A general formulation of tree structures is the hierarchical structuring. Ariyo et al. (2008) presents the hierarchical structuring as a technique to disassemble a (complex) product.

Within a product representation, not just the product structure as described has to be considered, but also dependencies and interrelations between a product, its components and the relevant assembly tasks. An extended product tree structure originally proposed by Zeng and Gu (1999) can describe these relationships. There are two types of nodes distinguished in an extended product tree. The assembly task node represents simplified assembly information included in the product structure while the component node represents a product or component. The connection between two component nodes is a parent-child relationship where a parent component (or assembly) consists of all its child components. A component is assembled by the appropriate assembly task that is signalled by the connection between the component and the assembly task. All nodes together form a recursive product structure tree (Deng et al. 2012). As long as functional requirements and cost-effectiveness persist, modules can be shared by different end products (Fujita 2002).

Another tree based approach for product representations has been realised in the tool suite OTD-NET (order-to-delivery and network simulator (cf. Wagenitz 2007). For different applications in logistics (e.g. process simulation, demand and capacity management or risk management) a hierarchical variant tree based on product descriptions with further enriched information is applied. Even sales information, technical rules and BOM rules are integrated into the hierarchy of product classes within the tree structure (Liebler 2013).

Another similar, but generalised form of product representations are graph structures, which – in contrast to tree structures – may be multidimensional. In mathematics, graphs are used to document pairwise relationships of features (Riggs and Hu 2013). In principle it is not necessary that a graph structure has just one "root" node. An illustration of relations between components can also be realised by liaison or connection graphs. Riggs and Hu (2013) used these graphs and developed a method to graphically illustrate the disassembly precedence relations among all components. Components or respectively parts are represented by nodes in liaison graphs, whereby the relation between those are depicted by edges (Hu et al. 2011). Within a precedence graph, the precedence order between components is documented instead of the physical connection (Riggs and Hu 2013). The disassembly precedence graph is a directed graph, illustrating the order of disassembly for the product in focus (see Figure 4).

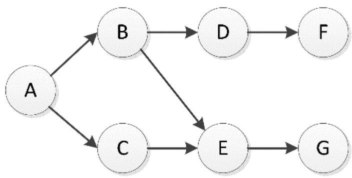


Figure 4: Example of a disassembly precedence graph (Riggs and Hu 2013)

Luo et al. (2016) describe another graph structure, the AND/OR graph, which is widely used in disassembly planning as product representation. These graphs consist of nodes and hyperarcs. Disassembly tasks are represented by hyperarcs, where nodes stand for components or subassemblies of a product (Viganò and Osorio Gómez 2013). A more complex subassembly can be formed by joining two or more components together (Li et al. 2002). Nodes in the graph are either AND or OR branches and form a hierarchical structure. AND relations are vertical links, where OR relations are nodes linked in the same level. In Homem de Mello and Sanderson (1990) the AND/OR graph has been applied to represent satellite equipment for an increased planning flexibility. A combination of weights and the AND/OR graph can be found in Min et al. (2010). This weighted AND/OR graph is used for disassembly planning and represents the product structure and element constraints.
The adjacent graph is another type of graph-based methods, which is used to represent component relationships of products (Song et al. 2010). Components or subassemblies are represented by nodes. Directed or undirected lines represent the relationships between connected components or subassemblies. Compared to AND/OR graphs the adjacent graph can include more information of component constraints for product structures (Luo et al. 2016).
There are some more innovative approaches in literature to depict product representations. These are based on ontologies and semantic networks.
An ontology is defined as a uniform vocabulary with the objective to exchange information in a particular field. The focus of ontologies is the description of real or intended things, whereby a consistency check for partial descriptions can be performed (Bock et al. 2010). It allows inter alia reuse and analysis of knowledge (Noy and McGuinness 2001).
A semantic network is, in contrast, a graphical representation of knowledge. These networks are realised with the aid of nodes and arcs (López-Morales and López-Ortega 2005). Nodes are used to represent objects, concepts or situations. The dependencies between the nodes can be deduced from the arcs (Yang et al. 2012). A simple example of a car as a semantic network has been illustrated in Figure 5, where, e.g. the relationship that a car is a ("ISa") vehicle with specific parts and components is expressed.

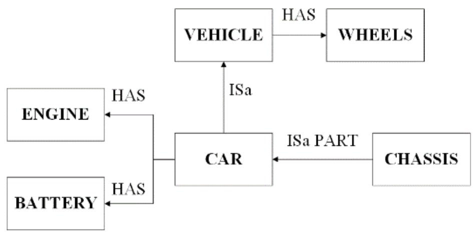


Figure 5: Example of car as a semantic network

Vegetti et al. (2011) present an example of an ontology-based semantic network where two hierarchies are applied to handle product variants from different angles. To efficiently deal with a high number of variants, the abstraction hierarchy allows to represent product data on various granularity levels. The organization of knowledge related to structural product information and to the BOM is obtained by the structural hierarchy.
"Design Structure Matrices" (DSMs) denote a compact representation of product element contexts, mainly used for a development perspective (Deng et al. 2012). This structures are suitable for models with many variant features as they allow a comprehensive presentation of information (elements of any type, i.e. components or process steps). To map the relationships of parameters between components the DSM is illustrated as a square matrix with the same columns and rows (see Figure 6) (Danilovic and Browning 2007). Only one type of connection (e.g. "…is linked to…") per DSM can be defined.

	element A	element B	element C	element D
element A		X		
element B			X	
element C		X		
element D	X			

Figure 6: Example of a DSM (Deng et al. 2012)

Nevertheless, it is difficult to keep an overview and ensure the manageability of the matrix representations for larger systems with several hundred elements (Kissel 2014).
Summarizing, relational databases, tree structures, (generalised) graph structures, ontologies, semantic networks and design structure matrices are the types of product representations which may be identified in literature. Based on the defined requirements, the next

section rates these forms of product representations in order to manage the challenges related to digitization in form of e-mobility and embedded systems.

4. FRAMEWORK FOR FUTURE PRODUCT REPRESENTATION – AN EVALUATION

The different concepts and approaches are now analysed in order to decide to what extent they match the requirements identified in section 2. The results are summed up in Table 1. The applied rating is:

- X : requirement fully met,
- O : requirement limitedly met,
- – : requirement not met.

The first requirement identified has been the **integration of new dependencies** like the compatibility of software versions. Relational data structures of course allow to integrate a multiplicity of dependencies by extension of the relational data schema (rating "X"), but at the expense of transparency. In contrast, tree structures, graph structures and ontologies allow to integrate new dependencies in a more structured way by adding new types of edges. By limiting the view on a different aspect, a transparent access to data can be guaranteed, thus leading to the rating "X". Different forms of dependencies can be deduced directly from the arcs in semantic networks, this also leads to the rating "X" here. The DSM in contrast (rating "-/O") depicts only information of one type of relationship. Thus, to depict another kind of dependency, another matrix needs to be generated. This would lead to a multiplicity of matrices which is not feasible in practice.

The second requirement, the **integration of cross-functional information**, is also fully (rating "X") met by all structures except the DSM (rating "–"). Similar to the integration of new dependencies, cross-functional information can be added in the same way.

Table 1 – Framework of rated product representations

Approach \ Criteria	Integration of new Dependencies	Integration of cross-functional Information	Modularity	Management of Comprehensive Data	Transparency
Relational Data Structure	X	X	X	O	-
Tree Structure	X	X	O/X	O/X	O
Graph Structure	X	X	X	X	X
Semantic Network	X	X	X	O	-/O
Ontology	X	X	X	O/X	O/X
DSM	-/O	-	X	-/O	O

The cross-functional characteristic of the new information can be explicitly documented. A DSM as presented e.g. by Deng et al. (2012) or Kashkoush and ElMaraghy (2016) on the other hand only allows to illustrate simple one-dimensional relationships and not to append additional cross-functional data.

Regarding the third requirement **modularity**, the tree structure approach is rated "O/X". The tree structures of Kesper (2012) and Schuh (1988, 2014) do not support modularity natively, but ElMaraghy et al. (2013) introduced an approach with evolving part/product families. Within the more general graph structures, two or more components or subassemblies can be joined together and form a more complicated (sub-) assembly (Luo et al. 2016). Therefore and in accordance with the ability to integrate a multiplicity of dependencies, modularity as described in section 3 is supported (rating "X"). Ontologies and semantic networks behave in a similar way as graph structures (rating "X"). By managing information over different data sources, relational data structures also fulfil the criteria of modularity (rating "X"). But it should be noted that it is difficult to remain transparency if the amount of data increases. DSMs with their simple direct connections between two components allow to replace components including their dependencies one by one. Though the level of information granularity is limited in DSMs they fulfil the requirements of this category in full (rating "X").

The next criteria to be evaluated is **the management of comprehensive and complex information**. Product representations based on tree structures and graph structures meet this requirement as the examples in section 3 show. But it should be noted that tree structures may become very complex in terms of system size (rating "O/X") (Kesper 2012). Graph structures fulfil this criteria better, because of their more general layout, which allows to structure the graph more flexibly according in dynamically changing environments (rating "X"). Especially the multidimensionality and the integration of several "root" nodes if need support this argument. In general, ontology based approaches have an average ability to manage complex systems (Lim et al. 2010). However, the approach of Vegetti et al. (2011) with its hierarchical concept shows a promising development and leads to a rating of "O/X". A DSM does not allow to illustrate complex multidimensional and cross-functional automotive data (rating "–/O"). Its tabular structure very quickly becomes intransparent (Kissel 2014). Semantic networks and relational data structures meet this requirement limitedly (rating "O"). Both approaches allow to manage complex and comprehensive automotive product data but transparency decreases and typically redundancies increase steadily with increasing data complexity.

The last requirement is **transparency**. The rating "–" is given to relational data structures due to the arguments already given above. The huge variety of distributed automotive data within the given relational data structures quickly leads to poor transparency and redundancies (Bockholt 2012). As stated before, the DSM can only depict one-to-one connections and is strictly limited in the size of the model. Under this limitation of data the transparency is high, but it requires multiple matrices for more complex cross-functional data sets. This results in the overall rating "O". In the case of semantic networks only similar relationships between two components are mapped, i.e. natively there is no kind of hierarchy or overall view in a semantic network (Yang et al. 2012), which limits the transparency radically in complex data environments leading to the rating "–/O". Graph structures limit the transparency eventually when being multidimensional, but allow to generate limited views on the structure (rating "X"). An especially enhanced graph structure is given by Riggs and Hu (2013) by presenting a disassembly precedence graph. The evaluated approaches based on ontologies are in their form similar to the hierarchical tree or graph structures, thus resulting in a rating "O/X".

All discussed ratings are summed up in Table 1 forming a framework of rated product representations.

5. KEY INSIGHTS AND OUTLOOK

In order to guarantee the availability of components and minimize obsolescence risks, it is essential that the product representation applied in logistics depicts dependencies and provides transparent holistic information. Especially, dependencies of technical and electronical components and the compatibility between hardware and software components have to be considered.

The research presented here is an important starting point, as it specified the five essential requirements on the automotive product representation originating from e-mobility and the integration of more embedded systems. A systematic literature research led only to a total of 27 relevant papers from the last decade which indicates the research necessary in this field.

The identified papers have been analysed thoroughly and the core types of product structures have been identified in result. Based on the beforehand deduced requirements, the state of research has been rated and promising approaches have been identified.

This structured discussion showed that DSMs are a valuable tool in development, but are not suited for complex cross-functional information as required by logistics. Regarding the complexity within the automotive industry, especially semantic networks, tree structures and generalised graph structures have proven to be promising candidates for a new generation of product representations. Especially, graph structures fulfil nearly all of the new requirements and may give answers to the challenges of parallel component development within the automotive industry.

To conclude, a variety of types of product representations is available today, but due to the findings of the literature review, an approach based on graph structures should be elaborated for logistics in the near future. Nevertheless, ontologies and semantic networks offer a new and different perspective on the integration of additional especially cross-functional information. Consequently, the combination of approaches may hold additional value and should not be neglected.

ACKNOWLEDGMENTS

Special thanks to the German "Bundesministerium für Forschung und Bildung" (BMBF) for making this study possible by providing funding to the ILogTec project. The funding of BMBF was granted with Funding-IDs 03FH008IA5 and 03FH008IB5.

REFERENCES

Alford, D., Sackett, P. and Nelder, G., 2000. Mass customisation — an automotive perspective. International Journal of Production Economics vol. 65, pp. 99–110.

Ariyo, O. O., Eckert, C. M. and Clarkson, P. J., 2008. Hierarchical decompositions for complex product representation. 10th International Design Conference Design, pp. 737–744, May 19 - 22. Dubrovnik - Croatia.

Blees, C., 2011. Eine Methode zur Entwicklung modularer Produktfamilien. Thesis (PhD). Hamburg University of Technology.

Bock, C., Zha, X., Suh, H.-w. and Lee, J.-H., 2010. Ontological product modeling for collaborative design. Advanced Engineering Informatics vol. 24, pp. 510–524.

Bockholt, F., 2012. Operatives Störungsmanagement für globale Logistiknetzwerke: Ökonomie- und ökologieorientiertes Referenzmodell für den Einsatz in der Automobilindustrie. Thesis (PhD). TU Dortmund University.

Brière-Côté, A., Rivest, L. and Desrochers, A., 2010. Adaptive generic product structure modelling for design reuse in engineer-to-order products. Computers in Industry vol. 61, pp. 53–65.

Danilovic, M. and Browning, T. R., 2007. Managing complex product development projects with design structure matrices and domain mapping matrices. International Journal of Project Management vol. 25, pp. 300–314.

Deng, X., Huet, G., Tan, S. and Fortin, C., 2012. Product decomposition using design structure matrix for intellectual property protection in supply chain outsourcing. Computers in Industry vol. 63, pp. 632–641.

Dörmer, J., ed., 2013. Produktionsprogrammplanung bei variantenreicher Fließproduktion. Wiesbaden: Springer Fachmedien.

Ebel, B. and Hofer, M. B., eds., 2014. Automotive Management. Berlin: Springer.

Edmunds.com, 2017. Toyota Recalls - Toyota Safety & Defect Recall News[Online]. Available at: https://www.edmunds.com/recalls/toyota.html (Accessed 12 April 2017).

ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tseng, M. and Bernard, A., 2013. Product variety management. CIRP Annals - Manufacturing Technology vol. 62, pp. 629–652.

eVchain, 2014. Modellierung der zukünftigen elektromobilen Wertschöpfungskette und Ableitung von Handlungsempfehlungen zur Stärkung des Elektromobilitätsstandortes NRW: (EM1006 – eVchain.NRW).

Fujita, K., 2002. Product variety optimization under modular architecture. Computer-Aided Design vol. 34, pp. 953–965.

Grimm, K., 2003. Software technology in an automotive company - major challenges. 25th International Conference on Software Engineering, pp. 498–503. Portland, OR, USA.

Gu, P. and Sosale, S., 1999. Product modularization for life cycle engineering. Robotics and Computer-Integrated Manufacturing vol. 15, pp. 387–401.

Holweg, M. and Pil, F. K., 2004. The second century: Reconnecting customer and value chain through build-to-order: moving beyond mass and lean production in the auto industry. Cambridge, Mass: MIT Press.

Homem de Mello, L. S. and Sanderson, A. C., 1990. AND/OR graph representation of assembly plans. IEEE Transactions on Robotics and Automation vol. 6, pp. 188–199.

Hu, S. J., Ko, J., Weyand, L., ElMaraghy, H. A., Lien, T. K., Koren, Y., Bley, H., Chryssolouris, G., Nasr, N. and Shpitalni, M., 2011. Assembly system design and operations for product variety. CIRP Annals - Manufacturing Technology vol. 60, pp. 715–733.

Kampker, A., Deutskens, C., Heimes, H., Ordung, M. and Haunreiter, A., 2016. Using E-mobility as an Enabler for a Fast and Lean Product Development to Optimize the Return of Engineering with the Example of Lithium-ion Battery. Procedia CIRP vol. 50, pp. 166–172.

Kashkoush, M. and ElMaraghy, H., 2016. Optimum Overall Product Modularity. Procedia CIRP vol. 44, pp. 55–60.

Kesper, H., 2012. Gestaltung von Produktvariantenspektren mittels matrixbasierter Methoden. Thesis (PhD). Technical University of Munich.

Kissel, M., 2014. Mustererkennung in komplexen Produktportfolios. Thesis (PhD). Technical University of Munich.

Klug, F., 2010. Logistikmanagement in der Automobilindustrie. Berlin: Springer.

Krause, D., Beckmann, G., Eilmus, S., Gebhardt, N., Jonas, H. and Rettberg, R., 2014. Integrated Development of Modular Product Families: A Methods Toolkit. In: Simpson, T. W., Jiao, J., Siddique, Z. and Hölttä-Otto, K., eds. Advances in Product Family and Product Platform DesignMethods & Applications. New York: Springer, pp. 245–269.

Krog, E.-H. and Statkevich, K., 2008. Kundenorientierung und Integrationsfunktion der Logistik in der Supply Chain der Automobilindustrie. In: Baumgarten, H., ed. Das Beste der Logistik. Berlin, Heidelberg: Springer, pp. 185–195.

Krumm, S., Schopf, K. D. and Rennekamp, M., 2014. Komplexitätsmanagement in der Automobilindustrie – optimaler Fit von Vielfalt am Markt, Produktstruktur, Wertstrom und Ressourcen. In: Ebel, B. and Hofer, M. B., eds. Automotive Management. Berlin: Springer, pp. 189–205.

Lechner, A., Klingebiel, K. and Wagenitz, A., 2011. Evaluation of product variant-driven complexity costs and performance impacts in the automotive logistics with Variety-driven Activity-based Costing. International MultiConference of Engineers and Computer Scientists, pp. 1088–1096. Hong Kong.

Lee, J. H., Kim, S. H. and Lee, K., 2012. Integration of evolutional BOMs for design of ship outfitting equipment. Computer-Aided Design vol. 44, pp. 253–273.

Li, J. R., Khoo, L. P. and Tor, S. B., 2002. A Novel Representation Scheme for Disassembly Sequence Planning. The International Journal of Advanced Manufacturing Technology vol. 20, pp. 621–630.

Liebler, K. M., 2013. Eine prozess- und IT-gestützte Methode für die Produktionsplanung in der Automobilindustrie. Thesis (Phd). TU Dortmund University.

Lim, S. C. J., Liu, Y. and Lee, W. B., 2010. A methodology for building a semantically annotated multi-faceted ontology for product family modelling. Advanced Engineering Informatics vol. 25, pp. 147–161.

López-Morales, V. and López-Ortega, O., 2005. A distributed semantic network model for a collaborative intelligent system. Journal of Intelligent Manufacturing vol. 16, pp. 515–525.

Luo, Y., Peng, Q. and Gu, P., 2016. Integrated multi-layer representation and ant colony search for product selective disassembly planning. Computers in Industry vol. 75, pp. 13–26.

Marwedel, P., 2011. Embedded system design: Embedded systems foundations of cyber-physical systems. New York: Springer.

Meissner, S., 2009. Logistische Stabilität in der automobilen Variantenfließfertigung. Thesis (PhD). Technical University of Munich.

Meyr, H., 2004. Supply chain planning in the German automotive industry. OR Spectrum vol. 26, pp. 447–470.

Min, S.-s., Zhu, X.-j. and Zhu, X., 2010. Mechanical Product Disassembly and/or Graph Construction. International Conference on Measuring Technology and Mechatronics Automation, pp. 627–631. Changsha City, China.

Moynihan, T., 2014. Apple's New Car System Turns Your Dashboard Into an iPhone Accessory[Online]. Wired. Available at: https://www.wired.com/2014/03/apple-carplay-system/ (Accessed 5 April 2017).

Nagel, J., 2011. Risikoorientiertes Anlaufmanagement. Thesis (PhD). Brandenburg University of Technology.

Noy, N. and McGuinness, D. L., 2001. Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory, Stanford University.

Pawlikowski, K., Fruhner, D., Klingebiel, K., Toth, M. and Wagenitz, A., 2016. Benefits of an Integrated Hierarchical Data Structure for Automotive Demand and Capacity Management. The Eleventh International Multi-Conference on Computing in the Global Information Technology, pp. 20–25. Barcelona, Spain.

Rapp, T., 1999. Produktstrukturierung: Komplexitätsmanagement durch modulare Produktstrukturen und -plattformen. Thesis (PhD). University of St.Gallen.

Riggs, R. J. and Hu, S. J., 2013. Disassembly Liaison Graphs Inspired by Word Clouds. Procedia CIRP vol. 7, pp. 521–526.

Romberg, A. and Haas, M., 2005. Der Anlaufmanager: Effizient arbeiten mit Führungssystem und Workflow - von der Produktionsidee bis zur Serie. Stuttgart: LOG X.

Schuberthan, J. and Potrafke, S., 2007. Die Anforderungen des Kunden… In: Gehr, F. and Hellingrath, B., eds. Logistik in der Automobilindustrie: Innovatives Supply Chain Management für wettbewerbsfähige Zulieferstrukturen. Berlin: Springer, pp. 8–18.

Schuh, G., 1988. Gestaltung und Bewertung von Produktvarianten: Ein Beitrag zur systematischen Planung von Serienprodukten. Thesis (PhD). RWTH Aachen University.

Schuh, G., 2014. Produktkomplexität managen: Strategien - Methoden - Tools. München: Carl Hanser Fachbuchverlag.

Sharon O'Malley, 2016. Product Rrecalls. In: SAGE Business Researcher: SAGE Publishing.

Song, S., Hu, D., Gao, Y., Yang, M. and Zhang, L., 2010. Product disassembly sequence planning based on constraint satisfaction problems. China Mechanical Engineering vol. 17, pp. 2058–2062.

Stäblein, T., 2008. Integrierte Planung des Materialbedarfs bei kundenauftragsorientierter Fertigung von komplexen und variantenreichen Serienprodukten. Aachen: Shaker.

Trojan, A., 2007. …und die Auswirkungen auf den 1st-Tier-Lieferanten. In: Gehr, F. and Hellingrath, B., eds. Logistik in der Automobilindustrie: Innovatives Supply Chain Management für wettbewerbsfähige Zulieferstrukturen. Berlin: Springer, pp. 12–16.

Vegetti, M., Leone, H. and Henning, G., 2011. PRONTO: An ontology for comprehensive and consistent representation of product information. Engineering Applications of Artificial Intelligence vol. 24, pp. 1305–1327.

Viganò, R. and Osorio Gómez, G., 2013. Automatic assembly sequence exploration without precedence definition. International Journal on Interactive Design and Manufacturing (IJIDeM) vol. 7, pp. 79–89.

Wagenitz, A., 2007. Modellierungsmethode zur Auftragsabwicklung in der Automobilindustrie. Thesis (PhD). TU Dortmund University.

Weinzierl, J., 2006. Produktreifegrad-Management in unternehmensübergreifenden Entwicklungsnetzwerken: Ein ganzheitlicher Ansatz zur Entscheidungsunterstützung im strategischen Anlaufmanagement. Thesis (PhD). TU Dortmund University.

Yang, Q., Pan, X., Wei, D. and Wu, K., 2012. Research on Individualized Product Requirement Expression Based on Semantic Network. Physics Procedia vol. 25, pp. 1926–1933.

Zagel, M., 2006. Übergreifendes Konzept zur Strukturierung variantenreicher Produkte und Vorgehensweise zur iterativen Produktstruktur-Optimierung. Thesis (PhD). University of Kaiserslautern.

Zeng, Y. and Gu, P., 1999. A science-based approach to product design theory Part II: Formulation of design requirements and products. Robotics and Computer-Integrated Manufacturing vol. 15, pp. 341–352.

INFORMATION MANAGEMENT ON THE EXAMPLE OF FEATURES OF IT SYSTEM "SINDBAD"

Mariusz Nepelski[(a)], Dominik Hryszkiewicz[(b)], Grzegorz Gudzbeler[(c)]

[(a),(b),(c)]Affiliation Police Academy in Szczytno

[(a)]m.nepelski@wspol.edu.pl, [(b)]d.hryszkiewicz@wspol.edu.pl, [(c)]g.gudzbeler@wspol.edu.pl

ABSTRACT
The aim of the article is to present the results of research on the construction of IT tools as a communication forum for improving information management in hierarchical institutions, especially the Police. The research are carried out as part of the project entitled „Construction of an IT system supporting communication in the Police and in other formations subordinated to the Ministry of the Interior in aspects of internal security – SINDBAD", No. DOB-BIO7/03/01/2015. The project is co-financed by the National Centre for Research and Development in Warsaw, for which a grant in the amount of 5 million PLN has been obtained. This information exchange platform is being developed by a scientific-research consortium led by the Police Academy in Szczytno. The result of the project will be an anonymous IT modular platform for collecting and analysing data to monitor problem areas and plan for interventions at the organizational level, including organizational and technical solutions that ensure the anonymity of the research participants. Importantly, the developed IT system will allow maximum anonymity of speech and security against unauthorized access of third parties.

Keywords: information management, internal communication, new technologies, security

1. INTRODUCTION
The aim of the article is to present the results of research on the construction of a system supporting information management in the Police and other services responsible for internal security of the state (Hryszkiewicz and Nepelski 2015). SINDBAD System (Information System for Research and Decision Analysis) will be built on the basis of a detailed analysis of the needs of future users and will be equipped with modules for the creation and management of research, legal consultations, crisis analysis and statistical analysis. The system will allow for maximum anonymization, while protecting against unauthorized access. The research are based on a grant obtained for the development of the project entitled "Construction of an IT system supporting communication in the Police and in other formations subordinated to the Ministry of the Interior in aspects of internal security – SINDBAD", No. DOB-BIO7/03/01/2015. The project is co-financed by the National Centre for Research and Development in Warsaw, for which a grant in the amount of 5 million PLN has been obtained. The paper presents partial research results, the final form of which will be achieved by the end of 2017. Currently the platform being developed is subject to specialist tests. More specific results cannot be presented at moment of publishing this paper – final results will be available after the project is finished (December 2017).

2. INFORMATION MANAGEMENT
Pondering over the management of information in hierarchical organizations, especially the Police, we shall begin by explaining the concept of communication. The term is derived from the Latin word *communico*, *communicare*, *communication*, which means *to consult* (Trzcińska and Wiciak 2011). In modern languages the term appeared in the fourteenth century and meant entering the community or maintaining a relationship with someone. Only in the sixteenth century, as a result of the development of postal services and roads, it was given another meaning – transmission, communication. This dual sense of the concept of communication is still valid. In one situation communication is a synonym of public transport, in another – transmission of information between people (Griffin 2003). Further scientific inquiry must take into account internal communication, which is perceived by the process of exchanging information between participants in the organizational system and linking the various parts of it. Effective communication in an organization occurs when information within an organization reaches the destination efficiently and in a form consistent with the intention of the sender. The way of transmitting the information should allow the recipients to fully accept the content and use it in their daily work (Dobek-Ostrowska 2016).
Therefore, communication is the process of creating, transmitting, receiving and interpreting messages between people – the process of transmitting information between the sender and the receiver. It runs on many levels, some of which are conscious and others are not. In order to communicate information about our beliefs, sensations, feelings, judgments, needs or expectations, we use verbal, nonverbal and vocal forms of communication. The process of communication directly translates into a good corporate image, including a hierarchical one, and starts with the daily professional service of every police officer and is dependent on his/her professional, communicative and narrative competencies (Stawnicka 2015). Practically

every police officer functions within two dimensions – internal and external, and the quality of functioning in each of them depends on constant improvement of internal communication. Research show that the opinion of the interlocutor is made in the first 10-30 seconds of the meeting/conversation and it has a significant impact on further contact. The first impression may be positive or negative. In the first situation people tend to attribute more positive (rather than actual) characteristics to the person and the case, their place of work and the institution they represent. Negative first impression discourages the interest in the person and the place he/she works at. Undoubtedly, the impression of professionalism during first meeting is influenced by:

- Competence/professionalism – the extent of knowledge and skills, and a well-founded belief in the ability to do something. Professionalism is about developing the skills in which we are the best.
- Credibility – to look and to act as we really mean to do something, to say that we are specialists and to do what we say and we know the conventions of our profession and move freely within them.
- Self-control – proves maturity. By engaging in emotion, we lose control of the situation, and thus lose our competence and credibility. We respond to the situations, but we do not take it out on our interlocutor (abreact).
- Confidence – comes from within, but reveals itself on the outside. We take care of its physical symptoms such as straight posture, eye contact, relaxed breathing. We shape mental symptoms such as positive attitude, awareness of competence, the use of positive language.
- Persistence – the art of building a positive image consists in consistency of behaviour. We take care of physical signs that help to maintain constancy, such as clothing, work pace, standards of conduct, procedures, vocabulary, jokes.

Information management involves both mastering and exiting the difficult situation and preparing for the possibility of its occurrence. Considering potential difficult situations usually allows us to prevent or at least reduce their adverse effects. There is no 100% perfect way to prepare for a difficult situation, because we never know what its course will be. Despite this, we should be prepared and ready to act. It is only after mastering the management of information that an organization can deal with the management of knowledge that arises from structured information, so that it can know what is going on within and around the sector, why it is happening, how it can work, and finally who or what the organization is (Drelichowski, 2004). Information policy is aimed at the environment, both proximate and distant, to provide the user as quickly as possible with comprehensive, reliable and relevant information, in terms of form and content, answering a given question about details of a given organization. However, the overriding objective is to improve the management of the organization itself (Flakiewicz, 2002). As it has been established above, information is a key tool in every organization and has a major impact on its management.

Issues related to information management and consequently internal communication are difficult to investigate, as participants with no anonymity may tend not to disclose their true views and fail to address important issues. This challenge has been taken up by the project that will develop an IT tool at IX (highest) level of technology readiness – ready to be deployed in practice at any time, which can be developed dynamically, even after the end of the project, wherever updates of its modules are needed. It is worth noting that component techniques used to implement sensitive system modules will allow for easy adaptation and change of functionality. Likewise, components of the external and internal access module will be designed in such a way that the system can continue to evolve irrespective of changes in the structure and content of the previously processed sources. It will be possible to add new, yet unknown sources, and include some of the metadata. In conclusion, the use of components will allow for continuous system development and will allow for easy updating of the system, so that the process of “aging” technology will be significantly reduced.

3. METHODOLOGY OF RESEARCH

The reason for undertaking the research on information management in hierarchical organizations, especially in the Police, was the awareness of the importance of communicating as a management system factor, as poor internal communication has a negative impact on the management system. There are several reasons why these studies have been undertaken. Firstly, internal communication as an element of information management is perceived as a function of modern management in the Police. It was the inspiration for the project’s creators. Secondly, it is no less important to indicate the direction of actions to maintain a positive image of the Police by, among others, properly shaping the organizational culture. These tasks are included in the priorities of the National Police Commander-in-Chief’s activities for 2016-2018, among others:

- Priority 1 – increasing the effectiveness of Police efforts to strengthen cooperation with the public.
- Priority 6 – improving the quality and effectiveness of Police work by successively increasing the professional competence of police officers and employees.
- Priority 7 – improving the quality of work carried out by the Police and police officers by providing optimal conditions of service/work.

The essence of the research is the diagnosis of the area of internal communication within the Police and predictions in this area. The presented research is interdisciplinary with regard to aspects of management, as well as psychological, linguistic, sociological, cultural and historical aspects. They are, among others, the subjects of study in the field of contemporary research on discourse. The starting point is police practice examined through the prism of conversational strategies. Against this backdrop, the research covered, among others, the communication competence of police officers, i.e. the ability of pragmatic communication activities, so the area of interest is the communication behaviours in hierarchical organizations, such as the Police. The research on internal communication covered several areas (Hryszkiewicz and Nepelski 2015, Stawnicka 2015):

- Theoretical background in social communication.
- Analysis of approaches to research on internal communication.
- Determinants of communication behaviour in the organization.
- Strategies of vertical communication in the Police.
- Strategies of horizontal communication in the Police.
- Public relations activities of the Polish Police through the prism of critical discourse analysis.
- The role of internal debates in optimizing the flow of information within police internal communication.
- Significators of deontical modality in police acts.
- The authority of the supervisor in the Police.
- The use of narrative marketing and narrative management as part of the communication strategy.
- The importance of internal communication for the effectiveness of preventive actions.

The problem of information management, including internal communication in hierarchical formations, especially in the Police, has not been a subject of interdisciplinary research so far, including pragmalinguistics. Emerging monographs on information management about the principles of effective communication are theoretical, scientific or popular science.
These studies are both theoretical and empirical. The first dimension manifests itself in a number of multidisciplinary analyses in the fields of communication, management, risk management, new and best IT solutions, to name the few. At the same time, experimental and development research are being carried out. The undertaken research is intended to enable the creation of an optimal IT tool for the formation of an intra-organizational forum. As a consequence, it will improve internal communication in such formations as the Police, and thus its functioning.

4. DESCRIPTION OF THE PROJECT – OUTCOMES OF THE RESEARCH

The main objective of the project is to create a system to support research in the population of police officers and workers of hierarchical formation, to monitor and predict threats, improve communication efficiency, consult and analyse crisis situations in the scope important to formations subordinate to or supervised by the minister responsible for internal affairs (formations responsible for internal security of the state). The system must be equipped with diagnostic and analytical tools enabling the design of research, data collection and analysis of results. Specific objectives have been defined as follows (Hryszkiewicz and Nepelski 2015):

- Defining the technical and substantive assumptions of the system, defining the critical functions of the system and thus providing effective methods of obtaining data useful for ensuring internal security by improving communication between the Police and other services subordinate to the minister competent for internal affairs.
- Clarification of areas of diagnosis of risk factors and their conditions, monitoring and minimization, elaboration of crisis analysis tools, review of internal risk factors, legal consultations and improvement of internal communication skills in hierarchical formations.
- Development of modular tools for legal consultations and reporting of innovative solutions.
- Development of algorithms for analysis of data collected with the use of diagnostic tools, supporting problem solving in the area of risk management, internal communication, crisis analysis and prediction of hazards in hierarchical formations.
- Development of a set of training scenarios for the use of the system, based on actual actions of the services subordinated to the Ministry of Interior, deemed as controversial. Based on the data from the system, it will be possible to collect data, formulate guidelines, integrate information between departments, eliminate language errors – communication misunderstandings, e.g. when issuing commands resulting, for example, from the specificity of the service. The developed scenarios will demonstrate the ability of the system to analyse the so-called difficult/problematic situations.
- Implementation and verification of developed tools. Diagnostic tools should be implemented in the form of an application – a network platform running in a secure Internet

connection. The platform should provide users with comfortable anonymous logins and unfettered expressions of feedback. Analytical tools should be implemented in the form of software that works with the diagnostic tool platform. They will serve as a tool to optimize the flow of information and decision-making processes in the areas described above. Ultimately, a tool should be created to optimize the course of information flow processes in the execution of tasks in the Police structures.

- Optimization of diagnostic and analytical tools, completed with demonstration of full functionality and realization of design assumptions in real conditions (network implementation). Development of a detailed recommendation to increase communication efficiency with the use of the proposed system. It must be demonstrated to be ready to install in a secure Internet connection.

This information system supporting communication in the Police and other forms subordinated to or supervised by the minister competent for internal affairs will be an innovative solution on a national scale. At present, any similar system does not function in any formation subordinate to or supervised by the minister responsible for internal affairs. The essence of the solution is to create it in such a way to meet the expectations and capabilities of the particular formations. The more technologically advanced the scope of information systems applications, the more difficult it is to implement them, taking into account the specific needs of the users and their real competencies (Drelichowski 2005). Therefore expectations are high. During the construction of the indicated system, innovative IT solutions will be necessary. Implementation of the project will significantly improve the process of information management. The proposed system will be used to carry out investigations into important internal issues of an organization, identification and monitoring of threats (which may be concealed in traditional reporting systems), as well as consultations on draft legislation from the point of view of officers at various levels. It also aims to support free expression of opinions on the cooperation of various services. Different solutions to improve communication efficiency will be implemented. The system will include ready-made diagnostic tools for research, hazard monitoring, consultation and crisis analysis. The system will feature openness and the ability to attach new measuring tools in the future as needed. Diagnostic tools will allow anonymous logins and therefore will have individual logins for users. The analytical module will be the software that performs data-processing functions to support the decision-making process associated with creating and improving the flow of information in the service hierarchy, both in crisis and routine situations.

The project promotes the use of new technologies in information management. The new system will improve the effectiveness of monitoring of social security expectations and needs as it is dedicated to uniformed, hierarchical formations. The need to ensure data security will also be ensured respecting the environment through the use of database solutions that support automatic backup or replication without the need to store large amounts of documents. Developing new methods of effective internal communication for hierarchical formations can help to rationalize resources, in line with the principles of human capital management, that is, to make more effective use of organizational resources being also in line with the postulates of sustainable socio-economic development. In practice, this means more efficient planning of human and material resources in hierarchical formations, including primarily in-house training and retraining systems, and employee competence management. Newly identified areas of organizational activity in hierarchical formations may also determine the need to create new jobs and improve the quality of work in the future. What is more, the rationalized system of effective internal communication in the future will emerge to improve the quality of work of police officers. The new information management system will improve the exchange of knowledge and experience in hierarchical formations. This in turn will enrich the pragmatism of prosecuting and detecting crimes and offenses. In principle, an information campaign should be developed and implemented, disseminating knowledge among the officers about the project, its security and its use.

Within the framework of the project a tool will be developed, which will be on the IX (highest) level of technological readiness, ready to be implemented in practice at all times. However, it seems expedient to implement it as quickly as possible in order to improve the management of information, including internal communication in hierarchical formations.

The system will include ready-made diagnostic tools for research, hazard monitoring, consultation, and crisis analysis. The system should also be “open” and allow the inclusion of new measuring tools as needed in the future. It is assumed that in the described system it will not be possible to express opinions on personal matters. Diagnostic tools should enable anonymous logging and therefore individual logins should be provided for officers and staff members of formations supervised by the minister responsible for internal affairs, and the distribution method of logins will be developed. The system will be provided with a number of cooperating modules. The crisis analysis module will be used to obtain information from many formations, enabling analysis and improvement of the algorithms for interoperability. The analytical module should be the software that performs data-processing functions to support the decision-making process associated with creating and improving information flow in the service hierarchy, both in crisis and routine situations. Training

scenarios will be developed, based on the actual activities of formations subordinate to or supervised by the minister in charge of internal affairs, considered as controversial and difficult. On the basis of data from the system it will be possible to collect data, formulate guidelines, integrate information between the particular formations. The developed scenarios will demonstrate the ability of the system to analyse difficult situations.
Documentation of the system will consist of two parts: technical information and a part concerning the substantial information about the system. The technical part should describe how the system is built, installed and used. In particular, it will include the description of the method of anonymous data input and how it should be designed and implemented, a part of which will be the user manual (for persons taking the test) introducing the system. The part concerning the substance of the system should be a description of the theoretical basis for the construction of the implemented measuring tools (including the scientific data obtained in the process of developing the system).
The development of a variant based on an independent secure network connection, as well as the development of final decision on location of the system, will be undertaken by the end-user (compliance with the standard called Police Network of Data Transmission – PSTD, will be required). The police should be able to implement the system both in the internal network (PSTD) as well as in the external secure Internet connection. Other users should be able to use the system in a secure Internet connection, which will be limited. The system should provide the ability to integrate databases and reports from all users. System users will be:

- the structures of the interested formations subordinate to or supervised by the minister responsible for internal affairs, which scope of activities requires decision-making based on studies of officials and employees, as well as elaboration and implementation of strategic decisions.
- Research units carrying out research for services subordinated to the minister in charge of internal affairs.

The implementation of the system may also be potentially desired by all organizations responsible for internal security of the country, whose actions are necessary to gather honest and credible statements among their members (employees). The capabilities of anonymous threat reporting can also be useful for airlines, railways and large industrial plants. Therefore, the system can be used according to the project assumptions, i.e. by the state services, and it is possible to be used for future commercialization for other organizations.

5. ADDITIONAL POSSIBILITIES OF THE USE OF "SINDBAD" SYSTEM'S FEATURES IN THE POLICE

Another area of SINDBAD's functionality is the diagnosis and forecasting of training opportunities in Poland's largest uniformed formation, namely the Police.
According to the Directive of the Minister of Interior and Administration of 19 June 2007 on detailed conditions for vocational training and further training at the Police, it includes, in particular, vocational trainings and further trainings. Both identified forms of training are centrally implemented according to training and in-service training programmes approved by the Police Commander-in-Chief. As part of the further training in the Police, the adopted range of specialist courses and other undertakings is complementary to the knowledge and professional skills, in particular to acquire, update, extend or deepen the knowledge and professional skills of police officers required to perform their tasks and duties within the framework of individual Specialties, such as training programmes in the area of combating economic, drug or computer crimes, human trafficking, criminal investigation, forensic techniques, police operations and many others. There are currently three kinds of vocational trainings in the Police, i.e. central, local and external. The first one, is organized and implemented in the police schools and the Police Academy in Szczytno in the form of specialized courses – the programmes of which are introduced into service by decisions of the Police Commander-in-Chief, and also in the form of other ventures. Local trainings, the so-called "field trainings" are organized and implemented by police units, allowing for the free formulation of training contents, and consequently smooth modifications of programming content in response to quality changes in crime. The last type of training – external training, is organized by non-police entities, especially when the diagnosed training needs for various reasons cannot be delivered in the framework of central or local trainings.
The Directive of the Minister of Interior of 16 April 2015 amending the regulation on detailed conditions for vocational training and further training in the Police indicates that the Police Academy in Szczytno takes over the task of recognizing training needs, which before was the responsibility of the National Police Headquarters. Previously, recognition of training needs was held twice a year, which resulted in the situation when the same police officers were being directed for each course and training, and this did not reflect the actual needs. Currently, the diagnosis of needs is held once a year, i.e. by 20 October each year, taking into account the needs identified by 30 September this same year. After this date, only adjustments to current demands are possible.
A well-conducted training needs assessment allows us to decide on which training area to focus, how long the training should last for participants to acquire new competences, learn motivations, attitudes and

behaviours of participants, explore current knowledge and skills of future participants, so the training is most effective (e.g. by selecting the appropriate methods).

The process of diagnosing training needs should first of all enable problems to be identified, of course from the point of view of the purpose of the formation, to indicate whether they can be removed (solved) through training. The simplest way of identifying problems is to study the accomplished tasks (error identification), examine new tasks/challenges the particular formation is facing at the moment, e.g. due to new legal regulations (identification of necessary competences), and the gathering of opinions of executives and officers themselves.

The described objectives of SINDBAD system are directly embedded in the area of diagnosing training needs for individual types of services, individuals, but also the entire formations. This is enabled by functionalities such as intra-organizational research and legal consultations, i.e. the main features of the system. The appropriate development of research tools (questionnaires) and directing them to the target group of respondents, coupled with the provision of anonymous research, will be an excellent support, for instance for the Police Academy in Szczytno to fulfil the provisions of the above-mentioned regulation, i.e. the diagnosis of training needs at the central level. It is also possible to use the built-in solutions for diagnosing and forecasting training needs in individual units or types of services, by commissioning appropriate research.

The use of SINDBAD system to diagnose, but also to forecast training needs can be implemented in two dimensions:

- Complementary
- Target.

The first dimension would only be the completion of the already implemented diagnostic process, and on this basis forecasting the training needs of the Police. The other one would replace the current mechanism (current collection of training needs via email/fax/written form) through appropriately designed tools (questionnaire surveys) distributed among, for example, targeted units/cells responsible for the training process in each police unit. On the basis of the collected data, the system would generate the required training needs lists for the current training offer and would suggest training areas not included in the training offer as well.

6. CONCLUSION

The SINDBAD project will enable the creation of an IT tool that will shape the discussion forum within hierarchical organizations, especially in the Police. This will improve the management of information and internal communication in such security-conscious formations as the aforementioned Police service. It will enable the speed, reliability and adequacy of conveying information. It will undoubtedly improve the communication skills of executives at all levels of management. This in turn will ensure consistency in the transfer of information within the organization. This is linked to feedback that is reflected in improving the quality of the work of officers. The final form – the discussed IT system – will be developed and implemented by the end of 2017.

REFERENCES

Allwood, J., Scroeder R., 2000. Intercultural Communication in a Virtual Environment. Journal of Intercultural Communication, November, issue 4. http://www.immi.se/intercultural/ (27.03.2017).

Blair, D. C., 2002. Knowledge Management: Hype, Hope, or Help? Journal of The american Society for Information Science and Technology, 53 (12), pp 1019-1028.

Dobek-Ostrowska B., 2016. Komunikowanie polityczne i publiczne, Warszawa, p. 61.

Drelichowski L., 2004. Podstawy inżynierii zarządzania wiedzą. Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą, p. 14.

Drelichowski L., 2005. Interakcje teorii i praktyki zarządzania wiedzą. Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą, pp 36 - 41.

Flakiewicz W., 2002. Systemy informacyjne w zarządzaniu. Uwarunkowania, technologie, rodzaje, Warszawa.

Trzcińska W., Wiciak I., 2011. Skuteczne komunikowanie w administracji publicznej, Szczytno, p. 60.

Griffin E., 2003. Podstawy komunikacji społecznej, Gdańsk.

Stawnicka J., 2015. Projekt naukowo-badawczy „Strategia komunikacji wewnętrznej w Policji", Katowice.

Tubbs, S.L., Moss S., 2003. Human communication: Principles and contexts. (9th ed.). New York.

Hryszkiewicz D., Nepelski M., 2015. Development project entitled „Construction of an IT system supporting communication in the Police and in other formations subordinated to the Ministry of the Interior in aspects of internal security", No. DOB-BIO7/03/01/2015.

AUTHORS BIOGRAPHY

Mariusz Nepelski, PhD, Dr habil. – Associate Professor specializing in the field of security sciences, professor extraordinary of the Police Academy in Szczytno. He carries out works related to the development project entitled „Construction of an IT system supporting communication in the Police and in other formations subordinated to the Ministry of the Interior in aspects of internal security", No. DOB-BIO7/03/01/2015.

Dominik Hryszkiewicz, PhD – specializes in the field of security sciences, he is the Director of the Institute of Social Sciences of the Department of Administration of the Police Academy in Szczytno. He is the manager of the development project entitled „Construction of an IT system supporting communication in the Police and in other formations subordinated to the Ministry of the Interior in aspects of internal security”, No. DOB-BIO7/03/01/2015.

Grzegorz Gudzbeler, PhD – specializes in the field of security sciences, he is the Director of the Institute for Coordination of Research and External Funds of the Police Academy in Szczytno. He carries out works related to the development project entitled „Construction of an IT system supporting communication in the Police and in other formations subordinated to the Ministry of the Interior in aspects of internal security”, No. DOB-BIO7/03/01/2015.

PARAMETER SELECTION FOR "PCMA*" USING SURROGATE-ASSISTED BLACK-BOX OPTIMIZATION AND MICROSCOPIC TRAFFIC SIMULATION

Bernhard Werth[(a)], Erik Pitzer[(b)], Michael Affenzeller[(c)], Christian Backfrieder [(d)]

[(a),(b),(c)] Heuristic and Evolutionary Algorithms Laboratory, University of Applied Sciences Upper Austria, Hagenberg
[(a),(c)] Institute for Formal Models and Verification, Johannes Kepler University, Linz, Austria
[(d)] Research Group Networks and Mobility, University of Applied Sciences Upper Austria, Hagenberg

[(a)]bernhard.werth@fh-hagenberg.at, [(b)]erik.pitzer@fh-hagenberg.at, [(c)]michael.affenzeller@fh-hagenberg.at
[(d)]christian.backfrieder@fh-hagenberg.at

ABSTRACT
Routing algorithms have become significantly more sophisticated in recent years as increasingly more data on road networks, vehicles and drivers becomes available. The correct parametrization of such algorithms is a difficult task since the underlying traffic model needs to reflect the complexity of the algorithm and should resemble reality closely. A major drawback of such models is that evaluating a certain parameter setting may become computationally expensive, preventing the use of conventional optimization algorithms. In this work a combination of surrogate assisted black box optimization and microscopic traffic simulation is used to optimize the parameters of a recently published routing algorithm.

Keywords: traffic simulation, surrogate assisted optimization, evolutionary algorithms, noisy optimization

1. INTRODUCTION
With the ever increasing number of vehicles especially in urban areas, the issue of traffic congestion becomes more severe. The Urban Mobility Report (Schrank, Eisele, and Lomax 2012) emphasizes the need for new solutions in battling the growing costs created by traffic jams in US American cities. Microscopic traffic simulation could help finding such solutions as it is a powerful tool to evaluate the impacts of many different influences on a road network. Traffic light controls, the percentage of automatic vehicles, probabilities of accidents or different routing algorithms could be examples of such influences.
Recently a new predictive routing algorithm called *"Predictive Congestion Minimization in Combination with an A*-based router"* (PCMA*) which is based on predictions about bottlenecks and traffic jams in road networks has been proposed (Backfrieder et al 2017). In their experiments the authors noted that correct calibration of the thresholds used in their algorithm is a rather complex task and might even be impossible in a global manner as these thresholds and parameters depend on the specific road network considered as well as on the distribution of vehicles and their start and end points. Since mathematical representations for microscopic traffic simulations are usually not available, the calibration of such parameters on a specific scenario has to be treated as a black-box optimization problem. However, the use of conventional optimization algorithms like *Genetic Algorithms*, *Particle Swarm Optimization* or *Tabu-Search* is hindered by the fact that they require a large number of function evaluations to converge, which is infeasible when a single simulation run for testing one parameter configuration takes several minutes or even hours depending on the scenario and the degree of parallelization. The use of cheap surrogate models to facilitate the search is therefore imperative.
The remainder of this work is structured as follows: Section 2 provides an overview of surrogate-assisted optimization techniques and which of those are of relevance in the scenario at hand. Section 3 describes the different classes of routing algorithms and very briefly outlines the characteristics of the PCMA* algorithm. In Section 4 the optimization problem with its parameters and bounds as well as the algorithm and surrogate model type are described. Section 5 contains computational results, followed by conclusions in Section 6.

2. SURROGATE-ASSISTED OPTIMIZATION
Using surrogate models in optimization tasks is an established method and often used when the evaluation of solution candidates is either time consuming or associated with additional costs. Surrogate-assisted optimization has been successfully applied to various simulation scenarios including Navier-Stokes flow solvers for aerodynamic simulations (Han 2013), finite element simulations for fluid dynamics (Forrester and Keane 2007), flowsheet simulations in manufacturing environments (Boukouvala and Ierapetritou 2013) and mesoscopic traffic simulation (He 2014).
The main idea of surrogate-assisted optimization is to use interpolation and statistical models that are considerably cheaper to evaluate than performing a full simulation. Therefore, surrogate-based optimization strategies should be able to find good solutions within a considerably smaller number of simulations than conventional optimization algorithms would require. Much alike the wide variety of application scenarios for surrogate-assisted optimization, a considerable number

of algorithms, algorithm extensions and modifications exist. *Particle Swarm Optimization* (Sun et al. 2015), various versions of *Evolutionary Strategies* (Loshchilov, Schoenauer, and Sebag2013) and even interactive *Genetic Algorithms* have been enhanced with surrogate models.
In this work we decided to focus on an adaptive sampling scheme similar to the one described in (Wang and Shan 2007), depicted in Figure 1. First, an initial set of samples is created and evaluated. In the next step a Gaussian process model is constructed, which can then be used to select a solution candidate that is optimal with respect to some infill criterion. This candidate is then evaluated on the expensive simulation. After obtaining a result value from the simulation, the process can be started anew until a certain computational budget is exhausted.

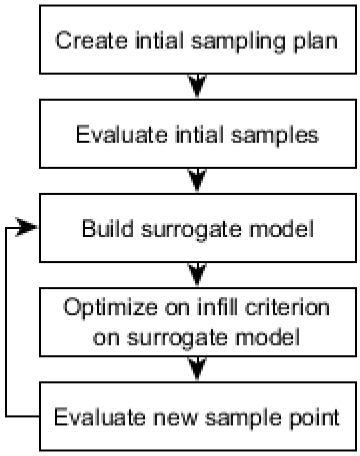


Figure 1: Adaptive Sampling Scheme

A variety of different types of models have been used in this context, with *polynomial regression*, *neural networks*, *support vector machines*, *radial basis function regression* and *Gaussian process regression* being among the most popular. Gaussian processes have the additional benefit of providing a measure for the models own uncertainty, which has been used in the *expected improvement* infill criterion in the widely used "Efficient Global Optimization" algorithm (EGO) by (Jones, Schonlau, and Welch 1998). The original expected improvement criterion was designed for deterministic black box functions without noise. Several new infill criteria for noisy optimization have since been proposed. A comprehensive overview and a benchmark comparison of such criteria for noisy optimization are given in (Picheny, Wagner, and Ginsbourger 2013). They are of special interest for this work since traffic systems are inherently influenced by many random effects ranging from drivers characteristics like attention span, reaction time, or willingness to take risks to outer influences like weather, slight variations in departure times or technical failures of vehicles and traffic control systems.

3. A PRIMER ON ROUTING ALGORITHMS

Nowadays, camera systems, traffic flow sensors, traffic control authorities and public transport agencies continuously produce large quantities of data concerning the condition of roads, vehicles, traffic lights and other infrastructure. Routing algorithms and automatic traffic guidance systems have evolved with this increased development in traffic surveillance infrastructure and can be roughly divided into three categories:
Dynamic Traffic Assignment systems incorporate theoretical traffic optimization solutions with the goal of achieving a suitable user equilibrium where no vehicle can achieve a shorter time to reach its destination by changing its route (Janson 1991). Major drawbacks of these methods are the lack of ability to react to unforeseen congestions and a high computational effort.
Reactive Routing Algorithms take the current traffic situation and condition of the road network into account to dynamically change the routes of vehicles in the vicinity of current congestion areas (Kesting, Treiber, and Helbing 2010). Unfortunately, some reactive algorithms provide the very same alternative route to all vehicles in a system, thereby effectively moving the congestion from one part of the road network to another rather than actually solving the problem.
Predictive Routing Algorithms use historical data to predict future congestions via a variety of models, which limited their application to highways or higher level interurban roads, as data for lower level roads was scarcely available. Recent developments like the rising number of trackable smart devices, may change this situation in favor of predictive routing systems.
The comparison of such algorithms can be done on traffic simulations with different levels of abstraction. In this work we chose microscopic traffic simulation as our means to evaluate and compare different algorithm parametrizations, as microscopic simulation not only captures reality more closely than a macroscopic model, but also permits to model aspects of reality that are not considered by the algorithm.
The PCMA* algorithm (Backfrieder et al 2017) uses both a predictive and a reactive routing strategy. It combines a routing algorithm that takes the current situation of the road network into account by weighting edges in the routing graph via the *Speed Average* or *Greenshield's* method (Backfrieder et al 2017, Greenshields 1935) with a congestion prediction method that is based around the calculation of so called "footprint" values for individual intersection. These footprints are influenced by information stemming from vehicular communication.
As rerouting all vehicles simultaneously just shifts congestions from one place to another, no vehicle should be treated exceptionally unfavorably and permanent rerouting of the same vehicle might upset the driver, PCMA* employs a strategy for selecting which vehicles to reroute when a congestion is either predicted or dynamically detected.
The predictive and reactive parts of the algorithm and the selection of rerouted vehicles use thresholds to decide

when a congestion occurs or a vehicle should be exempt from further rerouting maneuvers. The predictive component additionally has parameters specifying how far into the future predictions are considered relevant. These parameters need to be fine-tuned to the specific scenarios and road networks in order to achieve optimal results.

4. COMPUTATIONAL EXPERIMENTS

The goal of the optimization task will always be the minimization of the cumulative time spent by all vehicles on the road network. The seven real-valued parameters listed in Table 1 should be minimized and therefore constitute a solution candidate. Evaluation is performed via a full simulation in the microscopic traffic simulation framework TraffSim (Backfrieder, Mecklenbräuker, and Ostermayer 2012).

Table 1: Parameter ranges for the PCMA*

Name	Min	Max
Congestion level	1	15
Speed average congestion threshold	0	1
Rerouting threshold distance (reactive)	0	10
Rerouting threshold distance (predictive)	0	10
Congestion threshold (Vehicles per second)	0.1	3
Footprint prediction time resolution	5000	60000
Footprint prediction maximal forecast	1200000	2600000

Congestion level is relevant for the reactive routing component only and defined as the approximate size of a traffic congestion and all vehicles within this many road segments will be subjected to rerouting.

When the mean speed of vehicles passing a traffic node is below the *speed average congestion threshold* a congestion is detected by the reactive part of the algorithm.

It is unwanted for a vehicle to switch to an unreasonably longer alternative route as this would be perceived as an unfair treatment. If the ratio of the new route's length to the old route's length is larger than the rerouting threshold distance, the alternative route is no longer considered feasible. This threshold can differ for the reactive and the predictive parts of the algorithm.

When the predicted number of vehicles that should pass a certain node of the traffic graph exceeds the threshold specified by the *Congestion threshold* the predictive component will predict a congestion.

Footprint prediction time resolution and *Footprint prediction maximal* forecast specify the time frames for footprint calculation and the limit at which point predictions into the future are no longer considered relevant. Both parameters only concern the predictive part of the PCMA*.

Tests are performed on two different road networks. One is a partial network of the northern part of the city of Linz and the other one is artificially generated with the goal to resemble American downtowns (Lindorfer et al. 2013). Start and end nodes of the vehicles are chosen so that most of the vehicles follow a general direction through the network as it would happen in a morning traffic peak where more traffic happens towards the industrial areas of a city.

As stated in Section 2, the infill criteria created for noisy black box optimization are especially interesting for the task at hand. Table 2 lists the infill criteria used in our experiments.

Table 2: Tested Infill Criteria

Name	Published by
(noiseless) Expected Improvement (EI)	Jones, Schonlau, Welch (1998)
Augmented Expected Improvement (AEI)	Huang et al. (2006)
Plugin Expected Improvement(PEI)	Osborne, Garnett, Roberts (2009)
Expected Quantile Improvement(EQI)	Picheny, Ginsbourger, Richet (2010)
Minimal Quantile Criterion(MQC)	Cox, John (1992)
Expected Quality(EQ)	
Neighbor Distance (ND)	

Additionally to the proposed criteria, the strategy of using the model prediction directly as an infill criterion (Expected Quality) and the approach of maximizing the distance to the nearest neighbor (Neighbor Distance) are used to generate baselines for maximal exploitation (EQ) and maximal exploration (ND) are employed.

The shape of the road network might not be the only factor influencing the optimal parameter selection for the PCMA*, therefore, the number of vehicles in each scenario is varied to be 500, 1000, 1500 and 2000. Also the distribution of departure times might have a considerable impact on the performance of collaborative routing algorithms which is why we chose two different distributions: a uniform distribution with start times in [t, t+30 min] and a normal distribution with $\mu = t$ and $\sigma = 15$ min. Since neither weather nor time of day are considered within the simulation the parameter t has no influence on the results and is set to an arbitrary but fixed value.

Initialization: For creating the initial sampling plan a set of 50 points that are almost-optimal with respect to the well-known maximin criterion (Johnson 1990), which is defined as maximizing the minimal distance between all points of a set, is created using an evolutionary strategy (Beyer and Schwefel 2002) with $\mu = 20$ and $\lambda = 70$ and 500 generations.

Model Building: The surrogate model used is a Gaussian process regression with a constant mean function and a rational quadratic covariance kernel that assumes a homoscedastic error. The hyper parameters of each new

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

model are selected by optimizing the log-likelihood with a gradient descent algorithm. Since the gradient descent is susceptible to becoming stuck in local optima, an additional model with the hyper parameters of the previous iteration is built and used if the gradient descent cannot provide a model with a smaller root-mean-squared error.

Optimize Infill Criterion: The infill criterion is optimized using a covariance matrix adaption evolution strategy with a population size of 50 (Hansen and Ostermeier 1996). The best found solution from this algorithm is then evaluated with a full TraffSim simulation.

5. RESULTS

In Figure 2 six convergence curves corresponding to the five infill criteria listed in Table 2 are plotted. (Minimal Distance was omitted due to visibility reasons). Two conclusions can be drawn from this figure: Firstly, the difference between most of the infill criteria is different to evaluate and could be influenced by random effects introduced both by the heuristic nature of the optimization and the stochasticity of the simulation. Secondly, several solutions with considerably worse-than-average-quality exist, the frequency of which is illustrated in a histogram (Figure 3) of 1000 uniformly randomly sampled Parameter Vectors. Here each vector was evaluated 20 times, as to remove the impact of simulation specific randomness.

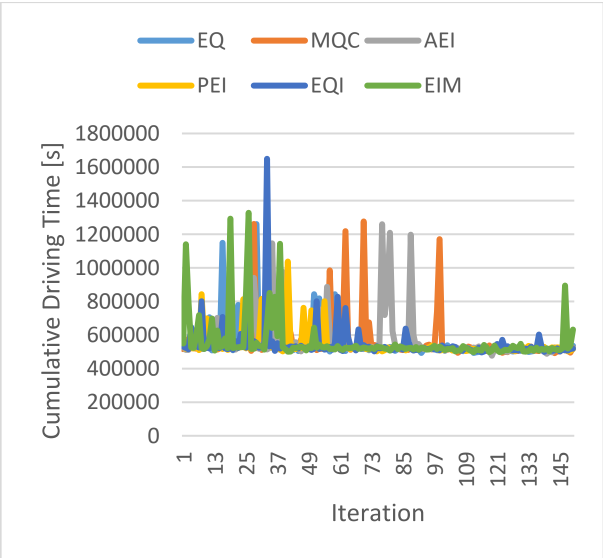


Figure 2: Quality curves of different infill criteria.

The impact of these outliers is twofold. Firstly, they are wasted evaluations that are not only bad in terms of quality but also in terms of execution time since the simulation takes longer to evaluate. Secondly, the strong difference in quality causes undesirable effects in the Gaussian process regression, as the model not only assumes an unjustifiably large variance but also leads the algorithm very closely to the edge of such outliers. Figure 4 shows a Gaussian process regression approximating a one dimensional function with a discrete jump in the fitness landscape. The regression model cannot handle the leap in fitness and oscillates in the vicinity of the jump. If the oscillation is larger than the potential fitness gains of the lower plateau, the algorithms can no longer rely on the regression model and at best fall back into random search and at worst are purposely misled by the model towards the fitness jump.

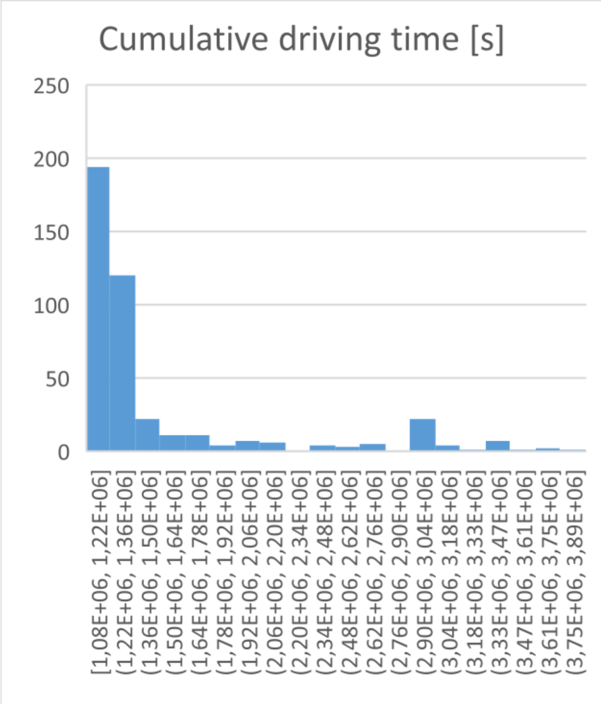


Figure 3: Histogram of cumulative driving times [s]

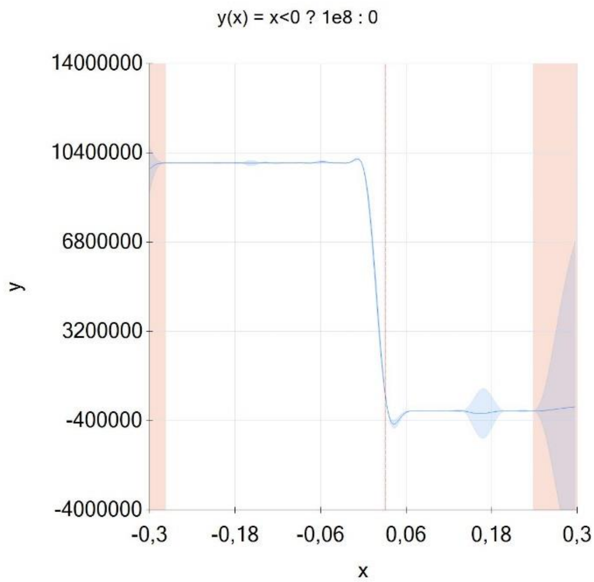


Figure 4: Gaussian process model of a fitness jump

It is therefore imperative to somehow handle these outliers. One option could be to log-transform the fitness landscape, however doing so would create asymmetrical (biased) errors and violate the assumptions of not only the Gaussian process but most regression models in

general. Further, the outliers cannot be attributed to random effects of the traffic simulation and are not distributed randomly in the search space.
In order to explain these outliers Figure 5 presents a projection of the reduced parameter space that was created via t-distributed stochastic neighborhood embedding *tSNE* (Maaten and Hinton 2008). (The most influential parameters were determined by symbolic regression and their impact values were used to weigh the Euclidean distance between points which were fed to the tSNE to determine neighborhood structures). As can be seen the outliers (orange large dots) are located at the edges of the search space. Increasing the lower bounds for the most relevant parameters (rerouting threshold distance (reactive), congestion threshold, congestion level and speed average congestion level) to (1.5, 0.535, 3 and 0.2) respectively removed the regions with extremely bad quality results.

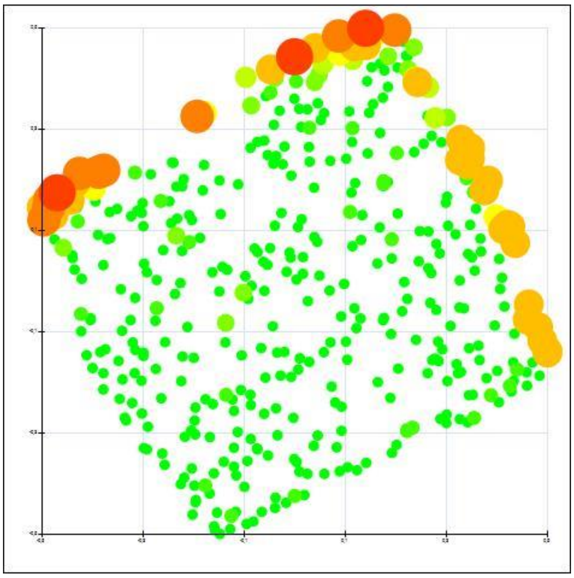


Figure 5: Weighted projection of random samples in the search space. Large red dots denote high objective values while small green denote small objective values

After applying the optimization algorithms to the reduced search space the convergence curves like the one seen in Figure 6 appear to be more similar to what one would expect from an optimization process on a stochastic minimization problem. There is a certain fluctuation that can be attributed to stochastic noise but also a drop in the objective values starting at the 50-samples-mark where the algorithm switches from random sampling to optimizing its infill criterion.
Figure 7 displays the box plots of the best qualities found by optimization algorithm with different infill criteria. Each algorithm configuration was run three times to offset the stochasticity introduced by both the simulation and the heuristic optimization. The Plugin Expected Improvement and the Augmented Expected Improvement appear to be the better infill criteria while the baseline criteria ExpectedQuality and NeighbourDistance are amongst the weaker. It is however important to notice that all achieved quality values lie roughly within the range of one standard deviation of the error measured for each specific problem instance (Standard deviation for the Linz road network with 1500 vehicles: 57629.78s; Standard deviation for the random road network with 1500 vehicles : 25041.30s)

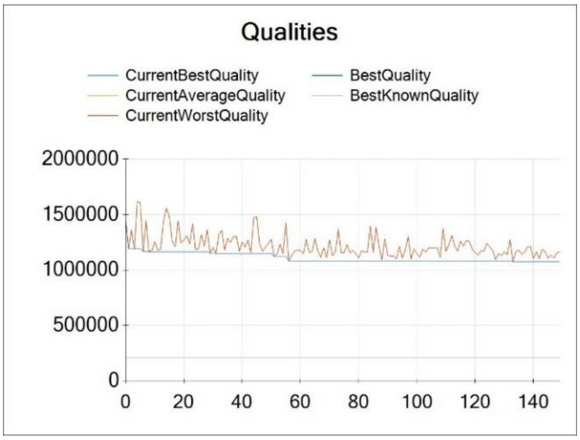


Figure 6: Convergence curve of EGO with Plugin Expected Improvement

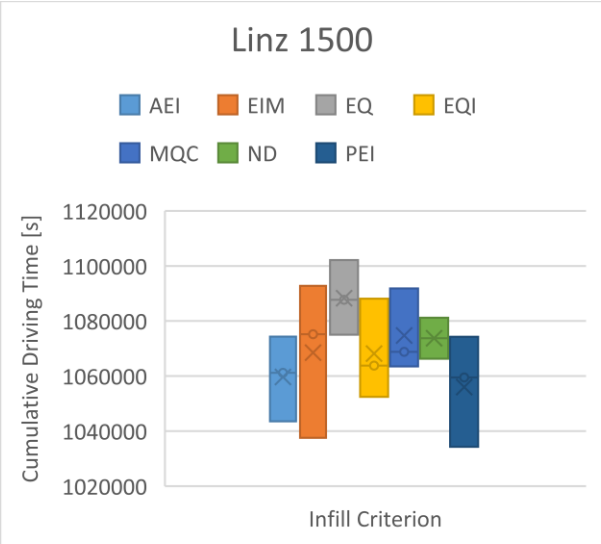


Figure 7: Achieved objective values on Linz with 1500 vehicles (reduced search space)

To gain insight on the almost equal performance of the different infill criteria all evaluated parameter settings from the comparison experiment were used in a linear projection of the search space which is displayed in Figure 8. As can be seen the reduced search space covers a relatively simple part of the fitness landscape where most of the best solutions lie in a valley that stretches diagonally across the projected space.
To gauge the impact of the specific distribution, the uniform distribution of start times was replaced with a normal distribution as described in Section 4. Figure 9 shows a mostly similar fitness landscape for the normal distribution as it was for the uniform distribution. The valley has not changed its position only the area in the lower left corner of the space has gone down in objective

values. Indicating that at least in this specific scenario peaks of high traffic might even be more forgiving to bad parametrization of the routing algorithm than continuous streams of traffic.

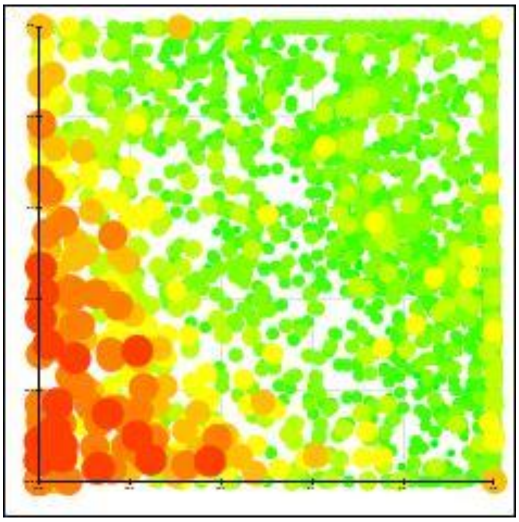


Figure 8: Linear projection of the reduced search space (Linz, 1500 vehicles)

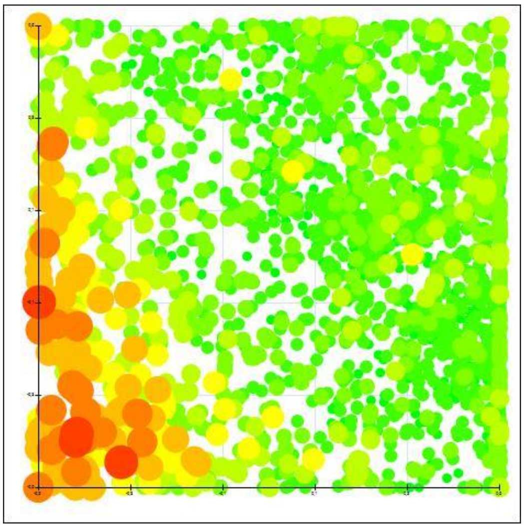


Figure 9: Reduced search space (Linz network, 1500 vehicles, normal distributed start times)

While all infill criteria produce fairly similar objective values, they all are in the lower 10% of the produced points. When looking at the best solution candidates produced by the optimization they all can be found in the valley region of the search space. Therefore the argument can be made that no infill criteria stands out in the boxplots, because the fitness landscape is simple enough that all variants of the EGO algorithm manage to converge towards the favorable region and the ability of an algorithm to converge very precisely is of less significance due to the noise in the fitness function.
Figure 10 uses the same weights as Figure 8 but is created using the samples evaluated while applying the different algorithm configurations to the randomly generated road network. While the same valley structure can be seen, it is important to notice that the valley itself moved a bit, while the red "hill" of unfavorable parameter settings encompasses more area. Some parameter configurations for PCMA* that were optimal for the road map of Linz are now amongst the worse solutions of the search space. Additionally the valley is no longer as well defined as before which can be explained by measuring the different level of noise on both scenarios. Figure 11 shows the best objective values found by all EGO-variants. Again no clear ranking of performance can be established, except for the neighbor distance criterion which seem to perform worst, potentially indicating that more exploitation than in the other scenario is required to find the bottom of the valley.

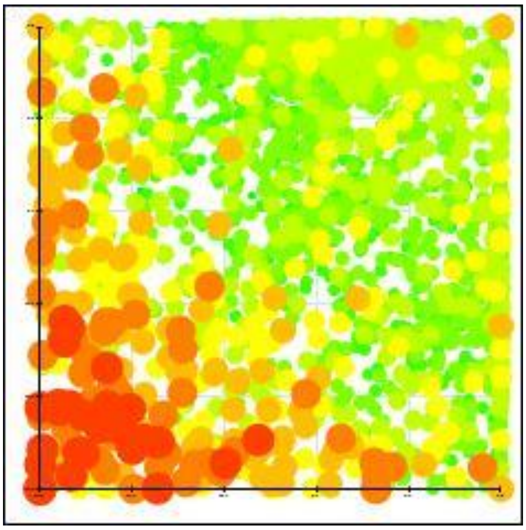


Figure 10: Reduced search space (Random network, 1500 vehicles)

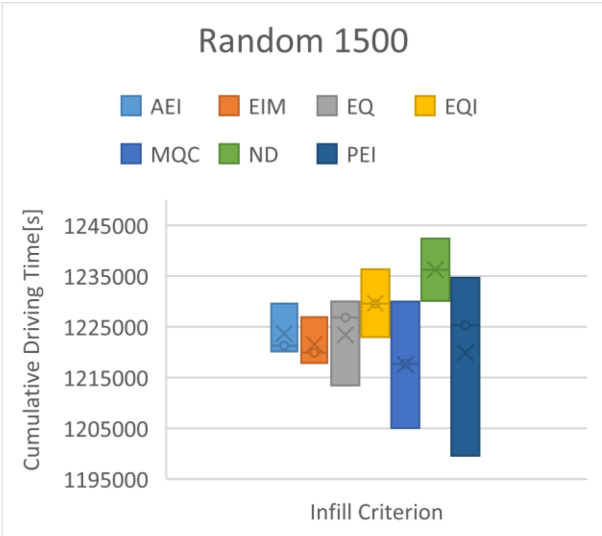


Figure 11: Achieved objective values on the generated road network with 1500 vehicles (reduced search space)

In order to measure the impact of vehicle density on the optimal parameter selection of the routing algorithm the number of vehicles were varied. Figure 12 compares the results achieved on the Linz network for 500 vehicles.

Figure 12: Achieved objective values on Linz with 500 vehicles

As expected, the variance in objective values diminishes with lower numbers of vehicles on the network. The differences in quality by the different infill criteria is neigh irrelevant for scenarios with few vehicles as not as much rerouting needs to happen as can be seen in Table 3 which shows the median cumulative driving times (CDT) for random parameters selections, the best found selection and the CDT achieved without any routing. The column “range” contains the range of all best CDTs achieved by each algorithm run, illustrating that while no essential difference between the performance criteria emerges, all EGO-variants managed to achieve substantial improvement, which increases with the number of vehicles on the network. Interesting to note is that for 500 vehicles on the Linz scenario several parameter settings are even disadvantageous so that the median CDT for random parametrization is effectively worse than disabling the routing algorithm altogether.

Table 3: Comparison of different achieved cumulative driving times

Scenario	Median CDT [s]	Best CDT[s]	Unrouted CDT[s]	Range [s]
Linz 1500	1294988	1034292	3426733	67847
Linz 1000	511211	481152	1243908	16322
Linz 500	220712	206649	208502	927
Random 1500	129967	1199615	4959762	42771

6. CONCLUSIONS

Several conclusions could be drawn from the results presented in the previous section.

- Firstly, that surrogate-assisted algorithms can suffer from unhandled outliers or sharp cliffs in the fitness landscape depending on the type of models that are used.
- Secondly that when it comes to parameter tuning for algorithms, it is rare that all parameters have the same level of impact on the overall performance. It might be a promising extension of existing surrogate-assisted optimization algorithms to include some form of dynamic feature selection.
- Thirdly, an argument can be made that well designed algorithms, be that routing algorithms or otherwise, should have at least somewhat predictable trends in their parameter settings and if the fitness landscape created by these parameters were riddled with local optima, tuning these parameter by hand would be neigh impossible.
- Fourthly it should be noted, that visual inspection of this landscape based on feature selection and dimensionality reduction techniques like principal component analysis, neighborhood component analysis, tSNE or similar methods, can provide significant insight concerning the performance of surrogate-assisted optimization algorithms and the difficulties they face on different problems.
- Fifthly, that when the optimization problem is very noisy, finding an algorithm that outperforms existing ones by a few percent might be fairly inconsequential as these improvements could be offset by small changes to the fitness landscape by choosing a different range or scaling for an input parameter.

The correct selection of parameters for PCMA* is more relevant and more difficult, the more vehicles are on a road network. The impact of the different road networks on the optimal parameter setting is difficult to estimate, but the results somewhat indicate that larger networks might increase the difficulty of selecting the best parameter settings as noise and uncertainty rise.
Future research in this area can be extended in different ways. The number of different road networks tested was fairly limited and larger benchmark instances of road networks and vehicle schedules are required to effectively compare different routing algorithms and their parametrizations. The application of predictive and reactive routing algorithms is geared towards fast decisions that should require considerably less computation time than dynamic traffic assignment, and while surrogate-assisted optimization of the parameters allegedly needs to happen only once per road network, a performance boost and additional insight might be gained from learning the dependence of optimal parameter settings on different geographical properties of a road network.

ACKNOWLEDGMENTS
This work was supported by the European Union through the European Regional Development Fund (EFRE); further information on IWB/EFRE is available at www.efre.gv.at)

REFERENCES

Backfrieder, C., Ostermayer, G., Mecklenbräuker, C. F. (2017). Increased Traffic Flow Through Node-Based Bottleneck Prediction and V2X Communication. IEEE Transactions on Intelligent Transportation Systems, 18(2), 349-363

Backfrieder, C., Mecklenbrauker, C. F., Ostermayer, G. (2013, November). TraffSim--A Traffic Simulator for Investigating Benefits Ensuing from Intelligent Traffic Management. European Modelling Symposium (EMS), 2013 (pp. 451-456). IEEE.

Beyer, H. G., Schwefel, H. P. (2002). Evolution strategies–A comprehensive introduction. Natural computing, 1(1), 3-52.

Boukouvala, F., Ierapetritou, M. G. (2013). Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing. Journal of Pharmaceutical Innovation, 8(2), 131-145.

Cox, D. D., John, S. (1992, October). A statistical method for global optimization. Systems, Man and Cybernetics, 1992, IEEE International Conference on (pp. 1241-1246).

Forrester, A. I., Keane, A. J. (2009). Recent advances in surrogate-based optimization. Progress in Aerospace Sciences, 45(1), 50-79.

Greenshields, B. D., Channing, W., Miller, H. (1935). A study of traffic capacity. In Highway research board proceedings (Vol. 1935). National Research Council (USA), Highway Research Board.

Han, Z., Zhang, K., Song, W., Liu, J. (2013). Surrogate-based aerodynamic shape optimization with application to wind turbine airfoils. 51st AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition (p. 1108).

Hansen, N., Ostermeier, A. (1996, May). Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. Evolutionary Computation, 1996., Proceedings of IEEE International Conference on (pp. 312-317). IEEE.

He, X. (2014). Simulation-based optimization of transportation systems: Theory, surrogate models, and applications Doctoral dissertation. University of Maryland

Huang, D., Allen, T. T., Notz, W. I., Zeng, N. (2006). Global optimization of stochastic black-box systems via sequential kriging meta-models. Journal of global optimization, 34(3), 441-466.

Kesting, A., Treiber, M., Helbing, D. (2010). Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368(1928), 4585-4605.

Loshchilov, I., Schoenauer, M., Sebag, M. (2013, July). Intensive surrogate model exploitation in self-adaptive surrogate-assisted cma-es (saacm-es). Proceedings of the 15th annual conference on Genetic and evolutionary computation (pp. 439-446). ACM.

Janson, B. N. (1991). Dynamic traffic assignment for urban road networks. Transportation Research Part B: Methodological, 25(2), 143-161.

Johnson, M. E., Moore, L. M., Ylvisaker, D. (1990). Minimax and maximin distance designs. Journal of statistical planning and inference, 26(2), 131-148.

Jones, D. R., Schonlau, M., Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global optimization, 13(4), 455-492.

Lindorfer, M., Backfrieder, C., Kieslich, C., Krosche, J., Ostermayer, G. (2013, November). Environmental-Sensitive Generation of Street Networks for Traffic Simulations. European Modelling Symposium (EMS), 2013 (pp. 457-462). IEEE.

Maaten, L. V. D., Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(Nov), 2579-2605.

Osborne, M. A., Garnett, R., Roberts, S. J. (2009, January). Gaussian processes for global optimization. 3rd international conference on learning and intelligent optimization (LION3), 1-15.

Picheny, V., Ginsbourger, D., Richet, Y. (2010). Noisy expected improvement and on-line computation time allocation for the optimization of simulators with tunable fidelity.

Picheny, V., Wagner, T., Ginsbourger, D. (2013). A benchmark of kriging-based infill criteria for noisy optimization. Structural and Multidisciplinary Optimization, 48(3), 607-626.

Schrank, D., Eisele, B., Lomax, T. (2012). TTI's 2012 urban mobility report. Texas A&M Transportation Institute. The Texas A&M University System.

Sun, C., Jin, Y., Zeng, J., Yu, Y. (2015). A two-layer surrogate-assisted particle swarm optimization algorithm. Soft computing, 19(6), 1461-1475.

Wang, G. G., Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. Journal of Mechanical design, 129(4), 370-380.

AGENT-BASED MICRO SIMULATION TO ASSESS THE PERFORMANCE OF ROUNDABOUTS CONSIDERING DIFFERENT VARIABLES AND PERFORMANCE INDICATORS

António A. C. Vieira[(a)], Luís M. S. Dias[(b)], Guilherme A. B. Pereira[(c)] and José A. Oliveira [(d)]

[(a),(b),(c), (d)]University of Minho, Campus Gualtar, 4710-057, Braga, Portugal

[(a), (b), (c), (d)]{antonio.vieira, lsd, gui, zan}@dps.uminho.pt

ABSTRACT
Traffic congestion problems in intersections are usually solved by building infrastructures such as roundabouts. Several variables influence its performance, e.g. geometry, size and driving behaviour. Thus, it becomes necessary to compare these variables. This paper proposes a simulation model, developed to compare the performance of roundabouts, employing the object and agent modelling paradigms of Simio, to model the individual behaviour of vehicles. The results indicate the optimum size of roundabouts is around 40 meters of diameter and that the driving style has a greater influence on the performance of the roundabout than its unbalancing. In addition, it was found that roundabouts considering unbalancing and human behaviour decreased: the flow of vehicles in 8%, the waiting time per vehicle in 3 minutes, the queue size in 90%, the number of stops per vehicle in 88% and vehicles spent three times more fuel, than the roundabouts that did not consider these variables.

Keywords: Roundabout, Micro simulation, Agent modelling, object paradigm, Simio.

1. INTRODUCTION
Since the motor vehicle became the main means of transport, we have been witnessing a growing number of vehicles circulating on traffic lanes, which results in traffic congestion problems. To overcome them, usually the intersection is expanded through the construction of roundabouts. However, there are different geometries and different sizes that can be adopted, configuring several variables that can be parametrized by managers, including the number of lanes. Furthermore, the driving style also affects the performance of roundabouts. Furthermore, these geometric alterations may be limited to the site conditions, such as limited space, which may limit the geometries.
Simulation enables the visualization of the results from modifications made to a system, without making experiments in the real world. However, to the best of the knowledge of the authors, the traffic simulation packages available lack the of modelling not standardized concepts, such as the one hereby proposed. As such, discrete-event simulation was used for this work. From the simulation tools on the market, the choice was Simio, a tool that uses object and agent-oriented paradigms, which are essential for this project, since it becomes possible to model the individual behaviour of each vehicle.
In this sense, the purpose of this paper is to propose a general-purpose discrete simulation model that was developed to assess the performance of roundabouts of different sizes and to analyse the impact of a specific human behaviour on the system. This impact can be felt on several aspects. For instance, roundabouts with low performance result in longer times spent on queues by drivers, bigger traffic queues and even more pollution. Thus, this paper proposes an agent modelling approach to model vehicles travelling to access a roundabout, using Simio, a recently developed object oriented discrete simulation tool that also supports processes and events. The KPI (Key Performance Indicators) include the flow of vehicles, queue size, crossing time and fuel consumed and its gas emissions.
This document is organised in six sections. The main purpose of the next section is to make a review of the literature on discrete-event simulation tools. Section 3 is dedicated to the data gathering and validation process, also offering the required literature background on traffic situations related to roundabouts. In section 4, the simulation model is briefly described and section 5 is related to the simulation experiments conducted. Last section discusses some withdrawn conclusions.

2. LITERATURE REVIEW
Currently there are not many studies that use general-purpose discrete-event simulation models for modelling traffic problems. A possible justification for this is that most of the studies that use simulation in problems related to traffic, use packages of micro simulation tools like VISSIM or AIMSUN. The number of commercial tool options can be very high; thus, simulation tool comparison becomes a very important task.
Hlupic and Paul (1999), compare a simulation tools, distinguishing between users of software for educational purposes and users in industry. In his turn, Hlupic (2000) developed a survey to academic and industrial users on the use of simulation software, to discover how the users are satisfied with the simulation software they

use and how this software could be further improved. In Dias, Pereira and Rodrigues (2007), Pereira, Dias, Vik and Oliveira (2011) and Dias, Vieira, Pereira, and Oliveira (2016) a comparison of tools based on popularity on the internet, scientific publications, WSC (Winter Simulation Conference), social networks and other sources, was established. According to the authors, popularity should not be used as the only comparison indicator, otherwise new tools, better than existing ones, would never get market place. However, a positive correlation may exist between popularity and quality, since the best tools have a higher chance of being more popular. According to this ranking, the most popular tool is Arena, whilst the classification of the "newcomer" Simio is noteworthy. Vieira, Dias, Pereira and Oliveira (2014a) and Oueida, Char, Kadri and Ionescu (2016) compared both tools taking into consideration several factors.

Simio was created in 2007 from the same developers of Arena and is based on intelligent objects (Sturrock and Pegden 2010, Pegden 2007, Pegden and Sturrock 2008). Unlike other object-oriented tools, in Simio there is no need to write programing code, since the process of creating objects is completely graphic. The activity of building an object in Simio is identical to the activity of building a model. In fact, there is no difference between an object and a model (Pegden 2013). A vehicle, a costumer or any other agent of a system are examples of possible objects and, combining several of these, one can represent the components of the system in analysis. In other words, the user can use realistic representations of the objects that compose the real system being modelled and, thereafter, at a lower level, define additional logic to the model, through the development of processes for instance. This way, Simio complements the main object paradigm with other paradigms such as events, processes and agents. Since each entity can execute its own processes and thus make their own decisions, applied to the context of vehicles in a traffic system, the result is a simulation model, on which entities are modelled as agents.

Thus, a Simio model looks like the real system. This fact can be very useful, particularly while presenting the results to someone non-familiar to the concepts of simulation. In Simio the model logic and animation are built in a single step (Pegden and Sturrock 2008, Pegden 2007). This feature is very important, because it makes the modelling process very intuitive. Moreover, the animation can also be useful to reflect the changing state of the object. In addition to the usual 2D animation, Simio also supports 3D animation as a natural part of the modelling process. To switch between 2D and 3D views the user only needs to press the 2 and 3 keys of the keyboard. Moreover, Simio provides a direct link to Google Warehouse, a library of graphic symbols for animating 3D objects. Next section complements the literature review of this paper, by providing a review over literature focusing on traffic situations related to roundabouts.

3. DATA GATHERING AND VALIDATION

To build a model capable of representing the real system, the following data related traffic situations was gathered through literature collected and analysed:

- **Safety distances kept while driving:** Drivers that travel at a speed next to 50 km/h maintain a safety distance of about 16 meters Luo, Xun, Cao and Huang (2011).
- **Space occupied by a vehicle in a queue:** The analysed studies indicate that a stopped vehicle occupies a distance between 7.6 meters and 7.9 meters (Bonneson 1992, Messer and Fambro 1997, Zhu 2008, Herman, Lam, Rothery 1971).
- **Start-up acceleration:** Zhu (2008) analysed several studies regarding this matter. The author developed a polynomial acceleration model characterized by expression 1. Since in Simio it is not possible to implement the acceleration of entities, it was necessary to use the correspondent velocity expression 2. In addition, we have replaced t with $t-1$.

$$a = 2{,}46 - 0{,}24t + 0{,}006t^2 \tag{1}$$

$$v = 2{,}66 + 2{,}46t - 0{,}12t^2 + 0{,}002t^3 \tag{2}$$

- **Reaction time of drivers on roundabouts:** It is difficult to find in the literature and to measure in the filed the reaction time that drivers take to start accelerating, from a resting position, in a roundabout queue. This is because drivers are constantly trying to access a gap in the roundabout and many times they do not completely stop, which influences their start-up accelerating process. This does not happen, for instance, in signalized intersection, since drivers must wait for a red light that they do not know when it is going to change. Thus, the reaction time of drivers in the queues of signalized intersection was used. According to Bonneson (1992), the first vehicle of a queue takes 1 to 1.3 seconds. On the remaining positions of a queue, drivers take 2 seconds (Bonneson 1992, Messer and Fambro 1997), or1.5 to 2 seconds (Bonneson 1992, George and Heroy 1966).

These values were incorporated in Simio, adjusting them to have the reaction time of drivers being dependent on their distance to the one on the first position of the queue, as authors agree. Moreover, since the reaction time of drivers in roundabouts in lower than on signalized intersections, these values were calibrated. Figure 1 shows the reaction time of two samples of drivers from a modelled roundabout and a signalized intersection, in Simio.

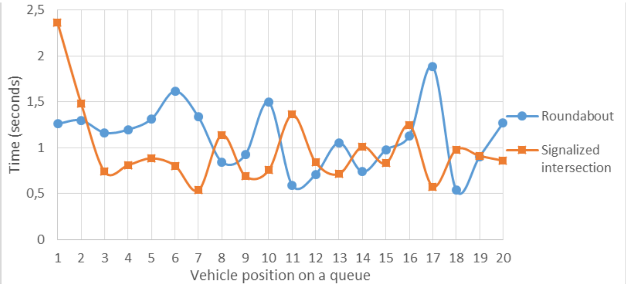


Figure 1: Average reaction time of drivers

As can be seen, the first vehicle of the queue on the signalized intersection took considerable more time than the vehicle on the same position of the queue of the roundabout. Concerning the reaction time of the vehicles on the remaining positions, their values decrease until an average of 1 second. After that, the average value is maintained.

- **Velocity while circulating inside the roundabout:** Skrodenis, Vingrys and Pashkevich, (2011) stated that speeds of vehicles, circulating inside roundabouts, of diameter varying between 16 to 45 meters, should be around 16-30 km/h. Furthermore, the speed of vehicles entering and circulating roundabouts tends to be higher for bigger roundabouts (Brilon 2005). Based on this and on numerous calibrations to the simulation model, it was considered that the vehicles could accelerate to a maximum speed of 30 km/h in roundabouts of similar size. For smaller roundabouts, the vehicles will only be able to speed up until 25 km/h. While circulating on roundabouts of 60 meters of diameter the vehicles will be able to speed up until 35 km/h and on roundabouts of 80 meters the vehicles will be able to speed up until 40 km/h. Thus, these speed differences also have an influence on the space gap required by the drivers to access the roundabouts of different sizes.
- **Space gap to access the roundabout:** While circulating a roundabout, the velocity of a vehicle affects the required space, or time, for a second vehicle to access the same roundabout. Since these values were modelled based on data collected from the literature, the authors empirically calibrated the required space gap, to minimize the occasions on which a vehicle decides to access a roundabout and, because of that, another vehicle, circulating on the roundabout, had to slowdown, since the available gap was too small for the other vehicle to access the roundabout. Thus, the space required for a vehicle, to access the roundabout was 17 meters for the roundabout of around 10 meters of radius, 22 meters for the roundabout with around 20 meters of radius, 33 meters for the roundabout of a radius of around 30 meters and 47 meters for the roundabout with around 40 meters of radius.
- **Instant speed when crossing the stop line of an intersection:** Bonneson (1992) stated that the velocity of each vehicle increases until the fourth or fifth vehicle. From that number, the velocity of the vehicles tends to stabilize.
- **Fuel consumption and emission rates:** Some of the models that estimate consumption rates and emissions include those based on the instant velocity of vehicles. Tong, Hung and Cheung (2000) established a formula for the fuel consumption of diesel vehicles in order of the instantaneous vehicle speed, whilst Chan et al. (2004) used a formula to estimate "the fuel consumption of petrol vehicles as a function of the instantaneous vehicle speed". Notwithstanding, there are models that consider other factors, such as the model proposed by Akçelik and Besley (2003), which considers the acceleration of the vehicle, its mass, instant speed, among other parameters. Akçelik (1983) also provided a model that expresses fuel consumption as a function of cruising, idling and stop-start manoeuvers. In its turn, Guo and Zhang (2014) indicated the formula currently being used by some traffic micro simulation tools (c.f. VISSIM, TRANSYT, and SYNCHRO).

Apart from formulas that estimate the consumption and emission rates, Coelho, Farias and Rouphail (2006) presented the emission factor of HC, NOx, CO2 and CO for several vehicle speed powers. In its turn, Tong, Hung and Cheung (2000) collected data related to vehicle speed, emission, and fuel consumption from four types of vehicles while they travel on different driving modes (i.e., idle, acceleration, cruise and deceleration). The authors presented the results in g/km, g/sec and g/kg fuel. Even though, there are more recent works that provide similar data, like the one Lau, Hung and Cheung (2011) conducted. These authors studied the CO, NO and HC emission rates, as well as the fuel consumption rates from four LPG taxis of different years, driven under urban traffic conditions. Notwithstanding, the data used in this study was the one collected by Tong, Hung and Cheung (2000), since it considers the time the drivers spend on each of the four driving modes. Thus, is consists on a simple, yet efficient, way to model the main consume patterns. The data provided by the authors and used on this study is presented in Table 1. To the best of the knowledge of the authors, this reference was the only one meeting the previously stated established requirements. Nowadays, all these values should be inferior, albeit at the same proportion.

Table 1 Modal emission and fuel consumption rates (Tong, Hung and Cheung 2000)

	Driving mode	**Modal emission rate (mg/sec)**			**Fuel Consumption**
		CO	**HC**	**NOx**	
Passenger Car	Acceleration	9.54	0.69	0.62	62.62
	Cruising	9.15	0.49	0.77	39.1
	Deceleration	9.96	0.58	0.69	28.11
	Idling	2.99	0.36	0.14	18.11
Petrol Van	Acceleration	15.14	1.85	1.96	67.29
	Cruising	14.52	1.70	1.81	52.14
	Deceleration	17.30	1.91	2.33	52.16
	Idling	8.39	1.88	0.81	12.71
Diesel Van	Acceleration	2.71	0.65	0.91	62.02
	Cruising	2.64	0.54	0.79	52.47
	Deceleration	2.67	0.65	0.89	56.01
	Idling	1.33	0.22	0.44	18.52

4. MODEL DEVELOPMENT

To enhance the animation of the simulation model, 3D models of road segments, vehicles and others were downloaded from Google Warehouse. Some sample videos of the model in execution were recorded and can be watched online at the following address: http://pessoais.dps.uminho.pt/lsd/pre_semaforos/. Figure 2 shows the modelled roundabout.

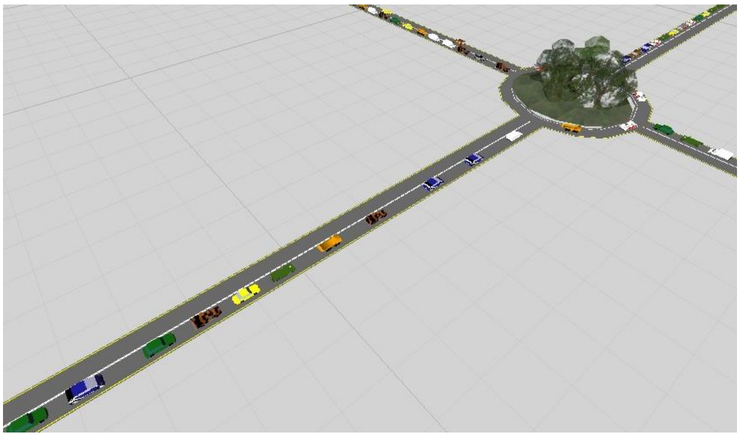


Figure 2: 3D view of the modelled roundabout

To model the behaviour of the vehicles on roundabouts, it was necessary to create many processes, functions, states among others, on the Simio software, to model all the traffic situations. Nonetheless, in this paper, only some of the processes will be illustrated. Figure 3 shows the process developed to have vehicles maintaining a safety distance between the vehicles of the model, while they are traveling. Figure 4 shows the process responsible for updating the fuel and emissions rates of the vehicles. To accurately calculate these rates, the 4 distinct operating modes of the vehicles (i.e. idle, acceleration, cruise and deceleration) had to be correctly defined.

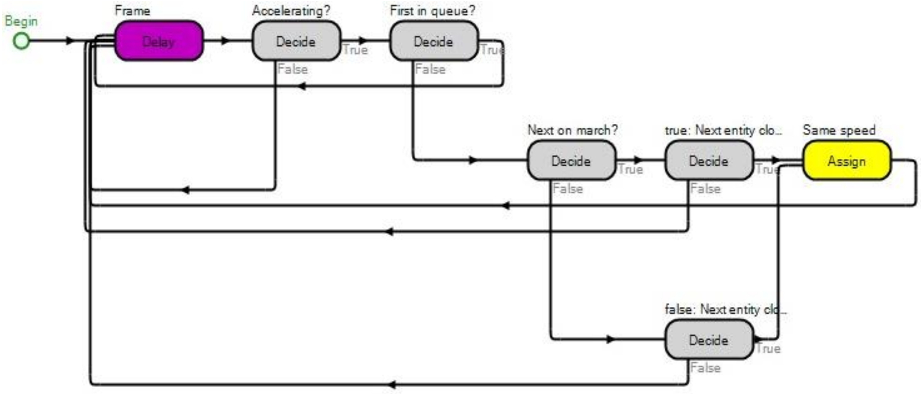


Figure 3: Process MaintainSafeDistance

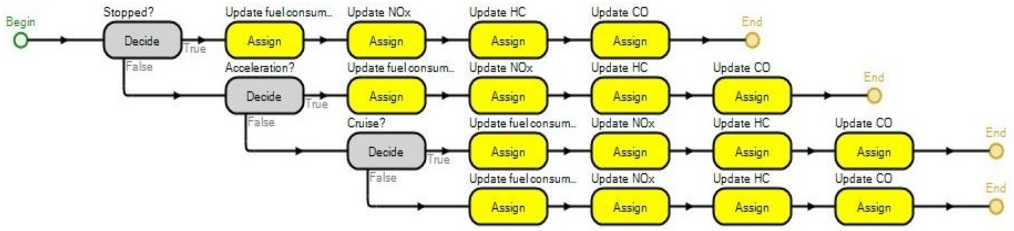


Figure 4: Process UpdateConsumption

The destination that the vehicles chose affects the system. The reason for this is that, in this type of intersection, all the vehicles compete for a gap to access the roundabout. Thus, when a vehicle arrives at the roundabout it decides whether it enters the roundabout or not, by evaluating the available gaps. While these times and distances are subjective to each driver, they are also influenced by the speed of the vehicles traveling in the roundabout and of the vehicle trying to enter it. Thus, in the developed simulation model, the speed of the vehicles approaching the entry lanes of the roundabout, and of the ones circulating inside the roundabout, is adjusted according to the size of the roundabout. The process that models the behaviour of each driver when evaluating if there is enough space in the roundabout to enter it is represented in Figure 5. In this process, each entity is actively deciding – agent modelling - if it can enter or not the roundabout, by analysing the distance to the closest cars at his left, on the roundabout.

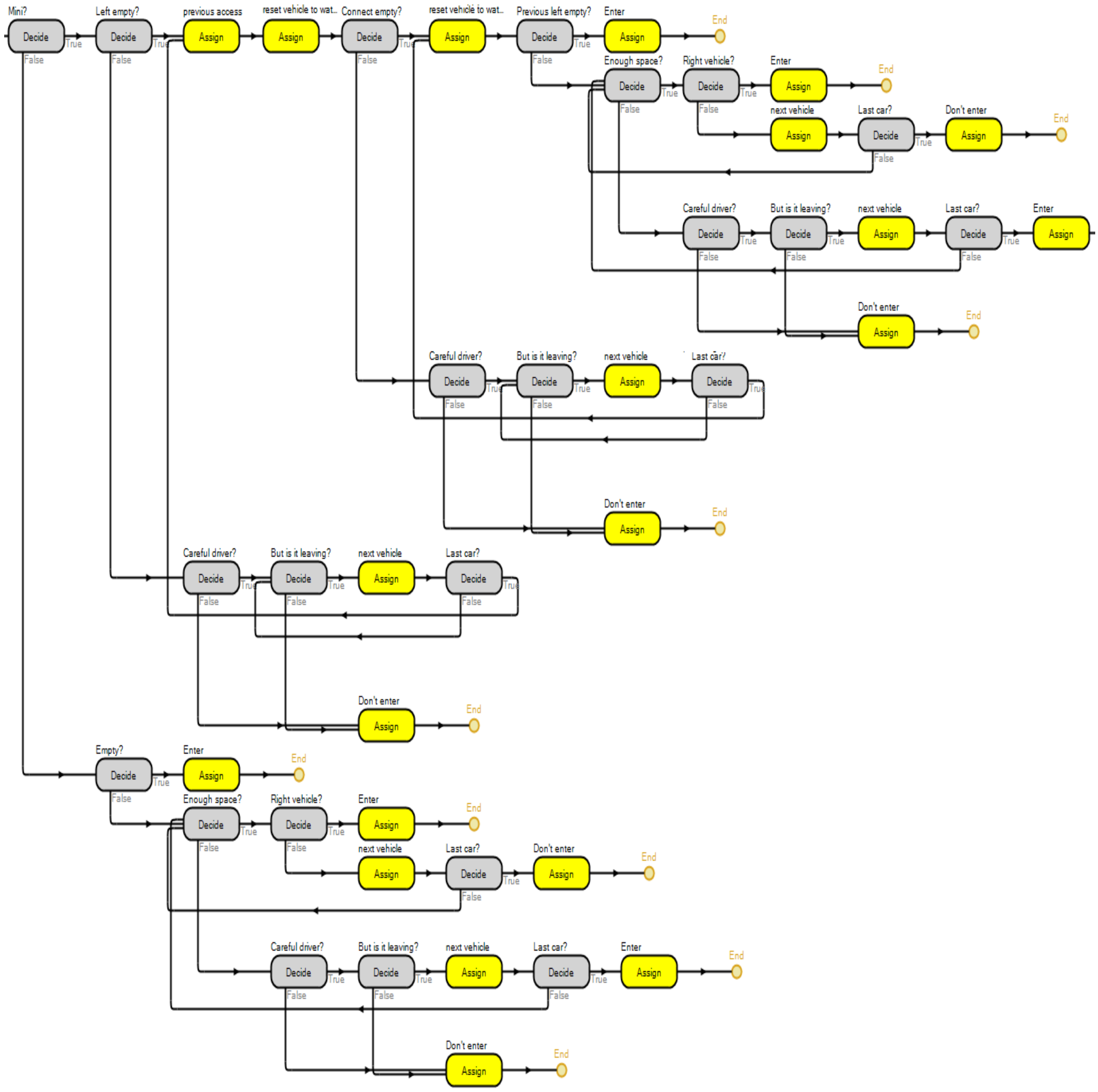


Figure 5: Agent process: “EnterOrNotTheRoundabout

5. SIMULATION EXPERIMENTS

For the present work, the authors considered the following properties, or parameters, for the conducted simulation experiments:

- the frequency with which the vehicles arrive to the system,
- the radius of the roundabout,
- the balancing of the roundabout, i.e., how balanced the outflow rates are, on the accesses of the roundabouts;
- the driver behaviour.

As KPI (Key Performance Indicators), the following were defined: average crossing time per vehicle in seconds, the average number of vehicles on the queues, the average flow of vehicles in vehicles/hour, the average total fuel consumed per vehicle in milligrams, the average total emissions of vehicles in milligrams (CO, HC and NOx) and the average number of stops per vehicle. Moreover, the values 4, 8, 13 and 50 seconds were considered, respectively, for the time interval that defines the creation of vehicles, and therefore the intensities very high, high, medium and low. Based on previous results (Vieira, Dias, Pereira and Oliveira 2014b), a warm-period of 360 seconds was used, along with a simulation time of 2 hours and 6 replications. Regarding these KPI, the following was considered:

- The time to cross an intersection is the elapsed time between when a vehicle is created and when it travels of 150 meters after having crossed the intersection.
- The number of vehicles on a queue is measured on every minute.
- The flow of vehicles is the inverse of the time interval between passages of vehicles through the intersection.
- The fuel consumption and its emissions rates start being accounted when vehicles are created and are updated every minute, until vehicles crosses the intersection.
- The average number of stops per vehicle recorded when one enters the roundabout.

The values considered for the radius of the central islands of the roundabouts were 10, 20, 30, 40 and 50 meters. Table 2 shows the obtained results.

Table 2: Comparing the modelled roundabouts

Traffic intensities	Very high				High				Medium				Low			
Radius (meters)	10	20	30	40	10	20	30	40	10	20	30	40	10	20	30	40
Flow rates (vehicles/hour)	1 537	1 974	1 766	1 636	1 525	1 799	1 752	1 595	1 108	1 109	1 108	1 108	289	289	289	289
Crossing time (minutes)	10,3	7,7	8,6	9,3	8,1	1,3	3,2	6,0	1,1	1,1	1,1	1,2	1,1	1,1	1,1	1,1
Queue size (number of vehicles)	66,0	61,5	62,1	61,8	51,2	1,9	16,9	37,0	0,3	0,1	0,2	0,4	0,1	0,0	0,0	0,0
Total fuel consumed per vehicle (g)	18,46	14,82	15,55	16,00	14,74	3,17	6,40	10,79	2,68	2,66	2,80	2,95	2,59	2,63	2,71	2,77
Total CO emissions per vehicle (g)	2,98	2,41	2,54	2,63	2,43	0,68	1,17	1,85	0,60	0,60	0,63	0,66	0,59	0,60	0,62	0,64
Total HC emissions per vehicle (g)	0,28	0,21	0,23	0,24	0,22	0,04	0,09	0,16	0,03	0,03	0,04	0,04	0,03	0,03	0,03	0,04
Total NOx emissions per vehicle (g)	0,17	0,14	0,15	0,15	0,14	0,05	0,08	0,11	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
Number of stops	34,3	29,4	28,8	27,1	26,9	1,7	9,5	17,6	0,6	0,3	0,5	0,6	0,3	0,2	0,2	0,1

As can be seen, in low and medium traffic intensities, regardless of the size, roundabout behaved similarly for all KPI. Yet, when traffic intensities increase (high or very high), the obtained results indicate that the roundabout performed better for radius of 20 meters, which is in accordance to previous studies (Oketch, Delsey and Robertson 2004). This can be explained by the more moderate speeds of the vehicles traveling in smaller roundabouts. Thus, the space gap for vehicles trying to enter the roundabout increases, conversely to bigger roundabouts, where vehicles, circulate at higher speeds and the space gap required also increases, as was also stated by previous studies (Fouladvand, Sadjadi and Shaebani 2004). Thus, based on these results, a roundabout of around 20 meters of radius seems to perform better than bigger or smaller ones.

The results analysed correspond to a scenario on which: **(i)** all the exits of the roundabout have an equal probability of being chosen by a vehicle to exit it; **(ii)** the human factor does not have influence on the performance of the intersection. However, in a realistic scenario this is not always the case, since on several cases, roundabout accesses may have different inflow and outflow rates. The human factor also has influence on the performance of the roundabout, for instance: when a driver signalizes he is going to exit the roundabout and a second driver trying to enter the roundabout through the same access decides to wait for the first driver to leave the roundabout, instead of entering while the exiting vehicle has not yet exited it; or when the first does not signalize its intention and thus the second must wait. These situations were modelled as percentages processes similar to the one represented on Figure 5. For both cases, a percentage of 50% was considered, albeit it can be adjusted. Thus, simulation experiments were conducted to analyse the impact of these factors on the performance of the intersection. Firstly, different probabilities were assigned to the roundabout destinies (40%, 30%, 20% and 10%). The results can be seen on Table 3.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Table 3: Comparing the modelled unbalanced roundabouts

Traffic intensities	Very high				High				Medium				Low			
Radius (meters)	10	20	30	40	10	20	30	40	10	20	30	40	10	20	30	40
Flow rates (vehicles/hour)	1 511	1 940	1 745	1 622	1 527	1 771	1 582	1 414	1 108	1 108	1 108	1 109	289	289	290	289
Crossing time (minutes)	10,5	7,8	8,6	9,2	7,7	2,4	5,0	5,9	1,1	1,1	1,2	1,3	1,1	1,1	1,1	1,1
Queue size (number of vehicles)	66,2	61,53	61,57	61,03	48,44	10,76	29,08	31,91	0,31	0,11	0,26	0,58	0,06	0,01	0,01	0,02
Total fuel consumed per vehicle (g)	18,6	14,8	15,2	15,5	14,0	5,0	9,1	9,9	2,7	2,7	2,8	3,0	2,6	2,6	2,7	2,8
Total CO emissions per vehicle (g)	3,01	2,42	2,50	2,56	2,32	0,96	1,59	1,74	0,60	0,60	0,64	0,67	0,59	0,60	0,62	0,64
Total HC emissions per vehicle (g)	0,28	0,22	0,23	0,24	0,21	0,07	0,13	0,15	0,03	0,03	0,04	0,04	0,03	0,03	0,03	0,03
Total NOx emissions per vehicle (g)	0,17	0,14	0,15	0,15	0,14	0,07	0,10	0,11	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
Number of stops	34,0	28,5	26,2	24,4	25,4	6,5	14,4	13,5	0,6	0,3	0,5	0,8	0,2	0,2	0,1	0,1

As can be seen, one of the main conclusion drawn from analysing Table 2, can also be observed here, i.e. the size of the roundabout where the best performance was achieved (20 meters of radius). However, it can also be seen that the performance of the roundabout decreased for all cases. In a second phase, the probabilities assigned to the destinies were reset to their default, but a probability of 50% was considered for the human impact factor. The results can be analysed on Table 4.

Table 4: Comparing the modelled roundabouts, considering human factor

Traffic intensities	Very high				High				Medium				Low			
Radius (meters)	10	20	30	40	10	20	30	40	10	20	30	40	10	20	30	40
Flow rates (vehicles/hour)	1 317	1 707	1 487	1 344	1 316	1 706	1 486	1 343	1 108	1 108	1 108	1 108	289	289	289	289
Crossing time (minutes)	12,2	9,1	10,5	11,5	10,7	4,5	8,4	9,7	1,2	1,1	1,2	1,3	1,1	1,1	1,1	1,2
Queue size (number of vehicles)	67,64	63,59	64,21	63,95	59,83	27,66	51,11	53,83	0,63	0,14	0,36	0,95	0,07	0,02	0,02	0,01
Total fuel consumed per vehicle (g)	20,5	16,7	17,9	19,0	18,1	8,8	14,7	16,2	2,8	2,7	2,9	3,1	2,6	2,6	2,7	2,8
Total CO emissions per vehicle (g)	3,31	2,71	2,93	3,11	2,96	1,53	2,44	2,68	0,62	0,61	0,64	0,69	0,59	0,60	0,62	0,64
Total HC emissions per vehicle (g)	0,32	0,25	0,27	0,30	0,28	0,13	0,22	0,25	0,04	0,03	0,04	0,04	0,03	0,03	0,03	0,04
Total NOx emissions per vehicle (g)	0,19	0,16	0,17	0,18	0,17	0,10	0,14	0,16	0,05	0,05	0,05	0,06	0,05	0,05	0,05	0,05
Number of stops	34,7	33,4	33,0	32,8	31,0	16,1	27,0	27,9	0,9	0,4	0,6	1,1	0,3	0,2	0,2	0,2

Once more, the main conclusion regarding the radius of the roundabout, on which the best performance was achieved is maintained (i.e. 20 meters). Nonetheless, by comparing Table 3 and Table 4, it can also be seen that the human factor has a greater influence on the performance of the roundabout than the different probabilities assigned to the roundabout exit lanes. Table 5 shows the results obtained for roundabouts with the two previous scenarios modelled.

Table 5: Comparing the modelled realistic roundabouts

Traffic intensities	Very high				High				Medium				Low			
Radius (meters)	10	20	30	40	10	20	30	40	10	20	30	40	10	20	30	40
Flow rates (vehicles/hour)	1 329	1 686	1 473	1 334	1 330	1 650	1 459	1 331	1 107	1 108	1 108	1 107	289	289	289	289
Crossing time (minutes)	12,1	9,2	10,5	11,5	10,5	4,5	7,0	8,5	1,2	1,1	1,2	1,5	1,1	1,1	1,1	1,1
Queue size (number of vehicles)	67,52	63,63	64,19	63,93	59,43	26,57	40,48	46,44	0,71	0,15	0,43	1,58	0,09	0,01	0,01	0,02
Total fuel consumed per vehicle (g)	20,3	16,7	18,0	18,9	17,9	8,6	12,1	14,2	2,8	2,7	2,9	3,3	2,6	2,6	2,7	2,8
Total CO emissions per vehicle (g)	3,29	2,72	2,94	3,1	2,92	1,51	2,06	2,39	0,62	0,61	0,65	0,72	0,59	0,6	0,62	0,64
Total HC emissions per vehicle (g)	0,31	0,25	0,28	0,3	0,27	0,12	0,18	0,22	0,04	0,03	0,04	0,04	0,03	0,03	0,03	0,03
Total NOx emissions per vehicle (g)	0,19	0,16	0,17	0,18	0,17	0,1	0,13	0,14	0,05	0,05	0,05	0,06	0,05	0,05	0,05	0,05
Number of stops	34,4	32,6	32,5	32,6	30,3	14,9	20,3	23,3	0,9	0,4	0,7	1,6	0,3	0,2	0,1	0,2

Considering both the human factor and the balancing of the roundabout affected more its performance than considering just one of the factors. These roundabouts will be referred as optimistic (Table 2) and realistic (Table 5). Comparing the two in low and medium intensities, it can be seen that there are no significative differences in the performance, for all KPI.

Focusing on the high and very high intensities, the average flow of vehicles can be decreased from 8 to 15% respectively, representing differences of 150 to 300 vehicles/hour. Regarding the crossing time and queue size, the differences are less significative for the highest traffic intensity. In its turn, for the high intensity the differences are more significative, which implies that the highest intensity is so high that both roundabouts could not properly handle these situations - the same conclusion can be withdrawn from the remaining KPI. In this sense, it can be concluded that roundabouts are not the most accurate solution for very saturated traffic situations, which is in accordance to previous studies (Fouladvand, Sadjadi and Shaebani 2003, Skrodenis, Vingrys and Pashkevich 2011). Thus, to accurately evaluate the performance difference between the optimistic and the realistic roundabout, the focus should be put on the high traffic intensity.

In the high traffic intensity, the crossing time per vehicle decreased more than 3 minutes per vehicle, resulting in a decrease in the average queue size of around 90%. This high difference is explained by the fact that only the vehicles that are stopped are accounted for this KPI. The remaining ones, even though they may be on the queue, they are not stopped, which further increases their fuel consumption and emissions. In fact, the average number of stops per vehicle increased up to 88%, culminating in an increase in the fuel consumption in up to 63% - vehicles spend three times more fuel. The respective emissions also increased in the same proportions.

6. CONCLUSIONS

The resolution of traffic congestion problems usually implies the construction of infrastructures such as roundabouts. However, these infrastructures have several decision variables. Thus, this paper proposed a general-purpose discrete-event traffic micro-simulation model that can compare different roundabouts, assessing their performance. The chosen simulation tool – Simio – offers the user the ability to use different simulation paradigms, such as: objective, agent, events, processes and others. Therefore, with some effort it was possible to develop and validate a simulation model in which entities were modelled as intelligent agents, in the sense that they can evaluate their surroundings and make decisions, similarly to what happens in the field.

The conducted simulation experiments concluded that the best size of roundabouts is 40 meters of diameter. The second set of experiments focused on evaluating the human factor in the driving behaviour and the unbalancing of the roundabout in its performance. Thus, a realistic roundabout – considering its unbalancing and the driving behaviour – and an optimistic roundabout were compared. The main conclusions from this analysis were that the human factor had more negative impact in the performance than the balancing did. In addition, it was concluded that on low, medium and on the highest traffic intensities these roundabouts achieved the same performance, which is in accordance to previous studies (Fouladvand, Sadjadi and Shaebani 2003, Skrodenis, Vingrys and Pashkevich 2011). For the remaining defined traffic intensity – where most significative differences were registered - it was found that the flow of vehicles decreased up to 8% when the optimistic roundabout was compared to the realistic one. It was also found that the scenario corresponding to the unbalanced roundabout and considering the human driving style resulted in a decrease in the waiting time per vehicle in 3 minutes, the queue size in up to 90% and the number of stops per vehicle in up to 88%. Furthermore, it also resulted in an increase in the fuel consumption in up to 63% - vehicles spent three times more fuel - and in the respective emissions.

For future development: (1) it would be interesting to adapt the developed model to handle roundabouts with multi lanes on the approaches, as well as inside the roundabout; (2) since agents are being modelled, it would be interesting to model different types of drivers – accelerate more, or less, requires respectively more, or less, space to enter the roundabout, among others.

REFERENCES

Akçelik R., 1983. Progress in Fuel Consumption Modelling for Urban Traffic Management. Australian Road Research Board.

Akçelik R. and Besley M., 2003. Operating cost, fuel consumption, and emission models in aaSIDRA and aaMOTION. 25th Conference of Australian Institutes of Transport Research (CAITR 2003).

Bonneson J.A., 1992. Modeling Queued Driver Behavior at Signalized Junctions. Transportation Research Record, No 1365, pp. 99–107.

Brilon W. 2005. Roundabouts: A state of the art in Germany. National Roundabout Conference, Vail, Colorado, vol. 16.

Chan T.L., Ning Z., Leung C.W., Cheung C.S., Hung W.T. and Dong G., 2004. On-road remote sensing of petrol vehicle emissions measurement and emission factors estimation in Hong Kong. Atmospheric Environment, 38, 2055-2066.

Coelho M.C., Farias T.L. and Rouphail N.M., 2006. Effect of roundabout operations on pollutant emissions. Transportation Research Part D: Transport and Environment, 11, 333-343.

Dias L.M.S., Pereira G.A.B. and Rodrigues G., 2007. A Shortlist of the Most Popular Discrete Simulation Tools. Simulation News Europe, 17, 33-36.

Dias L.M.S., Pereira G.A.B., Vik P. and Oliveira J.A., 2011. Discrete simulation tools ranking: a commercial software packages comparison based on popularity. Industrial Simulation Conference. Venice, Italy, 6-8 June: Eurosis.

Dias L.M.S., Vieira A.A.C., Pereira G.A.B. and Oliveira J.A., 2016. Discrete Simulation Software Ranking – a Top list of the Worldwide most Popular and Used Tools. Proceedings of the 2016 Winter Simulation Conference.

Fouladvand M.E., Sadjadi Z. and Shaebani M.R., 2003. Characteristics of vehicular traffic flow at a roundabout. Physical Review E 70.4 (2004): 046132.

George E.T. and Heroy F.M., 1966. Starting Response of Traffic at Signalized Intersections. Traffic Engineering 36.1Q, pp. 39-43.

Guo R. and Zhang Y., 2014. Exploration of correlation between environmental factors and mobility at signalized intersections. Transportation Research Part D: Transport and Environment, 32, 24-34.

Herman R., Lam T. and Rothery R.W., 1971. The Starting Characteristics of Automobile Platoons. Proc., 5th International Symposium on the Theory of Traffic Flow and Transportation, American Elsevier Publishing Co., New York, pp. 1-17.

Hlupic V., 2000. Simulation software: an Operational Research Society survey of academic and industrial users. Proceedings of the 32nd Winter Simulation Conference (pp. 1676-1683 vol. 2). Society for Computer Simulation International.

Hlupic V. and Paul R., 1999. Guidelines for selection of manufacturing simulation software. IIE Transactions, 31, 21-29.

Lau J., Hung W.T. and Cheung C.S., 2011. On-board gaseous emissions of LPG taxis and estimation of taxi fleet emissions. Science of The Total Environment, 409, 5292-5300.

Luo Q., Xun L., Cao Z. and Huang Y., 2011. Simulation analysis and study on car-following safety distance model based on braking process of leading vehicle. 9th World Congress on Intelligent Control and Automation (WCICA). pp. 740-743. IEEE.

Messer C.J. and Fambro D.B., 1997. Effects of Signal Phasing and Length of Left Turn Bay on Capacity, No. 644. 1977.

Oketch T., Delsey M. and Robertson D., 2004. Evaluation of Performance of Modern Roundabouts Using Paramics Microsimulation Model. TAC Conference.

Oueida S., Char P.A., Kadry S. and Ionescu S., 2016. Simulation Models for Enhancing the Health Care Systems. FAIMA Business & Management Journal, 4, 5.

Pegden C.D., 2007. Simio: A new simulation system based on intelligent objects. Proceedings of the 39th Winter Simulation Conference: 40 years! The best is yet to come, pp. 2293-2300, IEEE Press.

Pegden C.D., 2013. An Introduction to Simio for Arena Users. Simio. White paper. Available online at: http://www.simio.com/resources/white-papers/For-Arena-Users/An-Introduction-to-Simio-For-Arena-Users.htm.

Pegden C.D. and Sturrock D.T., 2008. Introduction to Simio. 2008 Winter Simulation Conference, pp 29-38, IEEE.

Skrodenis E., Vingrys S. and Pashkevich M. 2011., Lithuanian experience of implementation of roundabouts: the research of accidents, operation and efficiency. The 8th International Conference “Environmental Engineering”: Selected papers. Ed. by D. Čygas, KD Froehner, pp. 980-985.

Sturrock D.T. and Pegden C.D., 2010. Recent innovations in Simio. Proceedings - Winter Simulation Conference, pp. 52-62. Proceedings - Winter Simulation Conference, pp. 52-62.

Tong H.Y., Hung W.T. and Cheung C.S., 2000. On-Road Motor Vehicle Emissions and Fuel Consumption in Urban Driving Conditions. Journal of the Air & Waste Management Association, 50, 543-554.

Vieira A.A.C., Dias L.M.S., Pereira G.A.B. and Oliveira J.A., 2014a. Comparison of Simio and Arena simulation tools. ISC. University of Skovde, Skovde, Sweden.

Vieira A.A.C., Dias L.M.S., Pereira, G.A.B. and Oliveira J.A., 2014b. Micro Simulation to Evaluate the Impact of Introducing Pre-Signals in Traffic Intersections. University of Minho at Guimarães - Portugal: International Conference on Computational Science and its Application, pp. 722-745, Springer International Publishing.

Zhu H., 2008. Normal Acceleration Characteristics of the Leading Vehicle in a Queue at Signalized Intersections on Arterial Streets, Doctoral dissertation. Oregon State University.

AUTHORS BIOGRAPHY

António A. C. Vieira was born in 1989 in Vila Nova de Famalicão, Portugal. He graduated in Informatics Engineering in the University of Minho. He holds a MSc degree in Systems Engineering at the University of Minho. He is making his doctoral studies in Advanced Engineering Systems for Industry at the University of Minho. His main research interests are Simulation, Systems Performance, Operational Research and Programming Languages.

Luís M. S. Dias was born in 1970 in Vila Nova de Foz Côa, Portugal. He graduated in Computer Science and Systems Engineering in the University of Minho, Portugal. He holds an MSc degree in Informatics Engineering and a PhD degree in Production and Systems Engineering from the University of Minho, Portugal. His main research interests are Simulation, Systems Performance, Operational Research and Systems Visual Modeling.

Guilherme A. B. Pereira was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management in the University of Minho, Portugal. He holds an MSc degree in Operational Research and a PhD degree in Manufacturing and Mechanical Engineering from the University of Birmingham, UK. His main research interests are Operational Research and Simulation.

José A. Oliveira was born 1966 in Matosinhos, Portugal. He studied Mechanical Engineering at the University of Porto, Portugal. He graduated with a Ph.D. in Production and Systems Engineering at University of Minho, Portugal. His main research interests are Optimization with Heuristic Methods in Systems Engineering.

NESTED SIMULATIONS SUPPORTING TRAFFIC OPTIMIZATIONS RELATED TO RAILWAY STATIONS

Roman Diviš[a], Antonín Kavička[b]

[a],[b]University of Pardubice, Faculty of Electrical Engineering and Informatics

[a]roman.divis@upce.cz, [b]antonin.kavicka@upce.cz

ABSTRACT
Nested simulations present a general method suitable for use in realizing a multi-trajectory simulation or as a decision support in a simulator. The principle of nested simulation (as a decision support) is to find a solution to a problem using other time-limited simulations which verify alternative options. After the nested simulations have finished, the solutions of individual alternatives are assessed and the best solution is applied to the main simulation.
The aim of the article is to conduct a case study of using nested simulation as decision support in a mesoscopic simulator of rail transport. This study is based on a prototype railway station with mixed traffic of both passenger and cargo trains. The results of the simulation are compared with commonly used microscopic simulation tool Villon.

Keywords: nested simulations, decision support, rail transport

1. INTRODUCTION

The scope of examination is simulation of rail transport. At the moment, the focus is placed on realisation of own mesoscopic simulation tool called MesoRail (Diviš and Kavička 2015). The aim of the simulator is to allow the examination of both the deterministic and the stochastic operation, mainly within the scope of railway stations. Stochastic simulators deal with random events within the simulation, during which conflict situations, with possibly not known optimal solutions within defined rules, occur. When such events occur, it is possible to use a decision support subsystem, aim of which is to provide information for solving particular conflict states. There is a variety of methods and approaches suitable for realizing decision support. This article further focuses on using the method of nested simulations as decision support.

2. THE METHOD OF NESTED SIMULATIONS

The method of *nested/recursive simulations* presents a principle of using simulation inside a simulation in order to examine results of more alternative scenarios (or development) of the simulation. The original/main instance (one particular replication) of a simulation is cloned and individual clones have different parametrisation set for them. Conducting such nested simulation leads to simulating several alternative scenarios. The output of conducted nested simulations is a broader set of information about the given issue.

One possible use of nested simulations presents *decision support in a simulation*, the nested simulations run for a limited time and after they have finished, their results and outputs are assessed and the original simulation is again merged into a single instance (main simulation) and it can continue in a selected manner.

Another use of nested simulations can be realized as *a multi-trajectory simulation* - the simulation experiment is divided into nested simulations in individual points of decision and the simulation is gradually branching. Various scenarious are thus investigated and according to article (Gilmer and Sullivan 1999), such approach can be more efficient than using a large number of replications of a single simulation.

The following text deals with the first way of usage - as decision support in a simulation.

The use of nested simulations is also done by other authors. Use can be found, for example, in Bonté, Duboz, Quesnel and Muller (2009), Kindler (2010), Gordy and Juneja (2010). Individual publications are briefly presented in our previous article (Diviš and Kavička 2016).

2.1. The technique of a nested simulation

A nested simulation allows to use already existing simulation engine of a given software and a simulation experiment for own search of a solution of an issue. However, before the nested simulations can be conducted, it is necessary to list all steps solving a particular conflict (Diviš and Kavička 2016):

1) A conflict situation (an issue) requiring decision support is identified in the system.
2) Current instance of the main simulation (S^{main}) is interrupted in time *t*.
3) For the needs of nested simulation, it is necessary to set their parameters:
    a) the criterion of optimality (*CrOpt*),
    b) the duration of an outlook into the future for the nested trials (or rather the stopping condition, *StopCond*),

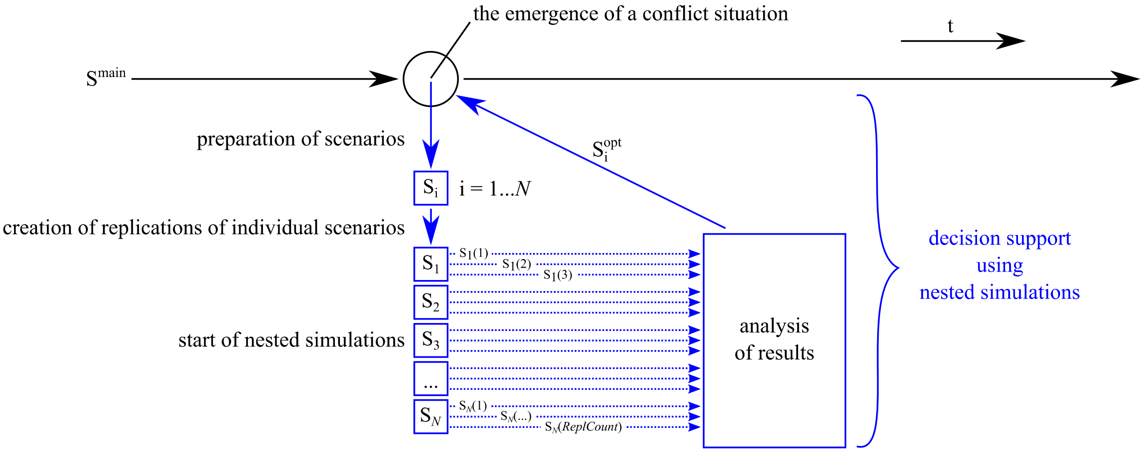


Figure 1: Illustration of occurrence and proceeding with solution of one conflict situation in one simulation replication

   c) the number of replications for all individual scenarios of the nested simulations (*ReplCount),*
   d) (maximal) number of alternative scenarios (*ScnCount),*
   e) generator of alternative scenarios (*ScnGen),*
   f) recursion limit of nested simulations (*RecLimit).*
4) N alternative scenarios for minor simulations are established.
5) The main simulation S^{main} is cloned and *ReplCount* of replications is created for each *i-th* scenario.
6) Individual replications $S_i(j)$ are started (for $i = 1 \ldots N, j = 1 \ldots ReplCount$).
7) Waiting for finishing all replications $S_i(j)$ (for $i = 1 \ldots N, j = 1 \ldots ReplCount$) to finish.
8) Assessing the results of individual scenarios from the replications $S_i(j)$ (for $i = 1 \ldots N$, $j = 1 \ldots ReplCount$) and then selecting the scenario with the best results according to *CrOpt.*
9) The main simulation S^{main} then continues with the selected scenario from the instant *t* of simulation time.

From the above described procedure (Figure 1) it is apparent that before the realization of nested simulations itself, it is necessary to solve several basic questions about how the nested simulations should even be parameterized and set up. A more detailed description of parameters of nested simulations can be found in article (Diviš and Kavička 2016).
A separate issue could be presented in how the alternative scenarios for solving conflict situations should be chosen, but such issue depends on particular application of a method and on the type of the conflict situation, and it can in the end present a non-trivial task.

2.2. Computational demands

Technical possibilities of nested simulation realization posed a separate and rather large issue for nested simulations. It could include a demanding computational task, complexity of which is given by the number of scenarios and replications and the length of simulated period of individual replications that need to be conducted.
Computational demand is influenced by several factors:

- the number of alternative scenarios,
- the number of replications of each scenario,
- the lenght of simulated period within one replication,
- the number of conflict states that occur in the original (main) simulation and that require a decision based on perspectives of the nested simulations,
- the number of replications of the original (main) simulation,

Apart from the abovementioned factors, it is also necessary to take into account the possibility of conflict situations occurring inside the nested simulations (example of recursive nested simulation is shown in figure 2). That could be cause by recursive launch of other nested simulations and thus cause an exponential growth in computation difficulty. One option how to avoid such an issue is to terminate the nested simulation exactly when a conflict situation occurs inside the nested simulation, or setting a maximal depth of recursion for nested simulation is a general solution.

3. DECISION SUPPORT WITHIN THE SCOPE OF SIMULATING RAIL TRANSPORT

In the general area of decision support, it is possible to use a variety of methods using even highly complex mathematical apparatus. From simple methods using priority lists to methods using artificial neural networks. The role of decision support and its possible use in the environment of simulating rail transport is described in the next chapter.
To realize decision support in the environment of rail transport, it is basically possible to use all commonly used techniques including, for example:

- the method of priority planning,

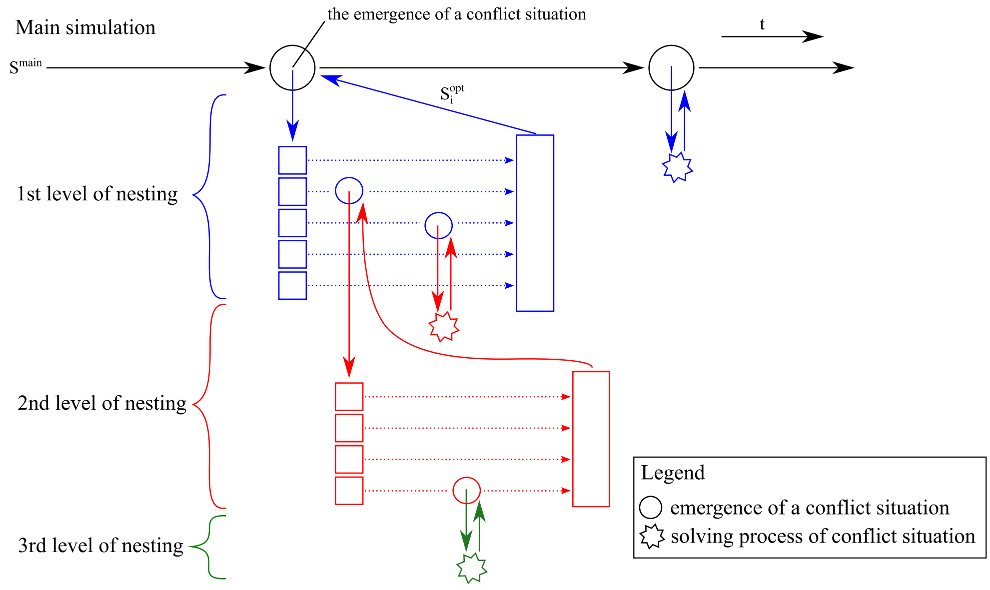


Figure 2: Illustration of a repeated occurrence of conflict situations during the main simulation and within nested simulations

- expert systems,
- methods of operational research,
- heuristic methods,
- methods of soft computing,
- the method of nested simulations,
- and others.

3.1. Methods used in simulators for rail transport

In the scope of microscopic simulators for rail transport, there are several different methods of decision support used.

Priority planning (or priority lists) is a rather common method of decision support. The principle of such method is to define a list of alternatives (e.g. train routes) that are ordered based on their priority. The decision is then done by choosing an option with the highest priority. Provided that such possibility is not permissible, the next item on the list is selected until all possibilities are exhausted. This method has been implemented in simulators OpenTrack and Villon.

More complex and more sophisticated system using the method of multi-criterion decision is newly implemented in the simulator Villon (Bažant and Kavička 2009).

3.2. Possible conflict situations, searching for alternative solutions.

Conflicts in stochastic simulations occur mainly as the result of individual entities (agents) competing with each other for limited resources. In the scope of the examined domain of rail transport, the identities are understood as trainsets competing for individual parts of infrastructure. Apart from the fact that there can be only one trainset present on one track, or rather a single train can be present on a single segment of switches, it is necessary to keep to defined rules of rail transport which are related to used safety device. Such rules also include safety time intervals with which it is not possible to allow another train to enter a track section in order to maintain safety and flow of traffic.

Specific examples of coflicts that can occur include, for example:

- Arrival of a train, requesting an already occupied station track, into the station.
- Gradual departure of more trains from the station (in a short time interval) while requiring the same line or part of the deviated tracks.
- Unscheduled train arrival into a stations with all platform lines occupied.

Searching for a solution depends on the aplication domain of the simulator. Within the scope of railway operation, it is possible to wait for clearing an occupied elements of the infrastructure. However, such solution usually is not suitable, it causes longer delays for the conflict train and it can cause other conflicts by occupying a line sector by the conflict train. An alternative solution is to move the train to another line or to another platform. It is impossible to easily predict whether such change would somehow affect the flow of traffic. In this case, nested simulations can analyze the impact of the given change on future traffic.

All solution possible to apply still need to keep to particular activities of safety device and all are limited by available infrastructure.

4. SIMULATION TOOL MESORAIL

The simulation tool *MesoRail* (Diviš and Kavička 2015) presents a mesocopic simulator of rail transport in development, which focuses on examining railway station capacity. The simulator specializes in support for

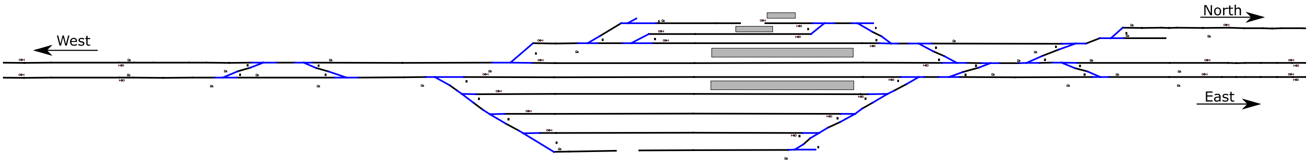


Figure 3: Infrastructure of center station in case study scenario

fast prototyping of rail infrastructure and fast setting of operational conditions specifications (including types and composition of trainsets, timetables, parameterization of random inputs, etc.). The aim of the simulator is to allow conducting simulation studies in shorter periods of time than it is possible with microscopic simulation tools. For these reasons, there are some abstractions implemented in *MesoRail* which sets it to mesoscopic level of abstraction. The main issue is finding a suitable compromise and level of abstraction for individual parts in such a way, that the required accuracy of result is achieved. To establish suitable abstractions, consultations with railway experts were conducted.

In the simulator the following abstractions were used - (i) schematic depiction of infrastructure, (ii) simplification of calculating train ride dynamics, (iii) neglect of mobile resources of operation (employees), (iv) neglect of some operational technological procedures. *MesoRail* can be applied for, for instance, determining the throughput of railway junctions, therefore it is necessary to meet the following requirements - (i) fast prototyping of infrastructure, (ii) real running features of trains, (iii) respecting the functionality of safety devices, (iv) applying station and track intervals, (v) simulating deterministic and stochastic train flows, (vi) animation outputs during the simulation, (vii) post-simulation statistics and time protocols in graphic form. The listed items is just a brief list of requirements expected form the simulator. Based on those requirements, the level of abstraction for individual components must be defined.

4.1. Architecture of simulation engine

Simulator *MesoRail* builds on hybrid architecture, which combines a discrete simulation engine with agent based simulation. The foundation is then a very simple engine using methods of event planning, which is, however, used in agent-oriented appoach.

4.2. Implementation of nested simulations

The method of nested simulations can make the impression that it is an easy and easily implemented task. However, its technical realization is rather challenging. For conducting of nested simulations, it is necessary to stop the original simulation, clone it, customize the clones' parameterization, and run the nested simulations for limited time.

Cloning a simulation means to create a complete copy of the current state of the simulation. For structured simulations deconstructed into agents (with inner state) that communicate among one another, it is necessary to create copies of agents, fill their state, and renew references among new instances of agents. If individual objects are realized as immutable, then creating a copy is easier. In *MesoRail* simulator, individual objects are - agents with inner state, and such state is changeable simulation, it is necessary to create a copy of a complex object graph.

After creating a copy, it is necessary to make available relevant agents and entities and to customize their parameterization for new simulation run. Launching the simulation as such is then a trivial task.

4.3. Technical realization

Because of above mentioned technical issues related to realizing nested simulations, general decision support for recording the state of object graph has been built in *MesoRail*. This system basically allows to arbitrarily go back in time of a finished simulation. Another function of this subsystem is creating a copy of an object graph and thus cloning the entire simulation.

Actual administration of nested simulation then efficiently uses other parts of the simulator for easy realization of a copy, parameter changing, and launching nested simulations.

4.4. Situations solved by nested simulations

Within the scope of realizing decision support, nested simulations are called in cases in which a train cannot allocate other line segments into its train path and stopping the train would be imminent. Such cases include, for instance, and occupied line at a station.

Individual trains have their train paths or groups of train paths defined, to which they keep druing the simulation. Nested simulations then use these alternative paths in group of train paths for finding a possible solution of a situation.

5. CASE STUDY

The following chapter describes realization of a case study using nested simulations as decision support in the scope of simulating a railway station. A simulation model contains one main prototype station (shown on figure 3), from which lead two double-lined (to the stations *West* and *East*) and one one-lined (to the station *North*) tracks. Those are finished by a simplified model of a railway station.

Total length of track from *West* to *East* is about 20 kilometers. The track contains also significant slope and arcs, track itself isn't completely artificial, but it is inspired by a several tracks in Czech Republic.

Table 1: Parameterisation of trains in the simulation model

Train type	Locomotive / vagons	Course	Interval between trains [h:mm:ss]	Total train count	Delay prob.	Delay mean time (exponential distribution) [s]
Express	1 / 7	West → Central → East	30:00	5	50%	420
Express	1 / 7	East → Central → West	30:00	5	50%	420
Passenger	2 / 4	West → Central → East	10:00	12	33%	270
Passenger	2 / 4	East → Central → West	10:00	12	33%	270
Passenger	1 / 2	West → Central → North	30:00	4	33%	270
Passenger	1 / 2	North → Central → West	30:00	4	33%	270
Cargo	1 / 22	West → East	1:00:00	2	50%	1800
Cargo	1 / 22	East → West	1:00:00	2	50%	1800

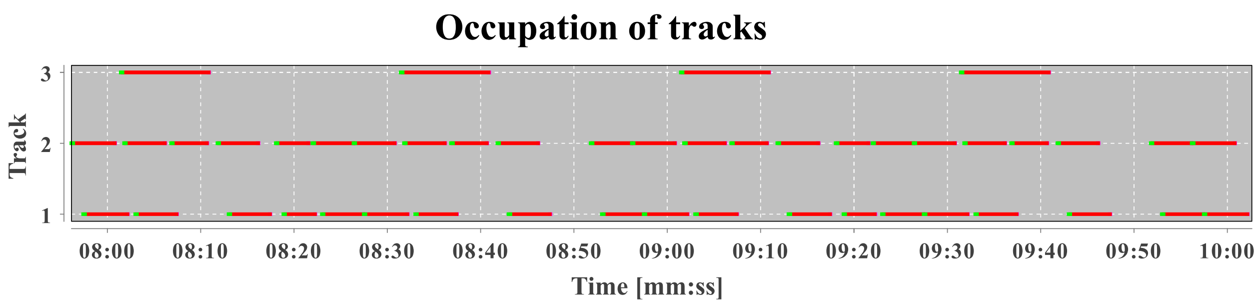


Figure 4: Occupation of station tracks during deterministic simulation

The traffic in station is mixed and it contains both passenger and cargo trains. It includes few express trains, dense traffic of passenger trains and several passing freight trains. A more detailed description of trains composition and parameterization is given in the table 1. The figure 4 shows the occupation of station tracks that are normally occupied by individual train sets under deterministic mode of operation.
Same simulation model was built also in the Villon for comparison of simulations results. The Villon represents microscopic simulation tool specializing in the construction of railway simulation models, but also supports road traffic and container logistics systems. The simulator works at a microscopic level of detail, and building a comparable model is more time consuming. In order to compare the results between Villon and MesoRail, it was first necessary to validate the results.

5.1. Validation of simulation models between MesoRail and Villon

After realization of the simulation models in both tools, deterministic simulations were performed without the use of random delays. During the simulations we recorded the moments of inputs and outputs of individual trains to or out of simulation. Based on these data, train travel times were calculated. The minimum travel time was 741 s, maximum 930 s, average 780 s. The results of the comparison of travel times are given in the table 2.

Table 2: Comparison of train travel times in MesoRail and Villon

	Relative difference [%]	Absolute difference $\lvert T_{MR} - T_V \rvert$ [hh:mm:ss]
Minimum	0,1285 %	00:00:01
Maximum	3,3735 %	00:00:31
Average	1,6200 %	00:00:13

The differences in the travel times of trains in both simulators is due to their different approach to train driving dynamics and minor deviations between both simulators and simulation models. The average size of difference is 1.62% and the maximum measured value is 3.37%. Since these values did not exceed the originally assumed limit of 5%, both models were considered to be identical and so it was proceeded to stochastic testing.

5.2. Stochastic simulations

When making stochastic simulations, every train entering the model may be delayed. First, using a random value generated with a uniform distribution, it is determined whether the train will be delayed or not. The amount of delay is then given by random number

Table 3: Evaluation of simulation results - statistic of change of train delays after simulation run

Type	Absolute delays [s]			Relative delays [%]		
	Villon	MesoRail 5 min	MesoRail 15 min	Villon	MesoRail 5min	MesoRail 15min
Express	40,61	14,79	14,27	5,16%	1,90%	**1,83%**
Passenger	40,56	24,42	26,93	5,25%	**3,16%**	3,49%
Cargo	35,50	6,96	11,59	3,95%	**0,76%**	1,27%
Total	40,13	20,81	22,84	5,11%	**2,68%**	2,94%

Table 4: Statistics of conflicts during one replication of the main simulation

Outlook duration	5 min			15 min		
	Min	Max	Average	Min	Max	Average
Conflicts in main simulation	3	15	7,88	3	18	8,04
Total number of conflicts	3	89	17,19	3	432	67,79
Total simulations ran	4	172	31,94	4	843	119,78

generator with exponential distribution (according to the type of train type). No further random effects (device faults, station delays, ...) are being used during the simulation.
Each train has a defined starting train path and three alternative train paths that use other station tracks. The strategy used to select a replacement train path is given by the simulator.
Villon uses the priority planning technique. If the station track is occupied, then based on the order of paths in priority list, first available alternative path is selected. Therefore, only the order of the defined routes is important for the selection of the alternate train path and whether or not they are available at the moment, there is no further analysis of the conflict situation.
MesoRail uses nested simulation techniques. In the event of a conflict situation, individual alternatives are simulated for a limited time period. The number of replications of nested simulations, the length of their run is described in the following section. To test the effect of the parameters, two sets of simulations were performed with different outlook values (the duration of the nested simulation).
The duration of one simulation replication is 2.5 hours, all the trains are scheduled to run during this period. In order to test different variants and to obtain data for processing statistics, 100 replications were performed in Villon and 100 replications in MesoRail.

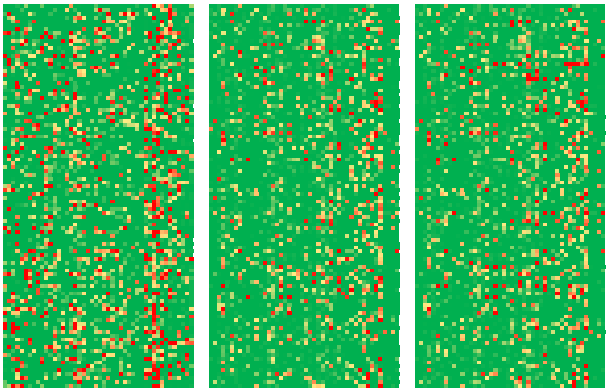
5.3. Parameterisation of decision support

Villon has defined paths according to the expected priorities appropriate for the type of train and the direction in which it is going.
Parameterization of nested simulations in MesoRail is as follows:

- Optimality criterion - the sum of weighted train delays, the weight represents the train's priority according to its type (express trains – 1.8, passenger – 1.0, cargo – 0.2);
- Length of outlook - 5/15 minutes of simulation time from the moment of conflict;
- Number of replications of nested simulations - 1 replication;
- Number of alternative scenarios - not limited;
- Alternative scenario generator - according to available alternative train paths;
- Recursion limit of nested simulations - up to 3 levels.

5.4. Simulation results

At the end of the simulations, the results were analyzed. As a basic indicator of the quality of the operation was chosen the change in the delay of the train (calculated as delay at the moment of leaving simulation minus train's input delay). Due to the fact that the train timing in both simulation models is not exactly the same, the results are evaluated not only in the form of the absolute value of the delay and also the relative delay given by the ratio of the absolute value of the delay to the travel time of the train. The aggregate results for each train type



(a) Villon (b) MR 5 min (c) MR 15 min

Figure 5: Heatmaps of relative delay changes (vertical axis – individual replication, horizontal axis – individual trains, green – no delay, yellow – 10% delay change, red – 30% delay change)

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

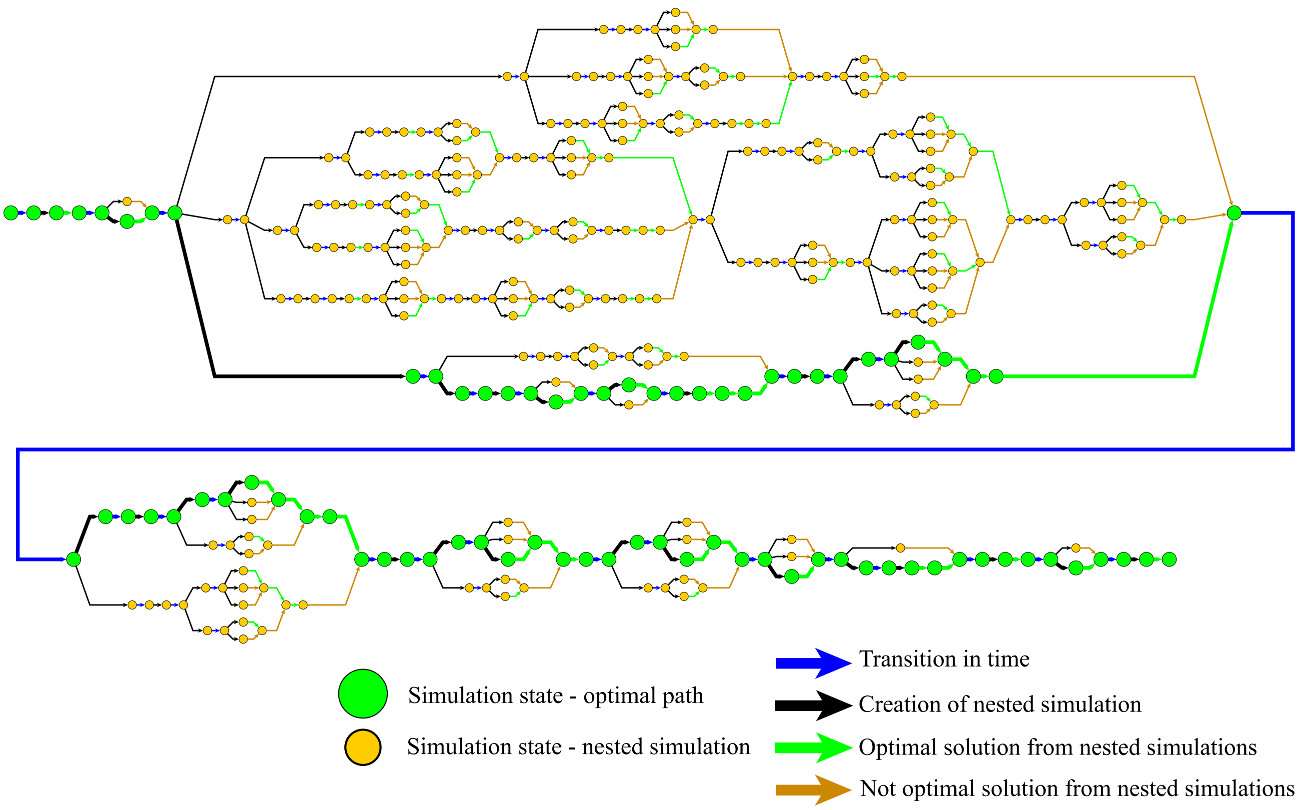


Figure 6: Simulation state diagram – main replication nr. 13, outlook length 5 min (11 conflicts in main simulation, 89 conflicts total, 172 simulations ran)

and the overall results of the results are presented in the table 3. Delays of individual trains by replication are shown in the image 5.

The results show that the technique of nested simulations achieves better results than the simple method of priority planning. Due to the differences between the two simulation models in Villon and MesoRail, it is not possible at this time to say that the improvement was exactly 50% in terms of delays of simulation trains. However, results from nested simulations suggest that use of individual tracks is more appropriate and leads to minimization of subsequent conflicts.

5.4.1. Analyze of nested simulations

Comparison of the two parameterisations of nested simulations with 5-minute and 15-minute outlooks did not entirely go as expected. A longer perspective gave slightly worse average results than a shorter variant. This is probably due to the higher priority of the express trains compared to other trains and the greater emphasis of the algorithm on the effort to optimize their travel times.

The statistics of the number of conflicts that occurred in the main simulation and in the nested simulations and the total amount of simulations performed within one simulation replication can be seen in the table 4. On the i7 quad-core test set, on average, one major replication with 5 minutes outlook needed 11 minutes (of real time) to complete the calculation, for a 15 minute outlook, the average calculation time was 35 minutes.

Figure 6 shows a state diagram of replication nr. 13 (5 minute outlook). Here are all the moments when the nested simulations were executed and the resulting optimal path is shown. This is an example of one of the more complex replications, with many recursive calculations. Simulation of this replication took about 20 minutes of real time.

Figure 7 shows the MesoRail simulator during calculation of replication nr. 79 (15-minute outlook). During the simulation, it is possible to display the status of each running simulation and the state diagram is automatically generated.

6. CONCLUSION

The method of nested simulations suitable for use as a tool for multi-trajectory simulation or as decision support in a simulator, with focus placed on the latter, has been presented.

The method has been implemented within the simulator MesoRail. It presents a mesoscopic simulator of rail transport with focus to test throughput of train nodes.

Nested simulations are tested on a complex case study of a railway station. The simulation results show that the nested simulation technique allows a better evaluation of the alternate station track rather than a simple priority planning method. Due to differences in the microscopic and mesoscopic approach, it can not be

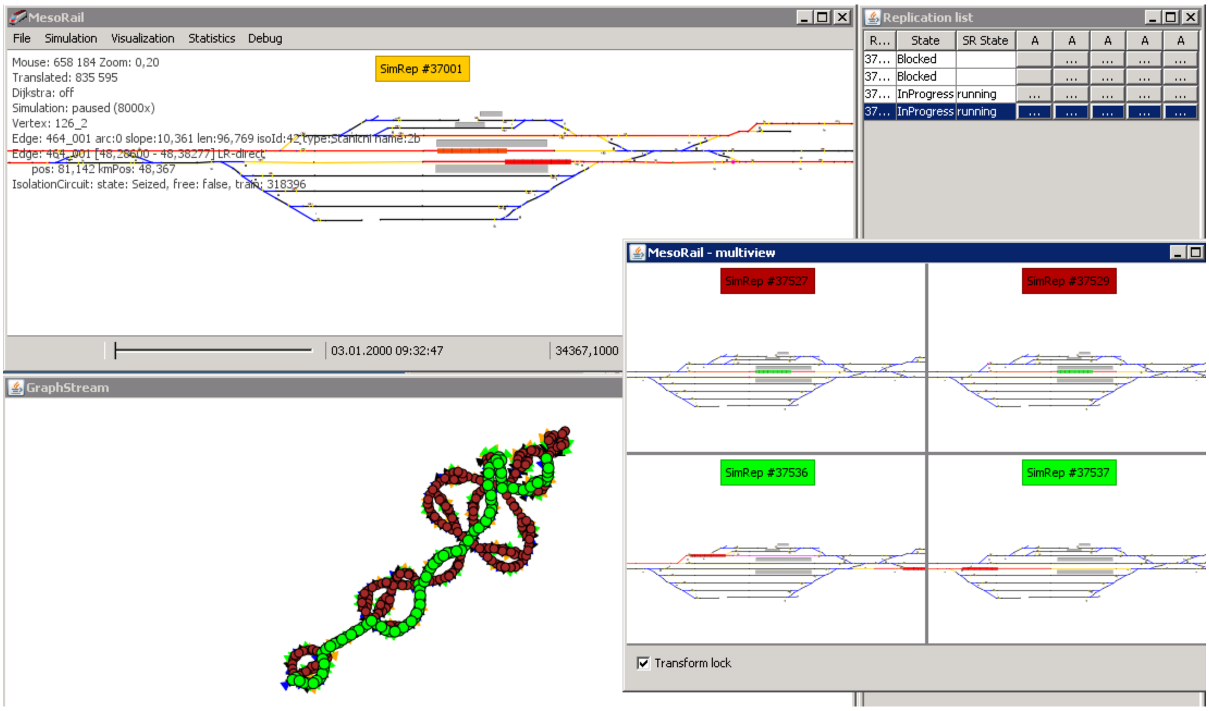


Figure 7: MesoRail tool during simulation (replication nr. 79, 15 minute outlook)

said that the improvement has reached a specific numerical value.

Techniques of nested simulations can be used as a general method of decision support within any application domain. However its implementation use is not quite trivial, and the computational demand of the method is also not negligible.

REFERENCES

Bažant M., Kavička A., 2009. Artificial neural network as a support of platform track assignment within simulation models reflecting passenger railway stations. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit Vol. 223(5) : 505-515.

Bonté B., Duboz R., Quesnel G., Muller J. P., 2009. Recursive simulation and experimental frame for multiscale simulation. In: Proceedings of the 2009 Summer Computer Simulation Conference, 164-172. July 13-16, Istanbul, Turkey.

Diviš R., Kavička A, 2015. Design and development of a mesoscopic simulator specialized in investigating capacities of railway nodes. Proceedings of the European Modeling and Simulation Symposium, 52-57. September 21-23, Bergeggi, Italy.

Diviš R., Kavička A, 2016. The method of nested simulations supporting decision-making process within a mesoscopic railway simulator. Proceedings of the European Modeling and Simulation Symposium, 100-106. September 26-28, Larnaca, Cyprus.

Gilmer J. B., Sullivan F. J., 1999. Multitrajectory simulation performance for varying scenario sizes [combat simulation]. In: WSC'99. 1999 Winter Simulation Conference Proceedings. 'Simulation - A Bridge to the Future' (Cat. No.99CH37038), 1137-1146. December 5-8, Phoenix, AZ, USA.

Gordy M. B., Juneja S., 2010. Nested Simulation in Protfolio Risk Measurement. Management Science 56:1833-1848.

Kindler E., 2010. Nested Models Implemented in Nested Theories. In: Proceedings of the 12th WSEAS International Conference on Automatic Control, Modelling & Simulation, 150-159. May 29-31, Catania, Sicily, Italy.

SIMULATION AND OPTIMIZATION WITH GRASP IN SUPPLY FUEL IN THE NORTH OF MEXICO CITY

Daniel Tello Gaete[(a)], Esther Segura Pérez[(b)].

[(a)] Universidad Austral de Chile
[(b)] School of Business and Economics, Universidad Anáhuac México, Av. Universidad Anáhuac No. 46 Col. Lomas Anáhuac, Huixquilucan, 52786, State of Mexico, Mexico

[(a)]dtellogaete@gmail.com, [(b)]esther.segura@anahuac.mx

ABSTRACT
In this paper, the fuel distribution network from fuel farms to a set of filling stations in the North of Mexico City is simulated and optimized. At present, the company supplies the fuel with a homogenous fleet and it is interested in including a heterogeneous fleet seeking to minimize the ordering, holding, and transportation costs, within constraints of inventory in filling stations, as well as the truck capacity. Data collection, demand analysis, and inventory and transport cost calculations are proposed for the simulation in the network; we propose a mathematical model with metaheuristics Greedy Randomized Adaptive Search Procedure (GRASP), with elements of linear programming. Finally, the results are analyzed getting 44.8 per cent in savings when compared with the current situation

Keywords: Mathematical model; metaheuristics; GRASP; inventory; transports; supply chain managements; fuels.

1. INTRODUCTION

Fuel supply has been studied since 1959, when Dantzig and Ramser, evaluated the optimization of gasoline-transporting vehicle routing from a terminal to different fuel stations.

Since then, a variety of bibliographical material on optimization and simulation of fuel supply has been published, most of them on optimization of the distribution from production sites to refineries, as well as from refineries to mayor storage terminals, mostly by pipelines.

Since then, a variety of bibliographical material on optimization and simulation of fuel supply has been published, most of which deals with optimization of the distribution from production sites to refineries, as well as from refineries to major storage terminals, mainly by pipelines.

Nowadays, the Supply Chain Management of petroleum fuel is very important for the development of human life in each community of the planet. Mexico City is one of the most populous cities in Latin America and the world; the supply of fuel in all of boroughs is not easily achieved. The goal of this work was to obtain a mathematical optimization of the quantity of each fuel station and the minimization inventory and transport costs.

2. PROBLEM STATEMENT

Mexico City has four fuel farms, that supply with three types of fuel: Fuel A (FA), Fuel B (FB) and Diesel (D) to 371 Filling Station (FS) in 16 boroughs and some suburban towns within the metropolitan area. The fuel farm has a homogenous fleet for supply with fuel to each FS.

In the FS, a continuous review and control inventory system and a weekly demand forecasting and scheduling method have been implemented, because it is necessary to have a good service level for the costumer.

Currently the company supplies the fuel from the fuel farm to the FS with a homogenous fleet of trucks with capacity of 20,000 cubic meters. The enterprise is interested in knowing if a heterogeneous fleet and their scheduling is convenient for the supply in the borough of Azcapotzalco in the North of Mexico City in order to minimize the inventory and transport costs. The studied region is shown in the Figure 1.

Figure 1: Study region and FS localization.
Source: Adapted of http://www.ubicalas.com

Now we will present a supply and management inventory proposal where we hope to decrease the inventory and transport cost, furthermore to determine the optimal type of truck for the supply. To validate this information the research of other authors like Coelho, Cordeau y Laporte (2012) was used as a reference, where they make a mathematical model that minimized inventory and transport cost of the vehicle fleet; Kasthuri y Seshaiah (2013) presented an inventory model with a lot of product with dependent demand; and Genedi y Zaki (2011) proposed a model for quality control in the inventories.

3. MATERIALS AND METHODS

3.1. Data collection

There is only data of monthly demand of FS1 and FS6 (between January 2014 and October 2015). To get the demand data of 16 FS, a geometric structure called the Voronoi Diagram (Guth & Klingel, 2012) was used to limit the supply area (Okabe, Boots & Sugihara 1992) of each FS and to get the estimated demand in relation to population. This data was obtained with tools like AutoCAD and the database of The National Insitute of Statistics and Geography (http://www.inegi.org.mx/).

3.2. Demand behavior

The analysis of demand behavior of fuel was made with a Coefficient of Variation (CV) developed with Peterson & Silver (1987); this is used to know if the demand is deterministic or probabilistic (Andrade et al., 2014). If the CV is less or equal to 0.2, the data is poorly dispersed in relation with the mean, because the demand is considered deterministic. The CV is the ratio between the variance and the mean squared like is in the Equation 1.

$$CV = \frac{\sigma^2}{DM^2} \tag{1}$$

3.3. Transport and inventory costs

The transport cost is calculated for the three types of different trucks; it is classified according to its capacity: 10, 20 and 60 cubic meters. This cost is proportional to distance between the fuel farm and the FS, the performance and the efficiency by type of fuel, and the cost per liter of fuel.

Table 1. Holding cost

	Holding Cost [$/m³/month]		
FS	**FA**	**FB**	**Diesel**
1	$0.09	$0.32	$0.15
2	$0.04	$0.12	$0.06
3	$0.08	$0.26	There is not
4	$0.04	$0.13	There is not
5	$0.04	$0.13	There is not
6	$0.05	$0.16	$0.08
7	$0.13	$0.41	There is not
8	$0.03	$0.09	$0.04
9	$0.02	$0.06	There is not
10	$0.06	$0.21	$0.10
11	$0.01	$0.05	$0.02
12	$0.02	$0.06	$0.03
13	$0.05	$0.18	$0.09
14	$0.04	$0.13	There is not
15	$0.03	$0.10	There is not
16	$0.04	$0.14	$0.07
17	$0.05	$0.16	$0.23
18	$0.09	$0.31	$0.15

Source: Elaborated by the author

The inventory cost is composed with the order and holding (Mora, 2008). The holding cost consider the series to storage fees with relation value of the land, opportunity cost, workforce, insurance, and energy consumption (Benítez, 2012). The order costs are the fixed costs involved in making the fuel order. In Table 1 the values of the cost are described.

Table 2. Transport cost

	Transport Cost [$/truck]		
FS	**Truck 10 m³**	**Truck 20 m³**	**Truck 40 m³**
1	$3.95	$4.18	$6.13
2	$4.58	$4.85	$7.10
3	$2.33	$2.47	$3.62
4	$3.05	$3.23	$4.73
5	$3.47	$3.68	$5.38
6	$4.55	$4.82	$7.05
7	$2.57	$2.73	$3.99
8	$4.82	$5.10	$7.47
9	$5.15	$5.45	$7.98
10	$3.17	$3.36	$4.92
11	$1.29	$1.36	$2.00
12	$2.99	$3.17	$4.64
13	$2.81	$2.98	$4.36
14	$2.27	$2.41	$3.53
15	$1.77	$1.87	$2.74
16	$2.69	$2.85	$4.18
17	$3.83	$4.06	$5.94
18	$5.12	$5.42	$7.94

Source: Elaborated by the author

3.4. Mathematical Model

Table 3: Parameters, index and decision variables.

	Item	Descriptions
Index	I	A set of fuel, i = FA, FB, D
	J	A set of truck, j = Truck 10 m^3, Truck 20 m^3, Truck 40 m^3,
	K	A set of FS, k = FS1, FS2, …, FS18
Parameters	D_{ik}	Demand of fuel i in the FS k [m^3/month].
	Ch_{ik}	Holding cost fuel i in the FS k [\$/m^3/month].
	Co_{ik}	Order cost fuel i in the FS k [\$/order].
	Ct_{jk}	Transport cost truck j in to supply fuel in the FS k [\$/ truck].
	Ctr_j	Capacity truck j [m^3].
	Ces_{ik}	Capacity FS [m^3].
Decision variables	Q_{ijk}	Quantity fuel i to order with truck j in the FS k [m^3].
	T_{ijk}	Numbers trucks j to supply fuel i in the FS k [numbers].

Source: Elaborated by the author

The variables involved in the design of the network of fuel distribution in Azcapotzalco is related with customer needs and the networks costs (Chopra, 2001). The index, parameters and decision variables are shown in Table 3.

For the mathematical formulation, a nonlinear programing mixed integer was developed, where the objective function is nonlinear and the constraints are linear. Equation 2 is the objective function that minimizes the holding, order and transport costs. Equation 3 says that the order quantity cannot exceed the inventory capacity in the FS; Equations 4 and 5 are the contrast about filling the truck between 90 and 95 per cent of its capacity; finally, Equation 6 is related with nature of variables like non-negative and the number of truck like integer.

$$Min\ CT = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{18} Ch_{jk}\cdot Q_{ijk} + \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{18} Co_{ik}\cdot \frac{D_{ik}}{Q_{ijk}} + \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{18} Ct_{jk}\cdot T_{ijk} \qquad (2)$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{18} Q_{ijk} \le Ces_{ik} \qquad \forall i,k \qquad (3)$$

$$Q_{ijk} \ge 90\% \cdot Ctr_j \cdot T_{ijk} \qquad \forall i,j,k \qquad (4)$$

$$Q_{ijk} \le 95\% \cdot Ctr_j \cdot T_{ijk} \qquad \forall i,j,k \qquad (5)$$

$$Q_{ijk} \ge 0, \quad T_{ijk} \ge 0 \qquad T_{ijk} \in \mathbb{Z} \qquad \forall i,j,k \qquad (6)$$

3.5. Simulation developed with GRASP Algorithm

The simulation model was based on imitation, with a mathematical model, a real situation, the study of the properties and operative characteristics (Heizer, 2008).
This model was developed using the metaheuristic GRASP; the search algorithm has two phases: building a feasible solution and a local search with iterations in the neighborhood where the solutions are found (Festa, 2009).
The building solution is developed in Microsoft Excel 2016 where 48 different feasible combinatories were found for transporting fuel from fuel farms to FS. Two variables were defined: α_{ijk} y $ß_{ijk.}$

Table 4. Variables α_{ijk} y $ß_{ijk.}$

Variables	Description
$\boldsymbol{\alpha_{ijk}}$	Alternative to supply fuel for fuel i, truck j and FS k. This variable is associated with T_{ijk} (numbers trucks).
$\boldsymbol{ß_{ijk.}}$	Percentage filling truck. It takes the values 90, 91, …, 95 per cent. It is defined to supply fuel for fuel i, truck j and FS k.

Source: Elaborated by the author

The variables α_{ijk} y $ß_{ijk}$ are necessary for the formation with variable Q_{ijk} (order quantity) represented in Equation 7 and then in Equation 8 that represented the total cost supply (CT_{ijk}). These variables are simulated in each iteration, while the better solution is found.

$$Q_{ijk} = T_{ijk}(\alpha_{ijk}) \cdot \beta_{ijk} \cdot Ctr_j \qquad \forall\ i,j,k \qquad (7)$$

$$CT_{ijk} = Ch_{jk} \cdot Q_{ijk} + Co_{ik} \cdot \frac{D_{ik}}{Q_{ijk}} + Ct_{jk} \cdot T_{ijk} \qquad (8)$$

The description about algorithm is explain below:

Algorithm

1: Set high value to variable CT_{ijk}*
2: *while* iterations < total iterations {
  set α_{ijk} y $ß_{ijk}$
    if ($CT_{ijk}' \le CT_{ijk}^*$ & CT_{ijk} !=0) {
    $CT_{ijk}^* = CT_{ijk}'$
    else
    $CT_{ijk}^* = CT_{ijk}^*$
    }
  }

3: $Min = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{18} Y_{ijk} CT_{ijk}^{*}$

$\sum_{j=1}^{3} Y_{ijk} \ge 1$

$\sum_{j=1}^{3} Y_{3jk} = 0 \qquad \forall\ k = 3, 4, 5, 7, 9, 14, 15.$

$Y_{ijk} \in \{0,1\}$

4: END

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The first step is the initialization phase, where variable CTijk setting value 999,999. The second stage is the local search when it stops the total iterations defined by the user.

In each iteration there are setting random variables permitted for αijk y ßijk with the proposal the to building a set of solutions CTijk'. If CTijk' is less than to CTijk* and non-equal to zero, this variable is saved like the new CTijk*, otherwise CTijk* is the same value.

When the iteration stops, it is applying a linear binary programming: the objective function minimizes the inventory and transport cost for to supply fuel i, truck j and FS k. The first constrains assures the supply fuel to everything all the FS. The second constrains says the FS is not suppliedy with Diesel. And Then the last constrains is are about the binary nature binary of decision variables.

4. RESULTS

The developmented of the model was compiled in Dev-C++ with language programming C++ making 100 million iterations in Windows 10, using an Intel Core i3 processor, Intel Core i3 CPU 2.40 GHz CPU for the stages 1 and 2. SThe stage 3 was run in Solver of Microsoft Excel 2016. The results are available in the Tables 5 and 6.

Table 5. Quantity of gas by FS.

FS	FA [m^3]	FB [m^3]	Diesel [m^3]
1	76	38	76
2	95	76	76
3	95	38	0
4	95	76	0
5	95	76	0
6	76	76	76
7	76	38	0
8	95	76	76
9	95	76	0
10	95	76	76
11	95	95	95
12	95	76	95
13	95	76	76
14	95	76	0
15	95	76	0
16	95	76	76
17	95	76	38
18	76	38	76
Average	90.8	68.6	46.4

Source: Elaborated by the author

Table 6. Number of truck and capacity [m^3].

FS	N° Truck FA/ Cap [m^3]		N° Truck FB/ Cap [m^3]		N° Truck Diesel/ Cap [m^3]	
1	2	40	1	40	2	40
2	5	20	2	40	2	40
3	5	20	1	40	0	0
4	5	20	2	40	0	0
5	5	20	2	40	0	0
6	2	40	2	40	2	40
7	2	40	1	40	0	0
8	5	20	2	40	2	40
9	5	20	2	40	0	0
10	5	20	2	40	2	40
11	5	20	5	20	5	20
12	5	20	2	40	5	20
13	5	20	2	40	2	40
14	5	20	2	40	0	0
15	5	20	2	40	0	0
16	5	20	2	40	2	40
17	5	20	2	40	1	40
18	2	40	1	40	2	40

Source: Elaborated by the author

In Table 7 it is shown that the saving holding costs and transport costs are increased on average 69.8 and 118.6 per cent respectively, while the order costs are reduced on average in 59.8 per cent. A relevant aspect to consider is that the increase in holding and transport costs are not as significant as the the order costs, mainly because the holding cost is cheaper than the order cost. In general, the total saving is 44.8 per cent, equivalent to $1,980 each month.

Finally, the saving total in a year is $23,770.61, in Figure 2 a graph of the current and the proposed costs is shown.

Table 7. Current situation versus proposal situation

	Actual situation monthly			Proposal situation monthly		
FS	**Hold Cost.**	**Order Cost**	**Trans Cost**	**Hold. Cost.**	**Order Cost**	**Trans Cost**
1	$9.9	$182.1	$12.5	$15.4	$51.8	$30.6
2	$5.0	$289.1	$24.2	$8.2	$112.6	$52.6
3	$5.6	$159.9	$4.95	$8.6	$41.3	$15.9
4	$3.2	$195.9	$9.7	$6.7	$68.8	$25.6
5	$3.6	$145.9	$14.7	$6.9	$66.1	$29.1
6	$5.5	$256.6	$19.2	$11.0	$89.7	$42.3
7	$8.9	$102.1	$5.45	$12.6	$30.5	$11.9

8	$4.4	$272.0	$35.7	$6.4	$141.3	$55.4
9	$2.1	$245.1	$32.7	$3.2	$144.6	$43.2
10	$6.9	$203.6	$13.4	$14.4	$63.0	$36.4
11	$3.3	$389.6	$13.6	$4.1	$240.0	$20.4
12	$3.8	$341.5	$25.3	$4.9	$187.2	$40.9
13	$6.0	$234.9	$11.9	$12.6	$72.8	$32.3
14	$3.5	$149.5	$9.6	$6.7	$67.7	$19.1
15	$3.0	$171.3	$9.3	$5.2	$90.7	$14.8
16	$4.6	$308.6	$11.4	$9.6	$95.6	$30.9
17	$8.2	$194.6	$16.2	$12.6	$73.6	$38.1
18	$9.5	$189.8	$16.2	$14.9	$53.9	$39.6
Avg	$5.4	$224.0	$15.9	$9.1	$94.0	$32.2

Source: Elaborated by the author

5. CONCLUSIONS

According to the response of GRASP algorithm there is a monthly saving of 44.8 per cent in cost, using a heterogeneous fleet. The company needs to implement a new system of supply fuel.

The GRASP is a metaheuristic technique, when it is possible to find a better solution in this problem with another algorithm.

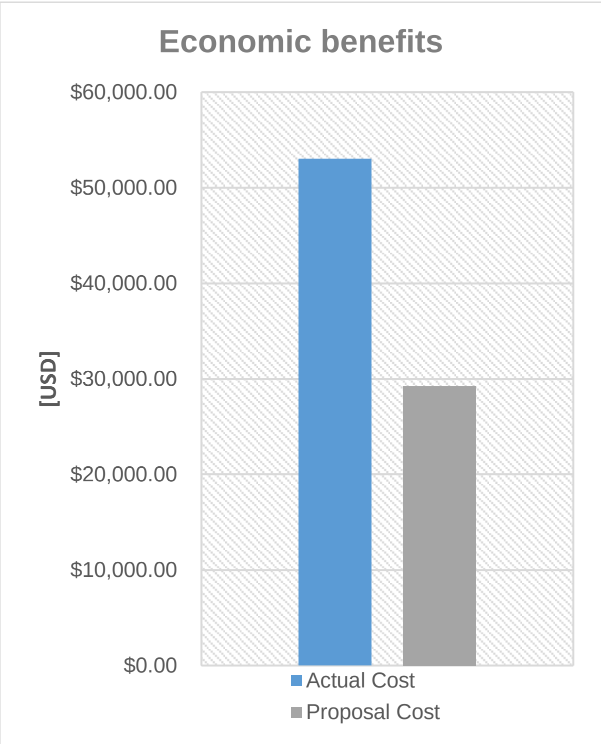


Figure 2: Economic benefits
Source: Elaborated by the author

REFERENCES

Andrade C., Tello D. & Romero O. 2014. Modelo matemático para la administración de inventario en una microempresa del rubro alimenticio. Caso de aplicación. En: VII Simposio Internacional de Ingeniería Industrial: Actualidad y Nuevas Tendencias Lima, Perú. 448p, 2014.

Benítez R. 2012. Influencia de los costos de mantenimiento en la toma de decisiones. La Habana: Centro de Inmunología Molecular.

Chopra S. 2001. Designing the Distribution Network in Supply Chain. Kellogg School of Management, Northwestern University. 1-10.

Coelho L., Cordeau, J. & Laporte, G. 2012. The inventory-routing problem with transshipment. Computers & Operation Research. 39 (2012) 2537-2548, 2012.

Dantzig, G., Ramser J. (1959). The Truck Dispatching Problem. Management Science, Vol. 6 No. 1, pp 80-91.

Dev-C++. Web site. Consulted in 2016. https://sourceforge.net/projects/orwelldevcpp/

Festa P. and Resende, M. 2009. An annotated bibliography of GRASP – Part I: Algorithms. International Transactions in Operational Research. N°. 16, pp 1-24.

Guth N., and Klingel, P. 2012. Demand Allocation in Water Distribution Network Modelling – A GIS-Based Approach Using Voronoi Diagrams with Constraints. Application of Geographic Information Systems, Dr. Bhuiyan Monwar Alam (Ed.), ISBN: 978-953-51-0824-5, InTech, DOI: 10.5772/50014.

Heizer J., Render, B., 2008. Dirección de la producción y operaciones. 8ª ed. Madrid, Pearson Education pp. 458-459.

INEGI: Instituto Nacional de Estadística y Geografía (INEGI). Web site: www.inegi.org.mx.

Kasthuri R., and Seshaiah C. 2013 Multi-item EOQ model with demand dependent on unit price. Applied and Computational Mathematics. Vol. 2, N°. 6, pp. 149-151.

Kotb K., Genedi H. & Zaki S. 2011. Quality Control for Probabilistic Single-Item EOQ Model with Zero Lead Time under Two Restriction: A Geometric Programming Approach. International Journal of Mathematical Archive, Mar -2011, pp. 335-338

Microsoft Excel 2016. Web site: https://www.office.com/.

Mora L. 2008. Gestión Logística Integral. Ecoe Ediciones. Bogotá.

Okabe A., Boots B. & Sugihara K. Spatial tessellations: concepts and applications of Voronoi diagrams. Willey New York.

Peterson, R. and Silver, E. 1985. Decision systems for inventory management and production planning. Willey New York.

Ubicalas.com. Web site. http://www.ubicalas.com/.

Winston W. 2004. Investigación de operaciones: Aplicaciones y algoritmos (4ª ed.). Thompson. México D. F.

AUTHORS BIOGRAPHY

Daniel Tello Gaete. Industrial Engineering, Universidad Austral de Chile. He studied a semester at the National Autonomous University of Mexico (UNAM). He has also done research about inventory, logistic and transporting related with operation research and mathematical models in papers presented in Peru, Colombia, Mexico, and Chile. Currently, he works in Solver Austral, a company dedicated to advice to SME in the south of Chile. (www.solveraustral.com).

Esther Segura Pérez. Is a PhD in Operations Research and Industrial Engineering by the National Autonomous University of Mexico. Currently Associate Professor "C" full time in the Department of Industrial Engineering in the area of Operations Research and applied statistics in the Division of Mechanical and Industrial Engineering (DIMEI) of the Faculty of Engineering of the UNAM. He is also an external consultant of the engineering institute by the same university.

Their contributions have been given in: distribution network design, optimization, analysis and design of algorithms and location of services.

SYNCHRONIZATION ALGORITHM FOR PEER-TO-PEER INTERACTIVE DISTRIBUTED SIMULATION IMPLEMENTED IN A SINGLE-THREADED WEB APPLICATION

Štěpán Karták

University of Pardubice, Faculty of Electrical Engineering and Informatics

stepan.kartak@student.upce.cz

ABSTRACT
This article presents a synchronization algorithm for distributed interactive peer-to-peer simulation in a web browser. It is a practical utilization of web browsers in combination with modern technologies for realization of computation- and network-heavy simulation tasks. Due to present limitations of web browsers, especially due to their limited computing power, a class of realizable tasks is defined. Such tasks can be successfully solved with web browsers. The article also covers the operating principles of web applications, with focus on JavaScript and problems arising from its concept. The text describes the algorithm and used network topologies of the logical processes, and their synchronization methods. Advantages and disadvantages of realizing a simulation with a web browser are described as well, as well as the reasons why the not-so-frequently used peer-to-peer simulation was used. To conclude, a use case for application scope testing is characterized, i.e. identifying the appropriate number of logical processes, frequency of interactive interventions, usable numbers of objects, etc., for which the presented algorithm is appropriate.

Keywords: Distributed Simulation, Web-based simulation, HTML5, WebRTC

1. INTRODUCTION

This article is focused on using a web browser to realize a user-friendly interactive distributed simulation. The goal is to design and create a relatively general algorithm (in terms of defined class and application scope), which would provide the user with basic functionality (simulation core, synchronization, interactive approach, …) to realize simple simulations. The user does not have to solve the basic implementation problems in the limiting environment of a web browser, and can focus on the implementation of the simulator behavior.

Today, web browsers are very well suited for such applications. Since the year 2012 (Karták 2014), web browsers have offered functionalities that allow realization without the use of third-party plugins. However, even in spite of the significant advances made by web browsers in the last few years, the realization of a distributed simulation is not possible without a number of compromises.

The core of the solution is based on the previous work on distributed web simulations (Kartak 2015, 2016), which focused on trainer applications, that is for applications for testing (examination, education) of workers / dispatchers and distributed web simulations in general. This solution served as a proof of concept, and the article expands the concept.

2. WEB APPLICATIONS AND TECHNOLOGIES

A web application from our point of view is considered to be distributed from the web server (in a certain configuration) through a computer network (local or global Internet network) into the client device, which is an instance of a web browser window. The application is distributed statically (an HTML description of the page structure is downloaded from the server, as well as additional information – uncompiled JavaScript code, CSS files, images, fonts etc.). After loading all the static parts listed above, the dynamic part of the web is launched – the program code written in JavaScript. This code is distributed to the client computer uncompiled, and is run according to the requirements of the author of the code.

The program of the webpage reacts to user interaction (typically mouse and keyboard) or external information (application state update by the server, information and events provided by the web browser itself).

All the actions stated above are called events (js event), and are executed sequentially, in the chronological order of their creation. The order or priority of their processing cannot be influenced on the browser level. This approach is suited for primarily static web pages with minimal amounts of program code or time-consuming calculations. When realizing heavily interactive applications, such as interactive distributed simulations, this approach is generally inappropriate (see chapter 4.4).

Web applications are always downloaded from a web server, which may or may not participate on further

client-side runtime. The server is often only a source of static content, and does not store any user states, or provides only elementary functionality, such as user identification, and is not informed of any further states of the client-side, or the server receives only the results of the algorithm that was run on the client-side, with no interaction or interruptions by the server after the static content is downloaded. Sending the results of user or script activity is typical for web applications.
Web applications can transfer workload to the client-side (web browser). From this point of view, a browser may be considered a thick client as well as thin client, depending on the used approach.

3. CHARACTERISTICS OF APPLICATIONS IN THE TARGET DOMAIN OF THE ALGORITHM

In this chapter, used terminology is stated, along with a basic introduction to the algorithm.

First, it must be stated that the algorithm assumes use of discrete distributed simulation, that is to say that the behavior of applications for which the algorithm is suitable must be determined by discrete events.

3.1. Basic terminology

As stated above, the simulation runs in web browsers. One specific instance of a web page (in a web browser) represents one logical process (LP). A group of logical processes forms a distributed simulation.

The algorithm primarily works with three elements:

- **Entity** ... in fact an object passing through the simulation,
- **Activity** ... an event handling procedure,
- **Simulation event** ... a planned discrete event of a concrete activity.

Interactive interventions are realized as interruptions of discrete planned activities.

3.2. Application aspect

Use of web browsers presents an inexpensive way of realizing distributed interactive simulation. A typical user might be a small company, requiring a trainer simulator or training software, that can be described by an algorithm of discrete events.

The introduced algorithm is designed for three application classes:

1. **Trainer simulator applications.** Assume a group of workers that forms a single team solving a problem or reacting to a chosen situation. Every worker/user works with a browser, where he or she observes a simulation scene, and interacts with the simulation runtime within the frame of the assigned logical process.
   Another example may be a railway station dispatcher in the scope of a region (where there are more dispatchers). Another example may be the simulation of a technological process / production, where every employee is responsible for a part of the process.
2. Realization of a **simple multiplayer game**, where logical processes represent the space for individual players, with implicitly shared state-space of the simulation (the game environment). This application class is covered by the use cases (chapter 7).
3. **Distributed space for data exchange** within a work group. This application class does not directly represent a simulation. Only the synchronization methods are used to keep the memory space up to date for all of the logical processes.

The primary application class can be generally classified as a distributed system requiring interactive approach, based on discrete events and with no complex calculations present.

3.3. Networking aspect and topology of the logical processes

This solution utilizes primarily web browsers, that contain a majority of the simulation calculations. The server part is not present in the calculations or logic, and serves only for undemanding secondary activities (initialization of the connection between clients, creation process of the simulation, etc.).

The solution is a purely peer-to-peer simulation.

This solution was chosen due to the fact that web browsers are commonly found on computers, and nothing prevents their participation in simulations. The opposing server architecture (for example the commonly used HLA architecture with federates running on servers) requires high-end (and expensive, or not widely accessible) servers (Kuhl et al. 2007).

Creating a peer-to-peer network of clients allows to tap into the potential of the client computers and, at the same time, requires no extra expenses such as powerful servers or software. With web browsers, there are usually no connectivity issues in terms of firewall limitations and similar problems, as web browsers activities are generally considered safe, due to the sandbox nature of the browsers themselves.

This solution however has certain disadvantages as well. The most significant disadvantage is the considerably lower performance in comparison to desktop applications. This is another reason why direct connection between the clients is beneficial, as opposed to communicating through a server. The sent message travels directly to the target client, instead of two messages being sent (client-server and server-client). If we consider a local network (low latency, high bandwidth), the theoretical time needed to deliver a message from client to client when using a server architecture is double the time needed in a peer-to-peer architecture.

3.4. Implementation scope of the synchronization algorithm

The introduced algorithm is a general algorithm that solves the synchronization problems of logical processes.

The implementation of the algorithm solves the following (details in chapters 6 and 7):

- discrete simulation core (primarily the event calendar queue),
- events,
- prototypes of logical processes,
- prototypes of discrete activities and auxiliary discrete activities (running in the background for algorithm needs),
- prototypes of entities,
- synchronization of logical processes,
- handling of interactive user input,
- elementary handling of entity collisions,
- (optional) rendering of a simple 2D scene,
- (optional) simulation interruption handling when waiting for user input.

The following is not solved by the algorithm:

- specific implementation of the target application,
- specific logical processes,
- specific simulation activities,
- specific entities,
- specific collisions (of entities) and exceptions of interactive user input handling.

The primary goal is to provide an algorithm that implements the necessary structures and solves the above stated problems for the user, allowing him or her to concentrate on the process of modelling the solved use case – the logic (activities) and objects (entities), and their interactions.

The aim of this work is not to create competition for extensive standards such as DIS, HLA, TENA etc. (IEEE 1278.1-2012; Kuhl et al. 2007), and similar, as that is, due to the limitations (see chapter 4.4) of web browsers, impossible.

3.5. Reusability of the solution

The aforementioned algorithm will be available as a JavaScript function library, which shall implement the following functionalities:

- Connection of a logical process into the administration interface,
- synchronization of a running simulation,
- basic functional support for realization of animated output.

4. USED TECHNOLOGIES

Web distributed simulation could not be realized without new technologies, collectively referred to as HTML5. These functions expand the capabilities of web browsers with functions that were formerly the domain of desktop or server applications (a typical example would be two-way network communication), and achieving the desired effect before HTML5 required use of third-party plugins (typically Java applets), or inefficient solutions (an example might client's periodical queries about state changes, instead of direct "state changed" notice sent directly from the server to the client).

An enumeration of the fundamental HTML5 technologies, on which this solution is based, of follows.

4.1. WebRTC

The WebRTC technology servers to connect clients (instances of browser windows) directly, without the need to use a server as a connecting link. This technology is primarily used for peer-to-peer sound and video transmission (typically videoconferences). However, pure data transmission is implemented as well, which can be used to send user data, and is fundamental for the algorithm's operation.

4.2. WebSocket

This network technology serves to create a permanent (until the browser window is closed) two-way server-client connection. The client may be informed of the server state changes directly by a message from the server, bypassing the need to periodically ask the server "Are there any news?". The second, equally important benefit is the persistent client-server connection. When sending messages, it is no longer necessary to create the connection every time. This can save up to tens of milliseconds, depending on how busy the server is.

4.3. Canvas

The HTML tag *<canvas />* defines area for 2D drawing. Thanks to this HTML element, a drawing area of any size can be created, and drawn upon using JavaScript.

However, the *<canvas />* tag does not allow the scene to be partially redrawn. This is a limiting factor. Depending on the size of the drawing area and the number of the rendered objects (generally graphic primitive types) the time needed to redraw the scene may increase significantly.

The WebRTC, WebSocket and Canvas technologies are the fundamental building stones of the web simulation realization.

4.4. Basic characteristics of JavaScript

The JavaScript programming language, generally used by web browsers to realize dynamic behavior of web pages, is a weakly-typed prototype language. JavaScript also contains several functions which significantly complicate optimization of compiled code, which in turn means that the resulting program performs significantly worse than desktop applications. However, this state improves with new versions of web browsers. Optimization libraries (such as the asm.js) exist, which compile C/C++ language code into strongly optimized JavaScript code (Voracek 2016). However, this approach does not solve the second, more significant, problem – the fact, that JavaScript has a "single-threaded" (this expression is not completely accurate, but captures the essence of the problem, which is why the expression will be used further in the text) approach.

A single-threaded event-based system of executing user code is a critical problem for a distributed simulation-type application. This property of JavaScript means that in practice, all operations are executed synchronously in a single thread. There are no concurrent multi-threading approaches available to the user (Processes that cannot be influenced by the user, such as data rendering, network communication etc. are, however, done asynchronously by the browser). This state does not present a problem to a large group of algorithms that are commonly realized in a browser, but generally complicates construction of algorithms for tasks that require parallel execution of operations (in terms of multi-threading and multi-processor execution), be it for effectiveness or individual tasks' time complexity reasons. Chapter 6.1 covers the specific reasons because of which this state presents a significant problem to realization of web (distributed) simulations.

5. AUXILIARY SOFTWARE SOLUTIONS

As stated before, the presented solution assumes the simulation runs only in a web browser (with no participation from the server), purely peer-to-peer.

However, no peer-to-peer solution is capable of bypassing the server side completely. At the very least, connection initialization must be solved, which is impossible without the participation of a sever element. If we want to conduct distributed simulation, we have to build it somewhere, or at least save the configuration (for example on a web server), from where it will be available to the clients. Due to the fact that the presented solution works with up to 40 connected clients, it is necessary to observe the state and behavior of the client computers.

There are 4 auxiliary server applications, connected into the Administration interface, which runs in the "background" of the simulation itself:

5.1. Model configuration

The fundamental part of the server-side Administration interface of this solution. Serves to register individual types of logical processes and consequentially use then when building the model of the distributed simulation.

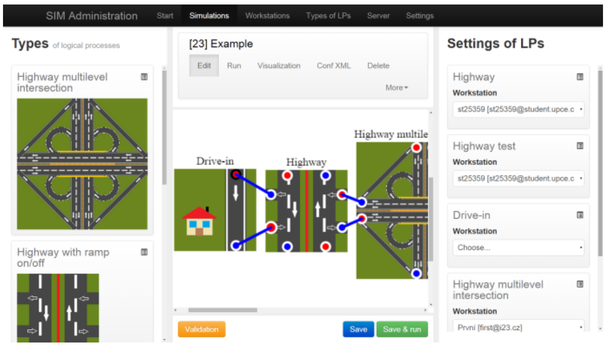


Figure 1: Administration web interface, visual editor; blue lines are network connections between logical processes (chapter 3.2 part 1)

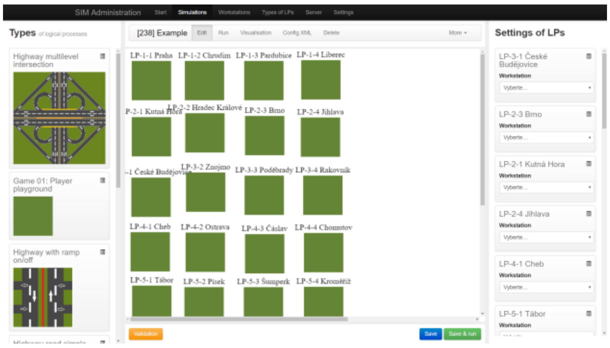


Figure 2: Administration web interface, visual editor; 20 logical processes represents 20 players (chapter 3.2 part 2 and use case – chapter 7)

5.2. Simulation control

A server-side application to which all the clients – logical processes – are connected. Through this application, it is possible to pass commands and instructions or gather runtime information from clients (in bulk). This application serves to control the simulation (initialization, start, pause, end, etc.) in a centralized fashion.

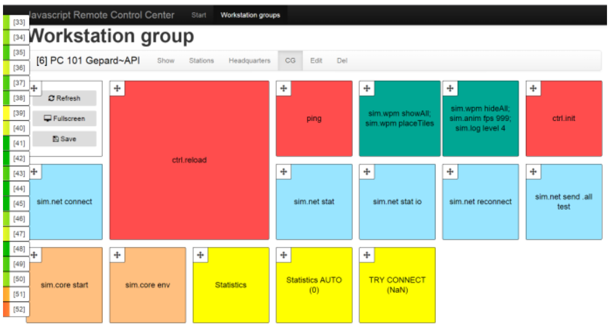


Figure 3: JSRC: Prepared command set, one square is user-defined command (or commands), prepared for touch-devices

5.3. Centralized visualization

This server part (realized as a component of the Administration interface) facilitates recording of the animation output. Screenshots for static preview (see figure 4) of the logical processes' state are captured, as well as the animation activities, making it possible to reconstruct the logical processes and simulation runtime.

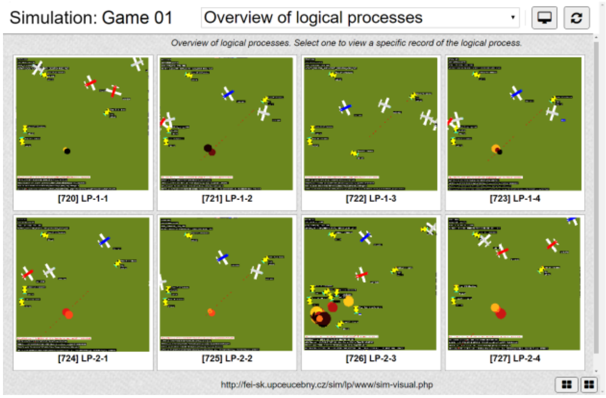


Figure 4: Overview of logical process in simulation in central visualisation, 8 LP/player, shared space/playground, differences between pictures are caused by the creation of a screenshot at different times

5.4. Initialization server

The initialization server serves to create a peer-to-peer connection between browsers. This component also graphically depicts the state of connections between the individual clients.

6. ALGORITHM CHARACTERISTICS

6.1. Simulator as a single-threaded application

A simulator faces four critical tasks, which are independent and running at all times: (i) simulation core, executing discrete events, (ii) network communication (sending and receiving messages) with other simulation participants, (iii) reactions to user input and (iv) animation output.
These four parts are commonly realized as parallel tasks in classic desktop applications. In JavaScript, this is not possible. This is why the simulator is realized as a series of cyclically repeated operations (only fundamental steps are listed):

1. Interpretation of incoming messages.
2. Interpretation of user input.
3. Synchronization of the logical process, based on steps 1 and 2 (see chapter 6.3).
4. Execution of available (especially in relation to synchronization of logical processes) discrete events.
5. Calculation of animation output:
   (a) entity position calculation,
   (b) calculation of collisions or other interactions between entities,
6. Broadcast data (local LP state information) to other logical processes.
7. Rendering of the situation onto the animation output.
8. Continue by step 1.

6.2. Basic structure of logical processes

A logical process is made up of 6 parts (for an UML diagram see Image 5):

1. Simulation core: operates simulation activities, ensures synchronisation. Includes:
   (a) Calendar: priority queue for simulation activity planning.
   (b) Environment: contains environment and state information related to the simulation (primarily activity handler).
   (c) Modules: any named data structure, usually auxiliary, available to all dependent parts (usually activity handler). Used, among others, for the text report of simulation states.
2. Simulation activity: specified the type of activity, time of execution and any other additional information
3. Activity Handler: execution of given activity type
4. ConnectionRegister: logical process communication realisation layer
5. Animation Activity: Described a graphic element for animation rendering. One of the modules of the simulation core.
6. AnimationManager: renders a scene based on the animation activities

Other program parts that are not critical for the execution of a logical process:

7. SettingsManager: contains a description of the simulation configuration.
8. EntityManager: contains information about entity types and individual entities.
9. ActionManager: describes interactions and eventual reactions of individual entity types.

The solution as a whole works under several basic premises:

- All simulation and animation activities can be serialized.
- All simulation and animation activities can be interrupted at any time (removed from the queue or scene).

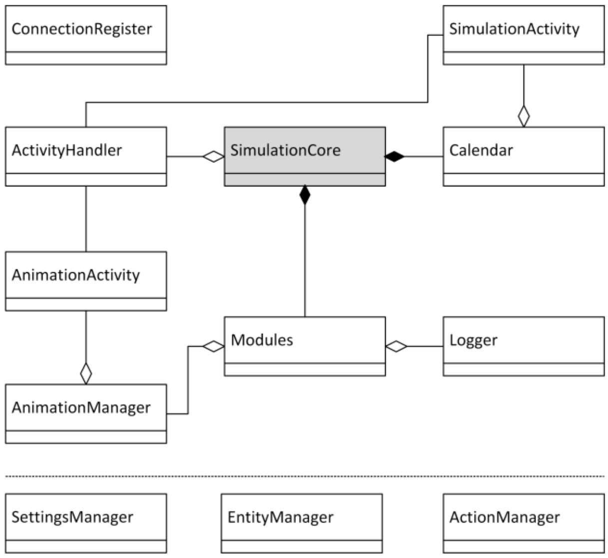


Figure 5: Basic UML schema of logical process (the simulator itself)

The algorithm stated above is only a simplified framework of the used solution, Provided for a basic understanding of the function.
Moreover, the problem presented by the event-based approach of JavaScript is not solved. That means that at any given time, the execution of the algorithm may be interrupted by executing different tasks (user input into the simulation, receiving messages, etc.). The event-based approach is not a problem in general, as the "main" algorithm will continue after the interrupting operation is finished. The problem is that these "unexpected" actions take some time to complete, and thus slow down the calculations and have a negative impact on the animation smoothness, which in the end means worse user experience.

6.3. Topology of connection between logical processes in the simulation model

A one-on-one connection is realized (through the WebRTC technology). During the initialization process of the simulation, a connection is made between each logical process. The connection is established primarily in order to maintain a global memory space. All state changes of a logical process are sent to all other logical processes (in fact a broadcast of state changes). Every receiving logical process then decides whether and how to process the received data. This process was inspired by the DIS standard.
This broadcast of changes between all logical processes is also used for synchronization purposes.

6.4. Logical process synchronization

Optimistic methods of synchronization are generally more suited for interactive simulation, as they do not require strict time synchronization of the logical processes runtime, which in turn means the calculations (and animations as well) are smoother (thanks to not having to wait for the "slow" logical processes). To ensure smooth operation (especially in terms of animation), the conservative approach is not effective, as it requires a short look-ahead (briefly: max. look-ahead must equals to delay of animation slides – animation FPS 25 required 40 ms between slides / 40 ms look-ahead) to ensure smooth animation, which increases communication load.

A "two-level" synchronization method was chosen:

1. For basic synchronization, the Conservative synchronization technique of sending null messages with a look-ahead (Chandy-Misra-Bryant Distributed Discrete-Event Simulation Algorithm, Fujimoto 2000) was used – the specific implementation can be found in the previous work (Kartak 2015). This method is utilized primarily in during the simulation initialization, and to capture above-average fluctuations (delays) in network communications.
2. For precise synchronization purposes, state information timestamp readings are used, as sent by other logical processes. The received times are compared to the actual system time of the client computer, and based on the differences of the other logical processes and the client logical process (and its anticipated behavior), the speed of the logical process is adjusted – and with it, the animation speed, as it is animation speed that determines the speed of the simulation.
   In this second level of synchronization, the strict time synchronization with conservation of local causality of time is *not* applied. We assume a deviation (depending on the scope of the simulation model) of up to 100 ms. We consider the deviations in this interval to be negligible, and (nearly) imperceptible by the user.

Due to the facts stated above, use on local (e.g. company) networks is presumed, where the latency of messages sent through the WebRTC is usually around 10 ms, which allows smooth runtime of the application.

6.5. Algorithm

We are in a web browser / JavaScript environment, a single thread event based approach.

Initialization

1. Create (WebRTC) connection between all LPs. n logical processes make
   $$n \times n - n$$
   full-duplex connections
2. Waiting until a every connection is established

Start of simulation

Every one logical process

1. Sync level 1: Conservative synchronization technique (see chapter 6.4)

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

2. Waiting until all logic processes are obstructed to run and initialization is not performed. (For example: waiting 5000 ms)

Running simulation (one logical process execution), inc. sync level 2

Every one logical process
SC ... Simulation core
A ... Animation
A.ActivityList ... list of animation activities
E ... simulation event
SC.E ... current (executing) event
SC[C] ... calendar of events/activities

```
DECLARE SC.run
BEGIN
   1. SC.executeUserAndNetworkEvents()
   2. isAnimTimeInFuture = A.time>=SC.time
   3. IF isAnimTimeInFuture
      THEN
      1. WHILE SC.time<=A.time
         THEN
         1. IF SC[C].isEmpty()
            THEN A.start(); follow step 1
            ELSE
            1. follow step 9 to 11

   2. IF SC[C].isEmpty()
      THEN A.start(); follow step 1
   4.
   5. IF SC.time>A.time
   6. THEN
         1. A.setTime(SC.time)
         2. A.setSpeedRatio(1)
      ELSE
         1. A.setSpeedRatio(0.98)
   7. A.start()
   8.
   9. SC.E = Shift first E from CS[C]
   10. SC.time = SC.E.time
   11. e.execute()
   12. IF SC[C].nextEvent().time == SC.time
       THEN follow step 3
       ELSE follow step 8

DECLARE A.start
BEGIN
SC.executeUserAndNetworkEvents()
stepStart = NOW.time
IF A.stepLastTime != 0
THEN timePlus = stepStart − A.stepLastTime
ELSE timePlus = 200 # magic constant for first step
timeAdd = timePlus*A.speedRatio*0.98 # 0.98 is a
constant defining a delay in the execution of the script
itself, experimental value
A.time += _timeAdd
A.stepLastTime = stepStart
WHILE A.ActivityList.hasNext()
1. AACurrent = A.ActivityList.next()
2. AACurrent.timePrepare(A.time) # Calc new
   position
SC.collisionCalculation()
WHILE A.ActivityList.hasNext()
1. IF AACurrent.getStartTime()>A.time
   THEN continue;
2. AACurrent.draw(A.time) # (re)draw activity to
   output buffer
A.outputFrame() # Render to screen
WHILE A.ActivityList.hasNext()
1. AACurrent = A.ActivityList.next()
2. IF AACurrent.isFinished(A.time)
3. THEN A.ActivityList.remove(AACurrent)

animNextDiff = 1000 / A.fps # requested FPS
timeAnimDuration = NOW.time − stepStart
timePlanPlus = animNextDiff − timeAnimDuration
IF timeNextAnimX<0
THEN timePlanPlus = 2
plan(A.start, timePlanPlus)
# plan A.start() by x ms, where timePlanPlus is
demanded delay between now and next output frame
(by requested FPS)
END

A.time  = SC[C]. nextEvent().time # Setup time of
animation output, example
SC.run()

SC.broadcastStateInfoInterval(ms=30) # Send state info
to all another LPs every 30 ms, this is realized as
standard E planned every x ms

SC.initMessageReceiver(  # Receive message event
   (scSender.name, listOfStateInfo) =>
      stateInfoEv = new E
      stateInfoEv.listOfUpdates = listOfStateInfo
      SC[C].addEvent(stateInfoEv, SC.time+1ms)
)
```

7. USE CASE AND TRACKED METRICS

Algorithm and implementation of distributed simulation in a web browser were tested on a game type program (see figures 6 and 7):

- Each LP contains a single user-controlled entity (UCE).
- One shared scene, representing the playing field (all logical processes share a single scene i.e. all users see the same).
- The playing field will be restricted by screen size (the area of the canvas is 1 MPx),
- The user controls the UCE with a keyboard (arrows allow movement in 4 basic directions, spacebar allows the user to shoot) and a mouse (click into the playing field represents a travel destination), figure 8.
- A shot (realized as an entity) travels with a limited speed, giving the remaining users time to react (figure 9).

- The collision of a shot with a soldier causes an action (frag count, unimportant for the use case), figure 10.

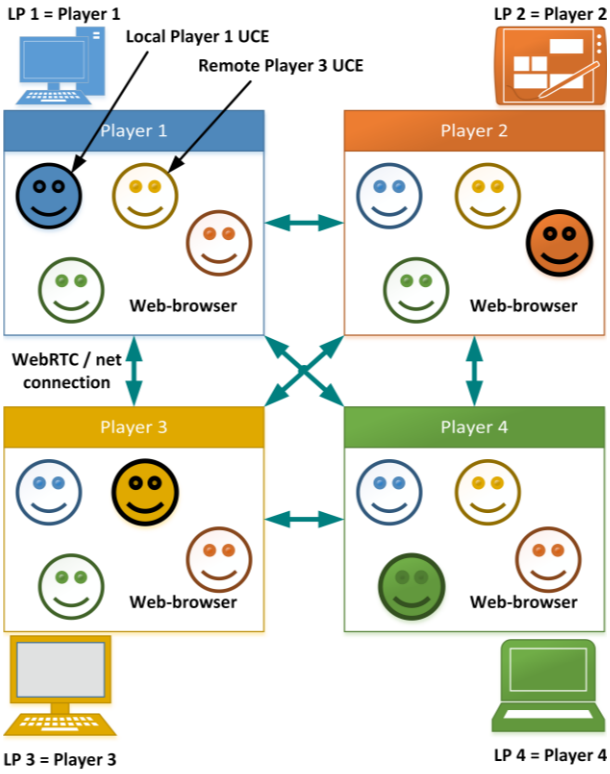


Figure 6: Conceptual picture of use case, LP topology

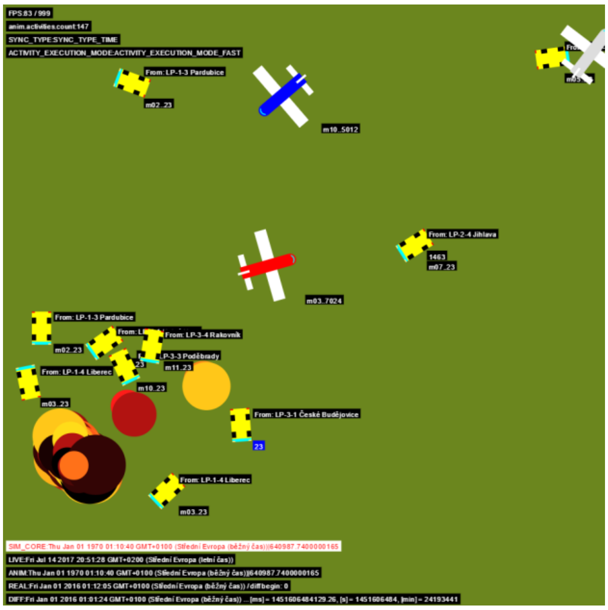


Figure 7: The screenshot from use case – a simple multiplayer game

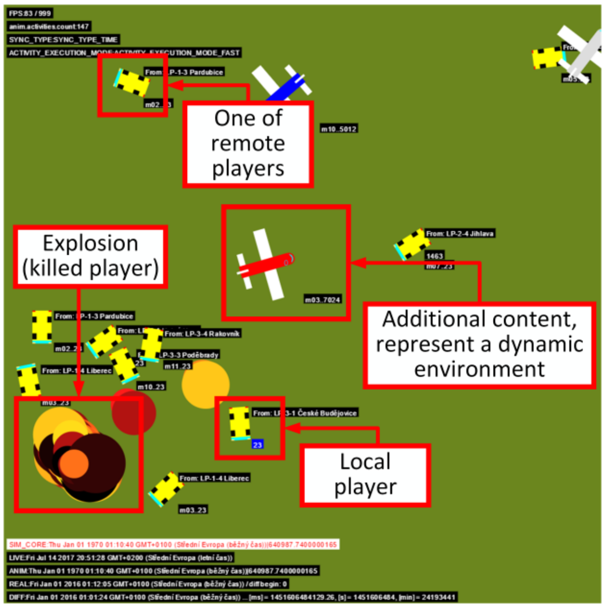


Figure 8: Detailed information about the use case

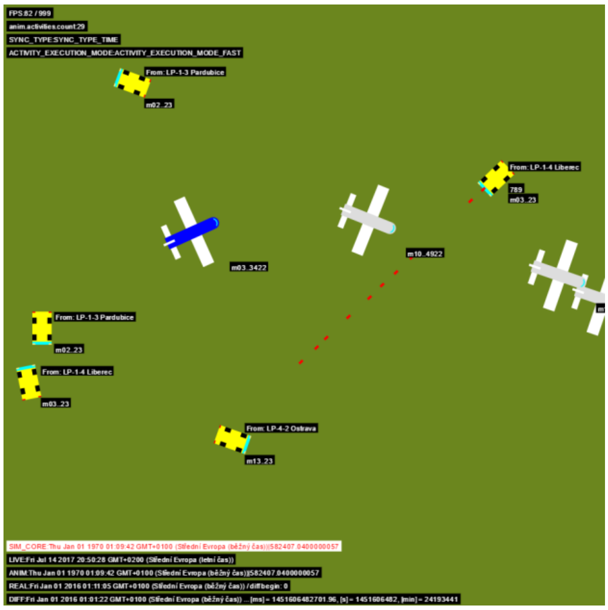


Figure 9: Example of interaction, player 1 shoots bullets

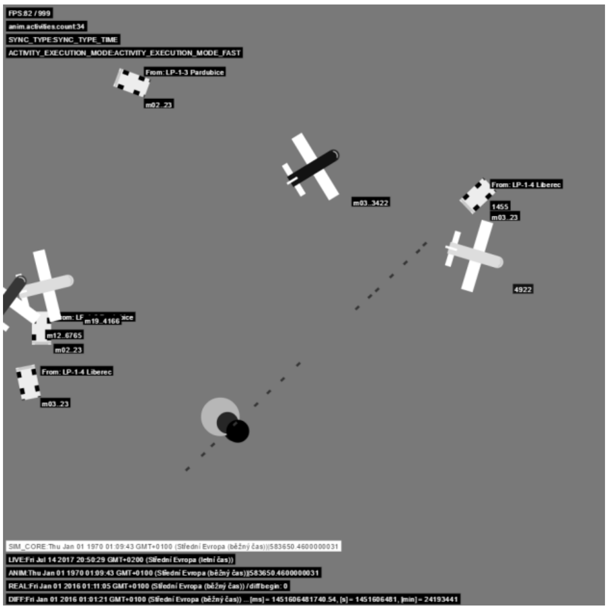


Figure 10: Example of interaction, player 2 was killed – player & bullet interaction and subsequent animation (explosion), gray indicates the end of the local player

The following properties of the simulation was tested:

User interactions (in relation to algorithm runtime interruptions) – event handler does not represent a measurable delay.

Entity-entity collisions (soldier-soldier, soldier-shot) – complexity is determined by the complexity of calculation and optimization of data structures, generally not a problem.

Entity-environment collisions (restriction of access to certain areas of the scene) – depending on the choice of the appropriate data structures and the required accuracy. For example, cross-border control of players' playgrounds was performed once every 100 ms (about 6 to 10 animation frames). This without a visible problem from the player's experience, an FTP increase of 1-2.

User interaction latency during the interaction of a local process with a remote one. The response delay is dependent on the frequency of the message forwarding, at a 30 ms status update, a response of up to 30 seconds for sending the static data, 5 ms for network transmission, max. 30 for sending the response back (sending the status data), 5 ms for network transmission. It follows that the response required after a remote logic process can take up to 70 ms (about 3 animations).

Number of synchronization messages and message latencies.

Scene rendering delays (FPS – max, min, avg) in relation to the number of animated objects – see table 1

Look-ahead for the first level of time synchronization is not relevant.

Differences between local simulation times and global average – the difference is on average several ms, the difference amount decreases with the increasing number of LPs because the increasing sync level 2 algorithm better compensates for the differences.

All the stated properties was tested on simulation model configurations of 8, 12, 20 and 40 logical processes. Measurements took place on identically configured PCs (Intel® Core™ i3-3240 CPU @ 3.40 GHz, 4 GB RAM, Windows 10 64bit, only one application running – Google Chrome browser, version 58). The best performance (FPS plus user experience) is when configuring 20 LPs.

Table 1: Results of tested use case, critical data for simulation run with focusing on animation output and interactive approach

Tested properties	#1	#2	#3	#4	#5	#6	#7	#8
LP count	8	8	12	12	20	20	40	40
User interaction & collision calc enabled	NO	YES	NO	YES	NO	YES	NO	YES
Animation FPS (AVG)	85	80	65	60	33	32	15	11
Sync request count	186	142	137	44	108	119	219	246
Animation 1 frame draw time (AVG) [ms]	34	32	42	42	61	60	92	95
Animation 1 frame draw time (MAX) [ms]	37	50	78	71	79	82	155	180
Animation activity count in animation scene	215	245	278	266	552	548	980	1005

Result notes (table 1):

- User interaction (mouse click, press key SPACE) was programmatically generated. Calculated by uniform distribution (min 400 ms, max 2000 ms) between simulated interactions.
- One animation activity is consists from 5 graphics base elements / primitives (real canvas draw elements) in average.
- Results were collected after 5 minutes (real time) run.

From the table 1, read that the limiting invoice is not the number of LPs (within the specified numbers), but (a) the complexity of the; calculations and (b) the scope (especially in terms of quantity) of the animation.

8. CONCLUSION

The primary motivation for use of web-based simulation is the availability of the runtime environment – web browser – on any computer or modern device connected to a computer network. JavaScript is very well supported by modern-day browsers, and is extensible and well known. This comfort of availability and simplicity is not without a cost – when compared to native applications, the scripts are slow. Web browser simulations can not be compared with native application (using standards like HLA, DIS, TENA, etc. or in general) due to the inequality between compiled languages, multi-threaded access, and graphical output – ie (relatively) direct access to the graphics card.

Altogether, the introduced solution is suited to solving small tasks that do not require complicated calculations and complicated graphical output, but require distributed space or operator workstation. A good example may be the training software operators of the (technological) process, where the distributed approach (different workplaces) is used and complex graphics output is not required – it is just an interactive diagram of the relevant technological process.

Regarding presented use case is at the edge of the technology possibilities. The only possibility of improvement seems to rewrite (source code) the graphics output to WebGL, which allows use graphics card to generate the output. There is the opportunity to reduce the load of current software approach to drawing the output, and use the new available (processor) time for more detailed calculations or larger scale (more entities, events, etc.) simulation itself.

REFERENCES

Fujimoto, Richard M. Parallel and distribution simulation systems. New York: Wiley, 2000. Print.

Kartak, Stepan, and Antonin Kavicka. "WebRTC Technology as a Solution for a Web-Based Distributed Simulation". Proceedings of the European Modeling and Simulation Symposium 2014. Genova: Università di Genova, 2014, s. 343-349. ISBN 978-88-97999-38-6.

Kartak, Stepan. "Web Simulation as a Platform for Training Software Application". Proceedings of the European Modeling and Simulation Symposium 2015. Genova: Università di Genova, 2015, s. 70-78. ISBN 978-88-97999-57-7.

Kartak, Stepan. "Web Simulation as a Platform for Training Software Application". Proceedings of the European Modeling and Simulation Symposium 2016. Genova: Università di Genova, 2016, s. 78-86. ISBN 978-88-97999-76-8.

Voracek, Jan. "Web Simulation as a Platform for Training Software Application". Proceedings of the European Modeling and Simulation Symposium 2016. Genova: Università di Genova, 2016, s. 73-77. ISBN 978-88-97999-76-8.

Kuhl, Frederick, Judith Dahmann, and Richard Weatherly. Creating computer simulation system: an introduction to the high level architecture. Upper Saddle River, NJ: Prentice Hall PTR, 2000. Print.

Tropper, Carl. Parallel and distributed discrete event simulation. New York: Nova Science, 2002. Print.

Kuhl, Frederick, Dahmann, Judith, Weatherly, Richard, Creating Computer Simulation Systems: An Introduction to the High Level Architecture, c2000, Upper Saddle River, NJ; Prentice Hall PTR. ISBN 01-302-2511-8.

Hridel, Jan, and Stepan Kartak. "Web-based simulation in teaching". The European Simulation and Modelling Conference 2013. EUROSIS-ETI, 2013. Print.

The Institute Of Electrical And Electronics Engineers, Inc, 2012, 1278.1-2012: IEEE Standard for Distributed Interactive Simulation - Application Protocols. New York; IEEE. 2012. ISBN 978-0-7381-7310-8.

The Institute Of Electrical And Electronics Engineers, Inc, 1996, 1278.2-1995: IEEE Standard for Distributed Interactive Simulation - Communication Services and Profiles. New York; IEEE. 2002. ISBN 0-7381-0994-0.

SIMULATION FOR TEACHING THE PRODUCTION CELL CONCEPT IN CIVIL PRODUCTION ENGINEERING

Daniela Matschulat Ely[(a)], Ana Carolina Brandão Taroni[(b)],

[(a)] Professor Dr[a]. at Centro Federal de Educação Tecnológica de Minas Geriais, Civil Construction, Belo Horizonte, Minas
[(b)] Civil Product Engineer

(a) daniela.ely@gmail.com, [(b)]anacarolinataroni@gmail.com

ABSTRACT

Simulations are instruments that can help the teaching process. The present study develop a simulation as an academic teaching tool and create a method for teaching the production cell concept for students of Civil Engineering course. Besides, the model seeks to teach about productivity, quality and waste on construction sites. The methodology is based on: (a) theoretical review about the use of simulation as a teaching didactics for engineering students; (b) prototype development for the simulation; and (c) a pilot study. The simulation consists on implementing a reduced physical model, in two dimensions, of a house using color paper cutting and collage. Three rounds of simulation are implemented, where they are taught the concepts in a continuous and gradual way. The use of three rounds has the purpose of allowing the comparison of data, productivity, quality and waste on the house construction highlighting its improvement with the introduction of the production cell use.

Keywords: cellular manufacturing, simulation, teaching

1. INTRODUCTION

The classic approach on education, based on traditional methods of knowledge transfer has clear efficiency limitation. The use of experimental active form of teaching that stimulates students to think creatively and act correctly brings more value to their education. (PURNUS; BODEA, 2015).

The teaching of multifaceted concepts such as Lean philosophy for the students and employees that had never had any contact with them can be a difficult task. The challenge when teaching students is to create an environment so they can imagine and understand why the Lean philosophy is important and how it can work (DUKOVSKA-POPOVSK et al, 2008). According to Romanel e Freitas (2009), to a better understanding of the Lean Construction, it is necessary to implement new learning tools. Games and simulations, which are still innovation for this purpose, fill that gap. Therefore, a field for game development is noticed, allowing content transmission in a didactic and effective way.

A very used tool on Lean manufacturing is a production cell. Its main purpose is to keep the production the closest as possible to the continuous flow. On construction, these tools applications are still rudimentary and they need more studies (MARIZ; PICCHI, 2014).

Therefore, this article has the purpose of showing a simulation development headed for Civil Production Engineer students that provides the production cell concept directed to a practical way of teaching.

2. REVIEW

2.1. Simulation

Simulation and games are tools that support the traditional and formal teaching, being a useful pedagogical resource and that has an educational intrinsic value. These tools bring to the training field for the classroom, practice and experience that are essential to the professional activities improvement (ROMANEL; FREITAS, 2011).

A game can simulate situations close to the reality what leads a participant to look for solutions using concepts and theories presented in class. Implementing didactic games, visualizing processes became easier and can make learning more effective (DEPEXE, 2010).

2.2. Production Cell

According to Rodrigues e Picchi (2010) studies, the Lean Thinking has been gaining more and more space on the Brazilian civil construction sector, and many cases are observed. However most Lean applications at construction flow are restricted.

Rodrigues e Picchi (2010) performed an experiment survey about the Lean Thinking at building construction and identified that the lean tools used more frequently on construction sites are: line of balance, last planner, process standardization, value stream mapping, small batches production, Kanban is a production cell.

According to Patussi e Heineck (2006) motivated and well-trained teams, with a clear responsibilities definition tend to lead their work in a better way. One way to do that is through a cell production formation.

For Hyer e Brown (1999) there must be equipment utilization planning on the production cell, from the similar process parts and the creation of a continuous work. Therefore the tasks and people who perform them should be close in terms of time, space and information.

3. METHODOLOGY

Based on the review about simulation and lean construction tools, a gap was identified on the simulation use for production cell teaching practice.

After identifying the gap, the destination target public was defined, Civil Production Engineer students or Civil engineer with a semester teaching system.

The next step to the simulation developing was the magnitude definition, in other words, the number of students for a team, game of the low cost and easy access material. Besides that it was necessary previously define how many interactions could be made in one semester, develop the simulation object, material and equipment to be used, time and rules for each interaction.

Finally, before using the simulation for undergraduate students teaching practice, it was defined a validating need. The simulation was tested through a pilot study observing the presence of clear information, instrument adequacy, time sizing to perform and the developed object complementation need.

The use of a reduced physical model in two dimensions, without the use of computers, was a choice of the researchers so that the simulation could be applied in environments with few resources.

4. GENERAL GUIDELINES

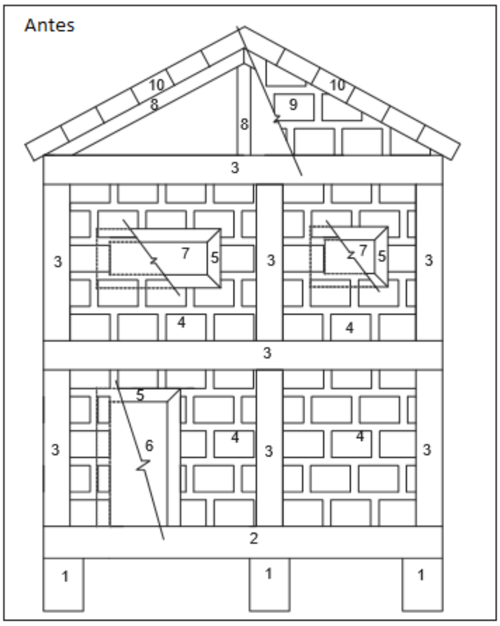
A simulation was developed with the purpose of teaching the Cell Production concept supported by the productivity concepts, quality and material waste.

The developed object consists in a two dimensions reduced physical model of a house, simulating a real execution. For this purpose color paper cutting and collage are used in a “building site”. The simulation includes ten activities execution, performed in a linear sequence where each interaction can be completed in a 100 minute class.

The house executive project (Figure 1 (a) and (b)) is a simple representation and its size is not proportional to the size of a real house. This is so that the parties that should be cut and glued can have ruler measurable dimensions and can be easily handled.

The simulation is consists in three iterations. In each iteration, the concepts are taught and applied in a progressive and continuous way, as described by Kaizen. This makes the student absorb the concepts better. In every iteration the same house is executed. By the end of each one, the performance is measured (execution time, quality and waste).

Design project



(a) Instruction project

Figure 1: The house executive project

The simulation is consists in three iterations. In each iteration, the concepts are taught and applied in a progressive and continuous way, as described by Kaizen. This makes the student absorb the concepts better. In every iteration the same house is executed. By the end of each one, the performance is measured (execution time, quality and waste).

The students who participate the simulation should have fulfilled a considerable part of the undergraduate subjects. Therefore, they already have technical knowledge about construction techniques and management. For the two first iterations the students are gathered in groups of four, three are responsible for the execution and one responsible for the execution time measurement and photographic record. On the third iteration, the students work forming a single team that work searching a good result in terms of productivity, waste and quality. At least eight students are necessary to perform the simulation so a competition can be established as a form of motivation.

On the first iteration the simulation rules are explained, but no concept is explained and each group should execute one house. The winning group is the fastest one to build the house. Thus, a construction site situation is simulated where the workers don't receive training and are under pressure to perform the tasks in short terms without a warning about the execution with quality and low waste.

On the second iteration, the students are taught about quality and waste. Now waste minimum levels are shown and quality achieved with a model house execution and executive project details are explained. Each group should execute one house. The winning group is the fastest to build the house with the lowest waste and the best quality.

On the first and second iterations the group organization is the same. Each group organizes the best way they think is the best for the house execution. They should follow the simulation rules, however there is no pre-established definition or training that establishes which function for each student in the group. The classroom organization on these iterations can be seen on Figure 2.

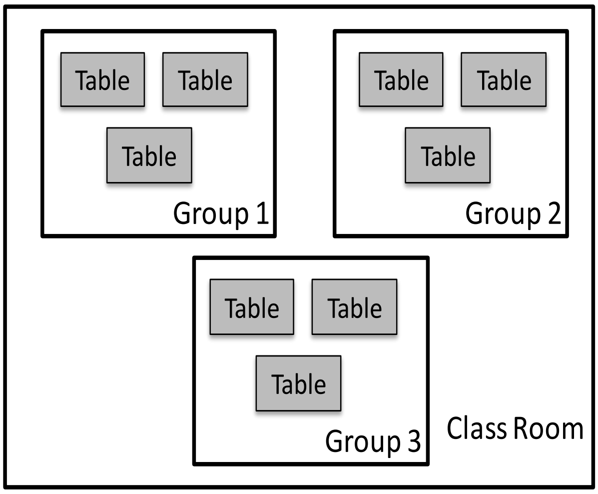


Figure 2: - Classroom organization on first and second iterations

On the third iteration, the students are taught about production cell. For the application of this concept the students are reorganized and trained to perform another function, thus favouring a continuous flow achievement (Figure 3). The number of houses executed is the same as the previous rounds so the results could be compared.

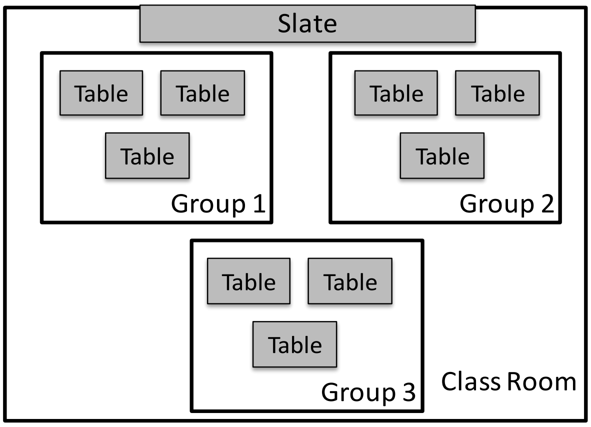


Figure 3: Classroom organization on third iteration

Besides that, the students begin to have function definition (pieces measurement on the executive project; marking the pieces dimensions on colored paper, cutting parts, gluing; marking the activities execution time and doing the photographic record of the activities execution) within the process and the tools designation that should be used. They are also instructed about other tasks so they have action knowledge to perform the activities and when they finish they are encouraged to collaborate with ongoing tasks and in this way they have a systemic view of the process and they work as a team.

The definition of the number of students that will perform each function seeks to allocate more students on the tasks that demand more time. In a classroom with 20 students that execute 5 houses, two students measure the colored paper parts, ten students mark the parts dimensions on the colored paper and cut pieces, three students glue the parts on the construction sites and five students record the houses execution time and take notes on the accomplishment of each activity. So the execution happen in an organized way, the students receive instructions on which house each one should work on first. The houses that should be worked on first are the closest.

The students are kept close to facilitate communication so it is possible to avoid misunderstandings and waiting time. Besides a visual communication tool is used to improve information. For this purpose, an executive project in a larger scale is fixed on the classroom board. The board is also used to take notes of information like for example, the parts dimensions, improving the form of communication. The desks are placed in front of the board allowing easy information view.

The students responsible for elements measurement are placed close to the board. They are the ones who write and explain the information on the board (they are the ones who write the explain information on the board). These students are instructed to inform the others the standards and parts repetition and making the marking and cutting tasks more effective. The standards are the paper colours of each activity and the hatching direction. After finishing their activities, these students help the others observing if the activities are being performed correctly and if the quality and waste standards are being followed.

In order to make the student pairs that mark and cut the pieces work together, some of the colored papers are cut in half. On the first and second iterations each house receives seven colored papers measuring 18 x 10 centimeters. On the third iteration each house receives seven colours of paper, two measuring 18 x 10 centimeters, two with 18 x 5 centimeters and four with 9 x 10 centimeters.

The number of students per task can be adapted based on the number of students participating the simulation. The important is that between rounds the house numbers are kept and that for each house executed, one student is responsible for recording and measuring the execution time.

In order to perform the iterations the students must observe the guidelines that follow:

- the house must be constructed according to the provided project
- the activities execution order should be in accordance with the project established;
- the activities should be performed in series (initiate a new activity only after the previous one has been finished, including the measuring services, dimension marking on colored paper, cutting and pasting);
- the students can only use the material provided to them
- follow each iteration specific rules, if that is the case.

For the house execution each group has at their disposal (figure 4):

- tables
- four chairs
- house executive project
- construction site
- colored paper
- two rulers
- two pencils
- an eraser
- scissor
- glue
- chronometer
- camera
- name tag to identify the student responsible for the execution time marking and photographic record of each activity
- container for depositing the remaining colored paper.

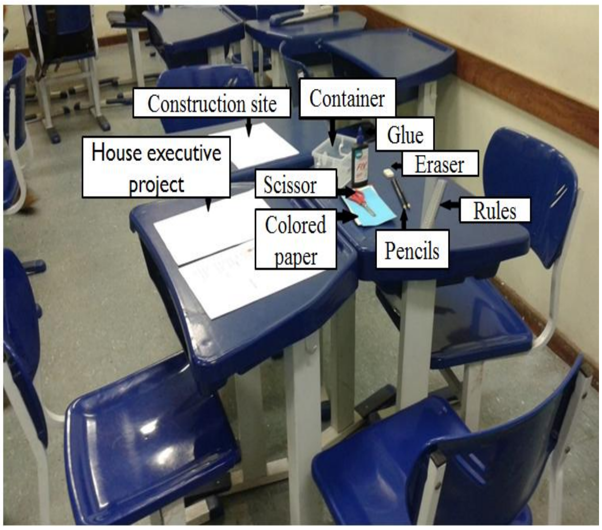


Figure 4: Location of the simulation

The execution sequence of each activity consists in four steps: measuring the project dimensions, mark the elements dimensions on colored paper, cut the elements and paste the elements. Each activity can be composed of more than one piece, with the same size or not. The activity steps can be executed in parallel.

The executive project is the graphic representation in two dimensions of a house cut. On the project the house elements are indicated, a legend indicating the activities execution order and the colored paper that should be used in each one. Also in the executive project there is a field for notes taken from the time spent on each activity. Besides that, the executive project serves as the house elements size source.

The construction site is fixed in one the tables simulating a real construction site situation, where workers move around a single fixed product. It consists in an A4 piece of paper with the house graphic representation with the same size as the executive project. The students should paste the colored paper on it.

The colored papers are provided in predetermined dimensions and on the same amount in order to execute each house. Each paper size should allow the evaluation of its use and its leftovers. Some papers have hatches so the material standardization care can be evaluated. Thus, it is observed if the respective colors are used on the correct activities and if the hatched papers are cut and pasted on the right direction.

The rulers are used to measure the piece's size from each house activity. This measurement should be performed on the executive project determined paper. The sizes measures should be marked on the colored paper with pencil and then cut with the scissors and glued on the construction site. The pencil should also be used to take notes of each activity time execution.

The camera is used to register the events along the iteration. Then, it is possible to observe the execution, group organization, material waste and the activities execution quality.

For each iteration, performance identification the execution time is computed, the quality is evaluated and the houses waste. This identification is performed after each iteration is finished. To evaluate the quality, the pictures taken during the iteration are compared to the model house (Figure 5). The waste is evaluated comparing colored paper from the executed house with the paper from the model house (Figure 6). The execution time is determined by the sum of the execution time from each activity, so each house has an execution time.

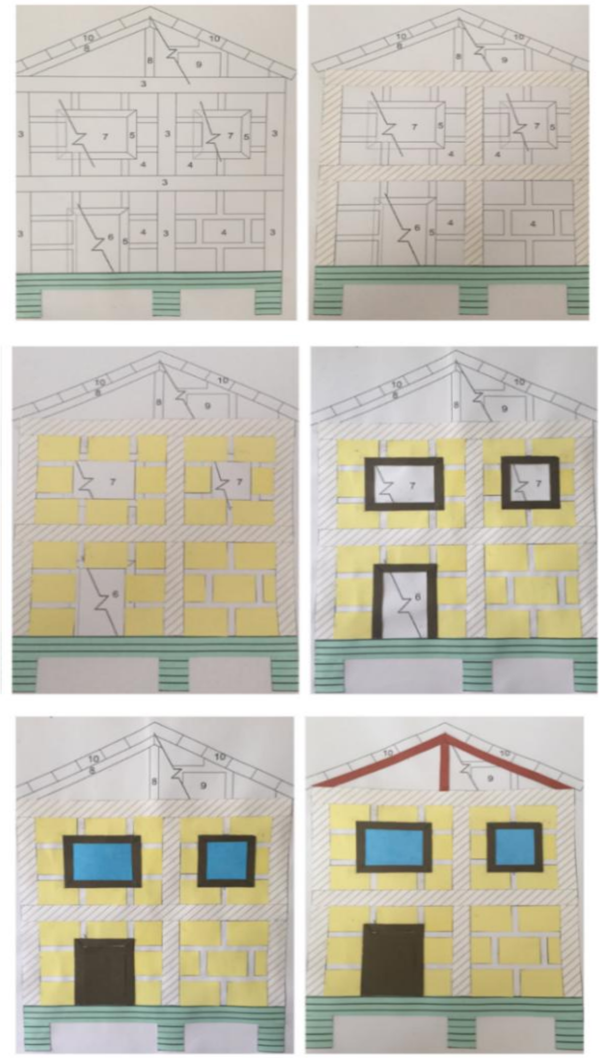


Figure 5: model house to evaluate the quality

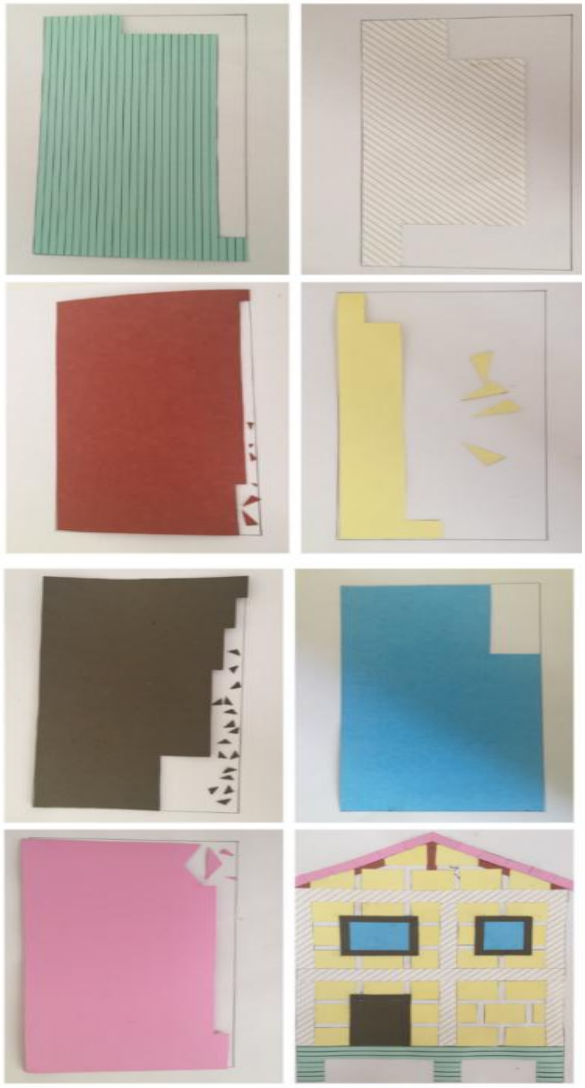


Figure 6: Minimum waste example and model house

4.1. Pilot Study

The pilot study consisted on the simulation application with a group of four Civil production engineer undergraduate students. On this study, the first iteration was applied along with a questionnaire. The questionnaire had the purpose of observing the clarity, gaps and doubts about the simulation.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The first iteration was chosen for the pilot study realization, because probably is the one that demands more time for its execution. This is because there is no training on the execution manner. With its implementation of the pilot study, it is possible to know approximately the longest time to execute a house and then adequate the simulation to its need.

At the pilot study beginning, the students received the material to perform the activities and were instructed about the simulation rules. One student was the inspector and the other three responsible for the house execution. After that, the students began the house construction according to what had been proposed.

With the application of the pilot study, it was possible to observe that the execution took longer than expected. The activity that took the most time to execute was the masonry of the 1st and 2nd floors, about 20 minutes. This occurred because it is the activity that has the highest number of measuring, cutting and gluing. To reduce execution time, the size of the blocks was changed from 1.50 x 1.0 centimeters to 2.3 x 1.7 centimeters, reducing the number of pieces (Figure 6).

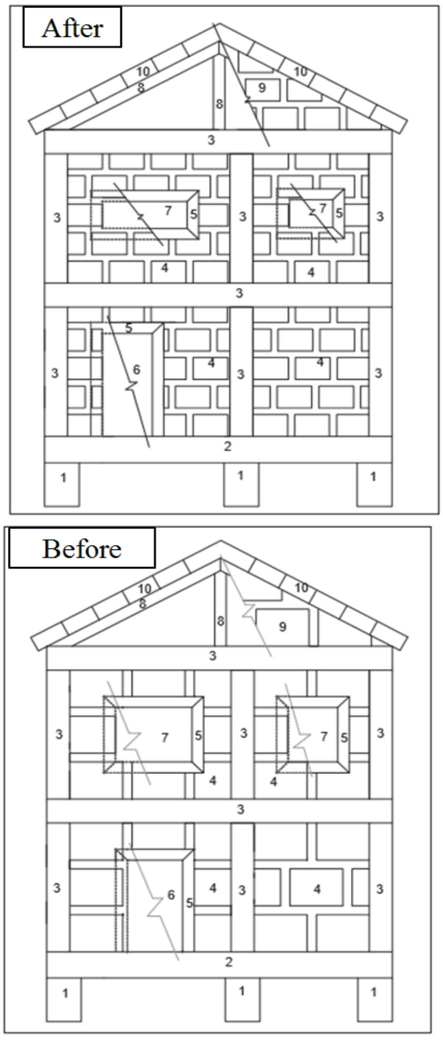


Figure 6: House before and after with adjustments after the Pilot Study

Another measure adopted to reduce the execution time of the simulation was the change in the dimensions of the windows. In the pilot study the windows had different heights and reduced sizes. During the simulation it can be seen that these characteristics made difficult its execution. For the final version of the game its dimension was increased and the heights equaled.

During the pilot study the students had difficulty understanding the cuts that indicated overlapping materials. In this way, for the final version of the game, the cut line was highlighted with a larger thickness and a dotted line was used as a projection form (Figure 7).

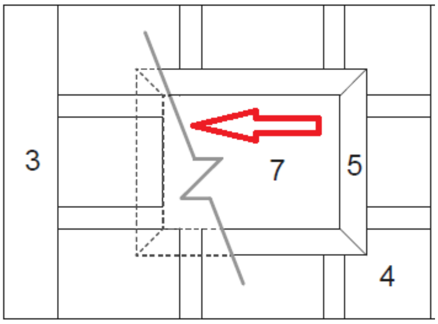


Figure 7: Cutting of doors and windows

Since there was a lot of colored paper left in the pilot study, it was decided to change the size of the material provided. They were 18.0 x 13.0 centimeters and went on to have 18.0 x 10.0 centimeters, reducing waste.

5. CONCLUSION

The use of simulation as a teaching practice brings students closer to reality. The developed game allows students to learn about production cell, quality and waste. In addition, with the use of simulation students have to be able to experience situations where they must make decisions in the search for a more efficient production, that is, they must apply project management knowledge.

In the creation of games, as a means to practice teaching, conducting a pilot study serves to perform tests on its functionality and clarity. After the observations, in conducting the pilot study of the game on production cell, small adjustments could be made and the game can be used teaching way.

The students who participated in the experiment answered the questionnaire. He revealed that 86% of students approve the simulation as teaching practice methodology and declare that they realized the link between staff training with quality and amount of waste.

For the development of this simulation the researchers chose not to use computational methods. Thus, the game can be used in places with few resources. The game can be adapted for teaching people who do not attend universities and can be adapted to run on a computer.

REFERENCES

DEPEXE, M. D. Simulação com Jogos de montar: um instrumento de ensino para o planejamento e programação de obras. Engevista, v.12, n.2, p.108-116, 2010.

DUKOVSKA-POPOVSKA, I.; HOVE-MADSEN V.; NIELSEN K. B. Teaching lean thinking through game: some challenges. Aalborg University, Denmark, 2008.

HYER, N. L.; BROWN, K. A. The discipline of real cells. Journal of Operations Management v.17,1999, p. 557- 574.

MARIZ, R. N.; PICCHI, F. A. Aplicação de célula de produção no serviço de execução de fachada: Um estudo de caso na Construção Civil. Revista Produção Online, Florianópolis, Santa Catarina, v.14, n. 2, p. 703- 719, 2014

PATTUSSI, F. A.; HEINNECK, L. F. A utilização de conceitos da Produção Enxuta na constituição de Células de Produção em obras de pequeno porte. Desempenho de Sistemas Construtivos, Unochapecó, 2006.

PURNUS, A.; BODEA, C. N. Educational simulation in construction project finalcial risks management. Creative Construction Conference, 2015.

RODRIGUES, M. R.; PICCHI, F. A. Análise de experiências de aplicações do Lean Thinking na construção de edificações no Brasil. XIII Encontro Nacional de Tecnologia do Ambiente Construído, Canela, Rio Grande do Sul, 2010.

ROMANEL, F. B.; FREITAS, M. do C. D. Jogo “Desafiando a Produção”: ensinando a construção enxuta na construção civil. 2011.

EVIDENCE-BASED EVALUATION OF PSYCHOSOCIAL RISK FACTORS AND THE INTERACTION OF THEIR STRESSORS USING SYSTEM DYNAMICS

Hossein Abaeian[(a)], Mohamed Al-Hussein[(b)], Osama Moselhi[(a)]

[(a)] Department of Building, Civil and Environmental Engineering, Concordia University, Canada
[(b)] Department of Civil and Environmental Engineering, University of Alberta, Canada

E-mail: abaeian@ualberta.ca, malhussein@ualberta.ca, moselhi@encs.concordia.ca

ABSTRACT
The experience of work-related stress caused by exposure to psychosocial risks can lead to poor physical and mental health, loss of productivity and ultimately results in worker injury. Considering that the impact of psychosocial risks may vary on an individual basis and the nature of the associated stressors varies widely among workplaces, this paper examines the latest research trend in this discipline by analyzing published psychosocial research in several journals in terms of contributions of institutions, adopted data collection and processing methods, and research interest. In addition, a database resulting from an evidence-based assessment of key psychosocial risk factors is provided and a conceptual System Dynamics (SD) model is created accordingly to illustrate the interaction between identified variables. The developed causal loop diagrams can better delineate the mechanism underlying the causal relationship between a given psychosocial factor and its adverse effects. The outputs are expected to assist project managers in gaining an in-depth understanding of the complexity and interaction of psychosocial factors and their outcomes.

Keywords: work-related musculoskeletal disorders, psychosocial risk factors, ergonomics, system dynamics.

1. INTRODUCTION

Workplace psychosocial factors are non-physical aspects of a workplace developed by the policies, culture, expectations, and social attitudes of the organization (CCOHS, 2012). Workers' behaviors are not random and occur due to external factors. In fact, psychosocial factors can be defined as factors associated with the way individuals react to the work environment and their job demands (Visagie, Swanepoel, and Ukpere, 2014). To implement effective, efficient and sustainable ergonomic interventions, it is important to understand how the way in which work is organized influences workers' biomechanical exposures, psychosocial stresses, and the relationships between these factors (Bao et al., 2015).

Previous studies identify a significant association between psychosocial work stressors and workers' work-related musculoskeletal disorder (WRMSD) symptoms (Eatough, Way and Chang 2012; Visagie, Swanepoel and Ukpere 2014; Thiese, et al. 2015; Bao, et al., 2015). For instance, factors such as job roles and responsibilities, control over work, and safety climate at work are shown to be related to increased risk of WRMSDs (Eatough, Way, and Chang, 2012). Similarly, factors such as job satisfaction, co-workers' support, physical and mental exhaustion, general health, and anxiety are shown to have a significant impact on the risk for the development of WRMSDs (Thiese et al., 2015; Widanarko et al., 2015). Other factors, including job insecurity and long and irregular working hours, may increase stress and consequently increase psychosocial risk.

Although occupational health and safety management systems address both health and safety in the workplace, several researchers argue that they focus mostly on safety rather than on workers' health. Also, there is a lack of understanding regarding the process mechanisms, such as how psychological stressors interact with one another, leading to adverse effects on employees' health. Therefore, this paper presents a complete integrated assessment for various working conditions from the perspective of reducing the risk of psychosocial factors and provides a list of psychosocial risk factors. First, a systematic literature review is conducted to ensure a comprehensive search of various databases to build an evidence-based assessment in the area of psychosocial risk factors. Then, a conceptual causal loop diagram is created to illustrate the interaction between psychosocial stressors based on research of modeling stresses and the review of various meta-models for stress.

System thinking, which is a compelling solution for many real-world problems, refers to the paradigm in which the world is seen as a complex system, in which everything is connected to everything else (Sterman, 2001). The deployment of system dynamics in this study allows for recognizing the problem of psychosocial risks in the workplace and describing its underlying mechanism through causal loop diagrams also known as the influence diagrams. The present study aims to advance the knowledge of the mechanisms through which job stressors influence safety outcomes, which can ultimately enhance the development of stress management and safety-related interventions. The findings are expected to assist project managers and job designers in the following ways: (i) estimate the impact of mental disorders on employees' health in order to set priorities and plan appropriate interventions in the workplace; (ii) explore the mindset which could trigger unwanted behavior within a psychosocial work environment; and (iii) investigate the mechanisms underlying the associations between psychological risk factors and their adverse outcomes through the developed causal links.

2. A REVIEW OF THE RESEARCH ON PSYCHOSOCIAL RISKS

In the interest of addressing psychosocial risks in the workplace, an impressive number of studies have been published by internationally renowned journals related to health and safety management in the construction industry. However, the absence of a holistic summary of the research developments in the discipline of psychosocial risk management is apparent. Therefore, a literature review is presented based on searches using electronic search engines and different online databases (e.g., Google Scholar, Science Direct, Web of Knowledge/Science, Concordia University's library catalogue) for relevant published articles that include the following key words: work-related stress, psychosocial risks, psychosocial risks assessment/management, musculoskeletal disorders, work stress/organization, and psychosocial risk mitigation/interventions. The articles from the search results are scanned to filter and retrieve related papers.

Accordingly, psychosocial-related literature can be categorized into the following areas: guidelines and standards, risk auditing approaches, single item-based studies, and meta-stress analysis researches. In the following paragraphs, different research traditions and study areas of psychosocial factors are reviewed. Figure 1 presents the academic journals through which the reviewed articles were published.

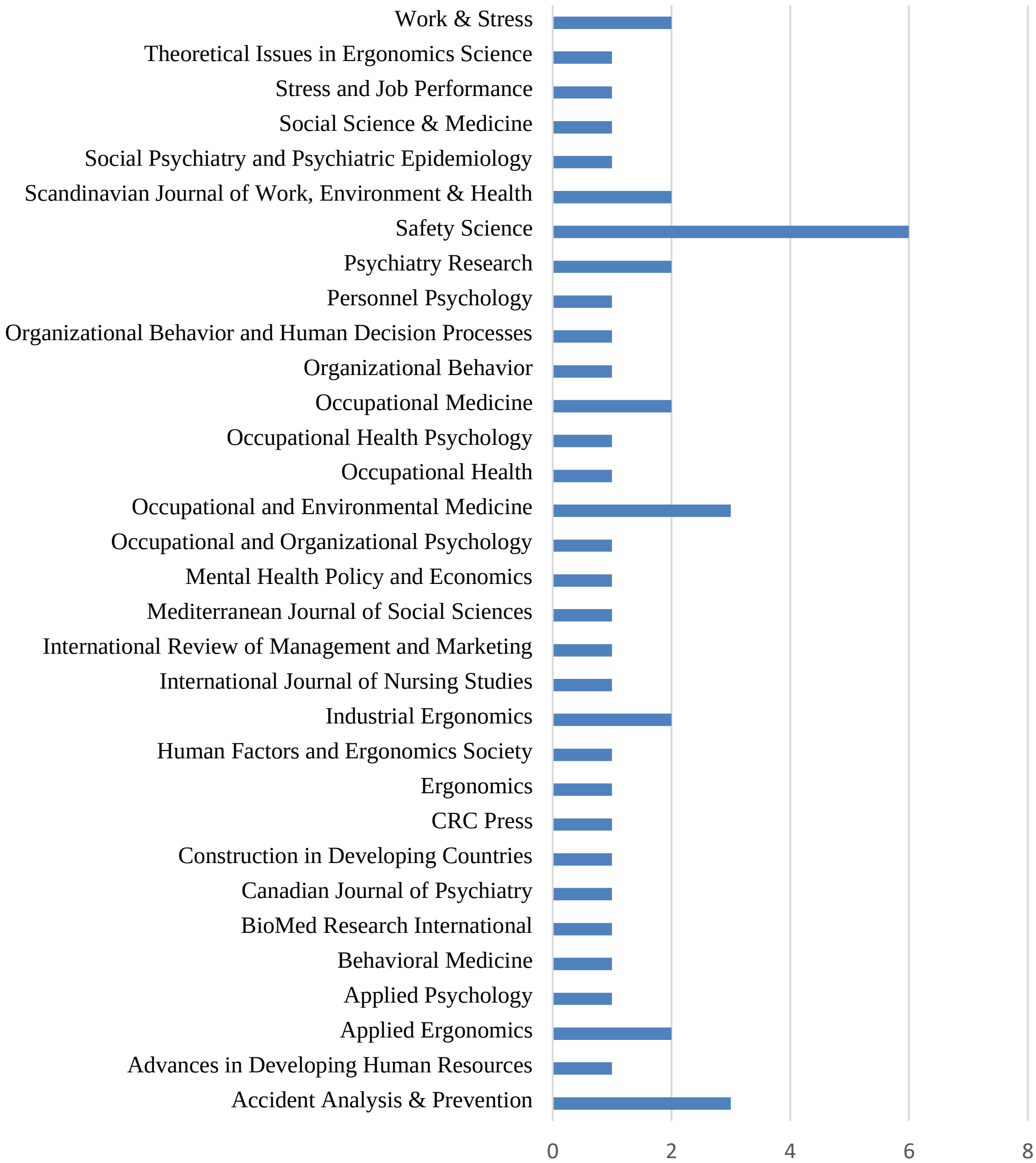


Figure 1. Psychosocial related articles reviewed in this study and their associated journals

To gain insight on the primary research stream in this domain through the reviewed articles, the contribution of each institution is quantitatively assessed using the score developed by (Howard, Cole, and Maxwell, 1987) by means of the following formula:

$$Score = \frac{1.5^{n-i}}{\sum 1.5^{n-i}} \qquad (1)$$

where each paper is assumed to have a score of one point, n is the total number of authors of the article and i is the ordinal position of the author of the article. For example, in a paper with two authors, the institution associated with the first author is given a score of 0.60 and the second author's institution is given 0.40. Table 1 presents the contribution score of each institution based on the reviewed articles.

In addition to the reviewed research on psychosocial risks, a number of guidelines and initiatives have been developed focusing on the management of the psychosocial work environment, including: the Canadian Standard for Psychological Health and Safety in the Workplace (National Standard of Canada, 2013), which is a suitable standard and provides guidelines to prevent psychological harm of workers across all sectors; World Health Organization (WHO, 2010), a global framework that combines evidence-based approaches and principles of health protection; The European framework for psychosocial risk management (PRIMA-EF) (Leka and Cox, 2008), and PAS1010, the British Standards Institution (BSI) guidelines that establishes the benchmark for the evaluation of psychosocial risks (BSI, 2011). Despite several occupational health and safety management systems, it can be argued that their primary focus is on workers' safety rather than workers' health (Bergh et al. 2016; Hasle, 2011).

Psychosocial risk auditing approaches (Bergh, Hinna, Leka, and Zwetsloot, 2016), and the establishment of the psychosocial exposure indicator (Bergh et al., 2014) are effective tools to confirm compliance to requirements in the management system and to monitor the status of influential psychosocial risks in the workplace based on

Table 1. Institutions and their contribution scores based on the reviewed articles

Institution	***Score***
Albert Einstein College	0.32
Asian Institute of Technology, Pathumthani	1.00
Babson College	1.00
California Department of Health Services	0.12
Central Michigan University	0.68
Columbia University, USA	0.80
Finnish Institute of Occupational Health	0.40
Global Established Pharma Medical Affairs, Japan	1.00
Hebei Medical University, China	0.40
Hiroshima University	1.00
Indiana University	1.00
Institute of Psychiatry, London	0.40
Jichi Medical School	0.12
Keele University, UK	1.00
Keio University	0.18
Leicester University	0.32
Leiden University Medical Center	0.12
London School of Economics and Political Science	0.60
Louisiana State University	1.00
Massachusetts Institute of Technology	1.00
Massey University, New Zealand	0.47
Michigan State University	0.40
National Institute for Occupational Safety and Health	0.21
National Institute of Industrial Health, Japan	0.60
National Institute of Occupational Safety and Health, Japan	0.42
Netherlands Institute of Mental Health and Addiction	0.28
North-West University, South Africa	0.79
Queensland University of Technology, Australia	0.60
Saint Mary's University, Canada	1.00
Seoul National University	0.60
Shenzhen University	0.40
Statoil ASA, Stavanger, Norway	1.20
Swinburne University of Technology, Australia	0.60
The National Research Centre for the Working Environment, Denmark	1.47
TNO Institute of Preventive Health Care	1.00

Institution	***Score***
University Teknologi Petronas	1.00
University College London	0.21
University of Adelaide	0.28
University of Alberta	0.28
University of California	1.88
University of Cincinnati	1.47
University of Florida	1.40
University of Gävle	0.60
University of Hawaii at Manoa	0.60
University of Illinois at Urbana-Champaign	0.42
University of Indonesia	0.32
University of Johannesburg	0.21
University of Lleida	1.00
University of London	0.21
University of Michigan at Ann Arbor	0.18
University of Milan	0.40
University of New South Wales	0.80
University of North Carolina	1.00
University of Nottingham	0.80
University of Seoul	0.40
University of South Australia	0.60
University of South Florida	0.60
University of the Netherlands Antilles	0.18
University of Tokyo	0.28
University of Umeå	0.40
University of Utah	0.81
University of Wisconsin Oshkosh	0.32
University of Wisconsin-Madison	1.00
University of Wisconsin-Milwaukee	0.60
University of Wuppertal, Germany	0.60
Utrecht University	0.12
Vincent's hospital	0.60
Virginia Commonwealth University	0.32
VU University, Netherland	0.42
Washington State Department of Labor and Industries	0.99

Organizational requirements. However, it can also be argued that they are costly and do not necessarily cover the complexity of the work environment or the interaction between psychosocial factors (Hohnen and Hasle, 2011). Furthermore, their findings cannot be directly adapted to other company settings.

Various cross-sectional studies examine the bivariate relationships between psychosocial work stressors and WRMSDs symptoms (Bongers et al. 1993; Nielsen et al. 2008; Kelloway, Mullen and Francis 2006; Wang et al. 2007; Silverstein et al. 2010), as well as the bivariate relationship between biomechanical exposure and psychosocial outcomes (Thiese, et al., 2015). According to (Thiese et al., 2015), as physical exposure (duration, repetition, and force) increases, psychosocial responses worsen. Also increased repetition found to be associated with depression and forceful exertions are shown to be related to physical exhaustion. Factors such as perception of safety climate, and workers' ability to control the pace of the work are found to be correlated with lost work days due to injury (Abbe et al., 2011). Safety-specific leadership behavior, such as emphasizing the value of safe performance and rewarding safety-related compliance, contributes positively to reduction of occupational injuries (Kelloway, Mullen, and Francis, 2006) and have a relationship with mental exhaustion (Nielsen et al., 2008). Also, high job demands, one's perception of the level of workload (Goldenhar, Williams, and Swanson 2003) and job dissatisfaction are found to be significantly associated with arm/hand and lower back disorders (Silverstein et al. 2010; Widanarko et al. 2015). Goldenhar et al. (2003) identified the following stressors to be directly related to injury: job demands, job control, job certainty, job training, safety climate at work, skill under-utilization, responsibility for the safety of others, exposure hours, and job tenure. Also, according to Bongers et al. (1993), low job control and lack of social support by colleagues are positively associated with musculoskeletal disorders (MSDs). One of the limitations of the bivariate studies in addressing the relationship between psychosocial risks and MSDs is that they often include a small number of stressors, making it a challenge to compare the impact of different stressors on outcomes. Results from a study of a sample of workers exposed to different levels of physical and psychosocial risks (Widanarko et al. 2015) indicate that exposure to high physical/high psychosocial levels has more associations with MSDs and presenteeism, and exposure to low physical / high psychosocial levels are more closely related to lower back symptoms than those in the high physical and low psychosocial group. Despite their valuable findings, these studies are based on limited factors and provide conflicting evidence. Additionally, the studies are unable to evaluate the relative impacts of psychosocial stressors on MSDs as the effects of psychosocial stressors and physical exposure on MSD symptoms are often lumped together (Thiese et al., 2015; Widanarko et al., 2015; Smith et al. 2004) and psychosocial factors are treated as potential confounders without evaluation of their independent influence on MSDs. The interaction of physical and psychosocial factors increases the probability of MSDs greater than the sum of the magnitude of the individual effects (Widanarko et al., 2015). All these make it difficult to understand the mechanisms underlying the associations between the stressors and their adverse outcomes.

Exposure to psychosocial risk factors can also affect worker performance, and, if prolonged, may result in serious health problems (Bergh et al., 2016), which are found to have a significant association with presenteeism (Tsuchiya et al. 2012; Plaisier et al. 2012) rather than absenteeism. Asami et al. (2015) establish a direct relationship between strain symptoms and productivity loss, even among workers with undiagnosed depression, implying that workers suffering from stress may not take days off but remain at work, leading to impaired work performance. In Canada, costs associated with presenteeism are shown to be 2.7 times higher than those associated with absenteeism (Evans-Lacko and Knapp 2016). Loss of productivity caused by mental stress is approximately 2.77 to 4.17 days per month for Western workers (Lim, Sanderson, and Andrew 2000), and 28-30 days per year for Japanese (Tsuchiya et al., 2012). Risk factors such as high physiological demand, low decision latitude and job control, and low social support reportedly reduce activity and result in poor performance (Plaisier et al. 2012; Aasa et al., 2005).

According to the literature review, psychosocial data collection methods usually include electronic questionnaires, face-to-face interviews, and survey techniques, in which questions are designed based on different standard questionnaires such as NIOSH Generic Job Stress Questionnaire (Abbe et al. 2011; Goldenhar et al., 2003;Thiese et al., 2015), Job Content Questionnaire (Widanarko et al. 2015; Thiese et al., 2015; Blanch 2016), ERI Questionnaire (Widanarko et al., 2015), HSE Safety Climate Survey Tool (Visagie et al., 2014), NIOSH Management Commitment to Safety Scale (Abbe et al., 2011; Goldenhar et al., 2003), North-Western National Life Insurance Company Survey (Abbe et al., 2011), and the General Well-Being Questionnaire (GWBQ) (Bergh et al., 2014).

Future research directions can be derived based on what has already been carried out to date and what remains to be carried out in the domain of psychosocial risks management. Based on the reviewed sources, suggestions for future studies include: more longitudinal studies toward the analysis of interaction of physiological and psychological factors, and a definite conclusion regarding causal relationships between stressors, strains, and MSD symptoms; exploration of the impact of psychosocial risks exposure on workers performance; and investigation of factors, other than physical exposure, that influence the psychological state of workers. Also, a continuation of research is expected on the role of psychosocial factors as independent contributors to injuries and MSDs.

3. EVIDENCE-BASED LIST OF PRIMARY PSYCHOSOCIAL RISK FACTORS

As a result of the review of the literature retrieved from academic publications, reports, and standards, a total of 19 factors, reported as influential in the emergence and development of psychosocial risks, are identified and summarized in Figure 2. Psychosocial risks present an ongoing challenge related to occupational safety and health (Bergh et al., 2016). While there are various psychosocial stressors found in the literature, various categories are selected by researchers to classify those stressors for further assessments, such as "demands and control" and "social support" (Bongers et al., 1993), "work role", "job control", and "social characteristics and safety leadership" (Eatough et al., 2012; Rosen et al. 2010), or "job demands", "role and responsibilities", "job control", and "social support" (Bergh et al., 2016). In the present study, categories are selected as they represent relatively stable characteristics in the work environment, and, consequently, workers may have prolonged exposure to these psychosocial work stressors. Thus, the identified factors in Figure 2 are categorized into job control, job demand, social support, safety climate, and individual characteristics.

Figure 2. Influential factors in the development of psychosocial strain symptoms, identified in previous literature

4. SD MODEL DEVELOPMENT

System dynamics is a powerful graphical illustration tool (through its causal mapping), and is effective in representing complex interactions between factors and making the problem more readily understood (Wang, Zou, and Li 2016). It has been utilized for improving construction safety and understanding workers' behavior and safety attitudes and has been proven to provide better and less subjective understanding of the accident and injury mechanism (Wang et al., 2016; Shin et al. 2014; Leveson 2004; Abaeian et al. 2016). Accordingly, introducing the interactions underlying psychosocial risks into a causal feedback model would help to reveal the pathways by which psychosocial risks lead to adverse effects in workers. Thus, this section aims to develop an SD model for better understanding and analyzing the complex feedback mechanism surrounding workers' mental processes due to exposure to psychosocial risks in the workplace.

The deployment of SD allows a four-stage approach: (1) recognizing the problem and breaking it down into smaller systems without violating the holistic concept of SD; (2) describing the system by means of causal loop diagrams also known as influence diagrams; (3) qualitative analysis that involves closely analyzing the causal loops; (4) construction of a simulation model that encompasses the specification of the structure estimation of variables, and behavioral relationships.

4.1 Causal Loop Diagrams

Based on a review of previous research findings and the identified psychosocial factors presented in Figure 2 in addition to in-depth discussion with construction managers and wellness specialists, an SD model is developed using Vensim software, containing the feedback structure of psychosocial risks and their impacts on workers' behavior as well as the pathways by which they lead to MSDs and impaired work performance, as shown in Figure 3.

As represented in Figure. 3a social support is a consistent mediator in the association of job control with job strain and thus is positively related to job control (Goldenhar et al., 2003; Blanch 2016). According to Blanch (2016) 70% of the variance in job strain is due to the mediation effect of job control through social support in addition to the large direct effect of job demand. While the safety climate at work is positively related to a perceived level of social support (Goldenhar et al., 2003), an increased level of social support, and, specifically more effective contact with supervisor and more effective supervision can further result in a better perception of safety climate (Han et al. 2014).

Strain results from a mismatch of psychosocial risks placed upon workers (including job demand, level of control and social support, perception of safety climate and effort-reward imbalance) and their psychological capabilities to cope with them. However, a combination of high demands and a high level of control can lead workers to experience a heightened sense of "personal accomplishment & feeling of competence" (Dollard et al., 2000). The outcome of strain may also interact with the environment and respond by impacting the level of work-related stress, leading to exaggerated perceived symptoms (Goldenhar et al., 2003; Sauter & Swanson, 1996). According to Thiese, et al. (2015) individual factors can affect the psychosocial measures either through the perceived level of strain or capabilities to cope with strain. As it takes some time for workers to recognize the risk and change their subjective risk based on their psychological hardiness, "time perceiving coefficient" is represented in the model to reflect the time it would take for the individual to perceive the risk.

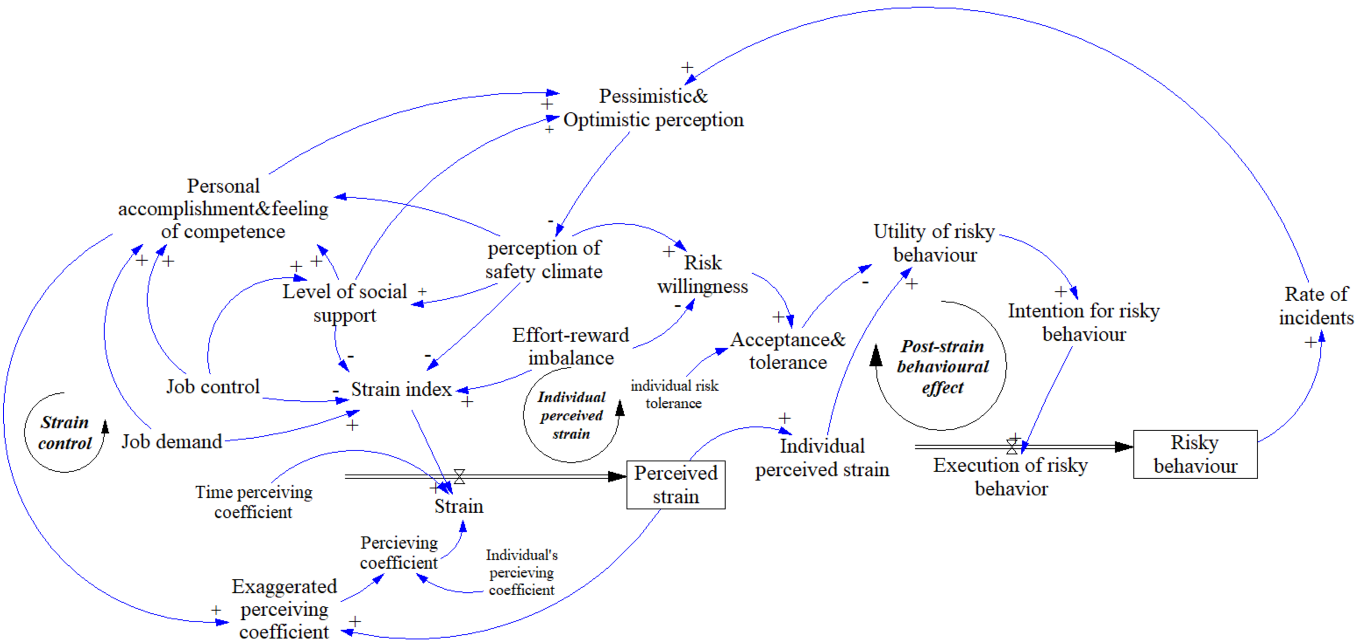


Figure 3a. SD model development

Proceedings of the European Modeling and Simulation Symposium, 2017
ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

This indicates that the better the individual hardiness, the longer it would take for the individual to perceive the strain. Also, previous studies suggest the exploration of how psychological dimensions of hardiness and conscientiousness, may moderate the relationship between psychosocial work -stressors and strain. Thus, "individual's perceiving coefficient" variable is represented in the model to explore this mediation and evaluate various perceived strains among different individuals based on workers' psychological capacity by buffering their stressful experiences. This implies that the better an individual's psychosocial hardiness, the lower their perceiving coefficient would be. Individual risk tolerance plays a key role during objective risk assessment and thus is represented in the model as a function of individual tolerance and risk willingness (whether the worker wants to take the risk). In addition, proactive effort-reward measures and better perception of safety climate can affect the risk willingness. Strain plays a mediating role in linking job stressors and unsafe behaviors, including execution of risky behavior which includes any activity or behavior that deviates from normal accepted safety procedure. Execution of risky behavior increases the likelihood of accidents, leading to lower perception of safety climate in the workplace. This is represented in the model, through "pessimistic & optimistic variable". When workers engage in risky behavior due to prolonged exposure to poor working conditions, the workers may establish a new habit and act less safely in performing the task, as modeled in Figure 3b.

The signs and symptoms of strain tend to progress through different phases. Fig 3b illustrates further stages in which bio-physiological reactions begin to develop, leading to an increase in the likelihood of fatigue and compounded level of job demand that workers are under when performing the task. Fatigue itself can lead to worker deterioration of cumulative experience on a site (Love and Edwards 2004) and can increase the progression of warning signs, leading to a greater chance of fatigue and MSDs.

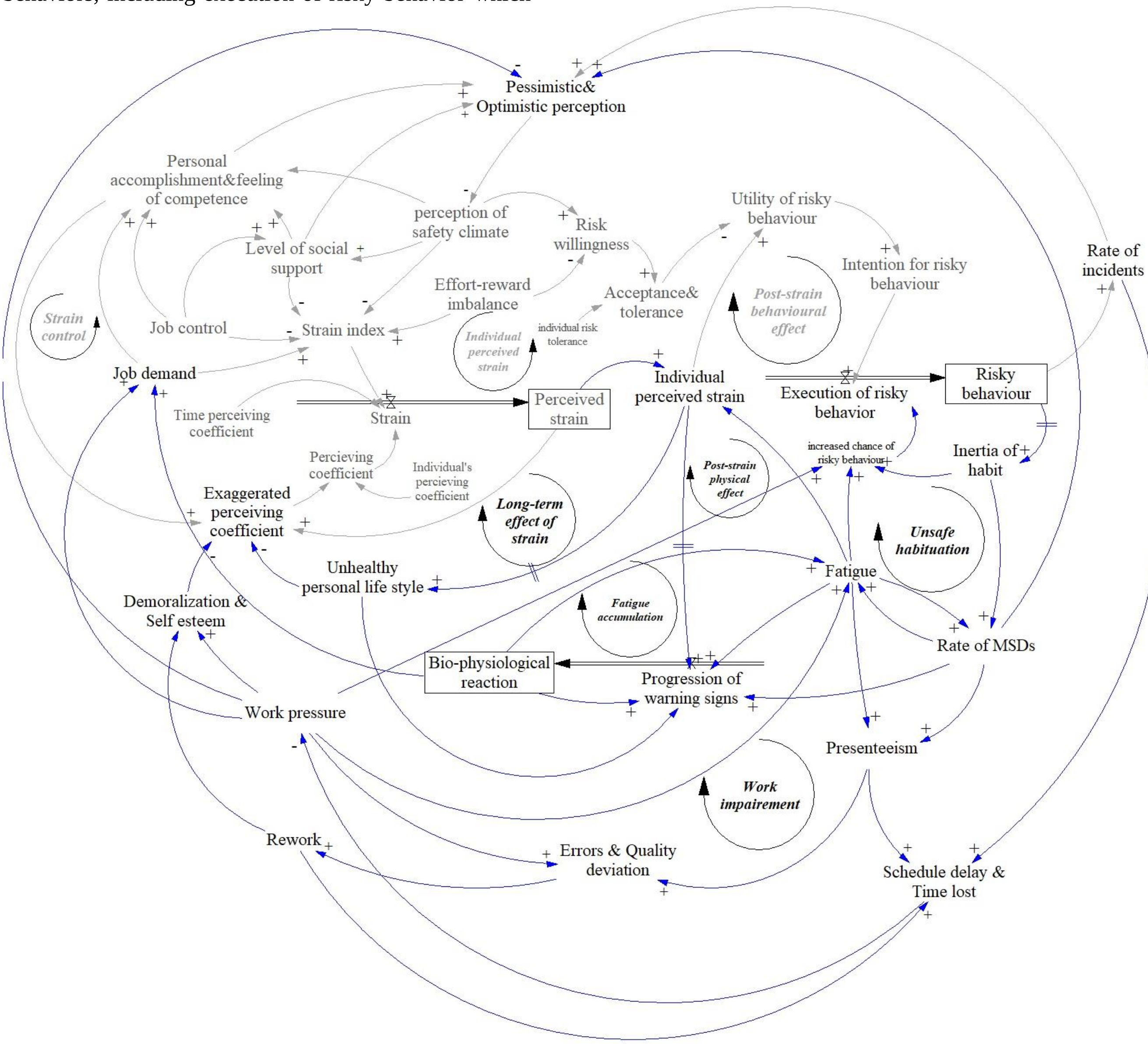


Figure 3b. SD model development

Psychosocial risk factors and their strain outcomes are shown to be more largely associated with presenteeism than absenteeism (Tsuchiya et al., 2012; Asami et al., 2015), thus the developed model focuses on the impaired work performance due to exposure to psychosocial risks in the workplace. According to CCOHS (2012), when presenteeism occurs, workers come to work, but are not fully functioning or mentally present due to stress or sickness, which in turn results in a greater chance of making errors, leading to rework and schedule delays, and consequently work pressure. Perceiving production pressure implies excessive workload, and higher levels of job demand, and also influences workers' perception of safety at work.
Eventually, if the early behavioral warning signs and physiological symptoms are ignored, workers' psychological well-being and their personal life style could be affected (change in sleeping routine, smoking habits, loss of sex drive, and alcohol abuse).

5. DISCCUSION & CONCLUSION

Psychosocial work stressors may have complex effects on strain and workers' health beyond simple bivariate relationships (Eatough et al., 2012; Bao et al., 2015). To gain an in-depth understanding of the research on psychosocial risk management, this study presents a systematic review of related articles and provides a list of psychosocial risk factors. Also, based on the reviewed articles, journal publications and the contributions of institutions on psychosocial research are analyzed using different methods of data collection, and the research trend as well as different categories for the discipline of psychosocial research are reviewed.

The literature review indicates that despite different stress process-based models, specific conclusions cannot be drawn in relation to the complexity of interaction and the mechanisms underlying the relationships between key psychosocial variables. Thus, an SD model is developed to illustrate the interaction between identified stressors and their adverse effects. The proposed SD model provides a useful tool for hypothesizing the structure that underpins the mechanism of psychosocial risk. This allows decision makers to evaluate the complex feedback process due to exposure to psychosocial risk factors. According to Coyle (1999), qualitative SD describes a system and the description "is in itself a useful thing to do which might lead to better understanding of the problem in question". As an initial effort to evaluate the impacts of psychosocial risks as well as their underlying interactions from a holistic view, this study pursues behavior prediction rather than point prediction. Future research will involve actual data collection to quantitatively assess the workers' mental processes due to exposure to psychosocial risks.

6. REFERENCES

Aasa, U., Barnekow, M., Angquist, K., & Brulin, C. (2005). Relationship between work-related factors and disorders in the neck-shoulder and low-back region among female and male ambulance personnel. *Journal of Occupational Health*, 481-489.

Abaeian, H., Inyang, N., Moselhi, O., & Al-Hussein, M. (2016). Ergonomic Assessment of Residential Construction Tasks Using System Dynamics. *33rd International Symposium on Automation and Robotics in Construction (ISARC).* Auburn, Alabama.

Abbe, O., Harvey, C., Ikuma, L., & Aghazadeh, F. (2011). Modeling the relationship between occupational stressors, psychosocial/physical symptoms and injuries in the construction industry. *International Journal of Industrial Ergonomics*, 106-117.

Asami, Y., Goren, A., & Okumura, Y. (2015). Work Productivity Loss With Depression, Diagnosed and Undiagnosed, Among Workers in an Internet-Based Survey Conducted in Japan. *Journal of Occupational and Environmental Medicine*, 105-110.

Bao, S., Kapellusch, J., Merryweather, A., Thiese, M., Garg, A., Hegmann, K., & Silverstein, B. (2015). Relationships between job organisational factors, biomechanical and psychosocial exposures. *Ergonomics*, 1-17.

Bergh, L., Hinna, S., Leka, S., & Jain, A. (2014). Developing a performance indicator for psychosocial risk in the oil and gas industry. *Safety Science*, 98-106.

Bergh, L., Hinna, S., Leka, S., & Zwetsloot, G. (2016). Developing and testing an internal audit tool of the psychosocial work environment in the oil and gas industry. *Safety Science*, 232-241.

Blanch, A. (2016). Social support as a mediator between job control and psychological strain. *Social Science & Medicine*, 148-155.

Bongers, P., de Winter, C., Kompier, M., & Hildebrandt, V. (1993). Psychosocial factors at work and musculoskeletal disease. *Scandinavian Journal of Work, Environment & Health*, 297-312.

BSI, British Standards Institution. (2011). *PAS1010;Guidance on the Managment of Psychosocial Risks in the Workplace* . London: British Standards Institution.

CCOHS. (2012). *Candaian Centre for Occupational Health and Safety.* Retrieved from http://www.ccohs.ca/oshanswers/psychosocial/musculoskeletal.html

Coyle, G. (1999). Qualitative modelling in system dynamics or what are the wise limits of quantification. *17th International Conference of the System Dynamics Society.* Wellington.

Dollard, M., Winefield, H., Winefield, A., & de Jonge, J. (2000). Psychosocial job strain and productivity in human service workers: A test of the demand-control-support model. *Journal of Occupational and Organizational Psychology*, 501-510.

Eatough, E., Way, J., & Chang, C.-H. (2012). Understanding the link between psychosocial work stressors and work-related

musculoskeletal complaints. *Applied Ergonomics*, 554-563.
Evans-Lacko, S., & Knapp, M. (2016). Global patterns of workplace productivity for people with depression: absenteeism and presenteesim costs across eight diverese countries. *Social Psychiatry and Psychiatric Epidemiology*, 1525-1537.
Goldenhar, L., Williams, L., & Swanson, N. (2003). Modelling relationships between job stressors and injury and near-miss outcomes for construction labourers. *Work & Stress*, 218-240.
Han, S., Saba, F., Lee, S., Mohamed, Y., & Pena-Mora, F. (2014). Toward an understanding of the impact of production pressureon safety performance in construction operations. *Accident Analysis and Prevention*, 106-116.
Hasle, P. (2011). Editorial: Occupational Health and Safety Management Systems: Issues and challenges. *Safety Science*, 961-963.
Hohnen, P., & Hasle, P. (2011). Making work environment auditable-A critical case study of certified occupational health and safety management systems in Denmark. *Safety Science*, 1022-1029.
Howard, G., Cole, D., & Maxwell, S. (1987). Research productivity in psychology based on publication in the journals of the American Psychological Association. *American Psychologist*, *42*, 975.
Kelloway, E., Mullen, J., & Francis, L. (2006). Divergent Effects of Transformational and Passive Leadership on Employee Safety. *Journal of Occupational Health Psychology*, 76-86.
Leka, S., & Cox, T. (2008). *PRIMA-EF; Guidance on the European Framework for Psychosocial Risk Management.* World Health Organization, PRIMA-EF Consortium.
Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, 237-270.
Lim, D., Sanderson, K., & Andrew, G. (2000). Lost productivity among full-time workers with mental disorders. *Journal of Mental Health Policy and Economics*, 139-146.
Love, P., & Edwards, D. (2004). Determinants of rework in building construction projects. *Engineering, Construction and Architectural Management*, 259-274.
National Standard of Canada. (2013). *Psychological health and safety in the workplace-Prevention, promotion, and guidance to staged implementation.* CSA Group and BNQ.
Nielsen, K., Randall, R., Yarker, J., & Brenner, S.-O. (2008). The effects of transformational leadership on followers' perceived work characteristics and psychological well-being: A longitudinal study. *Work & Stress*, 16-32.
Plaisier, I., De Graaf, R., de Bruijn, J., Smit, J., Dyck, R., Beekman, A., & Penninx, B. (2012). Depressive and anxiety disorders on-the-job: The importance of job characteristics for good work functioning in persons with depressive and anxiety disorders. *Psychiatry Research*, 382-388.
Rosen, C., Chang, C.-H., Djurdevic, E., & Eatough, E. (2010). Occupational Stressors and Job Performance: An Updated Review and Recommendations. *Stress and Job Performance*, 1-47.
Shin, M., Lee, H.-S., Park, M., Moon, M., & Han, S. (2014). A system dynamics approach for modeling construction workers' safety attitudes and behaviors. *Accident Analysis & Prevention*, 95-105.
Silverstein, B., Fan, Z., Bonauto, D., Bao, S., Smith, C., Howard, N., & Viikari-Juntura, E. (2010). The natural course of carpal tunnel syndrome in a working population. *Scandinavian Journal of Work, Environmental and Health*, 384-393.
Smith, D., Wei, N., Zhao, L., & Wang, R.-S. (2004). Musculoskeletal complaints and psychosocial risk factors among Chinese hospital nurses. *Occupational Medicine*, 579-582.
Sterman, J. (2001). Systems dynamics modeling: Tools for learning in a complex world. *California Management Review*, 8-25.
Thiese, M., Hegmann, K., Kapellusch, J., Merryweather, A., Bao, S., Silverstein, B., & Garg, A. (2015). Associations between Distal Upper Extremity Job Physical Factors and Psychosocial Measures in a Pooled Study. *BioMed Research International*, 1-10.
Tsuchiya, M., Kawakami, N., Ono, Y., Nakane, Y., Nakamura, Y., Fukao, A., . . . Uda, H. (2012). Impact of mental disorders on work performance in a community sample of workers in Japan: The World Mental Health Japan Survey 2002-2005. *Psychiatry Research*, 140-145.
Visagie, J., Swanepoel, J., & Ukpere, W. (2014). Exploration of Psychosocial Risk and the Handling of Unsafe Acts and Misconducts in the Workplace. *Mediterranean Journal of Social Sciences*, 997-1012.
Wang, J., Zou, P., & Li, P. (2016). Critical factors and paths influencing construction workers' safety risk tolerances. *Accident Analysis & Prevention*, 267-279.
Wang, P.-C., Rempel, D., Harrison, R., Chan, J., & Ritz, B. (2007). Work-organisational and personal factors associated with upper body musculoskeletal disorders among sewing machine operators. *Journal of Occupational & Environmental Medicine*, 1-8.
WHO, World Health Organization. (2010). *Healthy Workplace Framework and Model: Background and Supporting Literature and Practices.* Geneva: World Health Organization.

Widanarko, B., Legg, S., Devereux, J., & Stevenson, M. (2015). Interaction between physical and psychosocial work risk factors for low back symptoms and its consequences amongst Indonesian coal mining workers. *Applied Ergonomics*, 158-167.

DEEP LEARNING APPROACHES FOR SMALL DIMENSIONAL BIOMEDICAL DATA

Karin Proell [(a)]

(a) Department of Medical Informatics and Bioinformatics
Upper Austrian University of Applied Sciences Hagenberg, Softwarepark 11, Austria
(a) Karin.Proell@fh-hagenberg.at

ABSTRACT
In this paper we apply convolutional neuronal networks in different configurations to solve prediction tasks on medical data: Given 27 blood parameters obtained by labor blood examination the classes of tumor markers C153 and PSA should be predicted. Based on former work the results of trained Multi-Layer-Perceptrons (MLP) were moderate. Our major interest was now focused on the question if the prediction quality of CNN models outperforms MLPs. We had to transform the vector of input data into a two-dimensional pseudo image and augment it with different correlation values for increasing spatial structure. Various experiments with CNNs show that the prediction quality slightly increases compared to MLPs.

Keywords: deep learning, convolutional neural networks, Multi-Layer-Perceptron, transformation of vector into pseudo image

1. INTRODUCTION
Analyzing biomedical data is most times a classification task. Given a series of sample data a model should be provided which assigns each sample to one of several predefined output classes which can be used for prediction tasks.
Typically in labor blood examination numerous blood parameters such as HB, WBC, HKT, MCV, RBC, PLT, KREA, BUN, GT37, ALT, AST, TBIL, CRP, LD37, HS, CNEA, CMOA, CLYA, CEOA, CBAA, CHOL, HDL, CH37, FER, FE, BSG1, TF and tumor markers such as AFP, C125, C153, C199, C724, CEA, CYFRA, NSE, PSA, S100, SCC, TPS etc. are measured and used for diagnostic purposes (Djavan et al. 2002; Harrison et al. 2005; Jung at al. 2005). The value ranges of tumor markers are divided into four non-overlapping intervals, called classes. Class 1 includes all values less than *Normal Value* of marker, Class 2 includes all values between *Normal Value and Extreme Normal Value* of marker, Class 3 includes values between *Extreme Normal Value and Plausible Value* of marker and Class 4 includes all values *Greater than Plausible Value.*
The question is to find a model to predict the classes of each tumor marker separately using only the measurements of the blood examination as input.
One major problem of this task are missing values in the input data. It may be that a specific medical procedure was not considered necessary in a particular case or that the procedure was taken in a different laboratory with the values not available in the patient record, or that the measurement was taken but not recorded due to time constraints.
In previous work we focused on a variety of methods to handle missing data, including relatively simple approaches like discarding samples containing missing data values, replacing missing values with zero or applying mean imputation. We also applied different approaches for estimation of missing values in the input data: neural network based estimation of a specific marker value depending on existing values of a related marker and neural network based estimation of missing tumor markers depending on standard blood parameter measurements (Jacak et al. 2014; Markey et al. 2006, Liparini et al. 2005).
Additionally we trained Neural Networks (MLPs) with different configurations using the blood parameters as input and the tumor marker classes as output we could observe different prediction quality of the models for each marker type (Jacak et al. 2011; Jacak et al. 2010a; Jacak et al. 2010b).
In all experiments prediction quality of the models was moderate but the best results were obtained for models of tumor marker C153 and the worst for models of tumor marker PSA.
Our major interest in the current experiments was now focused on the question if the prediction quality of CNN models outperforms prediction quality of traditional MLPs. In our experiments we compared prediction quality of CNN Models (deep learning) to traditional MLPs (shallow approach) for C153 and PSA tumor markers. We did not apply sophisticated methods for missing value imputation but simply replaced missing value data with zero.
At our disposal examples of 27 blood parameters (see above) with known output classes of tumor marker C153 of approximately 6200 patients are provided.
So our data set comprises 27 blood parameters as input and C153 class values as output of approximately 6200 patients.
Secondly examples of 27 blood parameters with known output classes of tumor marker PSA of approximately 4300 patients are available.

1.1. State of the art:
Deep learning has in recent years set an exciting new trend in machine learning. The theoretical foundations

of deep learning are well rooted in the classical neural network (NN) literature. But different to more traditional use of NNs, deep learning accounts for the use of many hidden neurons and layers—typically more than two—as an architectural advantage combined with new training paradigms. While resorting to many neurons allows an extensive coverage of the raw data at hand, the layer-by-layer pipeline of nonlinear combination of their outputs generates a lower dimensional projection of the input space. Every lower-dimensional projection corresponds to a higher perceptual level. Provided that the network is optimally weighted, it results in an effective high-level abstraction of the raw data or images. This high level of abstraction render an automatic feature set, which otherwise would have required hand-crafted or bespoke features (Miotto et al. 2017; Schmidhuber 2015).

In domains such as health informatics, the generation of this automatic feature set without human intervention has many advantages:

Clinical Imaging: Following the success in computer vision, the first applications of deep learning to clinical data were on image processing, especially on the analysis of brain Magnetic Resonance Imaging (MRI) scans to predict Alzheimer disease and its variations. In other medical domains, CNNs were used to infer a hierarchical representation of low-field knee MRI scans to automatically segment cartilage and predict the risk of osteoarthritis. Deep learning was also applied to segment multiple sclerosis lesions in multi-channel 3D MRI and for the differential diagnosis of benign and malignant breast nodules from ultrasound images. More recently, CNNs were used to identify diabetic retinopathy in retinal fundus photographs, obtaining high sensitivity and specificity over about 10 000 test images with respect to certified ophthalmologist annotations. CNNs also obtained performances on classifying biopsy-proven clinical images of different types of skin cancer over a large data set of 130 000 images.

Electronic health record (EHR): More recently deep learning has been applied to process aggregated EHRs, including both structured (e.g. diagnosis, medications, laboratory tests) and unstructured (e.g. free-text clinical notes) data. The greatest part using deep architectures is applied for a specific, usually supervised, predictive clinical task. In particular, a common approach is to show that deep learning obtains better results than conventional machine learning models with respect to certain metrics, such as Area Under the Receiver Operating Characteristic Curve, accuracy and F-score. Most applications present end-to-end supervised networks, some works also propose unsupervised models to derive latent patient representations, which are then evaluated using shallow classifiers (e.g. random forests, logistic regression).
Several works applied deep learning to predict diseases from the patient clinical status. A four-layer CNN was used to predict congestive heart failure and chronic obstructive pulmonary disease and showed significant advantages over the baselines.
A comprehensive literature review about deep learning for healthcare, including a summary of articles can be found in (Miotto et al. 2017; Rav et al. 2017).

2. EXPERIMENTAL DATA

2.1. Initial Situation:

Compared to typical input data sets used in deep learning tasks we have to deal with the following problems in connection with our available data set:

- Input vectors comprise only 27 parameters (small dimension) per patient and the number of those sample vectors is limited to about 6200 for C153 and 4300 for PSA records.
- Many of the input vectors have a huge number of missing parameter values as blood examinations are expensive and therefore only a small subset of parameters is of interest for a specific diagnosis. The C153 data sets contains about 40 % and the PSA data set about 30% missing values in the blood examination data.
- The input vectors are not uniformly distributed among the four output classes. The major part of the input vectors is assigned to class value 1 which indicates no clinical evidence. We discarded samples of class 1 to obtain a proportion of not more than 50 % data sets of class 1 and 50 % in classes 2, 3 and 4.

2.2. Transformation of dimension:

The major problem for the CNN approach was the transformation of the one dimensional vector of 27 parameters of blood examination into a two dimensional augmented matrix without having additional information to be included. So the task was to transform the one dimensional vector of blood examination parameters of every patient into a two dimensional pseudo image. These images are considered as a matrix of pixel values with grey scaled values representing the values of blood examination. To increase the number of dimensions and to obtain a local spatial structure among every parameter value the following transformation methods for increasing the number of dimensions were applied:
Correlation was calculated

- between input parameters and
- between input parameter and tumor markers

and added as further dimensions to the pixel matrix.
This results in a non-quadratic matrix which was finally resized to a quadratic one with 28x28 dimensions. These pseudo images were used as input data to train MLPs and CNNs with different configurations.
In Table 1 correlation of blood parameters to tumor markers C153 and PSA are shown.

Table 1: Correlation of blood parameters to tumor markers C153 and PSA

Blood Parameter	Correlation to C153	Correlation to PSA
ALT	0,2	-0,1
AST	0,4	0,0
BSG1	0,2	0,2
BUN	0,1	0,1
CBAA	-0,1	-0,1
CEOA	-0,1	-0,1
CH37	-0,1	-0,1
CHOL	0,0	-0,1
CLYA	-0,2	-0,2
CMOA	0,0	0,0
CNEA	0,2	0,2
CRP	0,3	0,2
FE	-0,3	-0,1
FER	0,3	0,1
GT37	0,4	0,0
HB	-0,4	-0,2
HDL	-0,1	0,1
HKT	-0,4	-0,2
HS	0,1	-0,1
KREA	0,1	0,1
LD37	0,5	0,2
MCV	0,0	0,1
PLT	0,2	0,1
RBC	-0,3	-0,2
TBIL	0,0	0,0
TF	-0,3	-0,1
WBC	0,0	0,1

These pseudo images were used as input data to train MLPs and CNNs with different configurations. In Figure 1 und Figure 2 examples of such pseudo images are presented.

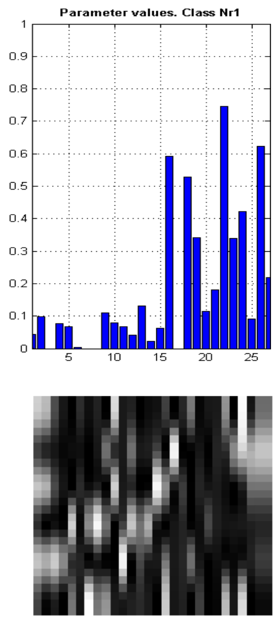


Figure 1: Example of a pseudo image for a Class 1 patient. Black fields indicate missing values.

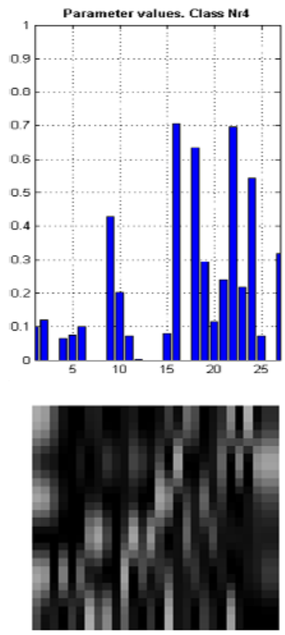


Figure 2: Example of a pseudo image for a Class 4 patient. Black fields indicate missing values.

2.3. Experiments

Our major interest in the experiments was focused on the prediction quality of CNN models in comparison to traditional MLPs. To perform the experiments we separated the data set with the 6200 and 3400 pseudo images into a collection of training and test set. The experiments with MLPs were conducted using the Neural Network Toolbox™ of MATLAB 2016a. The CNN experiments are based on the DeepLearnToolbox (Berg 2012) with slightly modified code.

After processing a training set its performance was measured on the corresponding test set. The following experiments were performed:

1. Experiments with MLPs: Traditional MLPs with various numbers of neurons were applied to train the pseudo images. The best result could be obtained using a two hidden layer MLP with 54 neurons.
2. Experiments with CNNs: For CNNs we used different configurations concerning:
   - Number of filters
   - Dimension of filters
   - Type of filters
   - Number of convolutional layers
   - Different scales of maxpooling

2.3.1. Experiments with tumor marker C153

For experiments with marker C153 we divided all samples (=6200) into 70% training data and 30% test data. Table 2 presents the different configurations for CNNs used for 10 experiments for marker type C153.

Table 2: Configuration of CNNs in experiments for tumor marker C153

	Number of filters	Dimension of filters	Number of convolutional layers	Scale of maxpooling	Types of filter
exp 1	1	5x5	1	4	average
exp 2	1	5x5	1	4	gaussian
exp 3	1	5x5	1	4	log
exp 4	1	5x5	1	4	prewitt
exp 5	1	5x5	1	4	random
exp 6	3	5x5	1	4	average, log, gaussian
exp 7	3	5x5	1	4	average, random, gaussian
exp 8	3	5x5	1	4	prewitt, log, random
exp 9	3	5x5	1	4	random
exp 10	3	5x5	1	4	prewitt, prewitt, log

The best result could be obtained using the configuration of CNN in experiment 9:

- Number of filters: 3
- Dimensions of filters: 5x5
- Types of filters: random
- Number of convolutional layers: 1
- Scale of maxpooling: 4

Performance in all experiments was measured by the following criteria:

- Percent of correct classification of samples (true positives)
- Distribution quality: Distribution of samples to correct classes
- Cohen's kappa coefficient

The results of the C153 experiments are shown in Figure 3.

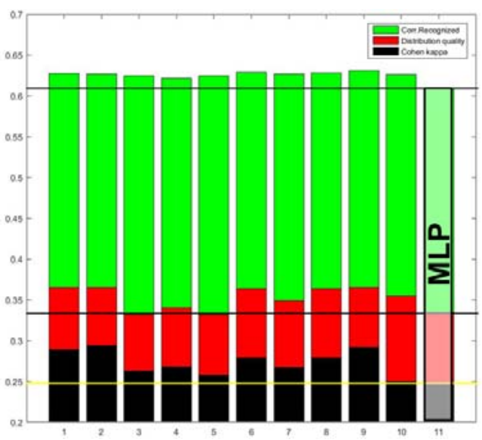


Figure 3: Comparison of results for marker C153 between ten different configuration setups of CNNs and one MLP based on the quality criteria: Percent of correct classification of samples, distribution quality among classes and Cohen's kappa coefficient

2.3.2. Experiments with tumor marker PSA

From previous work we know that the prediction quality of models for the PSA data is low. Therefore our experiments were focused on an increasing number of kernels working with random filters. The samples (=3412) were divided into 70% training data and 30% test data. Table 3 presents the different configurations for CNNs used for 4 different experiments for marker type PSA. The best result could be obtained using the configuration of CNN in experiment 3 with 7 random filters. Filter numbers greater 7 tend to over fit, the recognition rate of the test samples decreases.

Table 3: Configuration of CNNs in experiments for tumor marker PSA

	Number of filters	Dimension of filters	Number of convolutional layers	Scale of maxpooling	Types of filter
exp 1	3	5x5	1	4	random
exp 2	5	5x5	1	4	random
exp 3	7	5x5	1	4	random
exp 4	9	5x5	1	4	random

The best result could be obtained using the configuration of CNN in experiment 4:

- Number of filters: 7
- Dimensions of filters: 5x5
- Types of filters: Random
- Number of convolutional layers: 1
- Scale of maxpooling: 4

The results of the PSA experiments are presented in Figure 4.

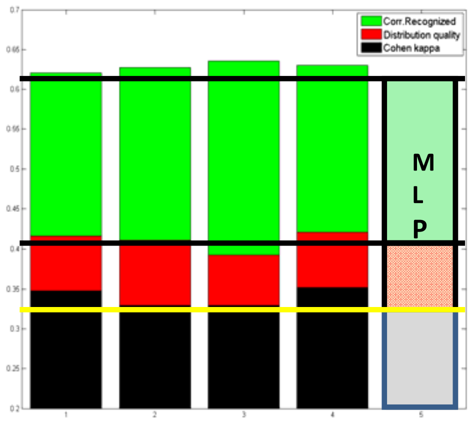


Figure 4: Comparison of results for marker PSA between four different configuration setups of CNNs and one MLP based on the quality criteria: Percent of correct classification of samples, distribution quality among classes and Cohen's kappa coefficient

3. RESULTS

We have achieved moderate increase of performance (about 3%) applying convolutional neural networks compared to MLPs trained on human blood parameters as input and C153 and PSA tumor marker as output.

The quality of the learned systems is primarily dependent of quality and size of the training sets. In the available blood examination data sets in our experiments we had to deal with about 40% missing values and input vectors are not uniformly distributed among the four output classes. We replaced missing values by imputation with zero value and discarded samples of class 1 to get a proportion of 50% class 1 samples and 50% samples of class 2, 3 and 4.

Standard neural networks are state-of-the-art classifiers that operate on vectors, without knowledge of the input topology. However, convolutional neural network exploit the knowledge that the inputs are not independent elements, but arise from a spatial structure.

Therefore we did several experiments with different configurations for CNNs augmenting the blood examination vector with correlation values between input parameter and tumor marker values and transformed it into a pseudo image.

In all experiments we found a CNN configuration which outperforms MLPs. Therefore we will apply further work in data preparation especially in context of missing value imputation and perform continue experiments with different CNN configurations on tumor markers other than C153 and PSA.

REFERENCES

Berg Palm R., 2012. Prediction as a candidate for learning deep hierarchical models of data, Available from: https://de.mathworks.com/matlabcentral/fileexchange/38310-deep-learning-toolbox

Djavan et al., 2002. Novel Artificial Neural Network for Early Detection of Prostate Cancer. Journal of Clinical Oncology, Vol 20, No 4, 921-929

Harrison, R.F., Kennedy, R.L., 2005, Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. Ann Emerg Med.; 46(5):431-9.

Jacak W., Proell K., Winkler, S., 2014. Neural Networks based Feature Selection in Biological Data Analysis. in: Klempous, R., J. Nikodem, W. Jacak, Z. Chacko, eds. Advanced Methods and Applications in Computational Intelligence, Springer, pp.79-93

Jacak W., Proell K., 2010a, Data Driven Tumor Marker Prediction System, Proceedings of EMSS 2010, Fes, Marokko

Jacak W., Proell K., 2010b, Neural Network Based Tumor Marker Prediction, Proceedings of BroadCom 2010, Malaga, Spain

Jacak W., Proell K, 2011. Neural Networks Based System for Cancer Diagnosis Support - Proceedings of International Conference on Computer Aided Systems Theory EUROCAST 2011, Las Palmas, Spanien, pp. 283-285

Jung I., Thapa D., Wang GN., 2005. Neural Network Based Algorithms for Diagnosis and Classification of Breast Cancer Tumor. In: Hao Y. et al. (eds) Computational Intelligence and Security. CIS 2005. Lecture Notes in Computer Science, vol 3801. Springer, Berlin, Heidelberg

Liparini A., Carvalho S., Belchior J.C., 2005. Analysis of the applicability of artificial neural networks for studying blood plasma: determination of magnesium on concentration as a case study. *Clin Chem Lab Med*.; 43(9):939-46

Markey M.K., Tourassi G.D., Margolis M., DeLong D., 2006. Impact of missing data in evaluating artificial neural networks trained on complete data. Computers in Biology and Medicine, 5, 516

Miotto R., Wang F., Wang S., Jiang X., Dudley J., 2017. Deep learning for healthcare: review. Briefings in Bioinformatics, 2017, 1–11

Rav D., Wong C., Deligianni F., Berthelo M., Andreu-Perez J., Lo B., Yang G., Fellow, 2017. Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, Vol 21, No. 1, January

Schmidhuber J., 2015. Deep learning in neural networks: An overview. Neural Networks, Volume 61, January 2015, Pages 85-117

UNITARY HYBRID MODEL OF RAILWAY TRAFFIC

Radek Novotný[a], Antonín Kavička[b]

[a],[b]Faculty of Electrical Engineering and Informatics, University of Pardubice

[a]radek.novotny3@student.upce.cz, [b]antonin.kavicka@upce.cz

The contribution deals with the methodology of building unitary hybrid simulation models reflecting railway traffic. Those models apply different level of abstraction (granularity) to diverse parts of a simulating system. Special attention is paid to different traffic models applied to microscopic and macroscopic elements of hybrid model. Because several different submodels coexist within a hybrid model, which are connected to different traffic indicators, it is necessary to solve the transformation of traffic flows on the interface of microscopic and macroscopic submodels.

Keywords: scalable simulation model, hybrid model, railway traffic, transformations of traffic flows

1. INTRODUCTION

Modelling a railway infrastructure and a corresponding railway traffic represents an important part of the research focused on railway system optimizations. For such purposes the researchers use the experimental research method of computer simulation in which the level of *granularity* (applied within simulators) plays an important role. It defines the level of details considered within an examined simulating system. The different levels of details required for investigating the railway traffic, can determine various simulation models utilizing different granularity. Thus, simulations can be classified according to their granularities as *microscopic, mesoscopic*, and *macroscopic* (Krivy and Kindler 2003; Burghout 2004).

For the needs of simulation studies with defined goals, it is essential to have at disposal appropriate specifications of simulating systems. It means that an appropriate level of details for the needs of relevant examinations has to be selected. Traditional approaches apply the same level of details for the entire simulator – i.e. that homogenous approach does not allow to combine microscopic and mesoscopic levels of details within one simulator. While applying investigations on the microscopic level, it is typical to explore in detail the interactions among individual mobile entities. On the other hand, the macroscopic level of details typically provides only rough features of traffic flows. There is a strong motivation for designers of traffic simulations to use methodologies for building scalable simulation models of railway traffic. Those models enable to combine and interconnect:

- various submodels of infrastructure built on different levels of details (Hansen and Pachl 2008; Cui, and Martin 2011; Novotny and Kavicka 2016) and
- different traffic submodels reflecting granularities of relevant infrastructural submodels – certainly relevant transformations of traffic flows are supposed to be carried out on the boundary between corresponding submodels.

The mentioned approach has to be supported by appropriate software tools (e.g. infrastructure editors and integrated simulation environments etc.).

2. UNITARY HYBRID SIMULATION MODEL

Our presented methodology is based on a hybrid simulation model implemented within one simulation tool (*unitary hybrid model*). That methodology supports combining submodels exploiting the microscopic and macroscopic levels of details. Microscopic simulation is connected to particular areas, within the frame of which important details about traffic (and infrastructure) are important for the experimenter. On the other hand, macroscopic simulation is utilized within those parts of the simulator for which rough operational/traffic observations are sufficient. Unitary hybrid model enables to adjust the *granularity* of selected parts of a simulator. The mentioned parts are connected to relevant traffic submodels operating over corresponding infrastructure submodels.

Overall computational demands of a unitary hybrid model are certainly lower than demands related to a corresponding model executing pure microscopic simulation (Novotny and Kavicka 2016).

3. MICROSCOPIC INFRASTRUCTURE MODEL

The methodology of building unitary hybrid models focuses primarily on the construction of a track infrastructure submodel. That submodel applies the highest level of details which can be required for the given part of the railway network. Within this context the editing tool *TrackEd* (Novotny and Kavicka 2015) can be utilized. The mentioned software is specialized in (i) quick constructions of track layouts with the help of prearranged prototypes of rail objects and (ii) subsequent schematic visualizations depicting track infrastructures. The resulting submodel is then represented by a data structure depicting a mathematical model based on an undirected graph. Each graph node encapsulates not only the position in a schematic plan, but also real geographic (kilometric) position/coordinates within the railway network. Within the editor it is possible to define topological, metric and slope characteristics related to all tracks or their parts. Hence, it is possible to carry out realistic calculations (during simulation trials) concerning the dynamics of train rides. The created microscopic infrastructure submodel considers: (i) tracks, (ii) switches, (iii) crossings, (iv) signal devices, (v) limit signs for train positions on tracks, (vi) platforms, (vii) isolated circuits, (viii) electrification and useful lengths of tracks, and finally (ix) speed limits valid for individual rail elements (Kubat 1999; Jirsak 1979).

4. HYBRID INFRASTRUCTURE MODEL

From the viewpoint of the resulting *hybrid submodel/layer of infrastructure* (e.g. built in editing tool *TrackEd*) it is necessary to distinguish between *micro-layer*, which corresponds to the above mentioned microscopic submodel and *hybrid-layer* composed of *micro-segments* and *macro-segments*. *Micro-segments* are represented by sub-graphs directly taken from the micro-layer. *Macro-segments* apply higher degree of granularity (i.e. lower level of details) to relevant disjoint connected sub-graphs from the *micro-layer* (Novotny and Kavicka 2016). Two types of macro-segments (*macro-nodes* and *macro-edges*) are distinguished within the presented methodology. *Macro-edges* typically encapsulate line sequences of edges from micro-layer. *Macro-nodes* can enclose a general connected sub-graph from the micro-layer. Constructions of *macro-segments* support creating variant configurations of hybrid submodels of railway infrastructure. It means in fact that different scenarios of simulation experiments can apply various levels of details (micro- or macroscopic) within an infrastructure submodel.

5. HYBRID TRAFFIC MODEL

Because of combining macro-segments with areas including microscopic elements within the hybrid infrastructure model, it is necessary to apply different traffic models implementing various levels of abstraction, which are connected to different traffic indicators (Cenek 2004). From the viewpoint of the implementation of the unitary hybrid model (Figure 1) it is necessary to distinguish between traffic models applied to microscopic and macroscopic infrastructure segments of hybrid model. *Microscopic traffic submodels* observe detailed riding features of individual trains. *Macroscopic traffic submodels* utilize the theory of traffic flow based on the analogy of flowing of liquids within the macroscopic areas.

Because several different traffic submodels coexist within a hybrid model, it is necessary to solve transformations of traffic flows, i.e. it is necessary to unambiguously define the information about railway traffic on the interface between each microscopic and macroscopic submodel in order to maintain consistency of data.

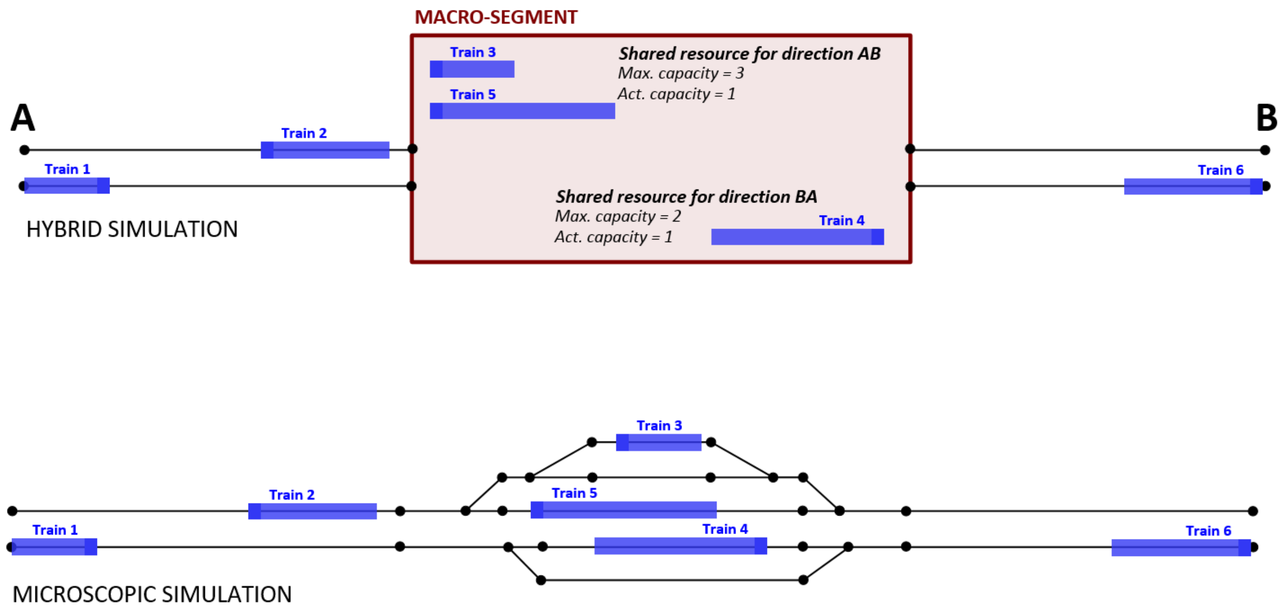


Figure 1 – Hybrid simulation model (analogy based on competition for shared resources)

.

From the implementation perspective, the transformations of traffic flows must distinguish between microscopic $\rightarrow$ macroscopic transformations and macroscopic $\rightarrow$ microscopic ones. When transferring the rolling stock from the microscopic to the macroscopic submodel, the main problem is the loss of a large amount of data items. Those items must be unambiguously defined for individual rolling stock which is later transferred from the macroscopic submodel back into the microscopic one (Burghout 2004; Novotny and Kavicka 2016). When transferring the rolling stock, the current status of traffic in the relevant surroundings must be considered.

6. MICROSCOPIC TRAFFIC SUBMODEL

Microscopic traffic submodel is based on applying realistic calculations connected with the train ride dynamics and train interactions (Divis and Kavicka 2015). Train rides computations use detailed information about railway infrastructure (real topology/metric and slope properties) defined in *micro*-layer of infrastructure model (Novotny and Kavicka 2015). For trains, it is necessary to maintain the traction characteristics, resistance, breaking and other train parameters. Some other parameters are abstracted (wind resistance in a tunnel or effects of transition curve on the train ride). Defining such parameters would prolong creating a hybrid simulation model in comparison with the insignificant effect on the riding time of trains.

The calculation of train driving dynamics is performed by numerical integration with an intermediate step of calculation for greater accuracy repeatedly according to a defined time frame due to the changing train position and track charakteristicks. In other words, the movement of trains on the infrastructure reflects the real driving characteristics with respect to the train and tracks properties. Within the above presented microscopic traffic submodel, which apply higher levels of abstraction to some aspects, is possible to achieve reliable results from simulation experiments without any significant loss of their informative value.

Within the integrated development environment of *TrackEd* is implemented simulation engine based on the method of discrete events planning using agent-oriented architecture. This method uses a calendar of events, from which events are removed in the correct chronological order and executed. *TrackEd* features online animation for easy and clear overview of the progress of a simulation and monitoring train position and speed during the simulation, especially in microscopic submodels. The simulator is capable of setting the speed of animation based on the user's needs or to pause the animation if a detailed inspection of the simulation status is needed.

7. MACROSCOPIC TRAFFIC SUBMODEL

According to the existence of *macro-segments* differences due to the encapsulation of different parts of microscopic infrastructure model (railway stations, tracks, etc.) within a hybrid model, there are different realizations of the traffic models in these macroscopic submodels.

7.1. Analogy of flowing of liquids

Movements of rolling stock through macroscopic submodels can be described analogously to the formalisms defining fluid flows. As well as the motion of fluid particles is affected by the movements of surrounding particles of liquid, the movements of individual pieces of rolling stock within *macro-segments* depend on surrounding train rides. This approach is directly taken from the domain of road traffic simulation (Daiheng 2011). Some of the original principles are not considered, because different variability of the rolling stock movement along the infrastructure exists (mainly determined by the train path and the assigned timetable). The proposed approach of transformation of traffic flow (characterized by a certain intensity φ) assumes an assignment of average speed $\overline{v}$ to each individual train entering into the macroscopic submodel. Assigned current average speed $\overline{v}(\rho_C)$ depends on the current density (ρ_C) of traffic flow in *macro-segment*. That density is determined for every *macro-segment* from the attributes (the number of isolated circuits, average length of train routes, etc.) defined in relevant part of *micro*-layer and the current number of trains being inside the relevant segment. According to the mathematical formula can be the average speed assigned to each train in *macro-segment* expressed determined by the maximum allowable speed v_{max}, the current density ρ and maximum density ρ_{max} (maximum number of trains in macro-segment):

$$\overline{v}_{(\rho)} = v_{max}(1 - \frac{\rho}{\rho_{max}})$$

Certainly, the average speeds of all trains in *macro-segment* are always recalculated (when a train approaches/leaves the segment) in order to maintain traffic consistency.

From the viewpoint of macroscopic traffic submodel discussed above, if the average speeds are known in the instants of time when the trains leave the macro-segment be pre-calculated (it is not necessary to observe individual trains during their temporary stay in the macroscopic submodel). It is essential convenient to consider a relevant time table as well (Burghout, Koutsopoulos and Andreasson 2006).

On the other hand, depending on the different characteristics of the *macro-segments* and maintaining consistency of information about traffic, it is also include to calculation information about train stopping on some of the encapsulated microscopic elements by the timetable (for example, station tracks).

7.2. Analogy based on competition for shared resources

Different approaches to modelling traffic within macro-segments are to be imagined as a shared resource. In which can be only a certain number of rolling stock.

The maximum number of trains within the submodel can be given by attributes defined in relevant part of micro-layer. A similar parallel can be seen in real railway traffic where the number of station tracks can be the factor determining the maximum number of trains allowed in the railway station the same time. Other trains are not allowed to enter the station and have to wait for the leaving of some station tracks.
The basic difference from the previous concept is that trains do not interactions on each other (i.e. trains do not change the density of traffic flow in submodel and it does not have an effect on the average speed assigned to the next entering train).
In other words, if the train can enter to the macroscopic submodel (from the rules defined above), it is defined the simulation time for individually train, when the current submodel leaves. Simulation time is calculated from the attributes of the microscopic elements, which consists of a train path through the submodel.
An important factor is the right simulation time to be added if the train stops in the macro-segment. This is the reason of braking or acceleration of the train in the case of microscopic simulation traffic.

7.3. Extension of competition for shared resources

In the case of a selected traffic model for macroscopic submodels based on competition for shared resource, the macro-segment can be blocked by train arriving from only one direction. Of course, leads to an increase the delay of trains in the opposite direction and the inconsistency of the hybrid simulation model compared with only microscopic one. In real traffic (in a railway station) a set of station tracks is always assigned to each direction. If we neglect the unusual situation of real traffic, when the station uses the same station tracks for different directions (dispatcher-controlled methodology), we can use a group of shared resources within each macro-segment. The advantage is the trains coming from different directions will affect each other only in cases, e.g. if railway stations use the same station tracks for different directions.

7.4. Summary

From approaches to modelling a macroscopic traffic submodel, it is not possible to use one global traffic model for all types of macro-segments. A general assumption is use a macroscopic traffic submodel based on competition for shared resources for macro-nodes (encapsulating railways station etc.) and for macro-edges (encapsulating open tracks etc.) traffic model based on the analogy of flowing of liquids.
In order to accept this fact, all presented concepts of macroscopic traffic models for both types of macro-segments must be implemented and validated by a series of simulation experiments.

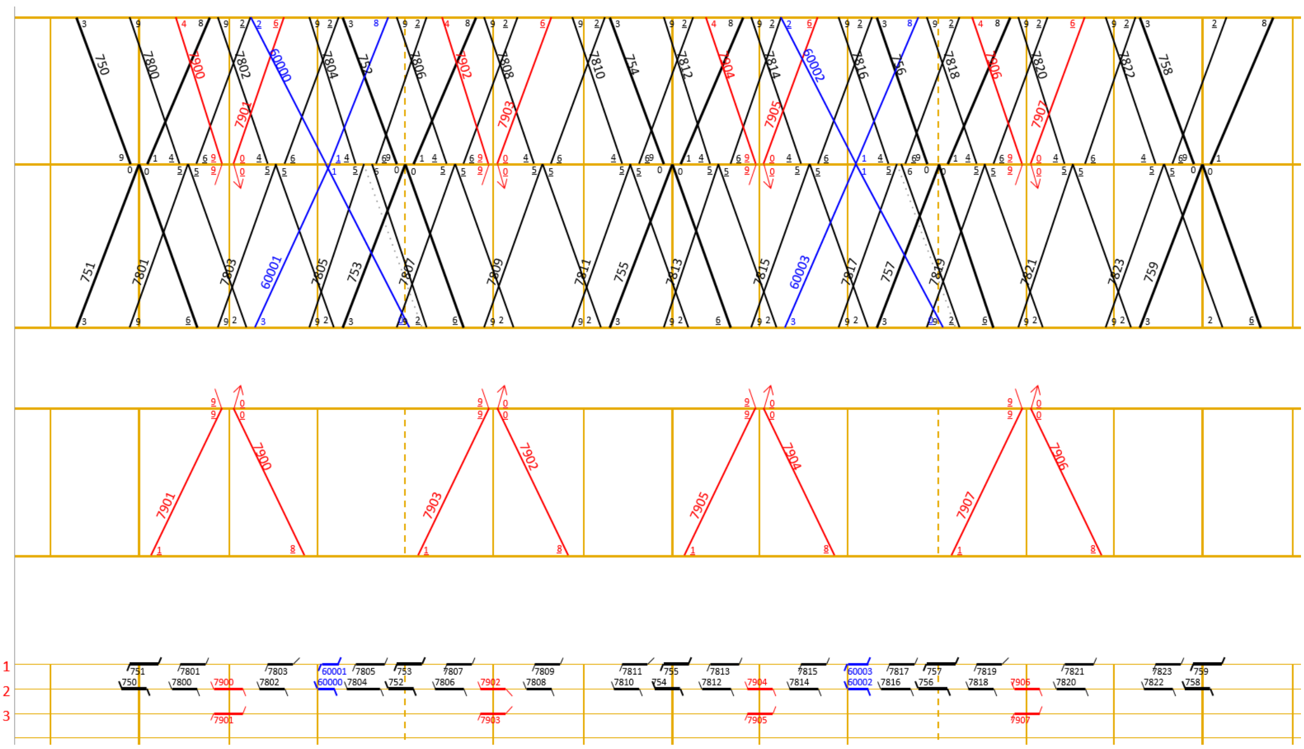


Figure 2 - Timetable of prototype station

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

8. VERIFICATION AND VALIDATION

In order for the unitary hybrid model of railway traffic (applying different level of abstraction to diverse parts of simulation system) to be used in practice, it is necessary for simulation results to be correct and real. To verify the correct functionality of hybrid simulation model, it is possible to identify validation methodology:

- deterministic simulation over the microscopic simulation model of railway traffic
- deterministic simulation over the hybrid simulation model of railway traffic
- hybrid simulation of railway traffic with various delays (stochastic simulation)

It was necessary to verify the microscopic model of railway traffic focused primarily on the proper implementation of the train driving dynamics in the initial phase. In other words, comparison of a tachogram obtained from the simulation results with a similar tachogram obtained from a real environment or other simulation tool (Divis and Kavicka 2015). Next, the series of simulation experiments were carried out on the case study of the railway traffic an appropriate area of the official railway network in Czech Republic used by *The Railway Infrastructure Administration* (SZDC). Other, more complex verification was compared with the results from the simulator Villon (representing a validated simulation tool widely accepted by experts from railway practice).

9. CASE STUDY

As a part of the initial testing, an infrastructure, which consists of a prototype station and several adjancent border stations. According to the real timetable of the prototype station (*Figure 2*), containing several trains in both directions, a series of deterministic simulation experiments was carried out. These steps have been replicated in the simulation tool Villon. Next, a series of simulation experiments was conducted. After the experiments were performed, statistical evaluations of the results were done and the differences in driving characteristics were compared. The difference of measured values is in the matter of units of percent and it is expected due to the apply higher levels of abstraction to some aspects of microscopic simulation. Implemented microscopic traffic submodel was accepted from the viewpoint of the deterministic simulation experiments.

Due to validation of the macroscopic traffic submodel, the prototype station was encapsulated into macro-node. Variant of hybrid traffic model was created, which apply different level of granularity in the simulator. The results of the deterministic simulation experiments of the hybrid model of railway traffic were compared with the results, when microscopic traffic submodel was used for the entire infrastructure. Verification of the correctness of the hybrid traffic model focused primarily on deviations of the train delays moves through the macro-node. In other words, the transformation of traffic flows on the interface between microscopic and macroscopic submodels of the hybrid model must have a minimal effect on the final train delay.

In the case of macro-node, a macroscopic traffic submodel based on analogy of flowing of liquids was tested. An increase in train delays in the order of tens of percent was detected. The reason may be that the prescribed timetable of the prototype station did not include, for example, train waiting at the station tracks in macro-node. The evaluation of macroscopic traffic submodel based on shared resource was better. Train delays showed little deviation according to the results of the microscopic simulation traffic.

Next step is implementation of other predefined macroscopic traffic submodels for defined hybrid model of infrastructure with macro-node. Similarly, the validation will proceed for hybrid model of infrastructure containing macro-edge. The results of simulation experiments will then determine the usability of macroscopic traffic submodels for different macro-segments and also to determine the complex validation of variant configurations of the unitary hybrid model.

10. PERSPECTIVES OF DEVELOPMENT

The next stage of development related to an implementation of traffic models for *macro*-edges and their evaluation. From the viewpoint of preserving consistency of traffic information, it will be necessary to clearly identify the implementation of macroscopic traffic models for *macro*-nodes and *macro*-edges. In the case of different macroscopic traffic submodels used for two types of macro-segments, to solve the transformation of traffic flows on their interfaces. In respect to the use of unitary hybrid model of railway traffic in practice.

Prospective option of further development can be propose a methodology of a general macroscopic traffic model applicable to any defined *macro-segments*.

Other prospective option of further development can be related to an extension of the unitary hybrid simulation model supporting constructions of mesoscopic segments/areas. Mesoscopic segments would enable more detailed traffic simulation than it is carried out within macroscopic areas. Or extending the editor by functionality, thanks to which macro-edges created above a multi open track are encapsulated into one aggregated macro-edge.

11. CONCLUSION

The article deals mainly with the explanation of basic phases of methodical approach to the construction of a hybrid model of railway traffic.

Hybrid model of railway traffic combining submodels applying different level of abstraction (microscopic and macroscopic). It is distinguished between *microscopic traffic submodel* applied to the microscopic elements of the hybrid model and *macroscopic traffic submodel* used in the case of macro-segments. Special attention is paid to explaining the methodology related to

transformation of traffic flows on the interface between microscopic and macroscopic submodel.
Microscopic traffic submodel is based on applying realistic calculations connected with the train ride dynamics. On the other side, apply higher levels of abstraction to some aspects, but still enables to achieve reliable results from simulation experiments without any significant loss of their informative value. There are two approaches to modelling railway traffic within macroscopic submodels (ie. analogy of flowing of liquids and analogy based on competition for shared resources) and their extended variants.
Within simulation experiments a hybrid model of infrastructure (containing macro-node) has been built over a real microscopic model of infrastructure. Both basic versions of the macroscopic traffic models have been implemented. Finally, the possibility of using them in a particular hybrid model was discussed. In addition, a validation methodology was proposed to evaluate the accuracy of the simulation results.
The scope of future development (in editor *TrackEd*) also considers the implementation of traffic models for macro-edges and their evaluation.

REFERENCES

Cui Y., and Martin U., 2011. Multi-scale Simulation in Railway Planning and Operation. PROMET Traffic. Available from: http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/186 [Accessed 13 July 2017].

Burghout W., 2004. Hybrid microscopic-mesoscopic traffic simulation. Doctoral Dissertation. Royal Institute of Technology.

Novotny R., and Kavicka A., 2015. Model of a railway infrastructure as a part of a mesoscopic traffic simulator. To appear in: The 27th European Modeling & Simulation Symposium. Bergeggi, Italy.

Novotny R., and Kavicka A., 2016. Scalable simulation models of railway traffic. To appear in: The 28th European Modeling & Simulation Symposium. Larnaca, Cyprus.

Krivy I., and Kindler E., 2003. Modeling and simulation I. Ostrava: PrF OU. [in Czech]

Cenek P., 2004. Management traffic processes in logistics systems. AT&P journal, 13-15. [in Czech]

Kubat Bohumil, Fliegel Tomas, 1999. Railway buildings 30. Prague: CVUT. [in Czech]

Jirsak Zbynek, 1979. Railway stations and junctions. Bratislava: Alfa. [in Czech]

Hansen Ingo, Pachl Jorn, 2008. Railway timetable and traffic. Hamburg: Euraipress.

Daiheng Ni, 2011. Multiscale modeling of traffic flow. Mathematica Aeterna, 27-54. University of Massachusetts.

Burghout W., Koutsopoulos H., and Andreasson I., 2006. A discrete-event mesoscopic traffic simulation model for hybrid traffic simulation. To appear in: Intelligent Transportation Systems Conference. Toronto.

Divis R., and Kavicka A., 2015. Design and development of a mesoscopic simulator specialized in investigating capacities of railway nodes. To appear in: The 27th European Modeling & Simulation Symposium. Bergeggi, Italy.

DES TECHNIQUES APPLIED TO THE DESIGN OF A TRAINING SESSION ABOUT LEAN METHODOLOGY IN A SHIPBUILDING PROCESS

José Antonio Muiña-Dono [(a)], Adolfo Lamas Rodríguez [(b)], Ángel Fernández Rodríguez [(c)], David Chas Álvarez[(d)]

[(a)] [(d)] UMI Navantia-UDC
[(b)] Universidade da Coruña, Navantia, UMI Navantia-UDC
[(c)] Universidade da Coruña

[(a)] jose.mdono@udc.es, [(d)] david.chas@udc.es
[(b)] alamas@udc.es
[(c)] angel.fernandez@udc.es

ABSTRACT

The aim of this work is the application of a discrete event simulation tool for the design of a learning sessions based on Lean Manufacturing techniques applied in a shipbuilding processes. This tool based on DES could serve us for teaching the possibilities of apply the lean manufacturing concepts in a shipyard where this type of techniques are unusual, except for a very specific purpose and on rare occasions.

To do that, an effort will be made to design the training session as feasible and realistic as possible by using DES techniques modelling a real shipbuilding process taking into account the existing restrictions related to the fabrication period and resources available.

Keywords: Lean, shipbuilding, discrete event simulation

1. INTRODUCTION

The shipbuilding sector has its own characteristics that make it different from the rest of the industry, such as the fabrication of a single or several units of one product for a particular customer and a specific purpose. This makes that the finished products have a high customization and a large technological component, in addition to having a high economic value and a very long manufacturing process that can even last for years.

The economic crisis in recent years and the current slow economic recovery result in an excess of productive capacity. In addition, a strong global competition and a very tightly cost strategy, focused on attracting new customers, put the shipyards in a critical situation where the majority of the orders had been awarded by Asian shipyards that absorb 83.3% of the world's contracting, 81.9% of the product portfolio and the 84.4% of production in 2011 (Suaz González 2012).

This skill of Asian shipyards to lead the shipbuilding industry has reduced the market share of European shipyards, passing from 25% of the world market in 1994 to 2.7% in 2009 (CC.OO. 2013). This situation has been particularly hard in Spain, which had been one of the main producers in the world, reaching almost 20% of the world orders in Europe, went through a difficult situation after the opening of the file and the subsequent problems happened in 2011 by the European Commission due to the tax breaks of shipowners.

This caused a delicate state in shipbuilding and based on the situation of the referents shipyards, Lean techniques are recommended for a more rational and efficient use of its resources, making our shipyards more competitive and attracting new customers.

However, although Lean techniques have not been fully implemented in any of the shipyards studied, they have been used by using tools designed only to solve specific problems.

This need for change the approach to apply Lean techniques to the entire manufacturing process has motivated this work, since most authors point out the need to actively involve all the staff in this process because they are the ones who really know the process and will coexist with these changes. Therefore, we must seek their participation, searching for the solutions so they must be involved and trained in this new philosophy of work.

Undoubtedly, another of the most important changes that will affect all industries is the use of new tools more in line with Industry 4.0. This new industrial revolution starts with the Smart Factory which it is defined as a "*manufacturing solution*" (Radziwon et al. 2014) that provide us a quick and flexible response to the changing market conditions. This new revolution will pass not only to a process automation, we are talking about a new cooperative factory between machines and humans, the interaction about real and virtual world…And a way to achieve this optimized factory is the use of simulation process and virtual tools in order to reduce waste (Turner, Hutabarat, and Oyekan 2016).

Therefore, in this work we try to facilitate this transformation by creating a tool based on simulation and Lean manufacturing techniques.

2. STATE OF THE ART

The aim of this work will be design Lean manufacturing training techniques tools based on the utilization of DES concepts.

(Padilla et al. 2016) study the need to create some tool for the training of employees in DES techniques, proposing the design of two games with a very different topic from the usual to attract their attention.

This rising trend to use games in the company is also shown in several articles of the prestigious magazine Forbes, highlighting the potential they have to involve the staff in the process. It is also possible observe the application of games in sectors with a varied objectives like the military (Raybourn 2014; Yildirim 2010), sanitary (Diehl et al. 2011; Ushaw, Eyre, and Morgan 2017) or safety (Silva et al. 2013). Even in the maritime sector you can find some example of the use of games and simulation as in (Longo et al. 2015) where a simulator is used to represent as complex situation as the arrival of a ship to the port and the interrelations between ship pilots and port traffic controllers.

In recent years we can observe an increment of the use of games and new technologies in classrooms too.

(Aldrich 2006) debates about the use of new techniques in teaching, such as simulation and video games, as well as the importance of "Learn by doing", demonstrating the importance of simulation to learn and understand complex processes that otherwise would be impossible.

(Page and Kreutzer 2006) treats the simulation as one of the main techniques of E-Learning, looking for understand very difficult and abstract ideas or scenarios with the use of these techniques.

A review of the use of simulations games at the Industrial Engineering universities was made in (Deshpande and Huang 2011), where the greatest disadvantage observed is the reticence to change by some trainers. The authors express the necessity to meet the needs of the new generation (the "*digital natives*"), aimed to solve the new problems of the industry. By this reason they show some real cases applied to any field of engineering (mechanics, electrical, logistics, etc.) highlighting as an important advantages: teamwork, better conceptualization and analytical thinking, etc.

(Liane Márcia, Marcel, and Subramanian 2007) apply the discrete event simulation for training engineering in production process, highlighting the possibilities for students to modify parameters and visually see what happens with their modifications.

(Standridge 2000) discusses about the advantages of simulation for the active training of engineers in a university through the development of a series of case studies prepared to represent the reality of their closest companies and skilling them for their next reality.

Another interesting work is (Van der Zee and Slomp 2005), where an explication of the complete creation process and the result obtained from the application of simulation and gaming to a real case of the manual assembly line was made. With this work they try to demonstrate the viability of the Lean implementation in a real company and the improvements produced in the motivation and the assimilation of concepts by the participants after some repetitions.

(Constantino Delago et al. 2017)proposes the creation of a virtual game based on the PBL (Problem Based Learning) methodology, using a DES 3D software for teaching the basics Lean concepts. Thanks to the use of these techniques, the authors express the possibility to analyse complex and realistic situations (the most important problem teaching engineering) and predict future situations, better than using traditional methods.

3. REAL PROCESS AND SIMULATED PROCESS

The complete process of shipbuilding is very complex with a large number of stages between the different processes. In addition, these stages differ significantly depending on the degree of evolution of each shipyard, such as the capacity of pre-outfitting before assembly that are able to reach.

We could say that the fabrication strategy usually applicable in shipbuilding is based on an integrated construction and has undergone no significant variations in the last 30 years, except some isolated cases that try to use robotic methods.

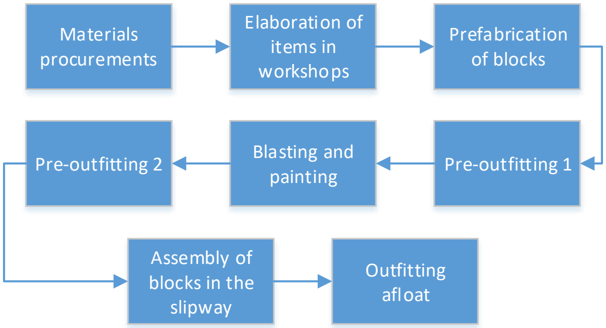


Figure 1: Usual Workflow in a Shipbuilding Process.

In the previous figure the main stages of a shipbuilding process have been represented, although these can be called differently depending on the shipyard to which we refer and, as we mentioned before, its degree of development.

The purpose of this work will not be to represent a process as complex as it is because we look for describing it in a training session, where the time and resources we have is significantly limited.

This is the reason why we have proceeded to search and define the main activities of each of these fabrication stages with the aim of showing, through a simplified representation, the main phases of the construction process of a ship, focusing in particular on a constructive process of an accommodation vessel.

In addition, to carry out this work, and to apply Lean techniques in a shipyard, we must discriminate between the elements where its application is most appropriate so, we distinguish the following elements:

Unique elements: are those products that go through an exclusive manufacturing process, which is not similar to

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

any other product. One example of this is the machinery room.
Repeatable elements: are those whose manufacturing process goes through the same stages although these stages do not have to be identical.
Experience has shown that around 35% of the ship's blocks could be defined as unique (singular) element; however, the remaining 65% could be classified as repeatable, distinguishing between different families.
Therefore, this work will deal with the construction process of repeatable elements as it represents the largest quantity of building blocks in a ship and is the natural way to implement Lean Manufacturing in the shipbuilding industry.
In addition, to facilitate this training for all the employees, regardless of their educational level, it has been proposed the creation of a physical model that allows all employees to interact with it and thus make easier the transition between the physical and the digital model without losing our goals.

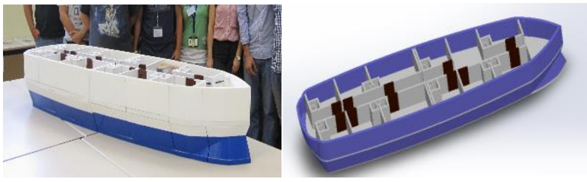


Figure 2: 3D Model Made for the Training.

These characteristics described in the paragraphs above, oblige us not only to clearly identify the repeatable elements and the main stages. In addition, thanks to the historical data of an important European shipyard and by consulting experts, we proceeded to perform a kind of scaling between reality and simulation to process times and their costs as well as the number of resources needed.

Table 1: Market Requirements.

Cost	Deadline
170 M€	26 weeks

Table 1 shows the assumed hypothesis regarding the term and cost demanded by the market, based also on historical data.
Therefore, these requirements will be the cost and maximum term that the shipowner would accept to this project, since another shipyard would be able to do it in those conditions. In consequence, if our cost exceeds this, we will not be competitive and the possibilities of our shipyard to award the contract will be reduced dramatically. On the other hand, a non-fulfilment in the delivery milestones failure to comply with the deadline will incur in penalties in addition to a poor image and reputation for our company.

4. MODEL DEVELOPMENT

Once the objectives of this work have been established, it has been decided to use a simulation software called ExtendSim whose versatility has led it to be used in various applications and even for the training of future professionals in the universities, as is the case of the University of A Coruña.

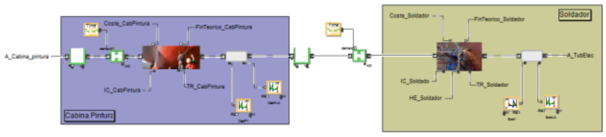


Figure 3: Part of the Model Made in ExtendSim.

Based on this, we have decided to use two models, one of them which represents the current manufacturing system called traditional method and, on the other hand, a more modern model where different Lean techniques are applied, adapted to the characteristics of the naval sector that has been named Lean method.
The different models will include, among others, the following parameters:

- Process time and its variability.
- Relationship between activities.
- Material and labour cost.
- Theoretical start and end time per activity.
- Real overtime limit:
  - Four hours during the week.
  - Twelve hours during the weekend.

4.1. Traditional method

Initially, we will have some operators to carry out this work according to the traditional method which is characterized because the activities will focus on the construction of large blocks, setting an individual objective to achieve our highest productivity which will lead to a search of our local optimum.
An easy example of this is the cutting process where the search for a local optimum would lead us to focus our work on the cutting of pieces according to the thickness, so we can cut as many pieces as possible trying to reduce the setup times of the cutting process.
Figure 4 shows the construction strategy based on large blocks and how this decision in the cutting stage affects the next operation (shipbuilder) that is lead to work with the resources available.

Figure 4: Training Session, Traditional Method.

Another characteristic of the traditional manufacturing method is the application of a push system where we try to perform our work in the best possible way without thinking in what is happening in the rest of the activities and producing as long as there is material and no matter how the entire manufacturing process goes downstream. In addition, it will be assumed that the quality of the processes will be evaluated only when the manufacturing process is completed.

4.2. Lean method

As an alternative to this traditional method we propose the possibility of applying a new culture of work in our manufacturing process. To do that, we will no longer focus our activity on the large blocks construction but we will focus on subblocks, proceeding therefore to the use of a reduction of batch size.
In addition, we will focus our activities on the fulfilment of the different requirements of all the stages, always taking into account the state of the workshop (queues, failures…). For this reason, we will have to include the concepts of internal customer and the quality integrated in the fabrication process named also jidoka (Lean concept).
We will also focus our activity on the search for the global optimum against the local as happened in the previous case.
If we focus now on the cutting activity, the difference would be that we would now cut the pieces according to the assembly needs of each subblock regardless of the thickness of the piece (ceiling panels, linings and bulkheads). This lead us to reduce the setup time in the tools which would be addressed for instances, through SMED techniques (a Lean concept).

Figure 5: Training Session, Lean Method.

Figure 5 shows how our activity is now focused on the fabrication of subblocks, also establishing a one-piece flow that all activities must respect.
In summary, the main differences between these two models are:

- Assume a certain versatility of workers.
- Search for global optimum.
- Waste reduction.
- Rational use of resources.
- Reduction of the batch size, from working by blocks to subblocks.

5. RESULSTS OBTAINED FROM THE MODELS

After doing all this work and creating our models with a discrete event software, we will now focus on the results obtained from both models, focusing especially on the fabrication period, customer´s milestones and costs.

5.1. Traditional method evaluation

To analyse the traditional method, it has been proposed the evaluation of different scenarios, assuming as the first hypothesis the impossibility to do overtime, allowing observing the real influence of overtime in this model.
This first hypothesis assumed would prevent us to achieve the markets goals established in terms of time and cost. If we assume the penalties for deviations in time, we would see that this would lead to a total cost of the project of 302.29M€, delivering our ship in week 36.02, something completely unviable.

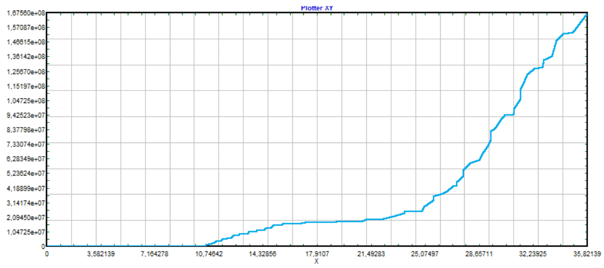


Figure 6: Accumulated Costs by Penalties.

After this situation, it was suggested the possibility of using overtime from an initial moment, assuming the limits established in the reality and commented previously. The results are shown below:

Table 2: Economic Result for the Traditional Method.

	Without overtime at the weekend	With overtime at the weekend
Material cost	80.686.760 €	80.686.760 €
Labour cost (without overtime)	67.404.942 €	67.259.853 €
Extra labour cost	39.069.472 €	27.793.817 €
Penalties cost	1.497.145 €	- €
Total cost	188.658.319 €	175.740.430 €

Table 2 shows our breakdown of costs allowing us to compare both situations, seeing that the main differences will be the amount of overtime required and the shipowner´s penalties but we do not obtain any benefit in any case, having a financial loss of 18.66M€ or 5.74M€ in the best case. As you can see, the best result is obtained by using as many overtime hours as possible, since in this way we would be able to achieve the milestones related to the delivery date of our accommodation vessel (without penalties) and reduce waiting times between activities and therefore those unnecessary costs.
In this case, thanks to the use of this type of DES software, it has been easy to identify the bottlenecks activities in our process (shipbuilder and bricklayer), which has led us to consider what would happen if we decide to increase our staff by doubling the employees available in these activities.

The simulation shows us that this increase in the personnel is not enough to achieve our objectives in terms of cost and time, requiring a reduction of 13% of our operating time to accomplish the requirements related to the fabrication period and delivery date.
In this case, we have significantly improved the workshop operations thanks mainly to the waiting times reduction between the activities which translates into a considerable cost reduction.
However, if we do not have the capacity to reduce our processing time, we would require using overtime to reach our goal.
In spite of this situation, the economic results improve considerably as we showing in the costs shown in the following figure:

Table 3: Economic Result for the Hiring Case.

	Without overtime	With overtime
Material cost	80.686.760 €	80.686.760 €
Labour cost (without overtime)	77.029.647 €	77.023.942 €
Extra labour cost	11.015.715 €	10.798.592 €
Penalties cost	2.859.227 €	- €
Total cost	171.591.349 €	168.509.294 €

With these results we can conclude that the non-use of overtime would mean a loss of 1.59M€, mainly due to the delivery date penalties.
However, the use of overtime would mean a considerable saving, achieving an estimated profit of around 1.5M€ in these two situations.
This case is especially significant because we do not need to use the temporary safety cushion we use through overtime on the weekends, so we will have enough time for temporary and economic contingences.

Table 4: Comparison Between the Use of Overtime.

	Without overtime at the weekend	With overtime at the weekend
Total cost	168.539.428 €	168.509.294 €

5.2. Lean method evaluation
As an alternative method, the use of Lean tools was proposed with the objective of reducing costs and lead-time and assuming the previously established hypothesis.
The simulation allow us to identify an underutilized resource as the painter who, through a brief training and the similarity between guilds, will become bricklayer. In addition, it will be assumed that the welder, considering his certifications, will be able to assist the shipbuilder when he needs it.
The simulation results show a significant cost reduction, mainly thanks to the reduction of the waiting time between activities which translates into a better overall economic result (Table 5), estimating the profit in this case in the 16.18M €.

Table 5: Economic Result for the Lean Method.

	Lean method
Material cost	80.686.760 €
Labour cost (without overtime)	72.632.651 €
Extra labour cost	497.900 €
Penalties cost	- €
Total cost	153.817.311 €

The extra labour cost shown in the Table 5 are due to the variability of the model and the consequent need to correct those deviations that are insignificant comparing them to the traditional method.
In total, a cost reduction of about 14.7M€ with respect to the case previously evaluated was achieved. In other words, we obtain a reduction of 20.1% in the labour costs which representing a reduction of 9.55% in the project cost.
These differences are even more significant if we evaluate the case in which we have the same number of employees, where we achieved a saving of 21.92M€.
In addition, in this case we will need to use a total of 24.9 weeks in our manufacturing process (4.6% time reduction) despite not having to employ overtime, so we will have an important safety cushion.

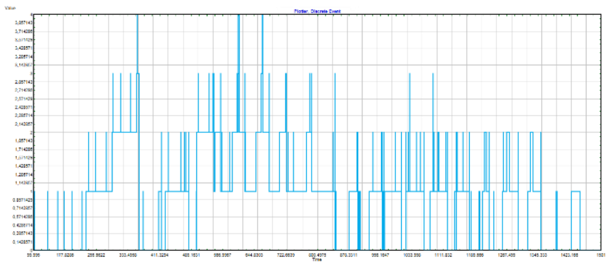


Figure 7: Quantity of Items in Buffers Waiting to be Processed.

In the previous figure, we show the items waiting to be processed in some stage. In this case, as can be observed, we would have a maximum of only four items waiting whereas in the traditional method it would be a total of six. This supposes a considerable reduction of the space required in the shipyard temporarily store blocks.

6. CONCLUSIONS
The use of the DES techniques for our purpose has given us significant benefits because, thanks to them, we have been able to create a realistic and viable Lean manufacturing training tool for all staff, regardless of the starting educational level that will show them the advantages of the process, involving them during their training.
On the one hand, for a higher level of training and with the intention of implementing simulation techniques in the company, they will serve as a starting point or as reference for employees to make their own models or

complete those exposed here. This is undoubtedly the great advantage of the creation of these models which can be modified and evolve according to the progress of the students.

In addition, it would be possible to create some interfaces that allow the modification of some predefined parameter and see how the system changes according to the results obtained by the simulation, considerably reducing the time necessary for the session.

If we try to do this by using the created physical scale model, obviously its realization would need a lot of time and it would not be possible to evaluate as many scenarios as in the simulation, because the limited time and available staff. Therefore, DES not only serve us to build the base of the scale model (planning, costs, etc.), but also this training tool could be complemented by the presentation of various scenarios studied thanks to the simulation, improving this training too.

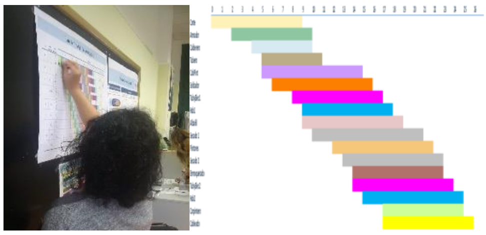


Figure 8: Monitoring the Progress of the Project During the Training Season (left) from a Planning Doing with the DES Simulation (right).

ACKNOWLEDGMENTS

The authors are thankful to Unidad Mixta de Investigación (UMI) Navantia-UDC for its valuable support.

REFERENCES

Aldrich, Clark. 2006. “Learning by Doing: A Compre-Hensive Guide to Simulations, Computer Games and Pedagogy in E-Learning and Other Educational Experiences.” *Educause Quarterly*, no. 3: 69–72.

CC.OO. 2013. “El Sector Naval. Situación Y Perspectivas.” Madrid. http://yoindustria.ccoo.es/comunes/recursos/99906/doc161240_El_sector_naval._Situacion_y_perspectivas.pdf.

Constantino Delago, Lucas, Flávio E.F.H.S Machado, Michael Oliveira De Brito, Gustavo Casarini Landgraf, Marcos De Andrade Schroeder, and Cristiano Torezzan. 2017. “Learning Lean Philosophy through 3D Game-Based Simulation.” In *Proceedings - Winter Simulation Conference*, 3385–92. Washington, DC, USA.

Deshpande, Amit A., and Samuel H. Huang. 2011. “Simulation Games in Engineering Education: A State-of-the-Art Review.” *Computer Applications in Engineering Education* 19 (3): 399–410.

Diehl, Leandro A., Eldon Lehmann, Rodrigo M. Souza, Juliano B. Alves, Roberto Z. Esteves, and Pedro A. Gordan. 2011. “A Serious Game Prototype for Education of Medical Doctors and Students on Insulin Management for Treatment of Diabetes Mellitus.” In *IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH).*, 1–4. Braga, Portugal.

Liane Márcia, Freitas E Silva, de Gois Pinto Marcel, and Anand Subramanian. 2007. “Utilizando O Software Arena Como Ferramenta de Apoio Ao Ensino Em Engenharia de Produção.” *XXVII Encontro Naciona de Engenharia de Produção*, 1–10. http://files.engenhariauniverso.webnode.com/200000119-16204169cb/ENEGEP2007_TR660482_9236.pdf.

Longo, Francesco, Alessandro Chiurco, Roberto Musmanno, and Letizia Nicoletti. 2015. “Operative and Procedural Cooperative Training in Marine Ports.” *Journal of Computational Science* 10. Elsevier B.V.: 97–107.

Padilla, Jose J., Christopher J. Lynch Saikou Y. Diallo Ross J. Gore, Anthony Barraco, Hamdi Kavak, and Bakari Jenkins. 2016. “Using Simulation Games for Teaching and Learning Discrete-Event Simulation.” In *Proceedings of the 2016 Winter Simulation Conference*, 3375–84. Washington, DC, USA.

Page, B, and W Kreutzer. 2006. “A Framework for Web-Based E-Learning of Discrete Event Simulation Concepts.” In *7th Simulation and Visualization (SimVis)*, 195–204. Magdeburg, Germany.

Radziwon, Agnieszka, Arne Bilberg, Marcel Bogers, and Erik Skov Madsen. 2014. “The Smart Factory: Exploring Adaptive and Flexible Manufacturing Solutions.” *Procedia Engineering* 69. Elsevier B.V.: 1184–90.

Raybourn, E. 2014. “A New Paradigm for Serious Games. Transmedia Learning for More Effective Training and Education.” *Journal of Computational Science* 5: 471–81.

Silva, Jose Fernando, Joao Emilio Almeida, Rosaldo J F Rossetti, and Antonio Leca Coelho. 2013. “A Serious Game for EVAcuation Training.” *IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH)*.

Standridge, Charles R. 2000. “Teaching Simulation Using Case Studies.” In *Proceedings of the 2000 Winter Simulation Conference*, 1630–34. Orlando.

Suaz González, Alfredo. 2012. “La Situación de La Construcción Naval En España.”

Turner, Christopher J, Windo Hutabarat, and John Oyekan. 2016. “Discrete Event Simulation and Virtual Reality Use in Industry: New Opportunities and Future Trends.” *IEEE Transactions on Human-Machine Systems* 46 (6): 882–94.

Ushaw, Gary, Janet Eyre, and Graham Morgan. 2017. “A Paradigm for the Development of Serious Games for Health as Benefit Delivery Systems.” In *IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*. Perth, WA,

Australia.
Yildirim, Sule. 2010. “Serious Game Design for Military Training.” In *Games: Design and Research Conference*. Volda, Møre og Romsdal, Norway.
Zee, D.J. Van der, and J. Slomp. 2005. “Simulation and Gaming as a Support Tool for Lean Manufacturing Systems - a Case Example from Industry.” In *Proceedings of the 2005 Winter Simulation Conference*, 2304–13. Orlando, FL, USA.

AUTHORS BIOGRAPHY

JOSÉ ANTONIO MUIÑA DONO holds a university degree in Industrial Technology since 2015 and an MSc in Industrial Engineering since 2017 from the University of Coruña, focusing his activity on the M&S of industrial processes.

ADOLFO LAMAS RODRÍGUEZ graduated from the University of Vigo in 1998. He holds an MSc and a PhD in Industrial Engineering. He combines his research activities in the researching group Grupo Integrado de Ingeniería and his position as a senior engineer and Project Manager in the Spanish leading shipbuilding company Navantia. He is also Associate Professor in the University of A Coruna since 2004 teaching in subjects related to manufacturing, simulation and Lean Manufacturing techniques He is the coordinator of one of the researching lines in the joint venture Navantia-University of Coruña (UMI) related to simulation and optimization models of industrial processes.

ÁNGEL FERNÁNDEZ RODRÍGUEZ graduated from the University of A Coruña in 1994. He holds an MSc and a PhD in Maritime Engineering. He is an Associate Professor in the University of A Coruna since 2010 teaching in subjects related to reliability, processes and lean manufacturing techniques. He combines his research activities in the University of A Coruña and his position as a senior engineer in lean manufacturing techniques applied to shipbuilding in the spanish leading shipbuilding company Navantia.

DAVID CHAS ÁLVAREZ holds an MSc in Industrial Engineering since 2015 and PhD student since 2016. He works as research engineer in the joint venture Navantia-University of Coruña (UMI) and he is mainly involved in the development of simulation and optimization models of industrial processes, especially in models of manufacture wind turbines foundations.

THEORY AND PRACTICE OF INFORMATION FUSION MODELS' QUALITY ESTIMATION AND MODELS' QUALITY CONTROL

Boris Sokolov[(a)], Stanislav Micony[(b)], Alevtina Ziuban[(c)], Vadim Burakov[(d)], Iliya Pimanov[(e)], Dmitri Ivanov[(f)]

[(a),(b) (c),(d),(e)]SPIIRAS— St. Petersburg Institute of Informatics and Automation, Russian Academy of Sciences, 14th line 39, St. Petersburg, 199178, Russia
[(a)]ITMO University— St. Petersburg National Research University of Information Technologies, Mechanics and Optics, 13 Gdanovskaya str., St. Petersburg, 197198, Russia
[(f)]Department of Business Administration, Berlin School of Economics and Law, Badensche Str. 50–51, 10825 Berlin, Germany

[(a)]sokolov_boris@, [(b)]smikoni@mail.ru, [(c)]alvz@yandex.ru, [(d)]burakov@eureca.ru, [(e)]burakov@eureca.ru, [(f)]dmitri.ivanov@mail.ru

ABSTRACT

At the moment, the theory, methods and techniques concerning the application of mathematical models are wide-used. Nevertheless such problems as a problem of multi-criteria models' quality estimation, a problem of analysis and arrangement of models' classes, a problem of justified selection of applied task-oriented models are not well investigated yet. The importance of the considered problem increases when the object of research is described not via a single model, but via a multiple-model complex, consisting of models related to different classes or combined models (for example, analytical-imitating, logical-algebraic, etc). Aforementioned problems are the primary objects of the theory of mathematical models' and multiple-model complexes' quality control. We proposed to implement this theory in the Information Fusion applied area. In the paper software prototype for Information Fusion quality estimation and models' quality control in the sphere of operational river flood forecasting is proposed.

Keywords: information fusion procedures, multi-criteria models' quality estimation, models' quality control, decision support system

1. INTRODUCTION

At the moment, Information Fusion (IF) modeling plays a role of universal instrumentality of knowledge, research and design in different areas of applications. According to JDL ("Joint Directors of Laboratories") model of multi-level information fusion is considered as multi-level process with the following hierarchy (Steinburg, Alan N., Bowman, Christopher L., 1998) : Level 0: Sensor fusion for sub-object assessment, Level 1: Data fusion for object assessment, Level 2: Information fusion for situation assessment, Level 3: Information fusion for impact assessment including for situation prediction, Level 4: Process refinement (In particular, it supposes actionable decision making for resource management), Level 5: Human assessment followed by dynamic re-planning of resources in context of top-level intent and respective requirements in order to manage resources, situation, objects and sensor-level strategy.

Now, the theory, methods and techniques concerning the application of IF mathematical models are wide-used. Nevertheless, such problems as a problem of IF multi-criteria models' quality estimation, a problem of analysis and arrangement of models' classes, a problem of justified selection of applied task-oriented models are not well investigated yet. The importance of the considered problem increases when the object of research is described not via a single model, but via a multiple-model complex, consisting of models related to different classes or combined models (f. e. analytical-imitating, logical-algebraic, etc). Aforementioned problems are the primary objects of the theory of mathematical models' and multiple-model complexes' quality control. The goal of this paper to present some elements of this theory (Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006.,).

2. THE RESULTS OF INFORMATION FUSION INVESTIGATION

Within *the first line of investigations* the following scientific and practical results have been obtained by now.

It was established that the change from an automated processing of measuring information to a computer-aided analysis of received materials involves semantic aspects of data representation in place of syntactic ones. Thus, the information about control objects should rather be regarded as a set of interrelated parameters jointly characterizing objects' technical state than a simple collection of measurements. This provided for a conclusion that the metric-space concepts, typically used in simple monitoring problems, are weak and not suitable for our purposes, hence more general constructions should be used.

It was proved that the parameters of objects' technical states can be described via a system of open sets forming a base of topology. It was assumed that the set of parameters has a topological structure. Thus a system of neighborhoods (meeting the axioms of topological spaces) was established for each element. The notion of a technical state was worked out. By the technical state we meant an abstract collection of data including whole information both about object's current attributes and the state of computations within the monitoring process. This view lets optimize computations in order to receive monitoring results in real time.
The following basic statements were proved: the whole set of technical-state parameters constructed trough the proposed model of knowledge representation is a lattice or a lattice ordered set; if the set of technical states have the greatest element and the least element (defining the initial data and the results correspondingly), then a complete lattice (an algebra over the set) can be formed via a construction of additive and multiplicative lattices; necessary and sufficient conditions for topology base existence were obtained for the set of technical parameters. The last result is very important, as the constructed topology is used for whole description of possible technical states and for planning of states analysis (for construction of computational scheme).
Moreover within the first line of investigations we have been obtained the following the results (Okhtilev M. Yu., Vasiliev I. Ye. 2004, Okhtilev M.Yu. 2001., Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006.): formal description of all possible kinds of controlled states (assessed situation) accounting for their adequacy to actual actions and processes on controlled object caused by application of different mathematical apparatus for various functional objects. Multi-model formalization intends for describe actions and processes on the controlled object; new integrated methods of program synthesis for automatic analysis (AA) of measuring information (MI) about complex technical object (CTO) states were worked out. These methods, as distinct from known ones, give an opportunity of, firstly, interactive intellectual processing of data and knowledge about CTO states for different physical properties (for example, functional parameters, range parameters, signal and code parameters, and integrated parameters) and for different forms of states description without reference to their physical features and, secondly, automatic generation of alternative program schemes for MI analysis according to the objectives of CTO control under the presence of changing environment; new algorithms of automatic synthesis of AA MI programs were proposed for poly-model description of monitoring processes via attribute grammars, discrete dynamic systems, and modified Petri nets. Applying of polytypic models resulted in adequate adaptation of the algorithms to different classes of CTO. Another distinguishing feature of the algorithms lied in application of underdetermined calculation and constraint-driven programming and provided that CTO states could be estimated rather adequately even if some parameters were omitted and the measuring information was incorrect and inaccurate; a general procedure of automatic (computer-aided) synthesis of CTO monitoring programs was developed. This procedure includes the following steps.
The 1st step. Description of conditions and constraints for the problem of AA MI programs synthesis via a special network model connecting input data with goals. An operator (he does not have to be a programmer) uses a special problem-oriented language to execute this step.
The 2nd step. Automatic existence analysis for a solution of AA MI problem that is defined via a formal attribute grammar.
The 3rd step. If the solution exists then the alternative schemes for AA MI programs are generated and implemented in a special operational environment (problem solver of the CTO monitoring system).
The main advantage and substance of the proposed procedure is simple modeling of MI sources (models generation) that can be performed by a non-programming operator in the shortest time and the real-time implementation of the intellectual methods and algorithms of MI processing and analysis for arbitrary structure of the measuring information.
The proposed methods of monitoring automation and modeling let switch from heuristic description of the telemetry analysis to a sequence of well-grounded stages of monitoring program construction and adaptation, from unique skills to unified technologies of software design. These methods are based on a conclusion that a functional description of monitoring process is much less complicated than detailed examination of software realizations. Consecutive specification of software functions is the ground of technologies to be used for creation of monitoring systems. The suggested technology of continuous design process includes such well-known phases as new proposal phase based on special operational environment.
Within the second line of investigations the following scientific and practical results have been obtained by now (Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006., Sokolov B.V. Yusupov R.M. 2002,2004.).
System analysis of the ways and means to formalize and solve the problem of the control over structure dynamics of monitoring system (MS) servicing CTO under changing environment was fulfilled. It was shown that the problems of structure-functional synthesis of monitoring systems and intellectual information technologies as applied to complex technical objects and the problems of CTO structure reconfiguration are a special case of structure-dynamics control problem. Other variants of structure-dynamics control processes in MS are: changing of MS objectives and means of operation; reallocation of functions, tasks, and control algorithms between MS levels; control of MS reserves; transposition of MS elements and subsystems.
The basic concepts and definitions for MS structure-dynamics control (SDC) were introduced. During our

investigations the main phases and steps of a program-construction procedure for optimal structure-dynamics control in MS were proposed. At the first phase forming (generation) of allowable multi-structural macro-states is being performed. In other words, a structure-functional synthesis of a new MS make-up should be fulfilled in accordance with an actual or forecasted situation. Here the first-phase problems come to MS structure-functional synthesis.

At the second phase a single multi-structural macro-state is being selected, and adaptive plans (programs) of MS transition to the selected macro-state are constructed. These plans should specify transition programs, as well as programs of stable MS operation in intermediate multi-structural macro-states. The second phase of program construction is aimed at a solution of multi-level multi-stage optimization problems.

One of the main opportunities of the proposed method of MS SDC program construction is that besides the vector of program control we receive a preferable multi-structural macro-state of MS at final time. This is the state of MS reliable operation in the current (forecasted) situation. The combined methods and algorithms of optimal program construction for structure-dynamics control in centralized and non-centralized modes of MS operation were developed too.

The main combined method was based on joint use of the successive approximations method and the "branch and bounds" method. A theorem characterizing properties of the relaxed problem of MS SDC optimal program construction was proved for a theoretical approval of the proposed method. An example was used to illustrate the main aspects of realization of the proposed combined method.

Algorithms of parametric and structural adaptation for MS SDC models were proposed. The algorithms were based on the methods of fuzzy clusterization, on the methods of hierarchy analysis, and on the methods of a joint use of analytical and simulation models

The SDC application software for structure-dynamics control in complex technical systems was developed too.

Within the third line of investigations the following scientific and practical results have been obtained by now the pilot versions of computer-aided monitoring system (CMS) for CTO states supervision (in space systems and atomics) work in network of IBM/PC-compatible computers; it uses special operational environment (Kalinin, V.N. and Reznikov, B.A., 1987., Okhtilev M. Yu., Vasiliev I. Ye. 2004, Okhtilev M.Yu. 2001., Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006), real-time database management system, multi-window interface, and programming language C/C++.

The prototypes of CMS belong under the class MMI/CACSD/SCADA/MAIS (man-machine interface/ computer-aided control system design/supervisory control and data acquisition/ multi-agent intellectual system).

Our investigation have shown that today a great number of information fusion models, methods, algorithms, and techniques are used for modeling the aforementioned objects and processes in application to the specified types of tasks. This set corresponds to different levels of information fusion related to the considered area of application (levels 0 – 5 of Information Fusion). The most used classes of models are the following ones: deterministic models, stochastic models, uncertain analytical-imitating models and multiple-model complexes, logical-algebraic models, logical-linguistic models and multiple-model complexes, combined classes of models. Nowadays an information fusion models plays the important role of universal instrumentality of knowledge, research and design in different areas of applications. For example, there are many approaches to model describing of information a knowledge fusion in Rome Laboratory (USA): a Bayesian techniques, Knowledge Based approaches, Artificial Neural Systems (Neural Networks), Fuzzy Logic, and Genetic Algorithms. In this case very interesting and important to develop information fusion models' quality estimation and models' quality control theory (Kalinin, V.N. and Reznikov, B.A., 1987., Merkuryeva, G., Merkuryev, Y. and Vanmaele, H., 2011., Peschel M., 1981, 1978., Polyak, Yu.G., 1971., Samarskii, A.A. and Mikhailov, A.P., 2001Norenkov I.P. 1998., Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006., Steinburg, Alan N., Bowman, Christopher L., 1998.).

3. METHODOLOGICAL AND METHODICAL FUNDAMENTALS OF THE INFORMATION FUSION MODELS' QUALITY ESTIMATION AND MODELS' QUALITY CONTROL THEORY

In this theory, we investigate two main research tasks. They are: elaboration of the methodological and methodical fundamentals of information fusion models and multiple-model complexes' qualimetry; development of the computer software prototypes implementing meta-models, methods, and algorithms of multi-criteria quality estimation and control in information fusion models and multiple-model complexes used for information fusion.

From our opinion, the proposed research theory includes the following main directions of the work: elaboration of the basic definitions, principles, and approaches used in the information fusion models' and multiple-model complexes' qualimetry; development of the hierarchy of conceptual models of developing situations, when the participants are the objects, the subjects, and the models being elaborated (used); classification and systematization of information fusion models and multiple-model complexes, determination of the interconnections and mutual associations of different types and kinds of information fusion models; classification and selection of the system of the parameters estimate the quality of the information fusion models and multiple-model complexes;

elaboration of the combined methods for estimation of parameters of quality that are presented by digital and non-digital scales in the information fusion models and multiple-model complexes; elaboration of the methods and algorithms for solving the problem of multi-criteria analysis, arrangement and selection of information fusion models and multiple-model complexes, and quality control of information fusion models; development of object-oriented specification of the software tool according to the developed model of developing situations. Working out detailed object-oriented specifications for the case of formulating and resolving the direct and inverse problems of performance efficiency estimation for MO CS. (levels 2, 3 of Information Fusion); development of the computer software prototype for solving the problems of estimation and quality control of information fusion models and multiple-model complexes based on direct problem of performance efficiency estimation for different types of complex systems (CS) which have artificial and natural background (CS).

The primary goal of our investigation is the elaboration of the theory of information fusion models' quality estimation and quality control, qualimetry of information fusion models and multiple-model complexes of CS. The main objects of the investigation are the characteristics of information fusion models and multiple-model complexes of CS. The information fusion models are here considered as models, developed on the basis of natural and artificial languages. The latter ones include all formal languages. The elaboration of the theory should be started from setting up the appropriate terminology and system of definitions that will be the basis for the further argumentations and conclusions. The basic definitions used in the theory are as follows: developing situations, gnoseological and ontological models' adequacy, certainty, completeness, accuracy, essential correctness, model's utility.

Methodological basic of the research consists of following items: concepts of system analysis and complex modeling; principles of program-aimed and situating control; principles of requisite variety; principles of exterior supplementation and embedding; subjective and objective, and integral approaches to the modeling of complex objects and processes.

The basics of the elaboration of conceptual meta-models of quality estimation and quality control of information fusion models' and multiple-model complexes are twofold (Aframchuk, E.F., et al., 1998., Mesarovich, M. and Takahara, I., 1978, 1975). The object of the modeling is not the actual (designed or abstract) object but the developing situation, that includes as participants the objects and the subjects of the modeling (persons responsible for the accepting the solutions (decision maker), persons responsible for the substantiation of the solution (decision), experts, and persons responsible for the execution of the solutions), and the designed or used models themselves. The specifics of the developing situation is that the space of conditions for all participants is time varying because of different objective, subjective, interior, exterior, and other factors. The process of modeling is here assumed as a control process of developing situation under uncertain conditions, caused by absence of information needed for forming the substantiated decisions. The aforementioned specifics of the conceptual representation of the investigated objects allow to apply for their formal description the mathematical structures being developed in the theory of control and knowledge engineering.

This step should be implemented via classification and systematization of information fusion models and multiple-model complexes. This step will be performed at the first stage of the research on the basis of theory of sets, mathematical structuring, and functor-category description. Particularly, in the case of functor-category descriptions the family of the models related to the same types are represented via a category consisting of the objects that conform to the models (Aframchuk, E.F., et al., 1998., Mesarovich, M. and Takahara, I., 1978, 1975, Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006). The interconnections between the models are characterized by category's morphisms. The interconnections between different families of the models (categories) are characterized by the functors. The main objective of these descriptions is the determination of interconnections and mutual associations between different types and kinds of information fusion models, detection of models' generalized specifics, and structuring of the investigated objects' space as well Figures 1–3 show different variants of classification of developing-situation models, methods, algorithms and technics of information fusion applied area. The analysis of these figures have shown that first focused item of the developing theory is the problem of multi-criteria estimation and quality control of information fusion models and multiple-model complexes. For this the classification of such models' parameters as adequateness, simplicity, accuracy, efficiency of the computer implementation, universality and scalability, multi-functionality and specificity, openness, cost, adaptability, flexibility, and intellectuality should be elaborated. Figure 3 presents as an example the basic elements of the theory of estimation and control of the quality of models developed recently, which are used in integration of data and knowledge (Information Fusion Models) (Okhtilev, Sokolov, Yusupov 2006; Kalinin and Reznikov 1987; Yusupov 1977; Ceany and Raiffa 1981). In our opinion, the development of qualimetry of models should be made in parallel in the main two lines of investigations that closely interact with each other. In the framework of the first line of investigations, the general problems based on the results obtained in solving applied problems of the theory of estimation and control of the model quality.

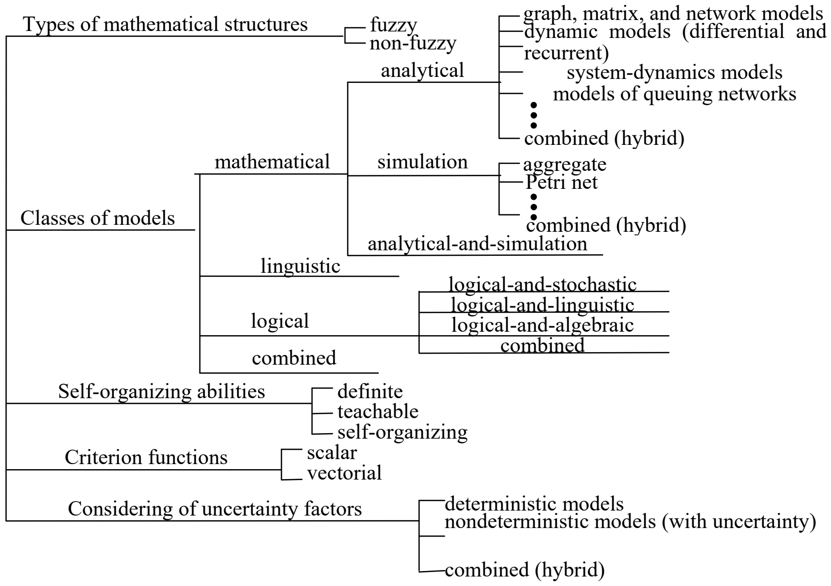


Figure 1: Classification of information fusion models for developing-situations analysis

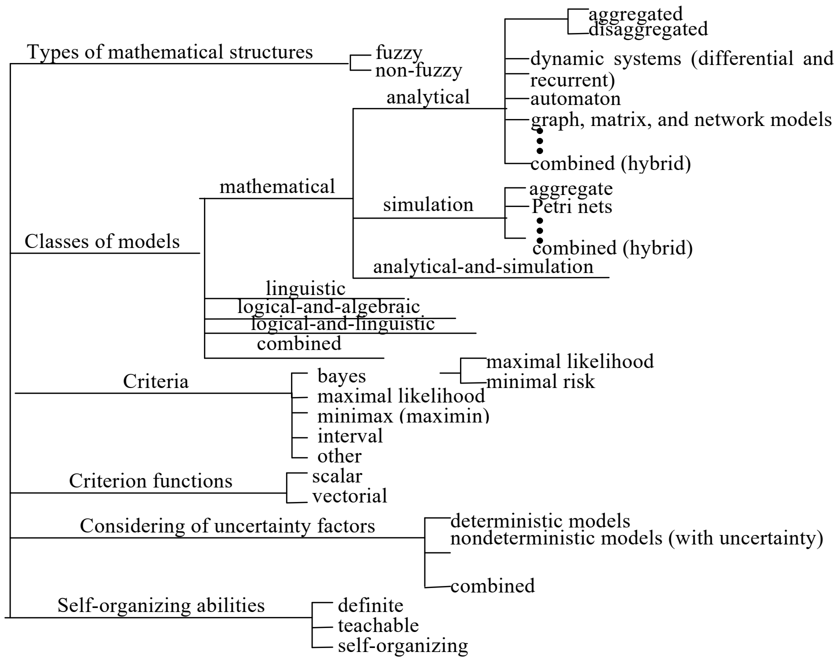


Figure 2: Classification of information fusion models for evaluation of states in developing situations

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

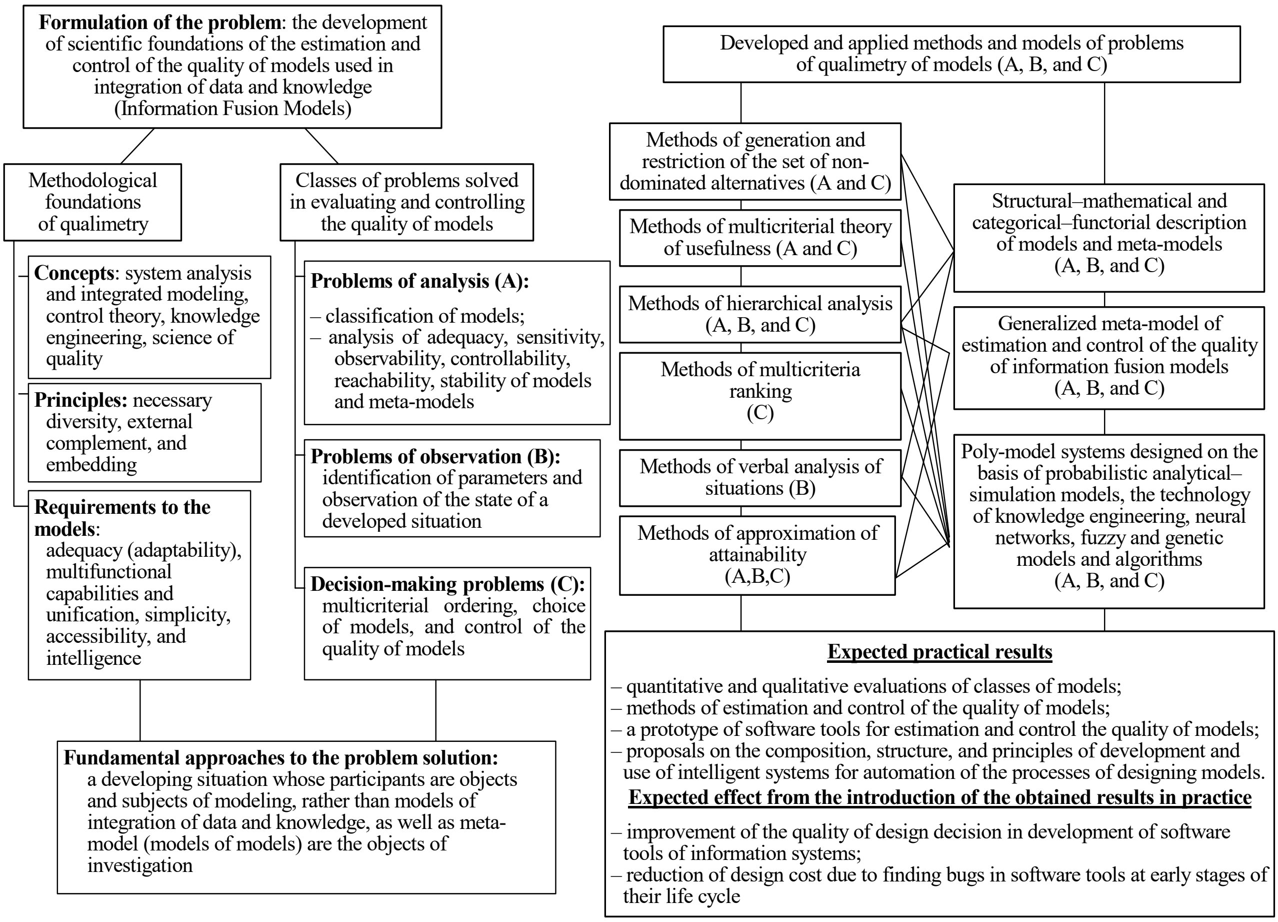


Figure 3: Classification of information fusion models for developing-situations analysis

The second item is the development of combined methods of multi-criteria estimation, analysis, arrangement, and selection of information fusion models. These methods are based on the methods of forming and reduction of set of non-dominated alternatives (Pareto sets), methods of Multi-Attribute Utility Theory, methods of Analytical Hierarchy Processes, methods of arranging of multi-criteria alternatives ELECTRE, methods of verbal analysis of situations (ZAPROS), methods of forming and approximation of domains of accessibility for dynamic systems (Aframchuk, E.F., et al., 1998. Larichev, O.I., 2000, Kalinin, V.N. and Reznikov, B.A., 1987., Val'kman, Yu.R., 1996., Ceany, R.L. and Raiffa, H., 1981)

Formulation and solving of the problem of information fusion models' quality control are performed based on their graph description. This description uses the dynamic alternative system graph that generalizes the known classes of structure models and causal-consequential models. After the additional parameters and structures are added to the information fusion models and multiple-model complexes according to the proposed techniques, the quality control allows to provide the desired (maximal) likelihood of the original object and its model. Moreover, in this case it is possible to adjust the model for the specific applied task The scientific and practical results of the research could be widely used for solving problems in different areas of applications. Developed techniques of multi-criteria information fusion models' quality estimation and control allow performing the well-substantiated and well-analyzed selection and design of the models of developing situations.

4. EXAMPLE OF PROPOSED THEORY IMPLEMENTATION

The design of operational flood forecasting automated systems implies, as was noted above, that no universal model exists to describe flood development processes at river sections that are different in length and configuration. When choosing hydrodynamic models to solve flood forecasting problems, it is advisable to implement the multiple-model approach: depending on the length of the monitored section and the presence of initial information, it is possible to choose between one-dimensional hydrodynamic models for river valley sections of 100–1000 km long and two dimensional models for sections shorter than 100 km with wide beds and floodplains, complex configurations, and various structures located in them.

Combined (hybrid, complex, multiscale) computing by one-dimensional and two-dimensional models can yield a significant effect in effort and fund economy when investigating and monitoring lengthy river areas. Therefore, a principal issue in designing information

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

systems for flood forecasting is to form a mechanism of choosing the most adequate hydrological model for specific conditions.

Within the theory of model qualimetry and polymodel complexes, which the authors of this article are developing (Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006), several technologies of their structural–parametric adaptation have been designed. One of them, applicable to operational flood forecasting, is given in Fig. 4.

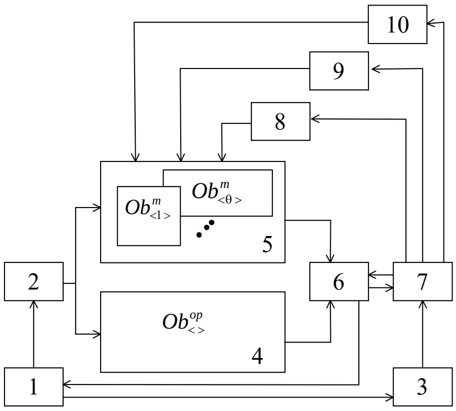


Figure 4: The generalized technique of estimation and control of the quality of models of the first class

In this figure, we take the following notation: 1, for forming the goals of functioning of $Ob^{op}_{<>}$; 2, for determination of input actions; 3, for setting goals of modeling; 4, for the modeled system (objects $Ob^{op}_{<>}$) of the first class; 5, for the model ($Ob^{m}_{<\theta>}$) of the investigated system $Ob^{op}_{<>}$; 6, for the estimation of the quality of a model (poly-model system); 7, for controlling the quality of models; 8, for controlling the parameters of models; 9, for controlling the structures of models; and 10, for changing the concept of model description. This technology was implemented in software prototype for information fusion quality estimation and models' quality control in the sphere of operational river flood forecasting. We are going to speak about it in the next subsection of our paper.

When building operational flood forecasting systems, this technology is implemented through the creation of an intelligent interface, which unites heterogeneous ground and space-based data and expert knowledge to be used later for forecasting. The intelligent interface's main purpose is to choose and to adapt a specific model to simulate water propagation and flood depths on the basis of contextual information (initial data accuracy, flood dynamics, operational result efficiency, etc.). The methods and algorithms that are currently being developed intensively within the theory of evolutionary modeling have been chosen as the scientific basis for the adaptation of hydrological models (http://litsam.ru).

Overall, the creation of an intelligent interface ensures a synergetic effect from the joint processing and use of heterogeneous ground–space data and the implementation of a model complex to assess possible flood areas.

The proposed architecture of intelligent information systems (IIS) is shown on fig. 5. It follows from the above described IIS composition that the IIS architecture should ensure flexible interaction of many components, including the existing and potential software modules based on hydrological models, input data processing modules, forecast result visualization modules, control modules, etc. In this respect, it is desirable to build IISs based on service oriented architecture (SOA). In this case, the proposed IIS should have three components: a control application, a service bus, and an intelligent interface. The control application is designed to implement the logic of the system's operation, which depends on the interrelationship of data acquisition, processing, publishing, and visualization operations. The service bus in this case is a software core that stores information about services available in the system. By services, we understand separate weakly interrelated software modules that realize technological operations, for example, data acquisition from external sensors, processing space images, and publishing a vector map on the geoserver.

The above described operational flood forecasting system was practically tested in several regions of Russia and abroad. The most demonstrative are the results obtained during the 2013 flood of the Daugava River (Latvia) near Daugavpils. In terms of estimating and forecasting flood areas and operational warning of the population and organizations, the best results were obtained for the 12–24 h forecasting interval. The St. Petersburg Institute for Informatics and Automation, RAS (SPIIRAS), and Riga Technical University jointly carried out this experiment within the ESTLATRUS cross border cooperation project [31]. Figure 6 shows the overall structure of this experiment.

5. CONCLUSION

In the present conditions, the development of methodological and methodical fundamentals of models' quality estimation and models' quality control, in other words the development of models' qualimetry theory is highly urgent. This theory belongs to the branch of science, known as quality control, and, in its turn, can be decomposed into many applied theories for estimation of models in different data domains. Thus, Fig. 4 shows main elements of the qualimetry theory for integration of data and knowledge (Information Fusion Models) (Okhtilev M.Yu. 2001, Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M).

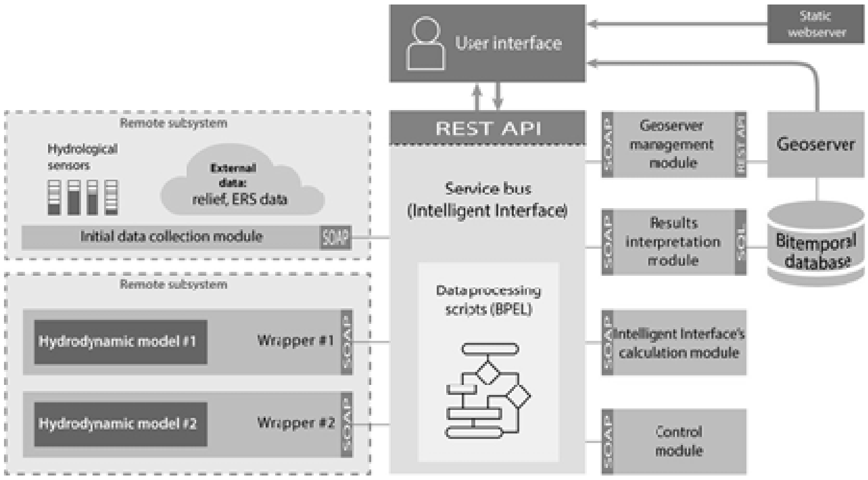


Figure 5: The generalized technique of estimation and control of the quality of models of the first class

It can be seen that the central problems of qualimetry, as applied to Information Fusion Models are, firstly, forming of qualitative and quantitative measures for estimation of models' characteristics, and, secondly, development of combined methods for multi-criterion estimation, analysis, classification, and selection of semiotic models (Steinburg, Alan N., Bowman, Christopher L., 1998, Kalinin, V.N. and Reznikov, B.A., 1987, Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006.). Figure 3 shows, as an example, methods and meta-models to be applied to estimation of the considered models.

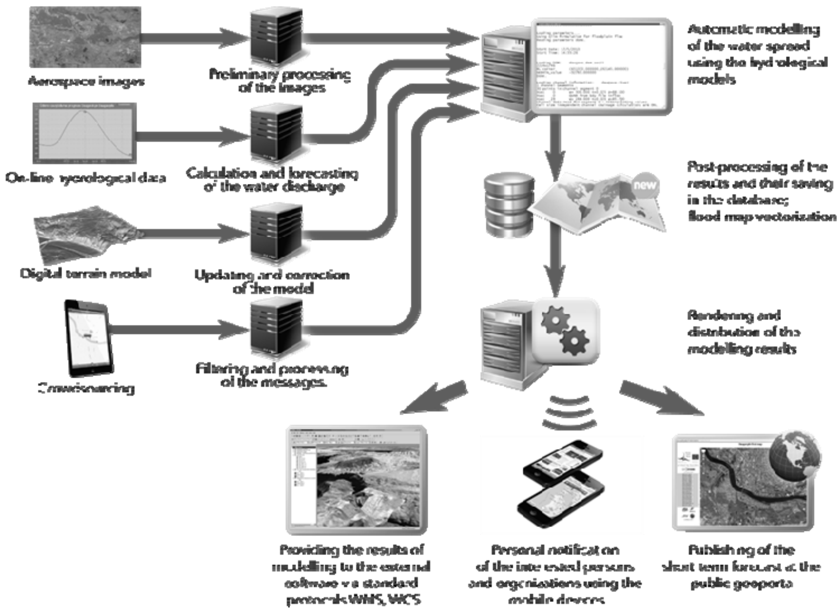


Figure 6: Scheme of the experiment for testing the operational flood forecasting system

From our point of view, models' qualimetry should be developed in two main mutually enriching directions. The first direction includes general matters of qualimetry and is based on the results of the second direction, that is the development of applied qualimetry theories for models' estimation and quality control in different data domains.

ACKNOWLEDGMENTS

The research described in this paper is partially supported by the Russian Foundation for Basic Research (grants 15-07-08391, 15-08-08459, 16-07-00779, 16-08-00510, 16-08-01277, 16-29-09482-ofi-i, 17-08-00797, 17-06-00108, 17-01-00139, 17-20-01214), grant 074-U01 (ITMO University), Russian Science Foundation (project № 17-11-01254), project 6.1.1 (Peter the Great St.Petersburg Politechnic University) supported by Government of Russian Federation, Program STC of Union State "Monitoring-SG" (project 1.4.1-1), state order of the Ministry of Education and Science of the Russian Federation №2.3135.2017/4.6, state research 0073–2014–0009, 0073–2015–0007, International project ERASMUS +, Capacity building in higher education, № 73751-EPP-1-2016-1-DE-EPPKA2-CBHE-JP, Innovative teaching and learning strategies in open modelling and

simulation environment for student-centered engineering education.

REFERENCES

Aframchuk, E.F., Vavilov, A.A., Emel'yanov S.V., et al., 1998. Technology of System Modeling. Emel'yanov S.V., ed. Moscow: Mashinostroenie. [in Russian].

Andrianov, Yu.M. and Subetto, A.I., 1990. Qualimetry in Engineering. Leningrad: Mashinostroenie. [in Russian].

Azgal'dov, G.G., 1982. The theory and Practice of Estimation of the Quality of Products: Foundations of Qualimetry. Moscow: Ekonomika. [in Russian].

Ceany, R.L. and Raiffa, H., 1981. Decision Making under Many Criteria: Preferences and Fillings. Moscow: Radio i Svyaz'. [in Russian].

International Standards ISO 9000 and 10000 for the Quality System: Versions of 1994, 1995. Moscow: Standarty. [in Russian].

Ivanov, D.A. and Sokolov B.V., 2010. Adaptive supply chain management. Berlin: Springer.

Ivanov, D.A., Sokolov, B.V., and Kaeschel, J., 2010. A multi-structural framework for adaptive supply chain planning and operation with structure dynamics considerations. European Journal of Operation Research, 200(2), 409-420.

Kalinin, V.N. and Reznikov, B.A., 1987. Systems Theory and Control Theory (Structural--Mathematical Approach). Leningrad: VIKI. [in Russian].

Krishans Z., Mutule A., Merkuryev Y. and Oleinikova I., 2011. Dynamic Management of Sustainable Development. Methods for Large Technical Systems. London: Springer-Verlag.

Larichev, O.I., 2000. The Theory and Methods of Decision Making. Moscow: Logos. [in Russian].

Lipaev, V.V., 2001. Estimation of the Quality of Products. Moscow: ERIS. [in Russian].

Merkuryeva, G., Merkuryev, Y. and Vanmaele, H., 2011. Simulation-Based Planning and Optimization in Multi-Echelon Supply Chains. Simulation: Transactions of The Society for Modeling and Simulation International. 8 (87), 698–713.

Mesarovich, M. and Takahara, I., 1978, 1975. General Systems Theory: Mathematical Foundations. Moscow: Mir, Moscow, 1978. New York: Academic.

Nemirovskii, A.S. and Yudin, D.B., 1979. The Complexity of Problems and Efficiency of Optimization Methods. Moscow: Nauka. [in Russian].

Norenkov I.P. The approaches to designing of automation systems //Information Technology (1998) Vol. 2. 2-9.

Okhtilev M. Yu., Vasiliev I. Ye. The data flow and distributed calculations intelligence information technology for decision support embedded system in real time // 16th IFAC Symposium on Automatic Control in Aerospace (2004), Saint-Petersburg, Russia, Vol. 2 235-239.

Okhtilev M.Yu. Specifics of technology for development of special computer-aided systems analizing information measured in real-time //Automatic Control and Computer Science, Allerton Press Inc., New York, Vol. 39, # 6 (2001).

Okhtilev, M.Y., Sokolov, B.V., Yusupov, R.M., 2006. Intelligent technologies of complex technical objects monitoring and structure dynamics control. Moscow: Nauka. [in Russian].

Pavlovskii, Yu.A., 2000. Simulation Models and Systems. Moscow: Fazis. [in Russian].

Peschel M., 1981, 1978. Modellbilding fur Signale und Systeme. Moscow: Mir. Berlin: Technik.

Polyak, Yu.G., 1971. Probabilistic Modeling on Computers. Moscow: Sovetskoe Radio, Moscow. [in Russian].

Samarskii, A.A. and Mikhailov, A.P., 2001. Mathematical Simulation: Ideas, Methods, and Examples. Moscow: Fizmatlit. [in Russian].

Sethi, S.P. and Thompson, G.L., 2006. Optimal control theory: applications to management science and economics. 2nd ed. Berlin: Springer.

Skurikhin, V.I., Zabrodskii, V.A. and Kopeichenko, Yu.V., 1989. Adaptive Control systems for Manufacturing. Moscow: Mashinostroenie. [in Russian].

Sokolov B.V. Yusupov R.M. Conceptual Foundations of Quality Estimation and Analysis for Models and Multiple-Model Systems // Journal of Computer and System Sciences International (2004) #6, 5-14.

Sokolov B.V., Yusupov R.M. Complex Simulation of Automated Control System of Navigation Spacecraft Operation// Journal of Automation and Information Sciences (2002), Vol. 34, #10, 19 –30.

Steinburg, Alan N., Bowman, Christopher L., White, Franklin E. Revisions to the JDL Data Fusion Model, presented at the Joint NATO/IRIS Conference, Quebec (1998).

Tyatyushkin A.I., 2003. Many-Method Technique for Computing Optimal Control. Izv. Ross. Akad. Nauk, Teor. Sist. Upr., 3 [Comp. Syst. Sci. 42 (3), 371 (2003)].

Val'kman, Yu.R., 1996. The Problem of "Alienating" Models of Studied Objects from Their Creators in Complex Object Design, Izv. Ross. Akad. Nauk, Teor. Sist. Upr., 3. [Comp. Syst. Sci. 35 (3), 487 (1996)].

Yusupov, R. M., 1977. Elements of the Theory of Testing and Control of Industrial System. Moscow: Energiya. [in Russian].

Yusupov, R.M. and Zabolotskii V.P., 2000. Scientific and Methodological Aspects of Informatization. St. Petersburg: Nauka. [in Russian].

AUTHORS BIOGRAPHY

BORIS 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 lead: optimal control theory for structure dynamics of complex systems. Re-search 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 multicriteria. He is the author and co-author of 9 books on systems and control theory and of more than 450 scientific papers. Professor B. Sokolov supervised more over 90 research and engineering projects. Web-page can be found at http://litsam.ru.

MIKONI STANISLAV — Ph.D., Dr. Sci., professor, professor of mathematics and modeling department, Petersburg State Transport University, leading researcher of information technologies in the system analysis and modeling laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS. Research interests: system analyses, decision making, intellect technologies. He is the (co)-author of more than 254.

ZUBAN ALEVTINA — PhD, associate professor; senior researcher of St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: innovations, computer networks, information technologies, database management systems. He is the (co)-author of more than 40.

BURAKOV VADIM — Ph.D., Dr. Sci., associate professor, leading researcher of information technologies in system analysis and modeling laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: technology of quality management software tools, software engineering, evaluation, and quality improvement programs. He is the (co)-author of more than 160.

ILIYA PIMANOV — Post graduate student, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: geoinformatics and geographical cartography, applied research in mathematical modelling, optimal control theory. Author of 15 publication.

DMITRY IVANOV — Prof. Dr. habil. Dr., professor for Supply Chain Management at Berlin School of Economics and Law (BSELЭ) His *research* explores supply chain structure dynamics and control, with an emphasis on global supply chain design with disruption consideration,distribution planning, and dynamic (re)-scheduling. He is (co)-author of structure dynamics control method for supply chain management. He applies mathematical programming, simulation and control theoretic methods. He is the (co)-author of more than 250 publications

AIR TRAFFIC SIMULATOR FOR PASSIVE ADS-B SURVEILLANCE SYSTEM

Jiří Kratochvíl[(a)], Karel Šimerda[(b)]

[(a),(b)]Faculty of Electrical Engineering and Informatics, University of Pardubice

[(a)]jiri.kratochvil@student.upce.cz, [(b)]karel.simerda@upce.cz

ABSTRACT
This contribution presents a concept of a hybrid air traffic simulator which will be used for ADS-B message validation algorithm testing. These validation algorithms will be used in a passive ADS-B surveillance system. In real life, it is not possible to experiment with data communication disruption of an air traffic by injecting spoof targets. It could lead to endangerment of people in aircrafts and even whole air traffic control. It is thus necessary to realize such attacks on data communication only in a simulated environment. Before attempting to create such environment, appropriate and thorough specification is needed. Simulation of an aircraft transponder can be very complex and may not entirely correspond to the behavior of real transponders. Because of that, the simulator will be using real captured data from ground stations in addition to simulated airborne targets.

Keywords: ADS-B communication, validation, simulator

1. INTRODUCTION

ADS-B (Automatic Dependent Surveillance – Broadcast) is a technology for unidirectional data transmission from an aircraft. It is a broadcast, which means that one aircraft is transmitting while an arbitrary number of receivers are receiving signal. Transmitters are not synchronized with each other and multiple aircrafts can transmit at once. Emission of ADS-B messages is automatic and is not dependent on any input from outside of aircraft. This is unlike other surveillance systems, e.g., primary or secondary surveillance radars, where radars placed on the ground are actively participating in a surveillance system's function. The ADS-B Airborne Position Message broadcast rate is 2 messages per second, more specifically the time interval between two consecutive Airborne Position Messages from the same aircraft is 0.4 – 0.6 of a second (EUROCAE 2009). Transmitted messages lack any encryption, so anyone can eavesdrop on the communication. The format of ADS-B messages is public and described in document from EUROCAE (European Organisation for Civil Aviation Equipment). This document contains Minimum Operational Performance Standards of airborne equipment for ADS-B (EUROCAE 2009). The ADS-B format description can be found on the internet for free (Sun 2017). Moreover, there are open source tools and libraries for ADS-B message decoding, such as *gr-air-modes* library for *GNU Radio* tool, which makes message decoding easy even for non-professionals.

1.1. Attacks on ADS-B

Due to many, sometimes even free, materials describing the ADS-B format, carrying out a passive attack on the ADS-B system is very easy. It is simply eavesdropping on broadcasted messages from aircrafts. That alone does not pose any threat to the air traffic, but this type of attack is usually the first step of an active attack (Schäfer et al. 2013).
Active attack means that adversary send messages. This can influence the ADS-B system in several ways from creating new spoof targets up to deletion of messages from real aircrafts. All of these attacks are composed from one of three basic actions or their combination. Basic actions are message insertion, message deletion and message modification (Schäfer et al. 2013). In addition to absence of encryption, ADS-B does not authenticate the sender, and thus identity of this transmitter cannot be verified from the received message only. To overcome this shortcoming, use of another technology, such as passive locator, multilateration or data fusion is needed. However, each one of these technologies contribute to greater complexity of the system for message reception and processing. In addition to that, it is necessary to use algorithms that can work with data from these multiple systems and thus validate received messages. The simulator designed in this paper will be used mainly for testing such algorithms for passive surveillance systems.

2. AIR TRAFFIC SIMULATION

The simulator designed in this paper is not a general air traffic simulator, although there are some shared ideas and concepts. It is of great importance to realistically represent target trajectories through the airspace. For example, an airborne target cannot suddenly change the direction of its motion to a perpendicular direction and should fly at a speed in some given range (greater than stall speed and lower than maximal speed). In addition to that, the radius of aircraft turns should be appropriate to its speed.
This simulator is focused on an air traffic simulation from the perspective of the passive locator and ADS-B

communication. Passive locator is an ADS-B message receiver with an ability to determine the direction from which an ADS-B message was received. The emulator of ADS-B transponder will be created, which will then periodically broadcast messages for each airborne target. Transponder is a device on board of an aircraft. This device emits a coded signal with information about the aircraft.

2.1. Airborne Target Trajectory

The question arises in the scope of the designed simulator as how to effectively work with airborne target trajectories. It is necessary to choose appropriate representation of trajectories that will enable both a programmatical generation and manual designing, depending on user's needs. There are four options for such representation:

- List of points in 3-D space.
- Curve in 3-D space.
- Waypoints.
- Aircraft vectoring.

Representing trajectory as a time-sorted list of points in 3-D space is advantageous form the viewpoint of programmatical generation of trajectories, but it is not appropriate when considering manual designing and memory efficiency. Because the simulator will be using recorded data from a real air traffic, this representation may be the most convenient one for this usage.

Better than the list of points in 3-D space, is to use a curve, which will describe required trajectory. Then it is enough for a program or for a user to set control points, which will determine the actual shape of the curve. As the most appropriate type of curve for aircraft trajectory representation seems to be B-spline curve. This curve belongs to the C2 differentiability class. Curves from the C2 differentiability class have continuous derivatives up to and including the second order (Delahaye et al. 2014). From the physical viewpoint, this means that an airborne target cannot abruptly change acceleration during its movement along the trajectory.

A very simple way how to design a trajectory by hand is a usage of waypoints. These waypoints are points in space (in this case it is three-dimensional space), which determine the shape of the aircraft's trajectory. Two types of waypoints are used in practice: fly-by waypoints and fly-over waypoints. A fly-by waypoint specifies that an aircraft should start a turn to change its current course to the next course before reaching this point. On the other hand, a fly-over waypoint specifies that an aircraft should fly through this point first before it starts a turn to change its course (Federal Aviation Administration 2015). Depiction of these types of waypoints is in the Figure 1.

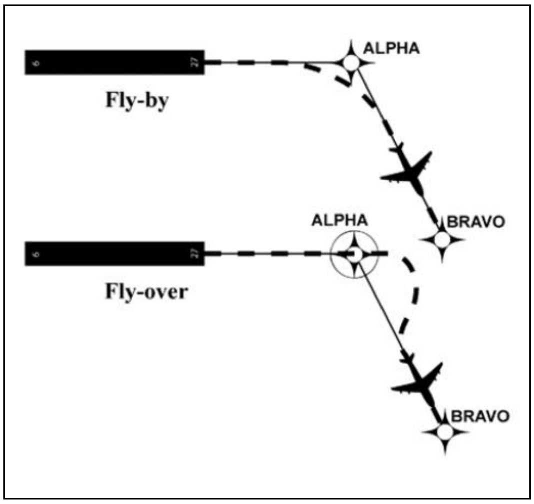


Figure 1: Types of Waypoints

The last mentioned option how to move an airborne target along a trajectory is to completely leave out the predefined trajectory and change aircraft's position using a user-given scenario. That corresponds to an aircraft vectoring; controlling aircraft's trajectory based on instructions for the aircraft's pilot, received from an air traffic control. Such instructions can have the form of this excerpt:

1. Fly 100 km at constant altitude without change of heading.
2. Perform +3° turn.
3. Fly 50 km at constant altitude without change of heading.
4. Descend at 3° angle for 10 km.

The main objective of the simulator is not a planning of efficient trajectories, but planning of such trajectories that would be indistinguishable from trajectories of real aircrafts, given that simulated airborne target should represent a real aircraft. Another case is spoof targets. Their trajectory can be realistic, but does not have to be. That depends on how much the attacker is sophisticated.

2.2. Target's Movement Along the Trajectory

When the trajectory is created, the target must fly along that trajectory during the simulation run. Its flight parameters need to be realistically represented, which is especially the case of velocity. To achieve this goal, it is possible to use the flight dynamics model, which represents an aircraft's behavior under different forces such as thrust, lift and other. Then there are multiple possibilities how to let aircraft fly along a specified trajectory. All of them are basically different forms of autopilot. Another, convenient way of representing a realistic movement of an aircraft is usage of recorded data from a real air traffic. This will obviously lead to the realistic representation of airborne targets.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

2.2.1. Computation of Model Inputs

Computation of model inputs is performed as a computation of appropriate equations, which describe the behavior of the aircraft. This model is represented as a system of ordinary differential equations. Through appropriate equivalent modifications of these equations, new equations are made where independent variables are required changes of trajectory. The result is the control input for given model of the aircraft (Quanbeck 1982). This way of controlling aircraft models is rather complicated and does not reckon with wind forces and dynamically changing environment, for example. Another and simpler solution is to use a system which will be able to pilot an aircraft model. For this purpose, a PSD controller or neural network could be used.

3. SIMULATOR DESIGN

In this paper, a design of the simulator for ADS-B message validation algorithm testing is described. This simulator provides a means for simulation of realistic air traffic on an appropriate level of abstraction from the viewpoint of passive ADS-B surveillance systems. It can also work with spoof targets. Since there are multiple coordinate systems used in the simulation and surveillance systems, it is an important feature of the simulator that it can convert coordinates between these coordinate systems. Simulation is performed in a 3-D cartesian coordinate system. However, positional data received from airborne targets contained in the ADS-B messages are in WGS-84 geodetic coordinate system (EUROCAE 2009). Moreover, passive locator reports only the angle at which the ADS-B message was received.

3.1. Simulator Requirements

Main simulator requirements are these:

1. Simulator must realistically represent behavior of airborne targets with emphasis on the following attributes:
   - Movement in airspace must take place with realistic parameters, such as smoothness of movement along trajectory (no abrupt changes), appropriate radii of turns, etc.
   - Aircrafts' speed must be appropriate at all points of trajectory.
   - Parameters of received ADS-B message must correspond to transmission from given position in 3-D space, where was transmitter placed at the time.
   - Appropriate representation of an aircraft transponder must be used. It will send messages in correct intervals, it will contaminate position measurement by error with given statistical properties, etc.
2. Simulator must permit simulation of spoof targets which emit messages with parameters that are not in accordance with a simulated situation.
3. Simulator must be able to generate appropriate random variables with realistic parameters.
4. For the performance testing of validation algorithms, the simulator must be able to simulate up to 400 targets at once.
5. Simulator must be able to convert position information between requested coordinate systems.

3.2. Sources of Nondeterminism

Several random variables or random processes will be used in the simulator. It is necessary to retrieve statistical parameters from the real modelled system for each one of them, and then apply them in the simulator accordingly. These sources of nondeterminism are:

- Occurrence of an observational error in airborne target position measurement by INS/GNSS.
- Time delay between performing position measurement and according ADS-B message emission.
- Occurrence of an observational error in passive locator measurement.
- Random errors during message transmission which leads to message dropping by receiver.

INS/GNSS position measurement is contaminated by an error depending on the source of position information. INS/GNSS is a term for systems that measure spatial position. INS (Inertial Navigation System) uses information about aircraft's rotation and acceleration from IMU (Inertial Measurement Unit), and thus can compute a new position relative to the last position information. GNSS (Global Navigation Satellite System) is based on different principles. For a position measurement, it uses satellites orbiting around the Earth. GNSS is thus an umbrella term for GPS, GLONASS, Galileo and other satellite-based systems.

For modelling of a GPS error, the Rayleigh probability distribution is commonly used. But individual measurements are, rather, realizations of random variables in a Gauss-Markov random process, because observational errors are time-correlated (Mohleji and Wang 2010).

Another source of a nondeterminism in the simulator is a delay between the time when aircraft's position is measured and the time when message with this position is emitted. If this delay is too long, position information becomes outdated since the aircraft has meantime traveled a non-negligible distance.

If this is not taken into account when creating validation algorithms with help from this simulator, it could lead to marking real aircraft as spoof target because of noncorresponding information from ADS-B messages and computed position from the passive locator or multilateration system.

Errors are also present in passive locator measurements. A passive locator measures direction from which the message from an aircraft or from an adversary was received. It is thus an angular error. Parameters of this random process are unknown. Simulation experiments will verify whether designed message validation algorithms are applicable for processing of information from a passive locator with a specific measurement error. This measurement error will be the subject of simulation experiments too. Which parameters should a passive locator have for practical use, will be tested.

In practice, it happens that received ADS-B message is damaged and cannot be repaired. If the number of such messages is low (in the order of units), the performance of the whole system will not significantly degrade. Performance evaluation of the simulated system with a different probability of the message dropping will be a subject of simulation experiments too.

3.3. Description of Design

The simulator will be based on a simulation core, using periodic activity sampling, which is a technique of continuous simulation. Trajectories of all airborne targets are recomputed with a very short period of time (in order of fractions of a second). But since there are also discrete activities present in the simulator, simulation will be discrete-continuous.

3.3.1. Entities

Four types of entities are present in simulator:

- *Airborne Target*
- *Spoof Target*
- *ADS-B Message*
- *Ground Receiver*

Airborne Targets are mobile temporary entities, whose position information are contained in ADS-B messages. They are representation of real aircrafts inside the simulator, whose movement along trajectories is modelled. *Airborne Targets* use a realistic model of movement.

Spoof Targets are temporary entities, which represent spoof targets created by some attacker. Their positions are also contained in ADS-B messages, which are broadcasted to other receivers. Movement of these *Spoof Targets* can be either realistic in the case of sophisticated attacker or simplified and not very realistic otherwise.

ADS-B Messages are endogenous temporary entities, which carry information about airborne targets or spoof targets. In the simulator, they move from the originating transmitter, which can be aircraft or adversary, to the corresponding receiver. Multiple message entities are created for each transmission. The number of entities is determined by the number of receivers, because one message is created for every pair of current transmitter and receiver. Duration of a message entity movement through space depends on the distance between the transmitter and the receiver.

Ground Receivers are endogenous permanent entities at fixed coordinates on the ground. Their purpose is to receive ADS-B messages and register times of their reception. If the receiver is also a passive locator, it will also determine the direction from where the message was sent (with appropriate measurement error).

3.3.2. Activities

Important activities present in the simulation model are:

- *Aircraft's position measurement by INS/GNSS system.*
- *Message assembly in aircraft's transponder.*
- *Message transmission from transmitter to receiver.*
- *Aircraft flight.*

Aircraft's position measurement by INS/GNSS system is a discrete activity. Measurement of the position has a non-zero duration. For example, GNSS systems typically output position information at the rate of 1 second, so the duration will be 1 s. At the end of this activity, event *Transponder position update* is scheduled. It is not an update of the aircraft's actual position in the simulated space. It is an update of the measured position inside the transponder in the same way as in a real aircraft.

Message assembly in aircraft's transponder is a discrete activity, which represents assembly of an ADS-B Airborne Position Message with a duration from the [400,600] ms interval. This duration variation leads to a random delay between the measurement of an aircraft's position and message emitting; which is implicitly specified by a delay between the last occurrence of event *Transponder position update* and next event *ADS-B message emission*. Activities *Aircraft's position measurement by INS/GNSS system* and *Message transmission from transmitter to receiver* are running in parallel, but independent of each other.

The position of an aircraft is encoded in the CPR format. This reduces the number of bits needed for encoding of an airborne target position from 45 bits to 35 bits. Outcome of this is that there are two types of messages – odd CPR format position messages and even CPR format position messages. Aircraft is then sending in turns even format and odd format messages. To be able to determine the global position of an aircraft unambiguously, one must receive both odd and even CPR format position messages within 10 seconds. It is specified that 2 messages should be transmitted in 1 s (EUROCAE 2009). In the best case, both (one odd and one even) messages should be received during 1 s.

At the end of a *Message assembly in aircraft's transponder* activity, event *ADS-B message emission* is scheduled. When the simulator engine is processing this event, next activity *Message assembly in aircraft's transponder* is started and so are activities *Message transmission from transmitter to receiver*, which means that transmission from transmitter to all of receivers in range is started. As mentioned before, it is multiple activities, one for each pair of transmitter – receiver.

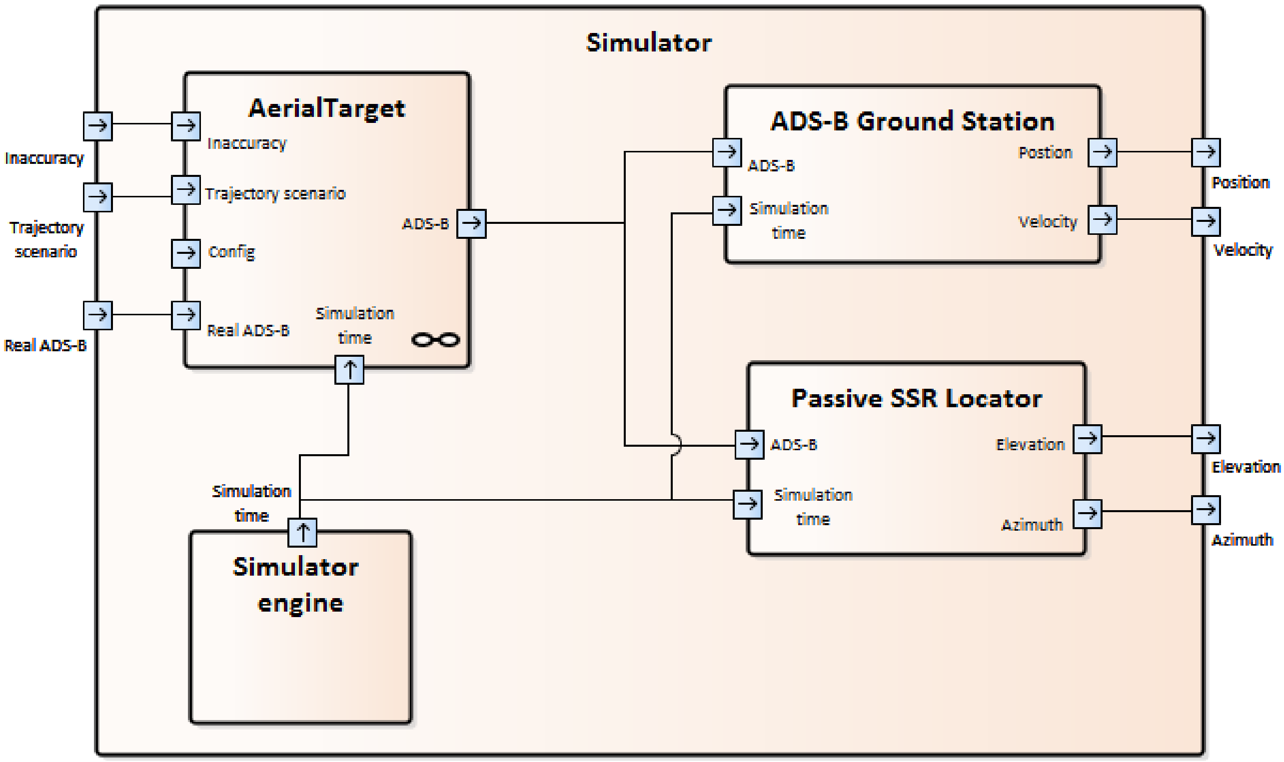


Figure 2: Modules of the Simulator

Message transmission from transmitter to receiver is a discrete activity. Duration of this activity and thus the duration of the message transmission is computed from the propagation speed of an electromagnetic signal through the airspace and the distance of the appropriate receiver from the transmitter. At the end of this activity, event *Message reception at ground station* is scheduled. *Aircraft flight* is a continuous activity, which represents a flight of an aircraft. It is periodically sampled by the simulator engine.

3.3.3. Events

In the scope of the simulator, these events are occurring:

- *Transponder position update*
  - This event occurs when the GNSS/INS module finishes measurement of a position. Current position is updated in the Transponder module.
- *ADS-B message emission*
  - This event occurs when the Transponder module finishes message assembly. Transfer to all receivers in range begins.
- *Message reception at ground station*
  - This event occurs upon message reception at the receiver. Receiver then reacts appropriately to the module type (ADS-B or passive locator). This behavior is described in the next chapter.

4. SIMULATOR ARCHITECTURE

From the viewpoint of the simulator creation and possible extension, it is convenient to create multiple standalone modules from which the simulator will comprise. Depiction of those modules and their interconnection is shown in the Figure 2. Modules themselves are described in next chapters together with their crucial outputs and inputs.

4.1. Simulator engine

The *Simulator engine* stands in the core of the whole simulator. It performs activity scanning, event scheduling and subsequent event processing. It initiates and coordinates functionality of individual modules. The engine in this simulator is based on the activity scanning method. The simulator engine increments the simulation time periodically with sufficiently small step Δt. The size of this step is dependent on the speed of the fastest simulated target, see chapter 3.2. In the Figure 3 can be seen what events occur in the simulator and at what time. These events are represented by boxes.

- PU is a pseudo-event, which occurs periodically with fixed time step during the continuous activity *Aircraft flight*. It is not an actual event in the simulation. It is the update of the aircrafts' position. In the picture, the time step is 0.1 s. However, concrete value is a matter of simulator's configuration. This time step is the Δt with which the simulator engine advances simulation time.

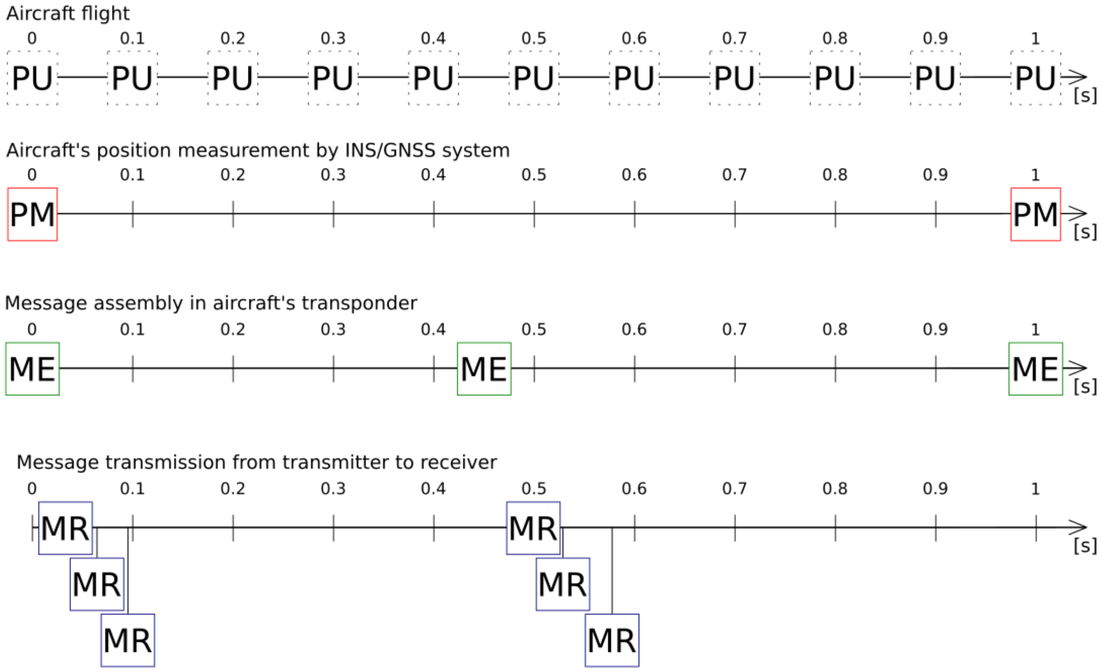


Figure 3: Events Occurence

- PM is the *Transponder position update* event, which occurs at the end of the *Aircraft's position measurement by INS/GNSS system* activity.
- ME is the *ADS-B message emission* event, which occurs at the end of the *Message assembly in aircraft's transponder*.
- MR is the *Message reception at ground station* event, which occurs at the end of the *Message transmission from transmitter to receiver* activity.

Due to the fact that the simulator engine is based on the activity scanning method, some events need more complex processing. As can be seen in the Figure 3, positions of all airborne targets in the simulated airspace are periodically updated at fixed time intervals. But some events can occur between two position updates and are dependent on a precise position information. It is therefore needed to temporarily compute current position at the given time of such event. But this updated position should be visible only for appropriate event, so the computation of the new position does not interfere with periodic updates of aircrafts' positions, which should be visible globally inside the simulator. This is true for events *Transponder position update* and *ADS-B message emission*. Both needs as precise information about current position as possible. Otherwise, if for example, an event *ADS-B message emission* is processed and uses current aircraft's position without computed correction, retrieved position information will not correspond to the position at the time of the event. It will be outdated.

- Input:
  - Events, which should be scheduled at some time in the future.
- Output:
  - Simulation time for modules in the scope of the simulator.

4.2. Trajectory module

This module deals with a representation of a target on its trajectory. In this simulator the module, which can load a user scenario, is realized. This scenario is in the form of instructions for a target movement and corresponds to the aircraft vectoring, mentioned in the chapter 2.1. More trajectory modules, which could for example work with 3D curves, are planned.

- Input:
  - Parameters for the specific trajectory. Currently, it means user scenario with vectoring instructions.
  - Simulation time from the *Simulator engine*.
- Output:
  - Precise target position at the current simulation time.

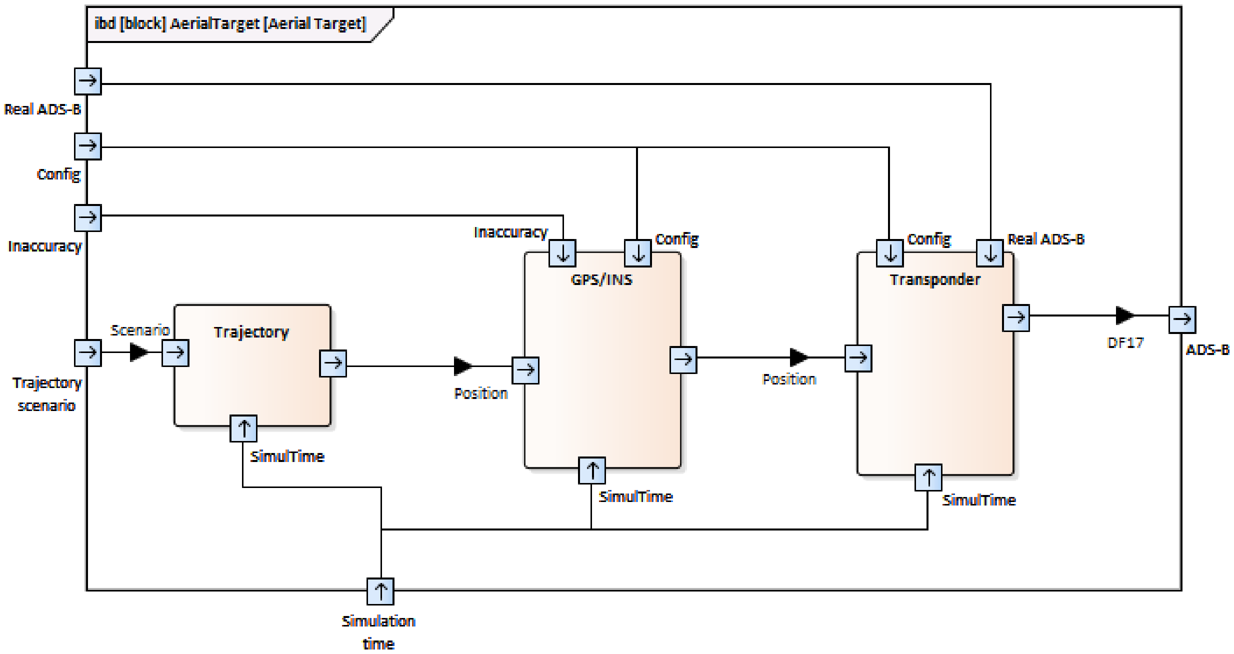


Figure 4: Inner Structure of the Aerial Module

4.3. GNSS/INS module

The *GNSS/INS module* generates measurement errors of an aircraft's GNSS/INS system. This error is generated in accordance with a user-chosen NAC. NAC stands for Navigation Accuracy Category and specifies precision of the positional information based on the type of position measurement source (EUROCAE 2009):

- Input:
  - Inaccuracy is chosen by user.
  - Chosen configuration of the position data source.
  - Position information from *Trajectory module*.
  - Simulation time from the *Simulation engine*.
- Output:
  - Target position with a measurement error added.

4.4. Transponder module

The *Transponder module* assembles and transmits messages from targets as mentioned in chapter 3.3.2 in the *Message assembly in aircraft's transponder* activity's description. Transponder periodically starts *Message assembly in aircraft's transponder* activity with appropriate duration. This module can also use already existing messages, which were obtained, for example, from real air traffic. A message assembly activity has the duration randomly generated and, in addition to that, it runs in parallel with an *Aircraft's position measurement by INS/GNSS system* activity. This leads to a new random variable, which is a random delay between the measurement of the position and the transmission of the message with that position.

- Input:
  - Position information contaminated by an error from the *GNSS/INS module*.
  - Real ADS-B data.
  - Simulation time from the *Simulation engine*.
  - User configuration.
- Output:
  - ADS-B messages in the DF-17 format together with coordinates of the message transmission origin.

4.5. Aerial Target module

The *Aerial Target module* is composed from the previous three modules; the *Trajectory module*, *GNSS/INS module* and the *Transponder module*, as can be seen in the Figure 4. This *Aerial Target module* represents individual aerial targets with a configuration given by these three modules. This module is also a base module for the *Spoof Target module*, which represents spoof targets. Due to the fact that this module's main objective is to group the three modules together, inputs of this module are made of inputs of inner modules. The same holds for outputs.

- Input:
  - Parameters for a trajectory (user scenario with vectoring instructions).
  - Chosen configuration of the position information source.
  - Real ADS-B data.
  - Simulation time from the *Simulation engine*.
- Output:
  - Output from the *Transponder module*.

4.6. Spoof Target module

This module is derived from the *Aerial Target module*. It contains information about real trajectory of the attacker's transceiver; in addition to the information from the original module. Frequent case of such trajectory will be a static point in the 3D space, which will correspond to the static attacker.

- Input:
  - All inputs from the *Aerial Target module*.
  - Trajectory scenario of the attacker's transceiver.
- Output:
  - Output of the *Aerial Target module*.

4.7. ADS-B Ground Station module

The *ADS-B Ground Station module* receives ADS-B messages. It stores the time of the message reception, and then sends the message data to the output for subsequent processing.

- Input:
  - Ground station coordinates.
  - ADS-B messages together with physical characteristics of these messages (position of the transceiver at the time of the transmission).
- Output:
  - Received data, especially position and velocity.

4.8. Passive Locator module

The *Passive Locator module* also receives ADS-B messages, but only processes the direction from where the message was received.

- Input:
  - Same as *ADS-B Ground Station module*.
- Output:
  - Azimuth of the incoming message signal.
  - Elevation of the incoming message signal.

4.9. MLAT module

The *MLAT module* is composed of multiple *ADS-B Ground Station* and/or *Passive Locator* modules. It combines functionality of those modules into one. Moreover, it computes time differences of the message arrivals at individual receivers. From these time differences of arrival (TDOAs), the origin of the transmission can be computed. This is the technique known as multilateration (MLAT).

- Input:
  - All inputs from multiple *ADS-B Ground Station* and/or *Passive Locator* modules.
- Output:
  - All outputs from the *ADS-B Ground Station* and/or *Passive Locator* modules.
  - Messages' time differences of arrival.

5. CONCLUSION

This article described problematics of an air traffic simulation from the viewpoint of a passive surveillance system. Brief description of the ADS-B technology was presented, together with its shortcomings and types of feasible attacks. Possible representations of airborne target trajectories and means of moving airborne targets along these trajectories were described. The last part dealt with the actual design of the simulator. It presented random values which will be generated during simulation and entities, activities and events that will be represented in the simulator. Also, separate modules that the simulator will comprise were described. The simulator is still in development and it is reckoned that some modules will be modified, so they will model real system more precisely. Simulator will be used to create ADS-B message validation algorithms for use in passive ADS-B systems made of passive locators.

REFERENCES

Sun, J., 2017. ADS-B Decoding Guide. Available from: https://adsb-decode-guide.readthedocs.io/ [accessed 6 March 2017].

Federal Aviation Administration, 2015. Aeronautical Information Manual: Official Guide to Basic Flight Information and ATC Procedures.

EUROCAE, 2009. ED-102A, Minimum Operational Performance Standards for 1090 MHz Extended Squitter Automatic Dependent Surveillance – Broadcast (ADS-B) and Traffic Information Services – Broadcast (TIS-B).

Schäfer M., Lenders V., Martinovic I., 2013. Experimental Analysis of Attacks on Next Generation Air Traffic Communication. In: Jacobson M., Locasto M., Mohassel P., Safavi-Naini R., eds. ACNS 2013: Applied Cryptography and Network Security. Lecture Notes in Computer Science, vol 7954. Berlin, Heidelberg: Springer, 253–271.

Delahaye D., Puechmorel S., Tsiotras P., Féron E., 2014. Mathematical models for aircraft trajectory design: A survey. In: Electronic Navigation Research Institute, eds. Air Traffic Management and Systems. Lecture Notes in Electrical Engineering, vol 290. Tokyo: Springer, 205–247.

Quanbeck, D. B., 1982. Methods for generating aircraft trajectories. Alexandria: Center for Naval Analyses.

Mohleji, S. C., Wang, G., 2010. Modeling ADS-B Position and Velocity Errors for Airborne Merging and Spacing in Interval Management Application. The MITRE Corporation. Available from: https://www.mitre.org/sites/default/files/pdf/10_3030.pdf [accessed 8 October 2016].

INVESTIGATIONS OF A RECONFIGURABLE PRESSING SETUP FOR MANUFACTURING BLADES OF LARGE FRANCIS TURBINES RUNNERS FROM VERY THICK PLATES

Zhengkun Feng [a✧], Henri Champliaud [b], Louis Mathieu [c]

[a, b] Mechanical Engineering Department, Ecole de technologie superieure, 1100 Notre-Dame West, Montreal, Quebec, H3C 1K3, Canada
[c] GE Hydro, 1350 St-Roch Road, Sorel-Tracy, Quebec, J3P 5P9, Canada

[b]henri.champliaud@etsmtl.ca, [c]louis.mathieu@power.alstom.com

ABSTRACT
The design of Francis turbines of hydropower plants is not standard, but different from one site to another due to hydraulic conditions. As a result, the hydraulic profile and the size of the blades of each Francis turbine are different. Therefore, the blades, one of the key components of Francis turbine runners, are produced in small batches and the setup of the dedicated punch and die increases significantly the unit production cost. In this paper, a flexible pressing process with reconfigurable punch and die for very thick plates is investigated. Firstly, a model for pressing very thick plates by conventional pressing process with matched punch and die is built. The simulation results agree well with the data obtained during a recent rehabilitation of a hydropower plant in Quebec. The results obtained from the model of the flexible pressing process are compared with the conventional pressing process.

Keywords: Pressing process; Multi-point forming; Finite element analysis; Very thick plate; Reconfigurable punch

1. INTRODUCTION

The pressing process has many advantages over the casting process and is widely used in automotive, shipbuilding, energy production and civil engineering. In casting process, gas entrapment may occur during mold filling (Wang et al. 2013) and an excessive amount of gases may result in gas porosity defect (Perzyk and Kochański, 2003). Consequently, the pressing process becomes an alternative of the casting process. For hydroelectricity, the blades of a Francis turbine runner (Fig. 1) can be produced by forming. This process reduces the manufacture time and cost. In addition, operators require less experience for the final machining (Casacci et al., 1977; Casacci and Caillot, 1983). Then a conventional pressing process with matched punch and die is actually used for manufacturing blades. Firstly, the pressing process provides a raw blade with constant thickness from a flat plate (Fig.2a). Then, after an operator machines this raw blade to a blade with the required size and hydro profile it is inspected (Fig. 2b) for geometry compliance.

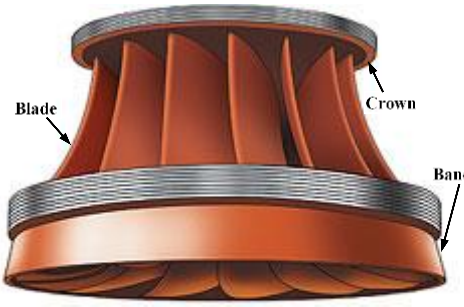
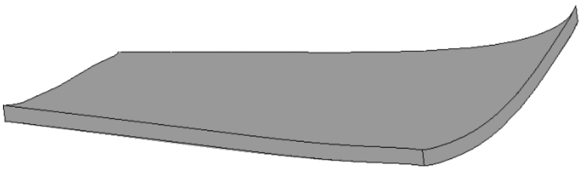


Fig. 1 Francis turbine runner composed of blade and other elements (Source from Hydro-Quebec)



(a) A raw blade;

(b) Inspection of a raw blade after machining

Fig. 2 Manufacture of a raw blade to a blade with hydro profiles from a raw blade.

However, the design of Francis turbines is not standard and may be very different from one site to another due to hydraulic conditions. As a result, the hydraulic profile, the size and the thickness of the blades of the runner are different from one Francis turbine to another. Therefore, the blades, one of the key components of a Francis turbine runner, are produced in small batches. The conventional pressing process with matched punch and die affects significantly the unit production cost. In addition, the trial and error technique that is practiced for thin plate and mass production becomes very costly for pressing of very thick plates and small production batches due to the intensive time and energy consuming. The flexible pressing process or multi-point pressing process with reconfigurable punch and die can save design and manufacturing time and attracts increasingly considerations for the manufacture of components with high thickness.

In the early 1980's, Hardt et al. (1981) investigated the multi-point forming process through various experiments. Until the late 1990's and the early 2000's, the focus was still on mechanical design and manufacturing of reconfigurable dies. Li et al. (1999) reported various multi-point forming processes. Closely packed pins to withstand the forming load were applied by Walczyk and Hardt (1998, 1999) and sectional multi-point forming was applied by Li et al. (2002). Papazian (2002) reported the very fast full reconfiguration of the die. With the rapid evolution of computer and computing technology, robust metal forming machines become possible. High efficient computational method and computing technique make the design of this kind of machine easier. Numerical simulations with explicit scheme (Li et al., 2002) and with implicit scheme (Chen et al., 2005) for sectional multi-point forming, with implicit scheme for multi-point forming (Cai and Li, 2005) were performed. Recently, the multi-point forming for cylindrical and spherical sections (Quan et al., 2011), double curved saddle section (Heo et al., 2010a, 2010b) were reported. Tests and numerical simulations on small scale and simple workpieces were carried out by Davoodi et al. (2014).

This study aims to investigate the pressing process for manufacturing the raw blades from very thick plates. The thickness of the plates is constant and so is the thickness of the raw blades obtained from the pressing process. The hydraulic profile with varied thickness is obtained by machining the pressed blank with a 5-axis CNC milling machine and by a final grinding stage (Sabourin et al., 2010) as shown in Fig.3.

Fig. 3 Robotic grinding of a Francis turbine runner blade (Sabourin et al., 2010).

In this paper, the modeling of the conventional pressing process with matched punch and die is given. The simulation results from this model are compared with the data obtained from the recent rehabilitation of a hydropower plant in Quebec, Canada. The proposed flexible pressing process with reconfigurable punch and die is investigated and the simulation results are compared with those obtained from the conventional pressing process with matched punch and die.

2. MODELING AND SIMULATION OF PRESSING PROCESS WITH MATCHED PUNCH AND DIE FOR VERY THICK PLATES

The numerical model of the pressing process with matched punch and die for thick plates is shown in Fig. 4. The model was composed of a flat blank, a matched punch, a matched die and pressing guides that are used to prevent the blank from moving out of the pressing machine. The material of the blank was the stainless steel ASTM A743 grade CA6NM. The elastic perfectly plastic material model at temperature of 800 °C was used for the blank. At this temperature the Young's modulus, the yield stress, the Poisson's ratio and the density were 92 GPa, 134 MPa, 0.3, 7850 kg/m^3, respectively. The punch, the die and the pressing guides were assumed as rigid bodies. The top and the bottom surfaces of the punch and die were identical to those of the desired raw blade. Therefore, the deformed shape of the raw blade was going to have the extrados on the top and intrados on the bottom. The blank, the punch and the die had thickness of 102 mm and were meshed with linear 8-node solid elements. Both top and bottom surfaces of the blank were more than 4 m^2. The thickness of the pressing guides was assumed to be very small and was meshed with linear 4-node shell elements with five integration points through the thickness. The static and dynamic friction coefficients between the blank and the punch, the die and the pressing guides were 0.6, respectively. The contact type between the blank and the punch and between the blank and the die was automatic surface to surface with the surfaces of the rigid bodies as contact surfaces and the surfaces of the blank as target surfaces. The contact type between the blank and the pressing guides was automatic node to surface contact with the nodes of the pressing guides as contact nodes and the surfaces of the blank as contact surfaces. As the thickness of the pressing guides was very small their nodes were very close to the blank surfaces. For the sake of CPU time, the simulation time was selected as 1 second. Figure 5 shows the shape of the blade obtained with the FE model. Figure 6 shows the strain distributions in six layers of the deformed blade that was divided by the six elements in the thickness. The layer order was from the bottom for the first layer to the top for the sixth layer. Figure 7 shows the time history of the applied force, i.e., the pressing force. Evidently, this force increased rapidly during the final stage of the pressing process. The red point indicated the maximum value of the pressing force of the pressing machine (see the following section).

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

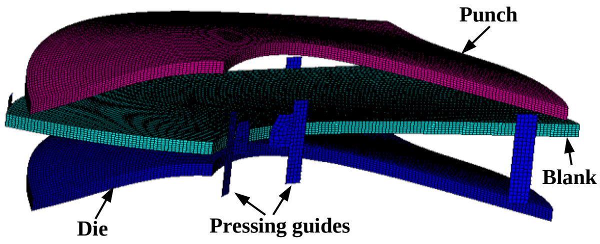


Fig. 4 Model of the pressing process with matched punch and die.

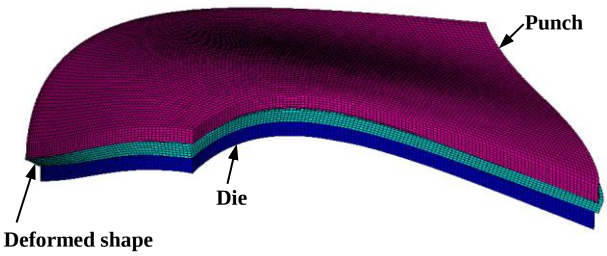


Fig. 5 Raw blade deformed by the matched punch and die.

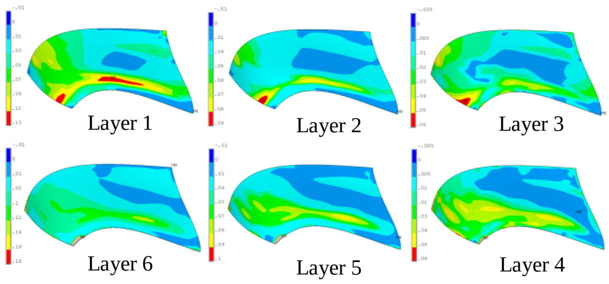


Fig. 6 Strain distributions at six layers of the deformed shape.

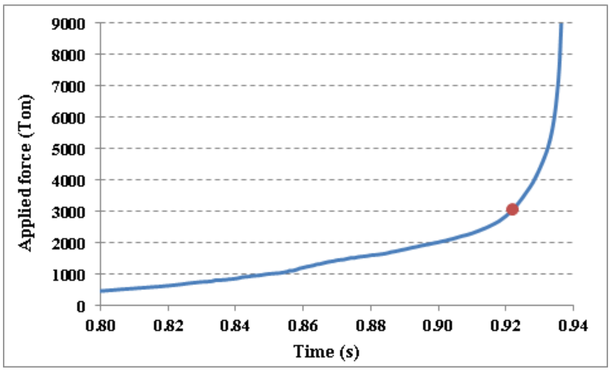
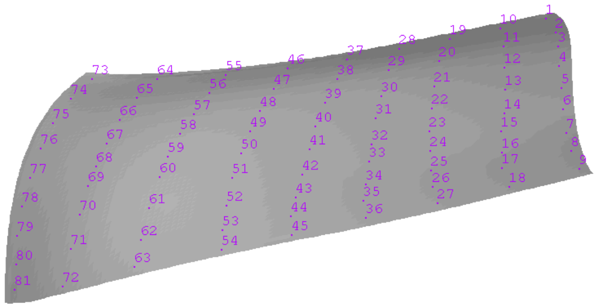


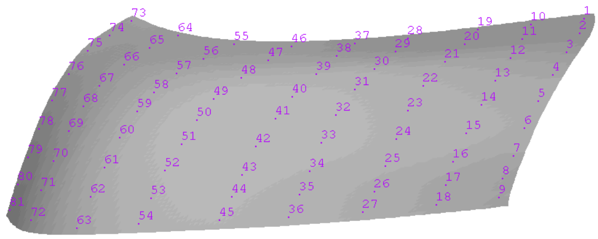
Fig. 7 Pressing force during hot blade pressing.

3. COMPARISON BETWEEN NUMERICAL RESULTS AND MANUFACTURING DATA

The application of the pressing process with matched punch and die was performed to manufacture the blades of Francis turbine runners used in the rehabilitation of Hydro-Quebec's historic Beauharnois power station. The maximum pressing force of the pressing machine was 3000 Tons (Fig. 7) and this value was reached at the final instant of the pressing process. The final displacements of two raw blades to the blanks were measured and were saved for the comparison with the numerical results obtained from the simulations as described in the previous sections. The locations of the measurement points on the extrados and intrados of the raw blades are illustrated in Fig. 8. There were 81 measurement points that are regularly spaced in 9 rows and 9 columns on both extrados and intrados of the raw blades. Although there were two points without available values on both extrados and intrados of the raw blades, the rest 79 measurement points allowed the performance of the comparisons.



(a) On the extrados;



(b) On the intrados;

Fig. 8 Measurement point locations on the raw blade obtained by the pressing machine

Figure 9 and Figure 10 show the differences between the simulation results and the manufacturer's data at the location of each measurement point on both extrados and intrados of the two raw blades. In these figures, a negative value was an insufficient displacement value of the simulation result at this point. For example, in Fig. 9, from point 10 to point 19 and from point 20 to point 27, there was not enough displacements, and from point 46 to point 54 and from point 55 to point 63, there was too much displacements. Tables 1-4 show the differences between the simulation results and the inspection data at each measurement point of the two raw blades on both extrados and intrados. The maximum values located at points 37, 46 and the minimum values located at points 9, 25, 26. The locations of these points were on or near the edges of the raw blades where the thickness was small for the required hydro profile of the blades. Therefore, these differences could not prevent from obtaining the blades with desired hydro profiles by machining the raw blades. Table 5 gives the maximum differences of these values on these two raw blades. Both the maximum and minimum values were less than 1 cm for a thickness of more than 10 cm. Therefore, the simulation results compared well with the manufacture data.

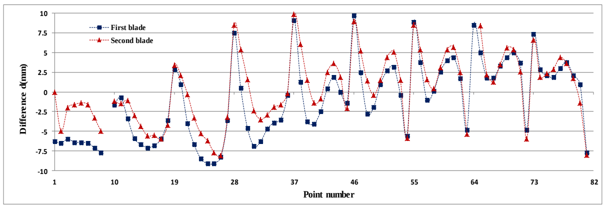


Fig. 9 Extrados difference between the simulation results and the manufacturer's data for the two blades.

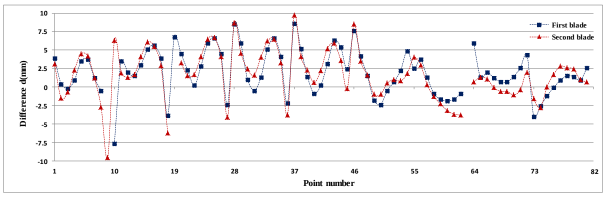


Fig. 10 Intrados difference between the simulation results and the manufacturer's data for the two blades.

Table 1: Position differences between simulations and test results for the extrados of the first raw blade

Point No	1	2	3	4	5	6	7	8	9
d(mm)	-6.3	-6.5	-6	-6.4	-6.4	-6.5	-7.1	-7.7	/
Point No	10	11	12	13	14	15	16	17	18
d(mm)	-1.6	-0.7	-3.4	-5.9	-6.7	-7.1	-6.8	-6	-3.6
Point No	19	20	21	22	23	24	25	26	27
d(mm)	2.9	1	-4	-6.7	-8.5	-9.1	-9.1	-8.3	-3.6
Point No	28	29	30	31	32	33	34	35	36
d(mm)	7.5	0.5	-4.6	-6.9	-6.3	-4.7	-3.9	-3.5	-0.4
Point No	37	38	39	40	41	42	43	44	45
d(mm)	9.1	1.3	-3.8	-4.1	-2.5	0.4	1.9	0	-1.4
Point No	46	47	48	49	50	51	52	53	54
d(mm)	9.7	2.5	-2.8	-1.9	1	2.7	3.2	-0.4	-5.6
Point No	55	56	57	58	59	60	61	62	63
d(mm)	8.9	3.8	-1	0.1	2.6	4	4.4	1.7	-4.8
Point No	64	65	66	67	68	69	70	71	72
d(mm)	8.5	5	1.8	1.8	3.3	4.4	5	3.7	-4.8
Point No	73	74	75	76	77	78	79	80	81
d(mm)	7.4	2.9	2.1	1.9	3	3.8	2.1	1	-7.7

Table 2: Position differences between simulations and test results for the intrados of the first raw blade

Point Nº	1	2	3	4	5	6	7	8	9
d(mm)	3.9	0.4	-0.2	0.9	3.5	3.7	1.2	-0.5	-7.7
Point Nº	10	11	12	13	14	15	16	17	18
d(mm)	/	3.5	2.0	1.5	3.0	5.1	5.6	3.9	-3.9
Point Nº	19	20	21	22	23	24	25	26	27
d(mm)	6.8	4.5	2.3	0.2	2.8	5.9	6.6	4.5	-2.4
Point Nº	28	29	30	31	32	33	34	35	36
d(mm)	8.5	5.9	1.0	-0.5	1.3	5.1	6.6	4.1	-2.2
Point Nº	37	38	39	40	41	42	43	44	45
d(mm)	8.6	5.2	1.4	-0.9	0.2	3.1	6.3	5.4	2.4
Point Nº	46	47	48	49	50	51	52	53	54
d(mm)	7.6	4.2	1.5	-1.8	-2.4	-0.5	0.7	2.2	4.9
Point Nº	55	56	57	58	59	60	61	62	63
d(mm)	2.5	3.7	1.3	-0.9	-1.7	-1.9	-1.7	-0.9	5.9
Point Nº	64	65	66	67	68	69	70	71	72
d(mm)	/	1.3	2	1.2	0.7	0.7	1.4	2.6	4.3
Point Nº	73	74	75	76	77	78	79	80	81
d(mm)	-4.0	-2.6	-1.2	-0.1	0.9	1.5	1.4	0.9	2.6

Table 3: Position differences between simulation and test results for the extrados of the second raw blade

Point Nº	1	2	3	4	5	6	7	8	9
d(mm)	0.0	-5.0	-2.0	-1.6	-1.4	-1.6	-3.3	-5.0	/
Point Nº	10	11	12	13	14	15	16	17	18
d(mm)	-1.2	-1.5	-1.1	-3.0	-4.4	-5.6	-5.5	-6.0	-4.2
Point Nº	19	20	21	22	23	24	25	26	27
d(mm)	3.4	2.1	-0.3	-3.3	-5.3	-6.2	-7.7	-8.1	-3.2
Point Nº	28	29	30	31	32	33	34	35	36
d(mm)	8.5	5.4	1.6	-2.4	-3.5	-2.9	-1.9	-1.6	-0.2
Point Nº	37	38	39	40	41	42	43	44	45
d(mm)	9.9	6.1	1.5	-1.4	-0.9	2.5	3.6	1.9	-2.1
Point Nº	46	47	48	49	50	51	52	53	54
d(mm)	9.0	5.2	1.4	-0.4	1.5	4.4	5.1	1.5	-5.9
Point Nº	55	56	57	58	59	60	61	62	63
d(mm)	8.5	5.4	1.6	0.4	3.1	5.4	5.7	2.5	-5.4
Point Nº	64	65	66	67	68	69	70	71	72
d(mm)	8.4	/	2.2	1.3	3.5	5.6	5.4	2.6	-6.0
Point Nº	73	74	75	76	77	78	79	80	81
d(mm)	6.6	1.9	2.3	2.9	4.4	3.6	1.7	-1.4	-8.0

Table 4: Position differences between simulations and test results for the intrados of the second raw blade

Point Nº	1	2	3	4	5	6	7	8	9
d(mm)	3.1	-1.5	-0.7	2.3	4.5	4.2	1.2	-2.7	-9.6
Point Nº	10	11	12	13	14	15	16	17	18
d(mm)	6.3	1.9	1.3	1.7	4.1	6.1	5.5	2.9	-6.2
Point Nº	19	20	21	22	23	24	25	26	27
d(mm)	/	3.3	1.5	1.7	4.1	6.5	6.7	4.1	-4.1
Point Nº	28	29	30	31	32	33	34	35	36
d(mm)	8.7	4.6	2.4	1.6	4.0	6.2	6.5	3.3	-3.8
Point Nº	37	38	39	40	41	42	43	44	45
d(mm)	9.7	4.1	2.3	0.6	2.2	5.2	5.9	3.6	-0.2
Point Nº	46	47	48	49	50	51	52	53	54
d(mm)	8.5	3.5	1.6	-1.0	-1.0	0.5	1.0	0.8	1.8
Point Nº	55	56	57	58	59	60	61	62	63
d(mm)	4.0	3.0	0.3	-1.3	-2.3	-3.2	-3.7	-3.8	0.7
Point Nº	64	65	66	67	68	69	70	71	72
d(mm)	/	1.3	1.1	-0.1	-0.6	-0.6	-1.1	-0.4	2.0
Point Nº	73	74	75	76	77	78	79	80	81
d(mm)	-1.6	-2.8	0.0	1.7	2.8	2.6	2.4	1.0	0.7

Table 5: Maximum position differences between simulations and test results

	First blade		Second blade	
	min d (mm)	max d (mm)	min d (mm)	max d (mm)
Extrados	-9.1	9.7	-8.1	9.9
Intrados	-7.7	8.6	-9.6	9.7

Proceedings of the European Modeling and Simulation Symposium, 2017
ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

4. GEOMETRY SET UP OF THE PRESSING PROCESS WITH RECONFIGURABLE PUNCH AND DIE

A multi-point pressing process consists mainly of a reconfigurable punch, a reconfigurable die and a workpiece, i.e., a blank. A simple example in two dimensions is shown in Fig. 11. The punch and die are composed of arrays of various pins with adjustable positions. During the pressing process the die is motionless and the punch applied a downward force on the workpiece through the contact point between each pin and the workpiece to deform it. The shape of a pin is a cylinder with a semi-sphere at one end that is used to deform the workpiece at the contact point. In comparison with the workpiece, the deformation of the tool is very small and can be neglected. Therefore, the cylindrical parts of the pins are not considered in the following text for simplicity and only the pin ends are reprensented in the model.

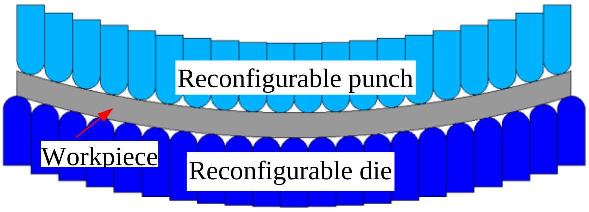


Fig. 11 Schematic diagram of a two-dimensional multi-point press process.

Figure 12 illustrates the geometry of the proposed multi-point pressing process with a reconfigurable punch and a reconfigurable die for producing raw blades. The upper matrix is the punch and the lower matrix is the die. The vertical direction is defined as the z direction in the geometry set up. Initially, the blank lies horizontally on the motionless die, i.e., in the xy plane. The top and bottom surfaces of the deformed shape correspond to the extrados and the intrados of the raw blade, respectively. The pressing zone for the punch and die are determined by the extrados and intrados of the raw blade. The pressing machine has an upper array of pins with the downward ends used for the punch and a lower array of pins with the upward ends used for the die. The positions and the number of pins used for the punch and die are determined by the shape of the desired shape, i.e., the raw shape of the blade. The positions of the pins of the reconfigurable punch are determined by the extrados of the raw blade and those of the reconfigurable die are determined by the intrados of the raw blade. The determination of the positions and the number of pins of the reconfigurable punch is given in the following subsections. The determination of the positions and the number of pins for the reconfigurable die is very similar and will not be repeated here.

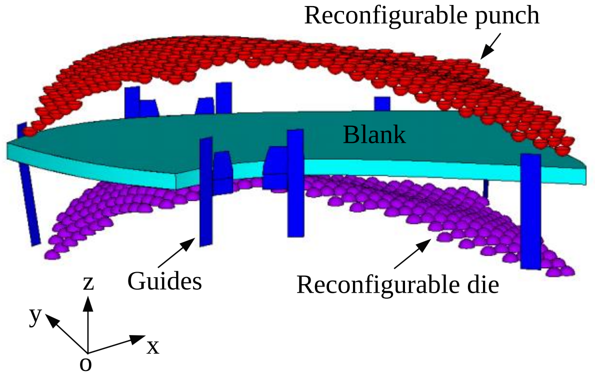


Fig. 12 Geometry of the multi-point pressing process.

4. 1 Determination of the number of pins

Figure 13 shows the upper array of 33x33 pins. Only the pins in the punch pressing zone are selected for building the reconfigurable punch. The relative positions of the pins to the extrados of the raw blade are illustrated in Fig. 14. However, the blank to be deformed is larger than the punch pressing zone Fig. 15. For ensuring the quality of the deformed shape, the pins of the punch pressing zone has to cover the extrados of the raw blank.

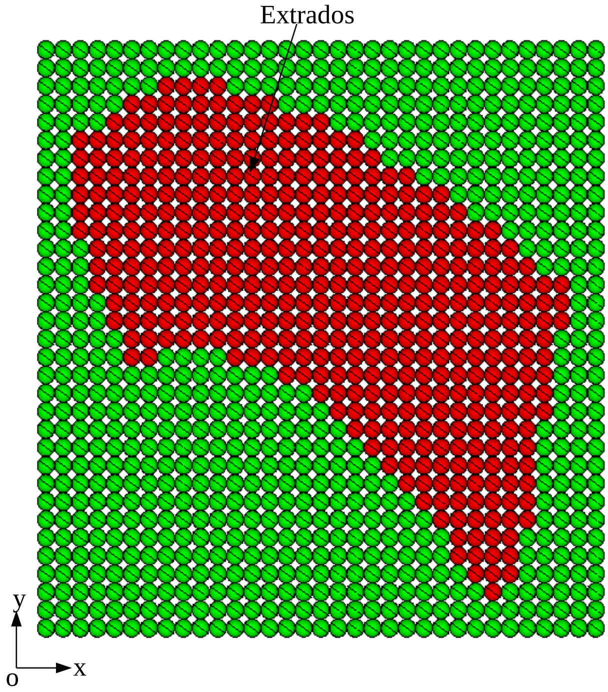


Fig. 13 Upper array of the pins of the pressing process

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

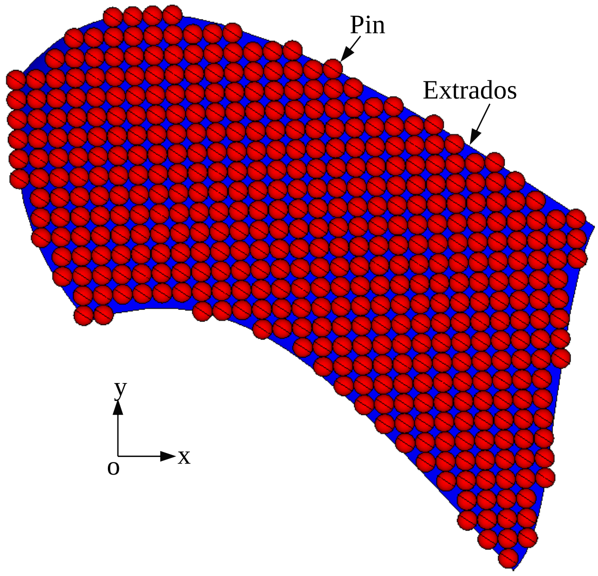


Fig. 14 Reconfigurable punch of the multi-point pressing process with the extrados of the desired raw blade.

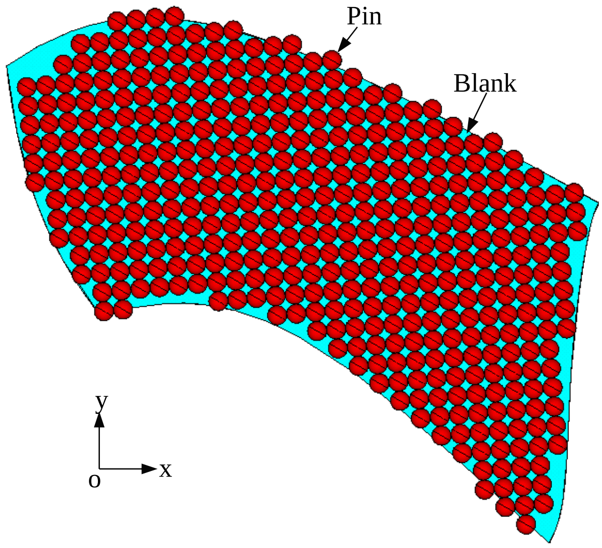


Fig. 15 Reconfigurable punch of the multi-point pressing process with the extrados of the blank to be deformed.

4.2 Determination of contact positions between the pins and the extrados of the raw blade

The positions of the pins of the punch are determined in the previous subsection. The x and y coordinates of the center of each pin are shown in Fig. 16 that is derived from Fig. 15. The coordinate in z direction (vertical direction) of each pin of the punch is determined in the following subsection.

Firstly, the extrados of the raw blade is divided by various triangles and these triangles are projected onto the xy plane as shown in Fig. 17. The center of the semi-sphere end of each pin locates in one of these triangles. Therefore, the x and y coordinates of these centers can be obtained from the positions of the centers of the semi-sphere ends of the pins in xy plane that are determined in the previous subsection.

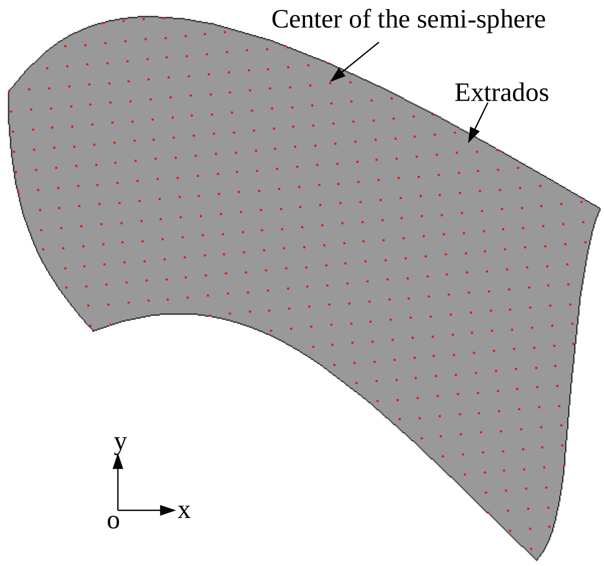


Fig. 16 Relative positions of the centers of the semi-sphere ends of the pins to the extrados of the raw blade.

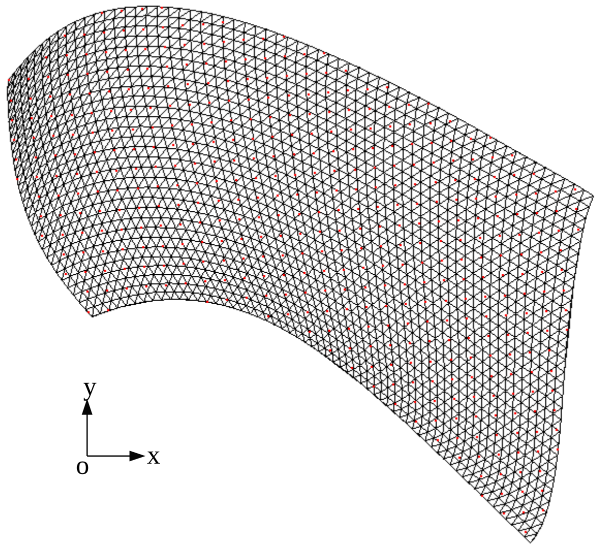


Fig. 17 Relative positions of the centers of the semi-sphere ends (red points) of the pins in the triangles that divided the extrados of the raw blade

For determining the z coordinate of the contact point of each pin, the following procedure can be performed. Firstly, select a center of the semi-sphere of a pin, following by selecting the triangles that the semi-sphere can cover with a margin to ensure the selection of all the triangles that among them one can be in contact with the semi-sphere as shown in Fig. 18 (from the top view, i.e., the view in the reverse z direction). Then, for each selected triangle, find the point in the plane of the triangle and verify if it locates in the triangle as shown in Fig. 19. In this figure the lines connecting points P_1, P_2 and P_3 represent the triangle. The line connecting P_4 and P_5 represents the axis of the pin, i.e., the pressing direction. Points P_6 and P_7 are the projected points of points P_4 and P_5 in the plane of the triangle, respectively. Note that points P_4, P_5, P_6 and P_7 are in the same plane. Finally, point P_8 is on the straight line between points P_6 and P_7 that has a distance of the radius of the pin to the axis of the pin can be found. If ξ, η and ζ calculated from eqs. (1) to (3) are positives this point locates in the triangle. This triangle can have a contact point with the pin. The coordinates are those in the local coordinate system with the xy plane defined by the triangle.

$$\xi = \frac{(x_8 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_8 - y_1)}{(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)} \quad (1)$$

$$\eta = \frac{(x_8 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_8 - y_1)}{(x_3 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_3 - y_1)} \quad (2)$$

$$\zeta = 1 - \xi - \eta \quad (3)$$

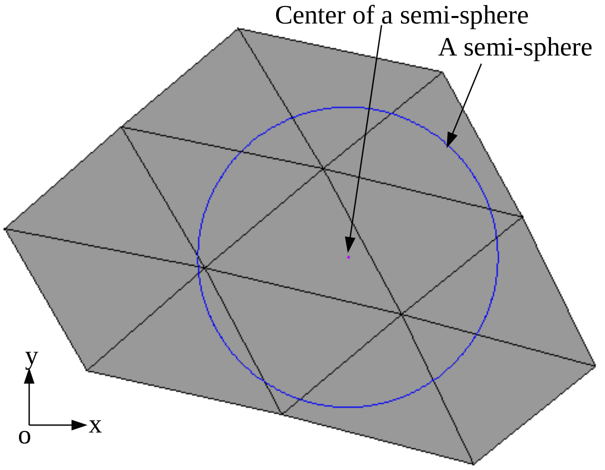


Fig. 18 Relative position of the centers of the semi-sphere end of a pin in a corresponding triangle.

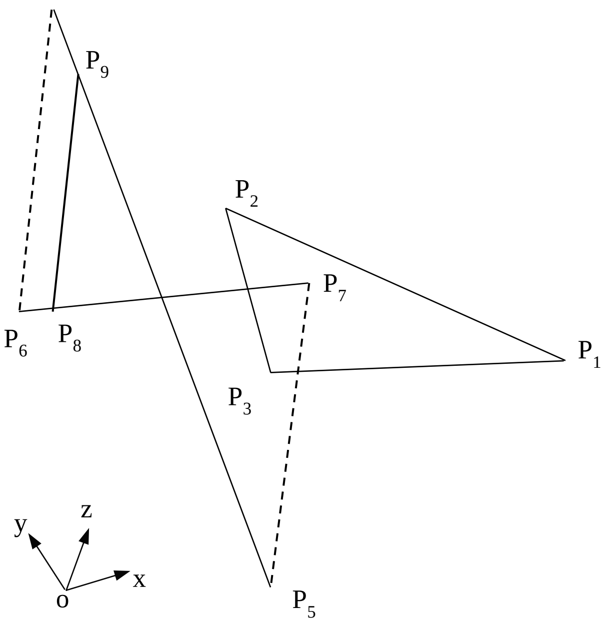


Fig. 19 Diagram for determining the distance of a triangle to the pin axis.

However, there may be more than one triangle that can have a contact point with the triangle. The one that has the biggest z coordinate in the global coordinate system (i.e., the pressing direction) is the real one. This triangle contacts with the pin when the punch moves downwards and this contact prevents the pin from contacting with the other triangles. The flow chart of the procedure to find the contact points between the extrados and the pins of the punch is illustrated in Fig. 20.

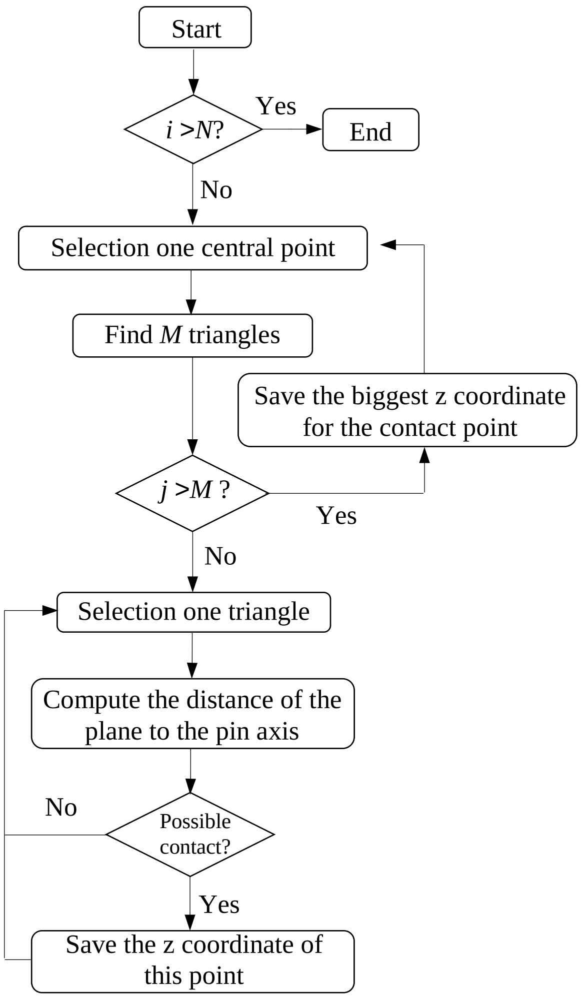


Fig. 20 Flow chart of finding the contact points between the extrados and the pins of the punch

In Fig. 20, the total number of pins is supposed to be N. The procedure starts with the first loop on the number of the pins. The center of the semi-sphere of the end of this pin is selected. Then the M triangles that have possibly contact points with the extrados are selected. The second loop on these selected triangles starts. Each triangle is checked to see if it is possibly in contact with the extrados. If possible, the z coordinate of the position of the center of the semi-sphere of the pin end is saved. The z coordinate of the position of the center of the semi-sphere is saved before the end of the second loop and before the first loop selected another pin. This value is the biggest z coordinate of the M triangles. Finally, the z coordinates of the positions of the semi-spheres of the

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

ends of all pins of the punch are obtained and the procedure ends. With the x and y coordinates of the semi-spheres of the ends of the pins are obtained in the previous subsection, the positions of the pins of the punch can then be computed in a similar way for the die.

5. MODELING AND SIMULATION OF THE PRESSING PROCESS WITH RECONFIGURABLE PUNCH AND DIE

The pressing process with configurable punch and die with the geometry set up described in the previous section was modeled with a finite element approach on the ANSYS/LS-DYNA platform. As described previously, the pins were assumed to be rigid bodies. As a result, only the semi-sphere ends of the pins were required to be modeled. In addition, the semi-spheres were assumed as hollow parts with very small thickness. The pins that did not locate in the pressing punch and die zones were not used in the model. The pins were meshed with 4-node shell elements and the numbers of elements and nodes were much less than a mesh with solid elements for solid pins. All pins were identical with a pin diameter of 10 cm. The pressing guides were considered as motionless walls and were also assumed as rigid bodies with very small thickness. Therefore, the pressing guides were also meshed with 4-node shell elements for reducing the numbers of elements and nodes of the model. For all of the shell elements, there were five integration points through the thickness of the element. The part of the model for the blank was the same as for the previous model (i.e., the same material properties, element type and total elements for the pressing process with continuous matched punch and die). The contact type between the blank and the rigid bodies (i.e., the punch, the die and the pressing guides) was automatic node-to-surface with the surfaces of the rigid bodies as contact surfaces and the surfaces of the blank as target surfaces. The static and dynamic friction coefficients on the contact surfaces were 0.6. The reconfigurable die was motionless as was the pressing guides during the process. The complete model is illustrated in Fig. 21.

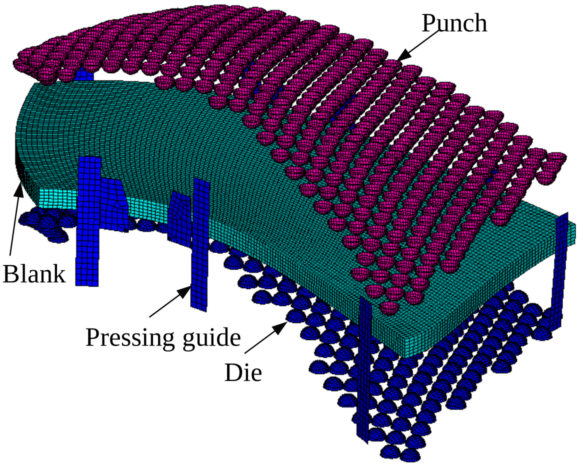


Fig. 21 Model of flexible pressing process with reconfigurable punch and die.

The obtained raw blade is shown in Fig. 22. Figure 23 shows the top surface of the raw blade compared with the one obtained from the simulation of the conventional pressing process.

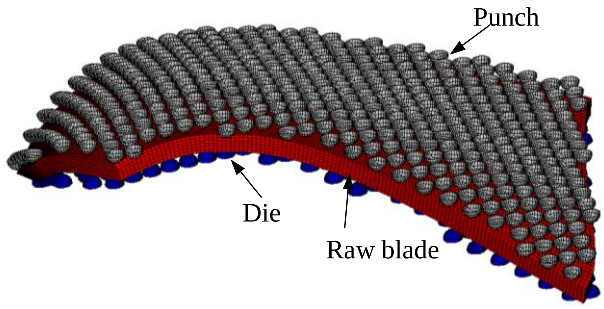


Fig. 22 Model of flexible pressing process with reconfigurable punch and die.

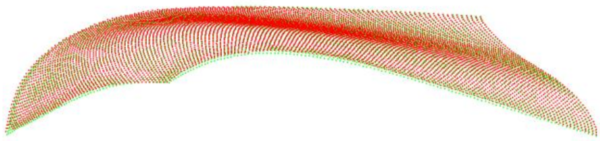


Fig. 23 Comparison between the tops of the raw blades obtained by conventional pressing process (Green points) and the flexible pressing process (Red points).

6. CONCLUSION

In this paper we described the simulation of the pressing process with reconfigurable punch and die. The dies are made of spherical pins assembled altogether in order to represent as close as possible the continuous shapes of conventional pressing matrices.

Comparison between manufacturer's data and both simulations, i.e. conventional punch and die and the reconfigurable ones, show that the reconfigurable option is very satisfactory and could be used for reducing the pressing costs of conventional press forming for small batch. Further investigations about setup frames of reconfigurable dies will be pursued as future works.

ACKNOWLEDGMENTS

The authors express their acknowledgements to the Natural Sciences and Engineering Research Council (NSERC) of Canada, Alstom Hydro Canada Inc. and Hydro Quebec for their financial supports to this research.

REFERENCES

Cai Z.Y., Li M.Z., 2005. Finite element simulation of multi-point sheet forming process based on implicit scheme. Journal of Materials Processing Technology, 161: 449-455.

Chen J.J., Li M.Z., Liu W., Wang C.T., 2005. Sectional multipoint forming technology for large-size sheet metal. International Journal of Advanced Manufacturing Technology, 25: 935-939.

Haas E., Schwarz R.C., Papazian J.M., 2002. Design and test of a reconfigurable forming die. Journal of Manufacturing Processes, 4: 77-85.

Hardt D.E., Olsen B.A., Allison B.T., Pasch K., 1981. Sheet Metal Forming with Discrete Die Surfaces. Manufacturing Engineering Transactions, 140-144.

Heo S.C., Seo Y.H., Park J.W., Ku T.W., Kim J., Kang B.S., 2010. Application of flexible forming process to hull structure forming. Journal of Mechanical Science and Technology, 24: 137-140.

Heo S.C., Seo Y.H., Ku T.W., Kang B.S., 2010. A study on thick plate forming using flexible forming process and its application to a simply curved plate. International Journal of Advanced Manufacturing Technology, 51: 103-115.

Li L., Seo Y.H., Heo S.C., Kang B.S., Kim J., 2010. Numerical simulations on reducing the unloading springback with multi-step multi-point forming technology. International Journal of Advanced Manufacturing Technology, 48: 45-61.

Li M., Liu Y., Su S., Li G., 1999. Multi-point forming: A flexible manufacturing method for a 3-d surface sheet. Journal of Materials Processing Technology, 87: 277-280.

Li M.Z., Cai Z.Y., Liu C.G., 2007. Flexible manufacturing of sheet metal parts based on digitized-die. Robotics and Computer-Integrated Manufacturing, 23: 107-115.

Li M.Z., Cai Z.Y., Sui Z., Yan Q.G., 2002. Multi-point forming technology for sheet metal. Journal of Materials Processing Technology, 129: 333-338.

Papazian J.M., 2002. Tools of change: Mechanical Engineering, 124: 52.

Perzyk M., Kochański A., 2003. Detection of causes of casting defects assisted by artificial neural networks. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 217(9): 1279-1284.

Quan G.Z., Ku T.W., Kang B.S., 2011. Improvement of formability for multi-point bending process of AZ31B sheet material using elastic cushion. International Journal of Precision Engineering and Manufacturing, 12: 1023-1030.

Socrate S., Boyce M.C., 2001. A finite element based die design algorithm for sheet-metal forming on reconfigurable tools. Journal of Engineering Materials and Technology, Transactions of the ASME, 123: 489-495.

Walczyk D.F., Hardt D.E., 1998. Design and Analysis of Reconfigurable Discrete Dies for Sheet Metal Forming. Journal of Manufacturing Systems, 17: 436-454.

Walczyk D.F., Hardt D.E., 1999. A comparison of rapid fabrication methods for sheet metal forming dies. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 121: 214-224.

Zhang Q., Dean T.A., Wang Z.R., 2006. Numerical simulation of deformation in multi-point sandwich forming. International Journal of Machine Tools and Manufacture, 46: 699-707.

✧AUTHORS' NOTE

In memory of “Jacques” Zhengkun Feng, our dear friend, who suddenly and very sadly passed away in December 2016.

RISK MANAGEMENT IN JACKETS MANUFACTURING PROJECTS USING DISCRETE EVENTS SIMULATION

Adolfo Lamas-Rodríguez[(a)], David Chas-Álvarez[(b)], José Antonio Muiña-Dono[(c)]

[(a)] Universidade da Coruña, Navantia, UMI Navantia-UDC
[(b)] UMI Navantia-UDC
[(c)] Universidade da Coruña

[(a)]alamas@udc.es
[(a)]david.chas@udc.es
[(a)]jose.mdono@udc.es

ABSTRACT

This paper presented an innovative parametric tool for quantifying project risks by applying Discrete Events Simulation (DES). This tool has been customized for wind turbine foundations manufacturing projects and will be applied to quantify the risk associated with delays in the supply chain and customer's AFC (Approved for Construction) drawings delays.

The difficult task of identifying risks, quantifying them and proposing mitigation plans to minimize their impact in this type of serial fabrication projects justifies the use of a tool based on DES. The short fabrication period and the high penalties related to the non-compliance of the delivery milestones, makes this tools very useful in project management.

This simulation tool has been validated by a real jacket manufacturing project and has been used in order to identify and quantify the project risks as well as to propose and check the effect of the mitigation plans associated with each project risk.

Keywords: jacket, risk management, mitigation plans, discrete events simulation

1. INTRODUCTION

In the last 3 years the wind energy in Europe has duplicated its installed capacity with a total installed capacity of 12,631 MW from 3.589 wind turbines along 81 wind farms in 10 European countries. As well as, at the end of 2016, 11 projects reached final investment decision, worth €18.2bn, will represented an increase of 4,948 MW in the total capacity installed. Offshore wind is one of the most dynamic renewable energy, in terms of installed capacity and technology development since 2010. (Carmen & Varela-v 2017).

According to (EWEA 2017b), by 2020, offshore wind is projected to grow to a total of installed capacity of 24,6 MW with trends to installing greater capacity wind turbines (4,8 MW average capacity in 2016)and situated in deeper and further away to shore areas (29m average deep and 44 km average distance to shore in wind farms completed or partially completed in 2016).

This trends means a promising future by the jackets market, representing 12 % of all foundations installed in 2016 and 6,6 % of cumulative installed foundations in Europe.

On the other hand, the wind industry has continuously reduced cost in order to make wind energy a competitive energy resource, making nowadays the onshore wind energy the cheapest new power generation in Europe. At the same time, in 2016, offshore wind energy has proven it can be in the same cost range (Ewea 2017a). Offshore wind is also undergoing an increasing cost reduction and technology improvements (Carmen & Varela-v 2017)

In search of this aim Jan Kjærsgaard (CEO of Bladt Industries) proposes the jacket design standardization to contribute to offshore wind cost reduction.

Another important aspect in jackets manufacturing project is the very demanding takt time established by the market which currently requires the delivery of 1 jacket per week.

This takt time implies that the manufacturer company needs to work with more than one supplier in the same project. This reason leads the fabrication company to make an exhaustive risk analysis on possible supply chain and engineering delays.

According to the described jacket manufacturing project needs, a Discrete Events Simulation model has been developed in order to quantify the risks arising from supply delays and to propose different mitigation plans depending on the likelihood associated with the risk. The use of simulation techniques is advised as an effective instrument for supply chain risk evaluation (Klimov & Merkuryev 2008).

Therefore, through the Simulation tool used, is possible to characterize the risk based in quantitative data and determine the appropriate level of detail grade of the mitigation plans depending on the probability of use them. Process risks should be modelled and assessed to account for the uncertainties and their consequences (Shah et al. 2017)

In addition, the simulation tool will perform an economic analysis based on the risk probability and the economic impact of the risk over the project.

2. STATE OF THE ART

The aim of this work will be develop a quantifying risks tool based on Discrete Events Simulation. Most of the literature use the Discrete Events Simulation in order to quantifying risks in supply chain, but applications that quantify risks with simulation models that combines the supply chain and manufacturing process are less frequent. The following paragraphs presents a brief summary of the most relevant studies related to the problem considered:

(Singh & Schmitt 2009), studied the impact on the customer service of disruptions in a customer products service using Monte-Carlo and discrete events simulation. They used their tool in order to assess the level of supply chain disruption risk, test different mitigation plans, can use the tool in case of a disruption in order to validate recovery steps before putting then into action and Identify redundancy in the system.

(Deleris 2005), designed a tool to assess uncertainty in supply networks based on Monte-Carlo Simulation. They have focused their tool on a method to estimate the losses in a supply network.

(Klimov & Merkuryev 2008), discussed the simulation-based risk evaluation in supply chain, and presented in their work a risk evaluation example with a simplified supply chain system.

(Shah et al. 2017), proposed a process-oriented quantitative risk assessment methodology in order to evaluate risk associated with processes using modelling, simulation and decision-making approaches.

(Ingalls 2014), based in his paper called The Value of Simulation In Modeling Supply Chains (Ingalls, 1998) presents the advantages and disadvantages of simulation use as analysis methodology to evaluate supply chains.

(Cube et al. 2016), designed a tool based on discrete events simulation for monetarily quantify risks independent of the depth of information and thus allow adjusting the model dependent on the use-case.

3. DESCRIPTION OF THE MANUFACTURING AND SUPPLY PROCESS IN A JACKET MANUFACTURING PLAN

To elaborate the study of this paper, we have used jackets whit three legs, which are differentiated into four principal parts (Transition Piece, Jacket Upper Block, Jacket Lower Block and Piles). The jacket decomposition is represented in the Figure 1.

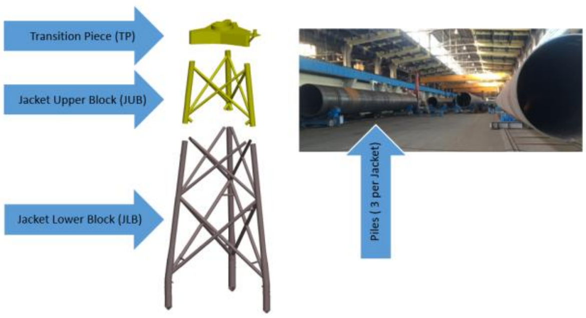


Figure 1: Main Parts of the Jacket.

The manufacture process used in this paper uses as input the four principal parts of the jacket. Figure 2 presents the manufacturing process flowchart.

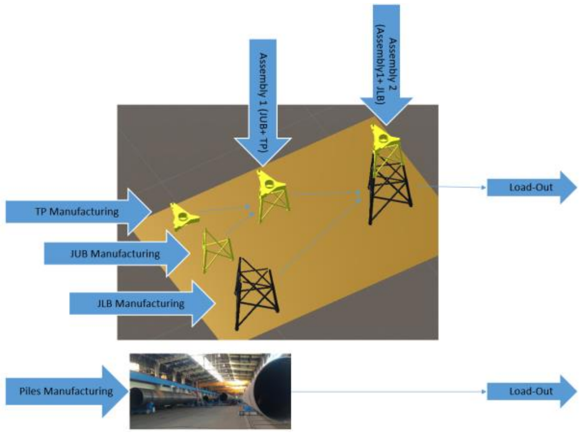


Figure 2: Flowchart of Jacket Manufacturing Process.

For simplify in this paper the experimental results, we go to use a manufacturing Piles Simulation model. This Simulation Model will represent the supply and manufacturing process in a piles manufacturing plant. This process it is not a trivial supply chain problem due to the first supplier is also the customer i.e., the customer marks the beginning and end of the project.

The supply and manufacturing process flowchart is represented in Figure 3.

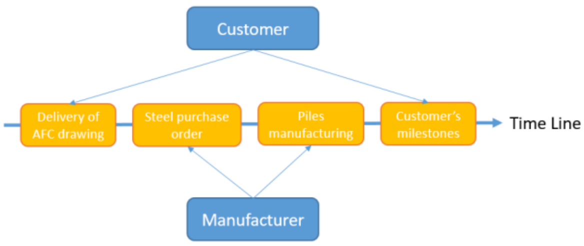


Figure 3: Supply and Manufacturing Process of the Piles.

At the same time, in jackets manufacturing projects is important to take into account the high penalties due to non-compliance customer milestones, which can be about 150.000,00 €/ (day and jacket).

4. PROBLEM DESCRIPTION

In the manufacturing jackets projects the risks quantifying analysis can be apply in the project planning phase or in the manufacture phase. If we do the risks analysis in the planning phase, the aim will be assess the level risk of break the customer milestones, and test the different mitigation plans proposed. On the other hand, if we do the risk analysis in the manufacture phase, our aim will be used the tool in case of appears the risk of break the customer milestones along the manufacturing plan.

In both cases, previously described, to elaborate the risks analysis, the project team needs to quantify the probability that the risk may exist and, at the same time, evaluate the impact of this risk in the project.

Depending on the risks analysis, the project team will develop with more or less detail the mitigation plans associated with each risk. At the same time, the parameters to elaborate the mitigation plans will be different if the risk is detected in the planning phase or in the manufacture phase.
In this point, and as main reason for the developed tool presented in this paper, we should take into account the difficulty of quantifying the impact of the risks in the project, and the importance of detecting the risk in the planning phase. i.e., a very likely risk could have not impact in the project and its mitigation plan should have a low detail or a less likely risk could be dangerous for the project and its mitigation plan should have a high detail. At the same time, a risk identified in the planning phase will be easier to solve than the same risk identified in the manufactured phase.

5. RISK MANAGEMENT METHODOLOGY

In this paper, we have based the risk management methodology in a company whose risk management process is represented by flowchart of the Figure 4.

Figure 4: Risk Management Process Flowchart

The simulation model developed in this paper will be used in the points two, three and four of the previous flowchart.

5.1. Risks Criteria

The risk quantification methodology used in this paper is based on the probability of occurrence of the risk-causing activity and the effects of the risk on cost and schedule.
To assessment the risk occurrence probability we are going to use a probability index (PI) represented in Table 1. This index is related to the occurrence of the risk.

Table 1: Probability Index

Probability Index (PI)	Denomination	Probability
1	Very low	0% < P ≤ 10%
2	Low	10% < P ≤ 30%
3	Medium	30% < P ≤ 60%
4	High	60% < P ≤ 90%
5	Very high	90% < P ≤ 100%

An assessment of the impact of a risk shall be performed in accordance with the Table 2.

Table 2: Impact Index

Impact Index (II)	Denomination
1	Very low
2	Low
3	Medium
4	High
5	Very high

To obtain the Impact Index (II) for a risk, we use as a reference the Table 3 and Table 4.

Table 3: Cost Thresholds

Cost Impact [€]				
Very low	Low	Medium	High	Very High
<50000	<75000	<150000	<300000	≥300000

Table 4: Schedule Thresholds

Schedule Impact [days]				
Very low	Low	Medium	High	Very High
<5	<15	<30	<50	≥50

The assessment of the Impact Index (II) for each risk shall be whichever is the highest of the values estimated for both criteria.

5.2. Risks Assessment Matrix

Once the Probability and Impact Indices have been defined, the Criticality Index shall be defined as the product of both:

$$CRITICAL\ INDEX\ (CI) = PI \cdot II. \quad (1)$$

Table 5 represents all possible values for the Criticality Index. At the same time, we have used a colour code in order to separate the thresholds in terms of mitigation or contingence actions.

Table 5: Critical Index Matrix

Critical Index (CI)		Impact				
		Very Low	Low	Medium	High	Very High
Probability	Very high	5	10	15	20	25
	High	4	8	12	16	20
	Medium	3	6	9	12	15
	Low	2	4	6	8	10
	Very Low	1	2	3	4	5

Risk Level (LOW): CI = 1, 2, 3, 4, 5, 6
Risk Level (MEDIUM-LOW): CI = 8, 9, 10, 12
Risk Level (MEDIUM-HIGH): CI = 15, 16, 20
Risk Level (HIGH): CI = 25

5.3. Simulation model development

The piles simulation model developed for this paper consist in five inputs (delivery of AFC drawings, customer milestones, plant capacity, construction

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

strategy and task simulation times) with the flowchart represented in Figure 5.

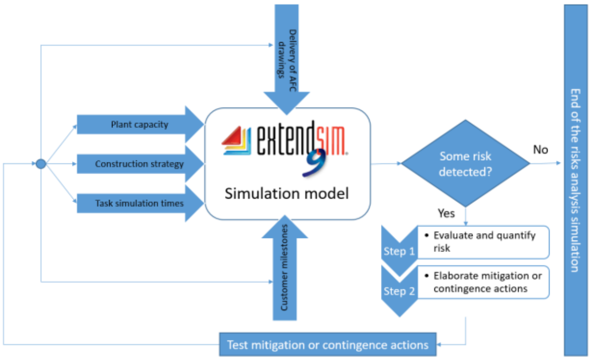


Figure 5: Simulation Model Flowchart

This simulation model uses the floating calculation in order to detect the risks, i.e. the model calculate the maximum delay in the project that allows compliance the milestones of the customer.
In piles manufacturing projects the model inputs (AFC drawings delivery and customer's milestones) are established by the customer and only can be modified if the risk analysis is elaborate in the planning phase. The impossibility of modified this inputs during the manufacturing phase limits the possibility of elaborate mitigation plans if the risk is identified during the manufacturing phase.
As to the piles construction strategy, in this paper we have used a strategy based on the tasks and the task overlap presents in the Table 6 where notation of task overlap used the Microsoft Project notation.

Table 6: Piles Construction Strategy

Task number	Task description	Task overlap
1	Plates cut and bevelling	
2	Bending of plates	1FC+1
3	Longitudinal Welding	2CC+1
4	Section assembly	3CC+2
5	Circular welding	4CC+2
6	Welding of beds	5FC
7	Non-destructive testing	6FC+1
8	Marking of section	7
9	Final inspection	8

The number of piles that can be manufacturing in parallel is a factor depending on the manufacturing plant, but in this study case we go to consider it a variable parameter, due to in this type of projects the market offer a great number of companies whit capacity for this type of works.
As a last input we have the tasks process time, for this practical work each task time has been approximated to a lognormal distribution. In this case we have used historical data of similar projects in order to elaborate de theoretical lognormal distribution.
This simulation model has been developed in the DES software ExtendSim 9.2 and the Figure 6 presents the model that capture the flow presented in Figure 5. The figure is too small to appreciate the blocks, but with the color-coded presents in the figure is possible identify the different parts.

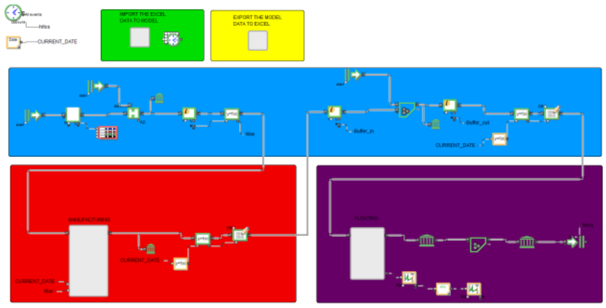


Figure 6: Simulation Model

- The green and the yellow block represents the exchange of data between the Simulation model and the Excel spreadsheet within which are represented the model changeful parameters. The green block occurs in time zero and feeds the simulation model and the yellow block occurs at the end of the simulation and feeds the risks analysis tables.
- The blue block controls the purchase orders, buffer dimensions and Load-Out dates, i.e., controls that the model does not break the restrictions.
- The red block represents the construction strategy and controls that the model does not exceed the real dimensions of the manufacturing plant.
- In the last place, the purple block is in charge of calculate the floating of the manufacturing project, and controls the model replications in order to obtain a floating value whose confidence interval keep a relative error less than 0.01.

6. RESULTS

The results of this paper has been organized in two different experiments. First experiment has been designed in order to detect the start project date with floating value equal to cero. Second experiment represents the risks analysis in the piles manufacturing project, within which presents a numerical example to risks quantification with Discrete Events Simulation, and a brief explain about the possible mitigation actions.

6.1. Start Date Analysis

When we speak of floating value equal to cero, we are referring to the start project date from which the project milestones will be broken.
In this first analysis we have used a range dates of five days up and down from the estimated AFC drawings delivery date.
In this point, we have not taken into account the quantification of risks, because we have focused in find

the latest start project date that assure us do not take early risks.
Table 7 presents the result for this analysis, providing the average floating, confidence interval, and the percentage in each threshold of floating value. In this analysis we have used 200 replications per scenario.

Table 7: Results of Start Date Analysis

Delivery of AFC drawings	Average floating	Confidence interval	Floating (f) [days]			
			<-1	-1≤f<0	0≤f≤1	>1
-5	1.295	0.1332	1.5%	0.0%	57.0%	41.5%
-4	1.220	0.1413	2.0%	0.0%	56.5%	41.5%
-3	0.110	0.2096	10.0%	0.5%	85.0%	4.5%
-2	-1.975	0.3067	45.5%	5.0%	49.5%	0.0%
-1	-1.820	0.3010	43.0%	4.5%	52.5%	0.0%
0	-2.045	0.3048	48.0%	5.5%	46.5%	0.0%
1	-4.480	0.1405	97.0%	1.0%	2.0%	0.0%
2	-5.535	0.1318	100.0%	0.0%	0.0%	0.0%
3	-6.810	0.1919	100.0%	0.0%	0.0%	0.0%
4	-8.815	0.2570	100.0%	0.0%	0.0%	0.0%
5	-10.460	0.1956	100.0%	0.0%	0.0%	0.0%

Interestingly, the real AFC drawings delivery date has a risk for the project whit a probability greater than 50% of break the customer milestones.
In addition, if the AFC drawings delivery date is advanced in 3 days the probability of compliance the customer milestones is about 90% and with an advance of 5 days the probability is increased to 99%.

6.2. Risks Quantification with Discrete Events Simulation

Taking advantage of the identified risk on the previous analysis, in this point we go to quantify the risk in accordance with the risk management methodology described in the point five of this work.
In first place, the simulation model is run again, in order to complete the schedule impact and the cost impact tables (Table 8 and Table 9).

Table 8: Example of Schedule Impact

Schedule Impact [days]						
Average floating	No impact	Very low	Low	Med	High	Very High
	>0	<5	<15	<30	<50	≥50
-2.180	43.5%	47.5%	9.0%	0.0%	0.0%	0.0%

Table 9: Example of Cost Impact

Cost Impact [€]						
Average cost	No impact	Very low	Low	Med	High	Very High
311250.0€	46%	0%	0%	0%	5%	49%

If we based on the schedule impact represented in Table 8, we have two possibilities in order to calculate the critical index (CI).

1. Schedule impact very low, whit an occurrence probability of 47.5%, then, consulting the tables 1, 2 we have a PI=3, II= 1, therefore, according to table 5 the critical index (CI) is equal to 3 (risk level low).
2. Schedule impact low, whit an occurrence probability of 9%, then, PI=1, II= 2, therefore, according to table 5 the critical index (CI) is equal to 2 (risk level low).

On the other hand, if we based the risks analysis by its cost impact, we have another two Critical Index (CI) possibilities.

1. Cost impact high, with an occurrence probability of 5%, then PI=1, II=4, therefore CI=4 (risk level low).
2. Cost impact very high, with an occurrence probability of 49%, then PI=3, II=5, therefore CI=15 (risk level medium high).

According to risk criteria, the mitigation plan shall be developed taking into account the highest CI, in this case is a risk level medium high.
In order to represents the analysis of mitigation plans with the developed tool, we go to study the effect of learning curve in the tasks times of the process. Based on historical data of similar projects we have observed that a reduction of 5% tasks times can be considered in the last third of the manufacture. Applying this reduction at the average value in the lognormal distribution that define the tasks time we obtain the results represented in the Table 10 and Table 11.

Table 10: Schedule Impact with the Learning Curve Effect

Schedule Impact [days]						
Average floating	No impact	Very low	Low	Med	High	Very High
	>0	<5	<15	<30	<50	≥50
0.39	91%	9%	0.5%	0.0%	0.0%	0.0%

Table 11: Cost Impact with the Learning Curve Effect

Cost Impact [€]						
Average cost	No impact	Very low	Low	Med	High	Very High
29250.0€	91.5%	0%	0%	0%	5.5%	3%

With the effect of the learning curve in the tasks times, is not possible eliminate the risk, but as is represented in Table 10 and Table 11 the max CI is reduced to 5, which means that the risk level is low.

7. CONCLUSIONS

Overall, the simulation model developed allows checking the possibility of non-compliance the customer milestones, and allows an exactly quantification as to schedule and cost impacts for each identified risk.
In addition, throughout the project we have observed the importance of detect the risks in the planning phase due to the greater flexibility for elaborate mitigation plans, and the importance of this kind of tools in order to presents the risks analysis to the project manager.
Finally, the simulation model results demonstrate the high economic impact of break the customer milestones, and the importance of fulfil the project schedule.

Likewise, the results of this paper leave an evidence the importance of analyse the risk under different aspects and as a good risk management methodology can help us to finish the project successfully.

ACKNOWLEDGMENTS

The authors are thankful to Unidad Mixta de Investigación (UMI) Navantia-UDC for its valuable support.

REFERENCES

Carmen, S, and Pedro Varela-v. 2017. "Estimation of the Potential Effects of Offshore Wind on the Spanish Economy." *Renewable Energy* 111: 815–24.

Cube, Philipp Von, Lasse Härtel, Robert Schmitt, Christophe Ponsard, Philippe Massonet, Renaud De Landtsheer, Gustavo Ospina, Stephan Printz, and Sabina Jeschke. 2016. "Monetary Quantification of Supply Risks of Manufacturing Enterprises - Discrete Event Simulation Based Approach." *Procedia CIRP* 57: 164–70.

Deleris, Léa A., and Feryal Erhun. 2005. "Risk Management in Supply Networks Using Monte-Carlo Simulation." In *Proceedings of the 2005 Winter Simulation Conference*, 1643–49. Hilton Downtown Disney Lake Buena Vista, FL, USA.

Ewea. 2017a. "Reviving Wind Markets and Delivering on Our 2030 Objetives." *The European Wind Industry's to the European Commision's Proposals on a Governance Regulation and a Renewable Energy Directive*. www.ewea.org. http://www.ewea.org/statistics/offshore-statistics/. 07/2017.

Ewea. 2017b. "The European Offshore Wind Industry." *Key Trends and Statistics 2016*. www.ewea.org. http://www.ewea.org/statistics/offshore-statistics/. 07/2017.

Ingalls, Ricki G. 2014. "Introduction to Supply Chain Simulation." In *Proceedings of the 2014 Winter Simulation Conference*, 36–50. Savannah, GA, USA.

Klimov, Ruslan, and Yuri Merkuryev. 2008. "Simulation Model for Supply Chain Reliability Evaluation." *Ukio Technologinis Ir Ekonominis Vystymas* 14:3: 300–311.

Liaqat A. Shah, Alain Etienne, Ali Siadat & François Vernadat. 2017. "Process-Oriented Risk Assessment Methodology for Manufacturing Process Evaluation." *International Journal of Production Research* 55:15 (June): 4516–29.

Singh, Mahender, and Amanda J Schmitt. 2009. "Quantifying Supply Chain Disruption Risk Using Monte Carlo and Discrete-Event Simulation." In *Proceedings of the 2009 Winter Simulation Conference*, 1237–48. Hilton Austin Hotel Austin, TX, USA.

AUTHORS BIOGRAPHY

ADOLFO LAMAS RODRÍGUEZ graduated from the University of Vigo in 1998. He holds an MSc and a PhD in Industrial Engineering. He combines his research activities in the researching group Grupo Integrado de Ingeniería and his position as a senior engineer and Project Manager in the Spanish leading shipbuilding company Navantia. He is also Associate Professor in the University of A Coruna since 2004 teaching in subjects related to manufacturing, simulation and Lean Manufacturing techniques He is the coordinator of one of the researching lines in the joint venture Navantia-University of Coruña (UMI) related to simulation and optimization models of industrial processes.

DAVID CHAS ÁLVAREZ holds an MSc in Industrial Engineering since 2015 and PhD student since 2016. He works as research engineer in the joint venture Navantia-University of Coruña (UMI) and he is mainly involved in the development of simulation and optimization models of industrial processes, especially in models of manufacture wind turbines foundations.

JOSÉ ANTONIO MUIÑA DONO holds a university degree in Industrial Technology since 2015 and an MSc in Industrial Engineering since 2017 from the University of Coruña, focusing his activity on the M&S of industrial processes.

USING SIMULATION AND RELIABILITY CONCEPTS TO SET STARTING TIMES IN MULTI-TASK PROJECTS WITH RANDOM DURATION AND A COMMON DEADLINE

Angel A. Juan[(a)], Laura Calvet[(a)], Carles Serrat[(b)], Sara Hatami[(a)]

[(a)]IN3-Computer Science Dept., Open University of Catalonia, Barcelona, Spain
[(b)]Dept. of Mathematics, Technical University of Catalonia, Barcelona, Spain

[(a)]{ajuanp, lcalvetl, shatami}@uoc.edu, [(b)]carles.serrat@upc.edu

ABSTRACT
In production and transportation logistics, it is frequent to have projects in which different tasks must be processed in parallel and completed before a common deadline. Usually, it is not desirable to assign scarce resources to a new project earlier than strictly necessary, since they might be currently occupied in other projects. When these tasks have random completion times, some natural questions arise: (*i*) how much can we delay the starting time of each task in the new project so that it can be completed by the deadline with a given probability? and (*ii*) how to compute these 'optimal' starting times when not all tasks need to be necessarily finished to consider the project as completed? This paper proposes a hybrid approach, combining reliability concepts with simulation and optimization techniques, to support decision makers in finding the optimal starting times for each task under the described uncertainty scenario.

Keywords: project scheduling, reliability analysis, simulation, optimization

1. INTRODUCTION
In a global economy, production and transportation logistics are becoming increasingly complex due to the existence of joint projects where different departments (or even companies) cooperate to manufacture a final product or provide a given service. A project can be frequently decomposed as a set of independent and parallel tasks (e.g., production lines), some of which need to be finished in order to consider the project as completed on a given deadline (Figure 1).

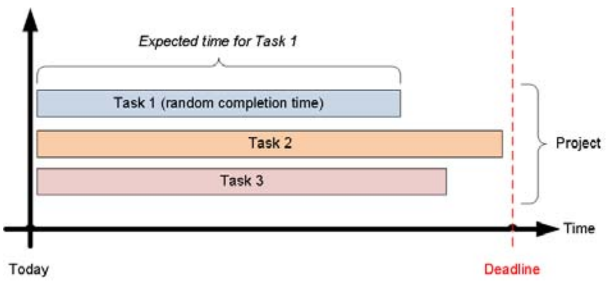


Figure 1: A Simple Project with Three Parallel Tasks

Notice that finishing all the tasks in a project might not be a necessary condition in order to consider the project as completed. In effect, depending on the final product or service specifications there might be a certain degree of redundancy in these tasks, which allows the project to be completed –at least partially– if some combinations of these tasks are satisfactorily concluded by the deadline. Thus, for instance Figure 2 shows a simple logical representation of a project ending condition: the associated project will not be terminated (i.e., it will "survive" the deadline) as far as any of the following combinations occur: (*i*) Tasks 1 and 2 are still not accomplished by the deadline; or (*ii*) Tasks 1 and 3 are still not finished by then. Thus, while in classical reliability analysis the goal is to increase the probability that a system survives a given target time, in our study the goal will be related to reducing the probability that a project survives a user-defined deadline.

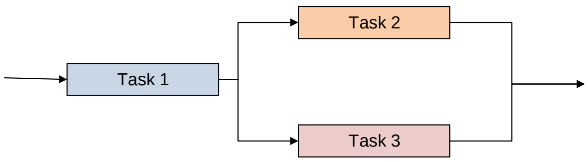


Figure 2: Logical Representation of Ending Condition

From a managerial perspective, it is usually not efficient or desirable to assign scarce production resources to a new project earlier than strictly necessary, especially if these resources are currently occupied in other projects. Thus, a manager might be interested in finding the starting time for each task in the new project that maximizes the total delay time while ensuring that the project will be completed by the deadline (Figure 3). Also, the manager might be interested in analyzing how these starting times might vary as different combinations of concluded tasks account for the project termination. In this paper, we will assume that each task is a process which duration can be modeled as a random variable. Since the problem becomes stochastic, the goal will be to maximize total delay time subject to achieving project competition by the deadline with a user-specified probability.
To support the manager during this complex decision-making process, we propose a hybrid approach in which simulation is combined with reliability analysis and optimization. One of the main benefits of our approach is that it does not make any assumption on the probability distributions employed to model the random

duration times associated with each task, i.e.: being based on simulation, there is no need to assume exponential or Normal times –instead, any empirical or theoretical distribution based on historical data can be employed to model the behavior of these times.

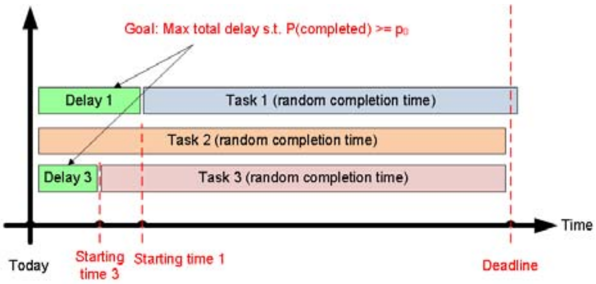


Figure 3: Determining Starting Times

The rest of the paper is organized as follows: Section 2 provides a review of related work on the use of simulation methods in reliability analysis. Section 3 describes the main ideas behind our hybrid simulation-optimization approach. An illustrative numerical example is discussed in Section 4, and the obtained results analyzed in Section 5. Finally, we conclude this paper by summarizing its main findings and highlighting future research lines in Section 6.

2. SIMULATION IN RELIABILITY ANALYSIS

Reliability or survival analysis of time-dependent systems is a research area with applications in engineering (Skrzypczak et al. 2017, Zheng and Chang 2017) as well as in experimental and social sciences (Topaloglu et al. 2016, Levin and Kalal 2003). Many works have discussed the benefits of maintenance policies in systems reliability. Some of these works highlight the fact that system management concepts, such as aging, repair obsolesce and renovation, are not easily captured by analytical models (Borgonovo et al. 2000).

As has been pointed out by many authors, when dealing with real-life complex systems only simulation techniques –such as Monte Carlo simulation (MCS) or discrete-event simulation (DES)–, can be useful to obtain credible predictions for system reliability and availability parameters (Billinton and Wang 1999). In fact, simulation has been revealed as a powerful tool in solving many engineering problems (Grasas et al. 2016, Sobie et al. 2018). This is due to the fact that simulation methods can model real-systems behavior with great detail (Dubi 2000). In addition, simulation methods can provide supplementary information about system internal behavior or about critical components from a reliability point of view. Applications of simulation techniques in the reliability field allow to model details such as component dependencies, dysfunctional behavior of components, etc. (Labeau and Zio 2000, Labeau and Zio 2002). Examples of works which deal with applications of simulation techniques in reliability are the ones of Barata et al. (2001) –who employs MCS in the modeling of components' degradation processes taking place in nuclear power plants–, or Barata et al. (2002), who use MCS in the modeling of repairable multi-component deteriorating systems (e.g., offshore structures and aerospace components). Juan and Vila (2002) proposed a MCS approach for determining the reliability of complex systems, while Faulin et al. (2007, 2008) proposed both MCS and DES algorithms for dealing with reliability and availability issues in complex telecommunication networks. Similarly, Cabrera et al. (2014) proposed a hybrid simulation-optimization approach for optimizing the availability of distributed computer systems. For additional examples of works linking simulation and reliability, the reader is addressed to Faulin et al. (2010).

3. OUR SIM-OPT SOLVING APPROACH

Following our discussion in the Introduction section, and under an uncertainty scenario in which the completion time associated with each task is a random variable, our goal is to maximize the total delay time –sum of tasks' delay times– subject to satisfying a user-specified probability of completing the project on or before a given deadline. In our approach, we assume that the specific probability distributions modeling each task's completion time is known. This could be achieved, for instance, by fitting historical data on the duration of each task by the most appropriate probability distribution –in case no theoretical distribution offers a reasonable fitting, then an empirical distribution could be used instead. Then, for each task, it is possible to plot its survival or reliability function based on the fitted probability distribution. Notice that the introduction of a delay in the starting time of task i shifts the associated survival function to the right, thus increasing the probability that task i cannot be terminated before the common due date (Figure 4).

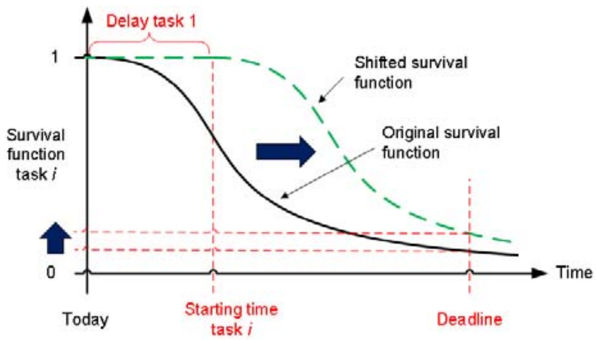


Figure 4: Effect of Delays on Task's Survival Functions

Thus, the question for the decision maker is how much can we extend these delays (i.e., increase the reliability level of each task at the deadline) without violating the probability constraint on the project completion. In order to provide an informed answer to this question, the following simulation-optimization method is proposed:

1. For each task i in the project ($i = 1, 2,\ldots, m$), use MCS to generate n random observations

(with n large enough) of its completion time, i.e., $t_{i1}, t_{i2}, \ldots, t_{in}$.

2. For each task i, consider the decision variable d_i, which represents the delay time associated with task i. Then, consider the n observations of the 'shifted' completion time for task i given by $d_i + t_{i1}, d_i + t_{i2}, \ldots, d_i + t_{in}$.
3. Use the shifted completion times from the previous step and the logical termination condition of the project to obtain n observations for the project completion time – which, at this stage, will be a function of the decision variables $d_{i1}, d_{i2}, \ldots, d_{im}$. Typically, the logical termination condition of the project is represented by a series of minimal paths that need to be terminated (all of them) before the project can be considered as completed. Each of these paths is composed of a series of tasks, and it terminates as soon as one of these tasks is finished (Faulin et al., 2008).
4. At this point, either an exact optimization solver (if m is reasonably low) or a metaheuristic algorithm is used to obtain (near-)optimal values for the decision variables. Notice that the probabilistic constraint makes the model non-linear, and thus linear programming solvers cannot be used.
5. Once (near-) optimal values for $d_{i1}, d_{i2}, \ldots, d_{im}$ have been obtained, it is convenient to perform several reliability-related studies, including: obtain statistics (average, variance, and quartiles) on the project's completion time, estimate the project survival function by measuring the probability that it might be finished at different target times, analyze how this survival function varies as different delay levels are considered, etc.

4. A NUMERICAL EXAMPLE

For the shake of illustrating the previous ideas with a numerical example, the following project was generated, solved and analyzed.

- Our project consists of 9 independent tasks to be carried in parallel. Each of these tasks has a random completion time which follows a probability distribution as depicted in Table 1, where the second column 'Alpha' refers to the shape of the Weibull distribution and the third column 'Beta' to the scale.
- For each task, a total of 10,000 random observations have been generated using the Minitab statistical software (any other statistical package could have been used instead). Then, an Excel model was constructed with these random observations and according to the procedure described in the previous section (Figure 5a).

Table 1: Probability Weibull Distributions for Each Task

Task	Alpha (shape)	Beta (scale)
1	2.8	1.8
2	2.7	1.7
3	2.6	1.6
4	2.5	1.6
5	2.4	1.4
6	2.2	1.2
7	2.3	1.3
8	2.1	1.1
9	2.0	1.0

	A	B	C	D	E	F	G	H	I	J	K
5			Tasks								
6		Weibull	1	2	3	4	5	6	7	8	9
7		alpha (shape)	2,8	2,7	2,6	2,5	2,4	2,2	2,3	2,1	2,0
8		beta (scale)	1,8	1,7	1,6	1,6	1,4	1,2	1,3	1,1	1,0
9		Observation	Task completing time -- We(alpha, beta)								
10		1	1,1367	0,7810	1,6554	0,9660	1,2363	1,1417	0,4388	1,3037	1,3375
11		2	1,2635	1,6395	1,1317	0,5584	1,9697	1,1294	1,4368	0,7213	0,6071

Figure 5: Simulated Random Completion Time for Each Task

- The delay time for each task is considered as decision variable. For each observation, the shifted completion time associated with delay time is calculated (Figure 6).

	L	M	N	O	P	Q	R	S	T
5	Tasks								
6	1	2	3	4	5	6	7	8	9
7	Delay time (decision variables) with 0 <= delay <= deadlinie								
8	0,5143	0,5182	0,5214	0,5153	0,5214	0,5142	0,5145	0,5176	0,5183
9	Delay time + Task completing time								
10	1,6510	1,2993	2,1768	1,4813	1,7577	1,6559	0,9534	1,8213	1,8558
11	1,7778	2,1578	1,6531	1,0737	2,4911	1,6436	1,9514	1,2389	1,1254

Figure 6: Shifted Completion Time for Each Task

- The project will be considered completed as soon as all the 7 paths (combination of tasks) represented in Table 2 are finished. For a path to be finished, it is enough that any of the tasks in it has been terminated. Therefore, the completion time for a given path is equal to the minimum shifted completion time of the task belonging to the given path. Thus, the project could be completed even if not all tasks have been finished by the deadline. For each observation, the maximum shifted completion time of the paths is considered as project finishing time (Figure 7, U9:AB11).

Table 2: Project-completion Paths

Paths of tasks		
1 – 4 – 7	2 – 4 – 7	3 – 6 – 9
1 – 4 – 8 – 9	2 – 4 – 8 – 9	N/A
1 – 5 – 6 – 9	2 – 5 – 6 – 9	N/A

- The common deadline was set to 2.0 time units (e.g., months, years, etc., in coherence with the units of the Weibull parameters), while the user-specified terminating probability was set to 0.90. The project termination situation is checked if it is completed before the determined deadline or no and consequently the probability of the project termination is calculated (Figure 7, AC3:AD7).

	U	V	W	X	Y	Z	AA	AB	AC	AD
1										
2										
3								E[time]	Sum delays:	4,66
4								1,5609	Deadline:	2,00
5	Paths (all need to be completed for the project to finish)							StDev	User Prob.:	0,90
6	1	2	3	4	5	6	7	0,3364	Observations:	10.000
7	Tasks to be finished for the path to be completed							Project finishing time	P(finished):	0,90
8	1v4v7	1v4v8v9	1v5v6v9	2v4v7	2v4v8v9	2v5v6v9	3v6v9			
9	Path completion time								Completed?	
10	0,9534	1,4813	1,6510	0,9534	1,2993	1,2993	1,6559	1,6559	1	
11	1,0737	1,0737	1,1254	1,0737	1,0737	1,1254	1,1254	1,1254	1	

Figure 7: Near-optimal Solution for the Example

In the proposed non-linear model, the objective is to minimize the summation of the delays. As constraints, *a*) the probability of the project termination should be more or equal than the user-specified probability, *b*) delay variable decisions should be less than deadline and finally *c*) these variables should be positive. Components of the constructed model (objective and constraints) in GRG Nonlinear solver are shown in Figure 8. The model was 'optimized' using the GRG Nonlinear solver integrated in the Excel Solver plug-in.

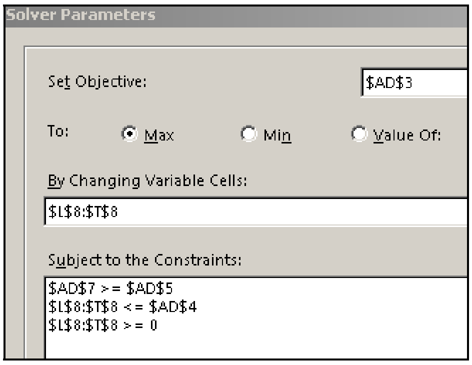


Figure 8: Non-linear solver parameters (objective and constraints)

The optimization process took about 2.5 minutes in a standard laptop computer (Intel Core i5 CPU @ 2.3 GHz, 8 GB RAM). Figure 6 (L8:T8) shows a partial screenshot of the Excel model, including the near-optimal delay values per task provided by the non-linear solver, which are the following ones: d_1 = 0.5143, d_2 = 0.5182, d_3 = 0.5214, d_4 = 0.5153, d_5 = 0.5214, d_6 = 0.5142, d_7 = 0.5145, d_8 = 0.5176, d_9 = 0.5183. These values were obtained by the non-linear solver after setting 0.5 as the initial value for all the tasks' delays (we have noticed that, probably due to convergence issues of the solver, the final solution provided might vary somewhat depending on these initial values).

5. ANALYSIS OF RESULTS

Figure 9 shows two survival functions for the project. These survival functions have been generated using the Kaplan-Meier estimator (Meeker and Escobar 1998) for non-censured data. In this case, the data refer to the observations provided by our method for the project completion times. One set of observations was obtained without assuming any delay in the tasks (i.e., all tasks were initiated from the very beginning), while the other set was obtained using the optimal delay values provided by the Excel solver.

Notice how the delayed survival function is shifted-to-the-right with respect the non-delayed one. Note also that, as it was anticipated, the delayed survival function intersects the deadline at 2.0 exactly when the survival probability reaches the value 0.1.

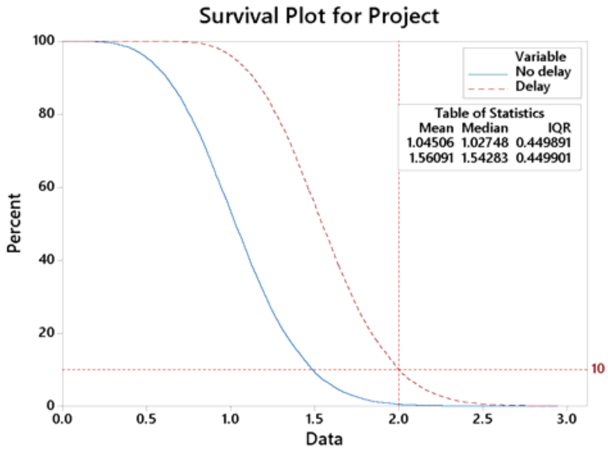


Figure 9: Survival Function for Project

Finally, Figure 10 shows a trade-off plot between total delay and probability of completing the project on or before the deadline. As expected, the higher the completion probability the lower the total delay that can be used –i.e., the sooner the resources must be assigned to the project.

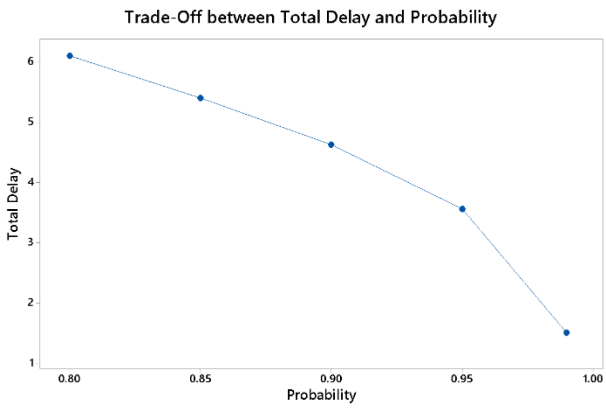


Figure 10: Trade-Off between Delay and Probability

6. CONCLUSIONS

Setting starting times for multi-task projects under uncertainty conditions can be a challenging problem for managers. This paper proposes a simulation-optimization method able to provide informed decisions on how to set these starting times, so that a given probability of project completion is achieved. Our method is combines simulation and optimization the best of our knowledge, it is the first time that such a relevant problem is discussed in the scientific literature and a solving approach is proposed.

Our results contribute to quantify how the survival function of the project evolves as more delay is included in the process. They also provide numerical values for quantifying the trade-off between the user specified probability of project termination and the maximum delay that can be considered without violating the probabilistic constraint on project completion.

Several research lines emerge from this preliminary work. First, a more robust non-linear solver should be used for larger cases. In fact, metaheuristic approaches could be a good alternative to exact methods for solving large-scale instances in short computing times. Also, other project topologies can be explored, including interdependencies among tasks. Finally, the tasks themselves could be complex processes, e.g., flow-shops, that also require some pre-optimization process.

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness and FEDER (TRA2013-48180-C3-P, TIN2015-66951-C2-2-R, MTM2015-64465-C2-1-R), and the Erasmus+ program (2016-1-ES01-KA108-023465).

REFERENCES

Barata J., Guedes-Soares C., Marseguerra M., and Zio E., 2001. Modelling components degradation processes of Monte-Carlo simulation. In: Zio E., Demichela M. and Piccinini N., editors. Safety and Reliability: towards a safer world. Proceedings of the European conference on safety and reliability ESREL 2001, 16-20 September, Torino, Italy, pp. 879–886.

Barata J., Guedes-Soares C., Marseguerra M. and Zio E., 2002. Simulation modelling of repairable multi-component deteriorating systems for ''on-condition'' maintenance optimisation. Reliability Engineering and System Safety, 77, 255–267.

Billinton R. and Wang P., 1999. Teaching distribution systems reliability evaluation using Monte Carlo simulation. IEEE Transaction on Power Systems, 14, 397–403.Blischke W.R. and Prabhakar-Murthy D.N., 2000. Reliability. Modeling, prediction, and optimization. New York: Wiley.

Borgonovo E., Marseguerra M. and Zio, E., 2000. A Monte Carlo methodological approach to plant availability modelling with maintenance, aging and obsolescence. Reliability Engineering and System Safety, 67, 61–73.

Cabrera G., Juan A., Lazaro D., Marques J. and Proskurnia I., 2014. A simulation-optimization approach to deploy Internet services in large-scale systems with user-provided resources. Simulation: Transactions of the Society for Modeling and Simulation International, 90 (6), 644–659.

Dubi A., 2000. Monte Carlo applications in systems engineering. New York: Wiley.

Faulin J., Juan, A., Martorell, S. and Ramirez-Marquez, J., 2010. Simulation Methods for Reliability and Availability of Complex Systems. Springer.

Faulin J., Juan A., Serrat C. and Bargueno V., 2007. Using Simulation to determine Reliability and Availability of Telecommunication Networks. European Journal of Industrial Engineering, 1 (2), 131-151.

Faulin J., Juan A., Serrat C. and Bargueno V., 2008. Predicting availability functions in time-dependent complex systems with SAEDES simulation algorithms. Reliability Engineering and System Safety, 93 (11): 1761–1771.

Grasas, A., Juan, A. and Ramalhinho, H., 2016. SimILS: A Simulation-based extension of the Iterated Local Search metaheuristic for Stochastic Combinatorial Optimization. Journal of Simulation, 10(1): 69-77.

Juan A. and Vila A., 2002. SREMS: System Reliability Using Monte Carlo Simulation with VBA and Excel. Quality Engineering, 15 (2) 333–340.

Labeau P.E. and Zio E., 2000. The mathematics of component-based and system-based Monte Carlo sampling in realistic systems engineering. Proceedings of the MRR2000, 2, 667–670.

Labeau P.E. and Zio E., 2002. Procedures of Monte-Carlo transport simulation for applications in system engineering. Reliability Engineering System Safety, 77, 217–228.

Levin M. and Kalal T., 2003. Improving product reliability: strategies and implementation. New York: Wiley.

Meeker W. and Escobar L., 1998. Statistical methods for reliability data. New York: Wiley.

Skrzypczak I., Słowik M. and Buda-Ożóg L., 2017. The Application of Reliability Analysis in Engineering Practice – Reinforced Concrete Foundation. Procedia Engineering, 193,144-151.

Sobie C., Freitas C. and Nicolai M., 2018. Simulation-driven machine learning: Bearing fault classification. Mechanical Systems and Signal Processing, 99, 403-419.

Topaloglu M., Caldibi E. and Oge G., 2016. The scale for the individual and social impact of students' social network use: The validity and reliability studies. Computers in Human Behavior, 61, 350-356.

Zheng L. and Chang Y., 2017. Risk assessment model of bottlenecks for urban expressways using survival analysis approach. Transportation Research Procedia, 25, 1544-1555.

AUTHORS BIOGRAPHY

Angel A. Juan is Associate Professor at the IN3-Open University of Catalonia. He holds a PhD in Industrial Engineering and a MSc in Mathematics. He completed a pre-doctoral internship at Harvard University and a postdoctoral internship at the MIT Center for

Transportation and Logistics. His research interests include applications of randomized algorithms and simheuristics in logistics, production, and Internet computing. He has published over 50 JCR-indexed articles in these fields.

Laura Calvet is Assistant Professor at the University of Barcelona and PhD student at the IN3-Open University of Catalonia. She holds a MSc in Statistics and Operations Research, a BSc in Economics and a BSc in Applied Statistics. Her work is related to the combination of statistical learning and simheuristics for solving complex combinatorial optimization problems under uncertainty.

Carles Serrat is Associate Professor of Applied Statistics in the Department of Mathematics at the Technical University of Catalonia (Barcelona, Spain). He holds a PhD and BS in Mathematics. Dr. Serrat develops his research in the context of the Institut d'Estadística i Matemàtica Aplicada a l'Edificació (IEMAE).

Sara Hatami is a postdoctoral researcher at the IN3-Open University of Catalonia. She holds a PhD and a MSc in Industrial Engineering. Her research focuses on the development of metaheuristics and simheuristics for industrial optimization problems, using traditional and modern Operations Research techniques and modeling complex real-life problems.

ESTIMATION OF EXPECTATIONS IN TWO-LEVEL NESTED SIMULATION EXPERIMENTS

David Fernando Muñoz

Departamento de Ingeniería Industrial y Operaciones, Instituto Tecnológico Autónomo de México, Río Hondo #1, 01080 Ciudad de México, México

davidm@itam.mx

ABSTRACT

Two-level nested simulation methods have been recently applied for the analysis of simulation experiments under parameter uncertainty. On the outer level of the nested run, we generate (n) observations of the parameters, while on the inner level; we fix the parameter on its corresponding value and generate (m) observations using a simulation model. In this paper, we focus on the output analysis of two-level stochastic simulation experiments for the case where the observations of the inner level are independent, showing how the variance of the simulated observations can be decomposed in the sum of parametric and stochastic components. Furthermore, we derive central limit theorems that allow us to compute asymptotic confidence intervals to assess the accuracy of the simulation-based estimators for the point forecast and the variance components. Theoretical results are validated through experiments using a forecasting model for sporadic demand, where we have obtained analytical expressions for the point forecast and the variance components.

Keywords: Bayesian forecasting, stochastic simulation, parameter uncertainty, two-level simulation

1. INTRODUCTION

Simulation is widely recognized as an effective technique for producing forecasts, evaluating risk, animating and illustrating the performance of a system over time (see, e.g., Kelton et al. 2012). When there is uncertainty in some components of a simulation model, these random components are modeled using probability distributions and/or stochastic processes that are generated during the simulation run, in order to produce a stochastic simulation.

The method that is usually applied to estimate a performance measure θ (e.g., an expectation) in transient simulation, consists in computing an estimator θ_n from the output of n independent replications of the simulation experiment. This estimator must be consistent, i.e., it must satisfy $\theta_n \Rightarrow \theta$, as $n \to \infty$ (where "$\Rightarrow$" denotes weak convergence of random variables). Consistency guarantees that the estimator approaches the parameter as the number of replications n increases, and the accuracy of the simulation-based estimator θ_n is typically assessed by an asymptotic confidence interval (ACI) for the parameter. The expression for an ACI is usually obtained through a Central Limit Theorem (CLT) for the estimator θ_n (see, for example, chapter III of Asmussen and Glynn 2007).

For $i = 1$ to n:
  For $j = 1$ to m:
    a. Generate (independently) a value θ_j by sampling from $p(\theta|x)$.
    b. Run a simulation experiment with $\Theta = \theta_j$ to obtain an independent replication W_{ij} of $W = g(Y(s), 0 \le s \le T; \Theta)$
  End Loop
End Loop
Compute $\hat{\alpha}_i = \frac{1}{m}\sum_{i=1}^{m} W_{ij}, i = 1, \ldots, m, \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i$.

Figure 1: Two level algorithm for calculating a point estimator under parameter uncertainty

In contrast to the estimation of performance measures, input parameters of a simulation experiment are usually estimated from real-data observations (x) and, while the majority of applications covered in the relevant literature assume that no uncertainty exists in the value of these parameters, the uncertainty can actually be significant when little data is available. In these cases, Bayesian statistics can be used to incorporate this uncertainty in the output analysis of simulation experiments via the use of a posterior distribution $p(\theta|x)$. A Bayesian approach using simulation as a forecasting tool has been reported in diverse areas; for example, healthcare (see, e.g., Santos et al. 2013) or software development (see, e.g., Lee et al. 2009). A methodology currently proposed for the analysis of simulation experiments under parameter uncertainty and, in particular, for the estimation of expected values, is a two-level nested simulation method (see, e.g., Zouaoui and Wilson 2003, L'Ecuyer 2009; Andradóttir and Glynn 2016). In the outer level, we simulate (n) observations for the parameters from a posterior distribution $p(\theta|x)$, while in the inner level we simulate (m) observations for the response variable with the parameter fixed at the value generated in the outer level (see Figure 1). In this paper, we focus on the output analysis of two-level simulation experiments, for the particular case when the observations of the inner level are independent, showing how the variance of a simulated observation can be decomposed into parametric and stochastic variance components. Afterwards, we derive a CLT for both the estimator of the point forecast and the estimators of the variance components. Our CLT´s allows us to compute an ACI for each estimator. Our results are validated through experiments with a forecasting model for sporadic demand reported in Muñoz and Muñoz (2011).

Following this introduction, we present the proposed methodology for the construction of an ACI for the point forecast and the variance components in a two level simulation experiment. Afterwards, we present an illustrative an example that has an analytical solution for the parameters of interest in this paper. This example is used in the following section to illustrate the application and validity of our proposed methodologies for the construction of an ACI. Finally, in the last section, we present conclusions and directions for future research.

2. METHODOLOGY

In order to identify the variance components in each observation W_{ij} of the algorithm illustrated in Figure 1, let $\mu(\Theta)=E[W_{11}|\Theta]$, and $\sigma^2(\Theta)=E\left[W_{11}^2|\Theta\right]-\mu(\Theta)^2$. Under this notation, the point forecast is $\alpha=E[\mu(\Theta)]$, and the variance of W_{ij} is

$$\begin{aligned} V\left[W_{ij}\right] &\overset{def}{=} E\left[W_{ij}^2\right]-E\left[W_{ij}\right]^2 \\ &= E\left[E\left[W_{ij}^2|\Theta\right]-\mu(\Theta)^2\right]+E\left[\mu(\Theta)^2\right]-E[\mu(\Theta)]^2 \\ &= \sigma_1^2+\sigma_2^2, \end{aligned} \tag{1}$$

for i = 1,…,n; j = 1,…,m, where $\sigma_1^2=E\left[\sigma^2(\Theta)\right]$, and $\sigma_2^2=V[\mu(\Theta)]\overset{def}{=}E\left[\mu(\Theta)^2\right]-E[\mu(\Theta)]^2$. It is worth mentioning that, in relevant literature, σ_1^2 is commonly referred to as stochastic variance and σ_2^2 as parametric variance.

2.1. Point Estimators

In this paper, we are interested in both the estimation of the point forecast $\alpha=E[\mu(\Theta)]$ and the estimation of the variance components of the observations generated by the algorithm of Figure 1 and defined in (1), thus we first define the natural point estimators

$$\hat{\alpha}(n)=\frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i, \tag{2}$$

$$\hat{\sigma}_T^2(n)=\frac{1}{n-1}\sum_{i=1}^{n}(\hat{\alpha}_i-\hat{\alpha}(n))^2, \hat{\sigma}_2^2(n)=\frac{1}{n}\sum_{i=1}^{n}S_i^2, \tag{3}$$

where $\hat{\alpha}_i=m^{-1}\left(\sum_{j=1}^{m}W_{ij}\right)/m$, and $S_i^2=\left[\sum_{j=1}^{m}(W_{ij}-\hat{\alpha}_i)^2\right]/(m-1)$, for i = 1,...,m. Note that the $\hat{\alpha}_i$ are independent and identically distributed (i.i.d.) with expectation $E[\hat{\alpha}_1]=\alpha$ and variance

$$\begin{aligned} V[\hat{\alpha}_1] &= E\left[(\hat{\alpha}_1-\alpha)^2\right] \\ &= m^{-2}\left(mE\left[(W_{11}-\alpha)^2\right]+m(m-1)E[(W_{11}-\alpha)(W_{12}-\alpha)]\right) \\ &= m^{-1}\left(\sigma_1^2+\sigma_2^2\right)+m^{-1}(m-1)\sigma_2^2=\sigma_2^2+m^{-1}\sigma_1^2 \end{aligned} \tag{4}$$

On the other hand, the S_i^2 are i.i.d. with expectation $E\left[S_1^2\right]=s_2^2$. Thus, the next proposition follows from the Law of Large Numbers (Theorem 5.4.2 of Chung 2001).

Proposition 1. Given $m\geq 1$, if $E[|W_{11}|]<\infty$ then $\hat{\alpha}(n)$ and $\hat{\sigma}_T^2(n)$ are unbiased and consistent estimators for α and $\sigma_2^2+m^{-1}\sigma_1^2$ (as $n\to\infty$), respectively. Furthermore, if $m\geq 2$, then $\hat{\sigma}_2^2(n)$ is an unbiased and consistent estimator for σ_2^2 (as $n\to\infty$).

2.2. Accuracy of the Point Estimators

As we established in the previous Section, the point estimators proposed in (2) and (3) are consistent, and thus converge to the corresponding parameters values as $n\to\infty$. Nonetheless, to establish the level of accuracy of these estimators, we must establish a CLT for each estimator that allows us to calculate the corresponding ACI. Note that both $\hat{\alpha}(n)$ and $\hat{\sigma}_2^2(n)$ are averages of i.i.d observations, thus the next proposition follows from the classic CLT for i.i.d. observations.

Proposition 2. Given $m\geq 1$, if $E\left[W_{11}^2\right]<\infty$ then

$$\frac{\sqrt{n}(\hat{\alpha}(n)-\alpha)}{\sqrt{V(\hat{\alpha}_1)}}\Rightarrow N(0,1),$$

as $n\to\infty$. Furthermore, if $m\geq 2$ and $E\left[W_{11}^4\right]<\infty$, then

$$\frac{\sqrt{n}\left(\hat{\sigma}_2^2(n)-\sigma_2^2\right)}{\sqrt{V\left(S_1^2\right)}}\Rightarrow N(0,1),$$

where $N(0,1)$ denotes the standard normal distribution, $\hat{\alpha}(n),\hat{\sigma}_2^2(n),\hat{\alpha}_1,S_1^2$ are defined in (2) and (3), $V[\hat{\alpha}_1]$ is defined in (4), and $V\left[S_1^2\right]=E\left[\left(S_1^2-\sigma_2^2\right)^2\right]$.

Since we have consistent estimators for $V[\hat{\alpha}_1]$ and $V\left[S_1^2\right]$, the next corollary follows from Proposition 1 and Slutsky's Theorem (see, e.g., Serfling 2009).

Corollary 1. Under the same notation and assumptions from Proposition 2, we have

$$\frac{\sqrt{n}(\hat{\alpha}(n)-\alpha)}{\sqrt{\hat{V}(\hat{\alpha}_1)}}\Rightarrow N(0,1), \text{ and } \frac{\sqrt{n}\left(\hat{\sigma}_2^2(n)-\sigma_2^2\right)}{\sqrt{\hat{V}\left(S_1^2\right)}}\Rightarrow N(0,1),$$

for $m \geq 1$ and $m \geq 2$, respectively, as $n \to \infty$, where

$$\hat{V}[\hat{\alpha}_1] = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\alpha}_i - \hat{\alpha}(n))^2 ,$$

$$\hat{V}\left[S_1^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}\left(S_i^2 - \bar{S}^2\right)^2, \bar{S}^2 = \frac{1}{n}\sum_{i=1}^{n} S_i^2 .$$

To obtain a CLT for $\hat{\sigma}_T^2(n)$, note that this estimator is the variance of a set of i.i.d. observations, thus we can use the following Lemma. We omit the proof of this Lemma, nonetheless, it can be proven by applying the Delta Method (see, e.g., Proposition 2 of Muñoz and Glynn 1997).

Lemma 1. If $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with $E\left[X_1^4\right] < \infty$, then

$$\frac{n^{1/2}\left(S^2(n) - \sigma^2\right)}{\sigma_S} \Rightarrow N(0,1),$$

as $n \to \infty$, where $\sigma^2 = \mu_2 - \mu_1^2$,

$$\sigma_S = \sqrt{8\mu_1^2\mu_2 - 4\mu_1^4 - 4\mu_1\mu_3 + \mu_4 - \mu_2^2} ,$$

$\mu_k = E\left[X_1^k\right]$, k = 1, 2, 3, 4;

$S^2(n) = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_1)^2$, $\hat{\mu}_1 = n^{-1}\sum_{i=1}^{n} X_i$.

Corollary 2. Under the same assumptions as in Lemma 1 we have

$$\frac{n^{1/2}\left(S^2(n) - \sigma^2\right)}{\hat{\sigma}_S} \Rightarrow N(0,1),$$

as $n \to \infty$, where

$$\hat{\sigma}_S = \sqrt{8\hat{\mu}_1^2\hat{\mu}_2 - 4\hat{\mu}_1^4 - 4\hat{\mu}_1\hat{\mu}_3 + \hat{\mu}_4 - \hat{\mu}_2^2} ,$$

$$\hat{\mu}_k = n^{-1}\sum_{i=1}^{n} X_i^k .$$

Corollary 2 follows from the fact that $\hat{\mu}_k$ is an unbiased and consistent estimator of μ_k, and the next corollary follows from the fact that $\hat{\sigma}_T^2(n)$ is the sample variance of the $\hat{\alpha}_i$.

Corollary 3. Given $m \geq 1$, if $E\left[W_{11}^4\right] < \infty$ then

$$\frac{\sqrt{n}\left(\hat{\sigma}_T^2(n) - \left(\sigma_2^2 + m^{-1}\sigma_1^2\right)\right)}{\sqrt{\hat{V}_T}} \Rightarrow N(0,1),$$

as $n \to \infty$, where

$$\hat{V}_T = 8\bar{\alpha}_1^2\bar{\alpha}_2 - 4\bar{\alpha}_1^4 - 4\bar{\alpha}_1\bar{\alpha}_3 + \bar{\alpha}_4 - \bar{\alpha}_2^2 ,$$

$$\bar{\alpha}_k = n^{-1}\sum_{i=1}^{n}\hat{\alpha}_i^k .$$

Let $0 < \beta < 1$, and using corollaries 1 and 3 we can establish a $100\beta\%$ ACI for the point forecast α, and variance components σ_2^2 and $\sigma_T^2 = \sigma_2^2 + m^{-1}\sigma_1^2$; each ACI is centered in the corresponding point estimator, $\hat{\alpha}(n), \hat{\sigma}_2^2(n)$ or $\hat{\sigma}_T^2(n)$, and the halfwidths are given by

$$H_n(\alpha) = z_\beta \frac{\sqrt{\hat{V}(\hat{\alpha}_1)}}{\sqrt{n}} , \tag{5}$$

$$H_n\left(\sigma_2^2\right) = z_\beta \frac{\sqrt{\hat{V}\left(S_1^2\right)}}{\sqrt{n}} , \tag{6}$$

$$H_n\left(\sigma_2^2\right) = z_\beta \frac{\sqrt{\hat{V}\left(S_1^2\right)}}{\sqrt{n}} , \tag{7}$$

for α, σ_2^2 and σ_T^2, respectively, where z_β is the $(1-\beta/2)-$quantile of a $N(0,1)$ distribution, $\hat{V}(\hat{\alpha}_1)$ and $\hat{V}\left(S_1^2\right)$ are defined in Corollary 1, and $\hat{V}_T$ is defined in Corollary 3.

Note that the ACI proposed in (5), (6) and (7) assume that the value of *m* from the algorithm of Figure 1 is fixed and the accuracy of the estimator improves as *n* (the number of observations in the outer level) increases (in turn, the halfwidth of the ACI gets smaller). Given that we can build a valid ACI for any value of *m*, a relevant question is how to find an adequate value of *m* to get an acceptable level of accuracy in a reasonable amount of running time. To answer this question for the case of the point estimator of α, let us fix the total number of iterations in the algorithm of Figure 1 to *k* = *nm*, and note from (3) that

$$n^{-1}V[\hat{\alpha}_1] = k^{-1}\left(m\sigma_2^2 + \sigma_1^2\right) \tag{8}$$

takes its minimal value for *m* = 1, suggesting that the point estimator $\hat{\alpha}(n)$ defined in (2) is more accurate as *m* approaches the value of 1. Note that for *m* = 1, a fixed number of iterations *k* = *nm* is convenient (from the point of view of running time), when the computation of W_{ij} requires the same or more computation time as Θ_i, as suggested in the relevant literature (see, for example, Andradóttir and Glynn 2016). Furthermore, if we allow *m* to increase with *n*, we can obtain the following proposition. We omit the proof, but mention that it follows from Lindeberg-Feller Theorem (Theorem 7.2.1 of Chung 2001).

Proposition 3. Given $0 < p \leq 1$, if $m = \lfloor n^{-1+1/p} \rfloor$ and $E\left[W_{11}^2\right] < \infty$ then

$$\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sqrt{V[\hat{\alpha}_1]}} \Rightarrow N(0,1),$$

as $n \to \infty$, where $V[\hat{\alpha}_1]$ is defined in (3).

If, once again, we set the total number of iterations in the algorithm of Figure 1 to *k* = *nm*, we let $n \approx k^p$, $m \approx k^{1-p}$ and *nm* = *k*, it follows from Proposition 3 that the asymptotic variance of $\hat{\alpha}(n)$ is $n^{-1}V[\hat{\alpha}_1]$ for every $0 < p \leq 1$. Note that this metric reaches its minimum value when $p = 1$, that is, when *n* = *k* and *m* = 1. However, note that we need $m \geq 2$ to estimate σ_2^2.

In the following section, we report some experiments that confirm our theoretical results.

3. AN EXAMPLE WITH ANALYTICAL SOLUTION

The following model (reported in Muñoz and Muñoz 2011) has been proposed to forecast sporadic demand. We can use this model to find analytical expressions for the parameters considered in this paper. This model is used in the following section to illustrate the validity of the ACI's proposed in the previous section.

Client arrivals for a particular item in a shop follows a Poisson process, yet there is uncertainty in the arrival rate , so that given , interarrival times between clients are i.i.d. with exponential density:

$$f(y|\theta_0) = \begin{cases} \theta_0 e^{-\theta_0 y}, & y > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where $\theta_0 \in S_{00} = (0, \infty)$. Every client can order j units of this item with probability Θ_{1j}, $j = 1, \ldots, q$, $q \geq 2$. Let $\Theta_1 = (\Theta_{11}, \ldots, \Theta_{1(q-1)})$ and $\Theta_{1q} = 1 - \sum_{j=1}^{q-1} \Theta_{1j}$, then $\Theta = (\Theta_0, \Theta_1)$ is the parameter vector, and $S_0 = S_{00} \otimes S_{01}$ is the parameter space, where total demand during a period with length T is

$$D = \begin{cases} \sum_{i=1}^{N(T)} U_i, & N(T) > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

where $N(s)$ is the number of client arrivals during the interval $[0, s]$, $s \geq 0$, and $U_1, U_2 \ldots$ are the individual demands (conditionally independent relative to Θ). The information about Θ consists of i.i.d. observations $v = (v_1, \ldots, v_r)$, $u = (u_1, \ldots, u_r)$ of past clients, where v_i is the interarrival time between client i and client (i - 1), and u_i is the number of units ordered by client i. By using Jeffrey's non-informative prior as the prior density for Θ, we obtain the posterior density (see Muñoz and Muñoz 2011 for details) $p(\theta|x) = p(\theta_0|v)p(\theta_1|u)$, where $x_i = (v_i, u_i)$, $i = 1, \ldots, r$, $x = (x_1, \ldots, x_r)$, $\theta = (\theta_0, \theta_1)$,

$$p(\theta_0|v) = \frac{\theta_0^{r-1} \left(\sum_{i=1}^{r} v_i \right)^r e^{-\theta_0 \sum_{i=1}^{r} v_i}}{(r-1)!}, \tag{11}$$

$$p(\theta_1|u) = \frac{\left(1 - \sum_{j=1}^{q-1} \theta_{1j} \right)^{c_q - 1/2} \prod_{j=1}^{q-1} \theta_{1j}^{c_j - 1/2}}{B(c_1 + 1/2, \ldots, c_q + 1/2)}, \tag{12}$$

where $c_j = \sum_{i=1}^{r} I[u_i = j]$, and

$$B(a_1, \ldots, a_q) = \prod_{j=1}^{q} \Gamma(a_j) / \Gamma\left(\sum_{j=1}^{q} a_j \right),$$

for $a_1, \ldots, a_q > 0$. Using this notation, we can prove that

$$\alpha = E[T\Theta_0] \sum_{j=1}^{q} j p_j, \tag{13}$$

$$\sigma_1^2 = \frac{E\left[T^2 \Theta_0^2\right]}{(q_0 + 1)} \sum_{j=1}^{q} j^2 p_j + \frac{E[T\Theta_0]^2 [(q_0 / n) - 1]}{(q_0 + 1)} \left(\sum_{j=1}^{q} j p_j \right)^2, \tag{14}$$

$$\sigma_2^2 = E[T\Theta_0] \sum_{j=1}^{q} j^2 p_j, \tag{15}$$

where $p_j = q_j / q_0, q_j = c_j + 1/2,\ j = 1, \ldots, q,\ q_0 = \sum_{j=1}^{q} q_j$

$$E[T\Theta_0] = Tr \left(\sum_{i=1}^{r} v_i \right)^{-1},\ E\left[T^2 \Theta_0^2\right] = T^2 r (1 + r) \left(\sum_{i=1}^{r} v_i \right)^{-2}.$$

4. EXPERIMENTAL RESULTS

To validate the ACI proposed in (5), (6) and (7), we conducted some experiments with the example from the previous section to illustrate the estimation of α, σ_2^2 and σ_T^2. We considered the values $T = 15$, $r = 20$, $\sum_{i=1}^{r} x_i = 10$, $q = 5$, $c_1 = 5$, $c_2 = 3$, $c_3 = 2$, $c_4 = 3$, $c_5 = 7$. With this data, the point forecast is $\alpha \approx 95.333$, and the variance components are $\sigma_1^2 \approx 380.667$, $\sigma_2^2 \approx 568.598$.

Table 1: Performance of the 90% ACI based on 1000 replications with m = 2

	n = 100, m = 2			n = 1000, m = 2			n = 10000, m = 2		
	α	σ_2^2	σ_T^2	α	σ_2^2	σ_T^2	α	σ_2^2	σ_T^2
Coverage	0.896	0.893	0.868	0.892	0.902	0.892	0.898	0.910	0.897
Mean HW	4.493	91.327	183.384	1.423	29.616	60.703	0.450	9.419	19.386
St Dev HW	0.352	18.481	45.769	0.036	2.092	5.606	0.004	0.215	0.613
R.M.S.E.	2.716	53.794	117.903	0.875	17.901	38.067	0.278	5.673	11.995
Bias	-0.015	-1.547	1.918	0.011	-0.326	-0.226	0.008	0.116	-0.048

In tables 1 and 2 we report the results of 1000 independent replications of the algorithm of Figure 1 for every 90% ACI, with values *nm* = 200, 2000, 20000, and *m* = 2, 5. Finally, with the objective of comparing the choice of *m* = 1 (which we consider optimal for the estimation of the point forecast), we also report the results of similar experiments with *nm* = 100, 10000, *m* = 1 and $m \approx (nm)^{1/3}$ (suggested by Andradóttir and Glynn, 2016, as an adequate choice for *m* in the case of biased estimators in the inner level of the algorithm of Figure 1). In Table 1, we present the coverage (fraction of ACI's that cover the parameter value), average, and standard deviation of the corresponding halfwidth defined in (5), (6) and (7), square root of the mean square error (R.M.S.E.) and empirical bias based on 1000 replications for each ACI. As observed in the table, the coverage are acceptable (very close to the nominal value of 0.9, even for m = 100). This validates the ACI defined in (5), (6) and (7). Furthermore, every performance measure of the ACI (average, standard deviation of the halfwidth, R.M.S.E., and bias) improves as the number of replications n increases.

Table 2: Performance of the 90% ACI based on 1000 replications with m = 5

	n = 40, m = 5			n = 400, m = 5			n = 4000, m = 5		
	α	σ_2^2	σ_T^2	α	σ_2^2	σ_T^2	α	σ_2^2	σ_T^2
Coverage	0.894	0.862	0.818	0.898	0.892	0.891	0.890	0.891	0.899
Halfwidth Mean	6.553	75.712	235.617	2.077	24.695	81.722	0.655	7.829	25.879
Halfwidth Std Dev	0.830	15.714	95.501	0.082	1.706	11.997	0.008	0.178	1.191
R.M.S.E.	3.978	48.081	164.462	1.248	15.262	50.276	0.405	4.911	15.557
Bias	0.134	-2.111	10.234	-0.028	-0.262	3.827	-0.024	-0.176	-0.049

In Table 2, we present the results with m = 5. We can observe that, while the coverage is close to the nominal value of 0.9, all of the performance measures for the ACI (average, standard deviation of the halfwidth, R.M.S.E., and bias) are worse (larger) tan the ones reported in Table 1, for the estimation of α and σ_T^2, and better for the estimation of σ_2^2, suggesting that, for the same number of observations nm, a lower value of m is better for estimating α and a greater value of m is better for estimating σ_2^2.

Table 3: Performance of the 90% ACI based on 1000 replications with m = 1 and $m \approx (nm)^{1/3}$

	nm = 100, m = 1		nm = 100, m = 5		nm = 10000, m = 1		nm = 10000, m = 20	
	α	σ_T^2	α	σ_T^2	α	σ_T^2	α	σ_T^2
Coverage	0.913	0.869	0.890	0.781	0.889	0.899	0.887	0.872
Halfwidth Mean	5.029	232.305	9.203	306.107	0.504	24.351	1.765	65.799
Halfwidth Std Dev	0.392	61.479	1.655	157.539	0.004	0.752	0.065	8.873
R.M.S.E.	2.933	147.120	5.507	234.267	0.308	14.570	1.115	42.196
Bias	0.020	1.367	0.004	11.555	-0.015	-0.194	0.034	-0.890

Finally, in Table 3 we show the results for the estimation of α and σ_T^2 for the cases m = 1 and $m \approx (nm)^{1/3}$. We again find that all performance measure for the ACI (average, standard deviation of the halfwidth, R.M.S.E., and bias) are worse (larger) for $m \approx (nm)^{1/3}$, confirming our finding that, for the same number of replications nm, m = 1 produces better point estimators for α than $m \approx (nm)^{1/3}$.

5. METHODOLOGY

In this paper, we propose methodologies to calculate point estimators (and their corresponding halfwidths), for both the point forecast and the variance components in two-level nested stochastic simulation experiments. This method can be applied to the construction of Bayesian forecasts using simulation models.

Both theoretical and experimental results confirm that the proposed point estimators and their corresponding halfwidths are asymptotically valid, i.e., the point estimators converge to the corresponding parameter values and the halfwidths converge to the nominal coverage as the number of replications (n) of the outer level increases.

Furthermore, given a fixed number of total observations (nm), we show that the choice of only one replication in the inner level (m = 1) provides more accurate estimators for both the point forecast (α), and the variance of the point forecast ($\sigma_1^2 + \sigma_2^2$). However, $m \geq 2$ is required for the estimation of σ_2^2.

Directions for future research on this topic includes experimentation with other point estimators, such as, quasi Monte Carlo or Simpson integration, with the objective of finding more accurate point estimators for the parameters considered in this paper.

ACKNOWLEDGMENTS

This research is supported by the Asociación Mexicana de Cultura AC.

REFERENCES

Andradóttir S. and Glynn P. W., 2016. Computing Bayesian means using simulation. ACM Transactions on Modeling and Computer Simulation, 26(2), paper 10.

Asmussen S. and Glynn, P. W., 2007. Stochastic Simulation Algorithms and Analysis. New York: Springer.

Chung K. L., 2001. A Course in Probability Theory. New York: Academic Press.

L'Ecuyer P., 2009. Quasi-Monte Carlo methods with applications in finance. Finance and Stochastics, 13(3), 307-349.

Kelton W. D., Smith J. S. and Sturrock D. T., 2013. Simio and Simulation: Modeling, Analysis, Aplications. 3rd ed. Sewickley: Simio LLC.

Lee S., Celik N. And Son Y. J., 2009. An integrated simulation modelling framework for decision aids in enterprise software development process. International Journal of Simulation and Process Modelling, 5(1), 62-76.

Muñoz D. F. and Glynn P. W. 1997. A batch means methodology for estimation of a nonlinear function of a steady-state mean. Management Science, 43(8), 1121-1135.

Muñoz D. F. and Muñoz D. F., 2011. Bayesian forecasting of spare parts using simulation. In: N. Altay and L. A. Litteral, eds. Service Parts Management: Demand Forecasting and Inventory Control. New York: Springer, 105-123.

Santos Jr E., Joo Kim K., Yu F., Li D. and Rosen J., 2013. Bayesian knowledge modelling for healthcare practices. International Journal of Simulation and Process Modelling, 8(1), 52-65.

Serfling R. J., 2009. Approximation Theorems of Mathematical Statistics. New York: John Wiley & Sons.

Zouaoui F. and J. R. Wilson, 2003. Accounting for parameter uncertainty in simulation input modeling. IIE Transactions, 35(9), 781-792.

AUTHOR BIOGRAPHY

DAVID F. MUÑOZ is Professor and Head of the Department of Industrial and Operations Engineering at the Instituto Tecnológico Autónomo de México. He holds a BS in Statistics from the National Agrarian University of Peru, a M.S. in Mathematics from Catholic University of Peru, and M.S. and Ph.D. degrees in Operations Research from Stanford University. He has been recognized as an “Edelman Laureate” for the Institute for Operations Research and the Management Sciences, for participating in the project "Indeval Develops a New Operating and Settlement System Using Operations Research", winner of the 2010 Edelman Prize. His research interests include input and output analysis for stochastic simulations and the applications of Operations Research.

ANALYSIS OF PROFITABILITY THROUGH WITH THE GENERATION OF L-SCENARIOS FROM A HYBRID METHOD BETWEEN ARTIFICIAL NEURAL NETWORK AND MONTE CARLO SIMULATION

Bermeo, J.[(a)], Castillo, H.[(b)], Serrano, S.[(c)], Arce, D.[(d)] , Bermeo, H.[(e)]

[(a),(b),(c) ,(d)]Universidad Politécnica Salesiana, Cuenca, Ecuador
[(e)]Universidad de Cuenca, Cuenca, Ecuador

[(a)]jbermeo@ups.edu.ec, [(b)]hcastillo@ups.edu.ec, [(c)]sserrano@ups.edu.ec
[(d)]darce@ups.edu.ec , [(e)]henry.bermeo@ucuenca.edu.ec

ABSTRACT
Business Intelligence analyze existing data, to create knowledge about environment, in this paper, the accounting and operating information is analyzed to generate L-scenarios from hybrid method between ANN and Monte Carlo Simulation (MCS), then analyze the profitability in a Collection center of Raw milk. Every scenario is generated into analysis period, and has information about purchases, sales, cost of goods, sales price, operative cost and opportunity cost, then the cash flow, Net Present Value NPV and Modified Internal Rate of Return MIRR is calculated in order to evaluate the profitability of each scenario. The statistics (with a 95% of confidence) shows that MIRR has a confidence interval between 18,8% and applying an expected rate of return of 20% results in the average NPV is positive, so it implies the project is profitable. Furthermore, the opportunity cost analysis suggests proposes to increase the plant size.

Keywords: Artificial Neural Network ANN, Monte Carlo simulation MCS, Business Intelligence BI, decision making, profitability, Net Present Value NPV, Internal Rate of Return IRR, Modified Internal Rate of Return MIRR.

1. INTRODUCTION

The reality is ruled by many variables, then variable groups would create some possibilities and constraints, which make scenarios. For decision making, business leaders must analyze each scenario with a validation criteria defined from aims of company, Figure 1 shows this scheme.
In this way, the aim of an investor, in the beginning of company, is to have a measurement about the profitability, the net present value (NPV) (Brealy, Myers, & Allen, 2013) or Internal Rate of Return (IRR) (Ross, Jordan, & Westerfield, 2014) so as to could estimate the anticipated profitability of project or investment, however, the real problem is the generation of reliable information to evaluate the NPV. New approaches of Business Intelligence (BI) are implemented with the use of Artificial Intelligence (AI) (Alexander & Jothivenkateswaran, 2016), (Kwon, Wu, & Zhang, 2016) in this paper, L-scenarios were created through a forecasting of purchases, sales, prices and cost are generated by AI and Monte Carlo Simulation, then all this information was used to estimate NPV and MIRR of the project in every scenario. Finally, profitability is evaluated by statistical analysis.

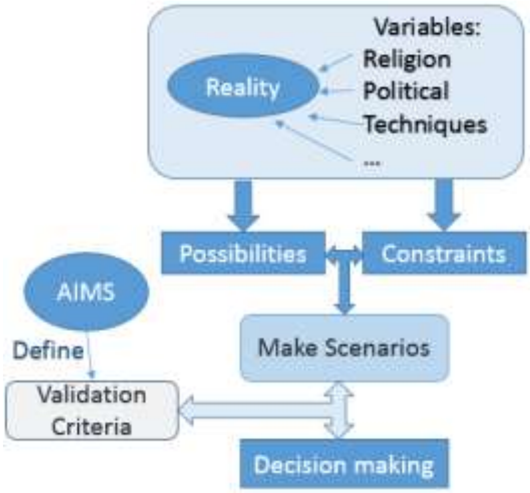


Figure 1: Scheme for decision making

For the forecasting of purchases by day, a hybrid method between Artificial Neural Network (ANN) and Monte Carlo Simulation (MCS), (Bermeo, y otros, 2017) was used. In addition, the generation of costs, sales and price were generated by Monte Carlo Simulation (Lebovka, Vygornitskii, Gigiberiya, & Tarasevich, 2016) from statistics of the historical accounting information of company.
Information from forecasting is applied to estimate yearly Gross profit for each scenario, then the Net profit was calculated considering the taxes and total operating expenses. Next, the system calculates the NPV and IRR, and make and statistics analysis to evaluate the scenarios. Figure 2 sketches the process for this paper. First, it begins with a little explanation of hybrid model between ANN and MCS. Second, L-scenarios were created by

hybrid model and MCS, where the hybrid model generates amount of purchases daily, and MCS generates values for costs, sales and price. Third, system calculates the yearly Net profit, NPV and MIRR, for every scenario. Finally, a statistics analysis supports the estimation about the profitability in all scenarios.

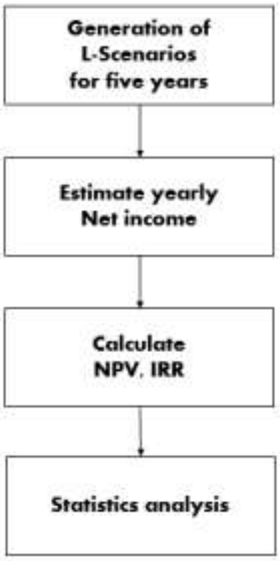


Figure 2: Flowchart for

2. HYBRID MODEL BETWEEN ANN AND MCS

The ANN is training with the historical information of purchases to make a forecasting, after the training, these predictions only need a limited group from historical information of purchases. The Monte Carlo Simulation applies accumulated probability density, estimated from historical records, to generate L-times a limited group of historical information. Next, every limited group is the input for trained ANN to generate a scenario, which is the forecasting of purchases, so at the end, they are L-scenarios, where each scenario is the amount daily purchase in desired period for the simulation. The figure 3 describes graphically the scheme of hybrid model.

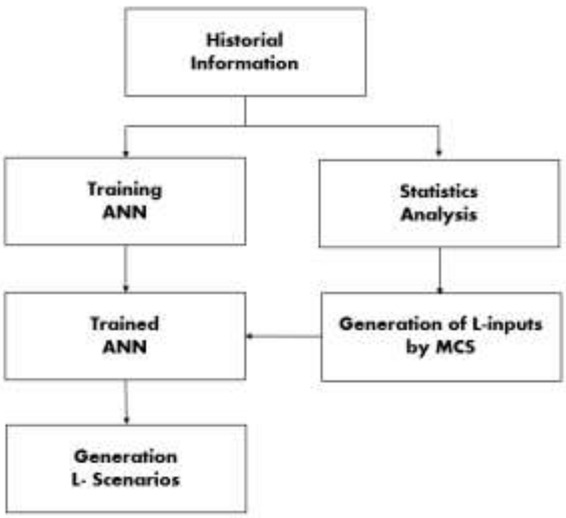


Figure 3: Scheme of hybrid model

2.1. Training ANN

ANN is training with the Historical information of company. In this case, the Inputs are thirty-three: year, month of prediction, amount of milk purchases in i-th day before the prediction day (P_1…P_30). The output is the amount of milk purchases in the prediction day (P_0).

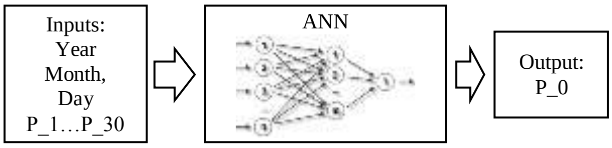


Figure 4: ANN implemented

2.2. Analysis of probability density function of daily purchases

First step is the analysis of probability density function (pdf) from milk purchases in this case. The figure 5 shows a comparative between relative frequency from real dates and a normal distribution applying the statistics information from table 1.

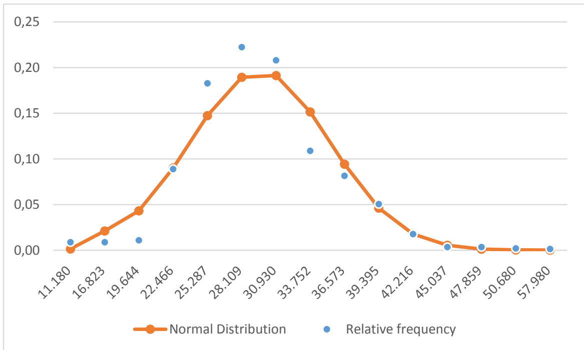


Figure 5: Analysis of Relative frequency of milk purchases

Table 1: Statistics of milk purchases

Factors			
Average	Standard Deviation	Minimum	Maximum
28.214,4	5.678,1	731	57.976,8

2.2.1. Chi-square test

The chart of relative frequency of milk purchases looks like normal distribution with average and standard deviation from table 1, however, the graphic is not enough to support that pdf of milk is a normal distribution, so it is necessary a Chi-square goodness of fit test (Balakrishnan, Voinov, & Nikulin, 2013). For this case, hypothesis Ho is defined by “frequency of milk purchases conforms a normal distribution”.
After the apply the test, table 2 shows the results from Chi-square test with 5% of significance level, in conclusion, test raise that Ho must be accepted, it means, the frequency of milk purchases conforms a normal distribution.

Table 2: Chi-square goodness of fit test

Factors: daily milk purchases			
Degrees of Freedom	χ Reference	χ Calculated	Ho
1460	1.372,27	246,34	Accepted

2.2.2. Estimate L

The normal accumulative probability density function can be used by MCS because Ho is accepted, so MCS is used to generate L-inputs for ANN. Next, outputs from neural network are used iteratively to complete the milk purchases in the desired period. In this case study, inequality (1) estimates the value of L, where a confidence interval of 95% (z=1,96), standard deviation from table 2, and an error close to one percent of average (error=282.14) are used, so L minimum is 1555,87. In this experiment L=1600 was used.

$$L \geq \left(\frac{z_{\alpha/2} \cdot S}{error}\right)^2 \quad (1)$$

2.3. Generation of L-inputs by MCS

Each ANN input has thirty-three records, so Monte Carlo Simulation with the normal accumulative pdf generates thirty-three random numbers for each input, the figure 6 shows L-inputs generated by MCS.

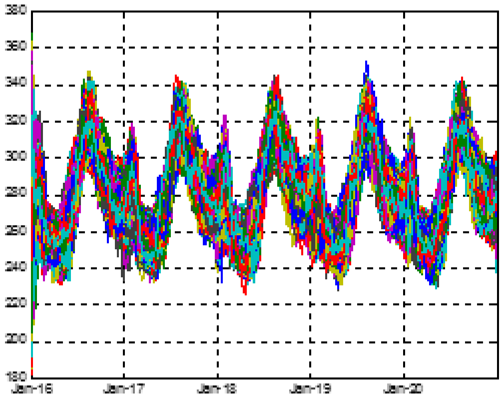


Figure 6: L-inputs generated by MCS

2.4. Generation of L-scenarios

Each input is entered to ANN to get prediction in one day, then a position is shifted in original input, and between the new prediction and shifted input a new input is created. This process is repeated until to get whole prediction in the desired period, all this information conforms a scenario, so L-inputs will generate L-scenarios. Figures 7 sketches L-scenarios generated for five years of milk purchases.

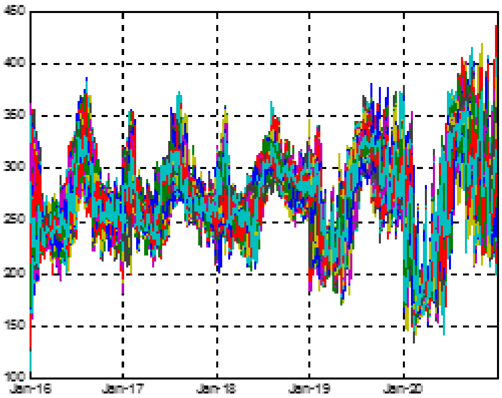


Figure 7: Hybrid method generates L-scenarios of Milk purchases

3. ESTIMATION OF NPV, IRR AND MIRR FROM L-SCENARIOS

3.1. Input information

Hybrid method generates daily milk purchases, and the similar way, MCS generates demand of monthly liters (D), cost per liter of raw milk (C), and sales price per liter (P) from historical accounting information. In addition, Opportunity cost per liter (O) and Retention cost per liter (R) even though are fixed for first month, the inflation rate is applied for next months to build a matrix of inflation. The value used for inflation rate was 8.83% (INEC, Instituto Nacional de estadística y censos, 2016), which is the biggest in the last decade. The table 3 shows the statistics values applied for simulation by MCS. Furthermore, the rate of population growth (1.31% (Central Intelligence Agency, CIA, 2017)) was applied to amount of monthly liters sold.

Table 3: Statistics of accounting information

Description	Average	Standard Deviation
Cost of Milk per liter	$ 0,4485	$ 0.0265
Monthly liters sold	879 883,4	257 397.9
Price of milk sold/lt.	$ 0,4831	$ 0.0203

3.2. Gross Profit (GP)

The data of milk purchases are daily, so they are consolidated to get monthly purchases, in the case of collection of raw milk, the 97% from all purchases will be the supply (Q) the 3% is spoiled. Then, for each month, the Demand (Di) and supply (Qi) of i-th month are compared. If Di is shorter than Qi, implies that the sales only could be a Di, and there are stocktaking for the next month equals to Qi-Di, in the other case when Di is larger than Qi, then sales are Qi and there are a deficit equals to Di-Qi. The retention cost (R_{CT}) is equal to stocktaking by retention cost per liter (RC), and Opportunity cost (O_T) is equal to deficit by Opportunity cost per liter. The cost of goods sold (CM) is equal to supply in i-th month (Qi) by cost per liter (C). Finally, the Gross profit (GP) is equal to Gross Sales (GS) minus cost of goods sold (CM), retention cost (R_{CT}) and opportunity cost (O_T). Figure 8 sketches the process to calculate the Gross profit.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

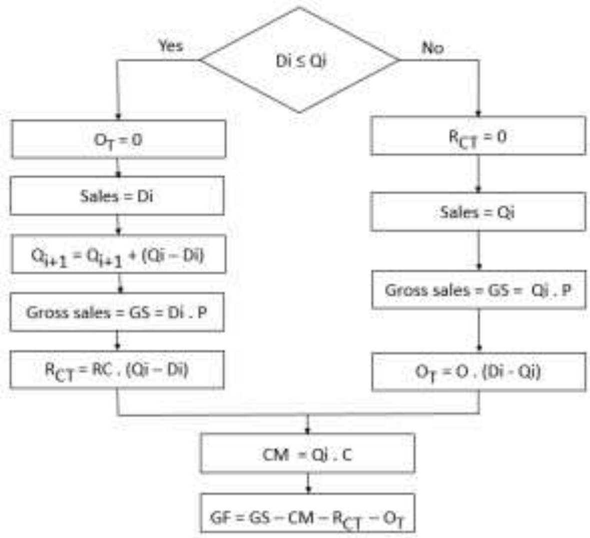


Figure 8: Calculation of Gross Profit

3.3. Retention Cost and Opportunity Cost

The retention cost define the retention capability, if firms make a good job then the Customer lifetime value (CLV) is increased (Zhang, y otros, 2015), in the case of Collection Center, where the rate of churn is almost zero, so the retention cost is only composed by the cost of storage to have enough supply for the customers. From historical values, the storage cost is $ 0.015 per liter, then matrix of inflation rate is applied to get the retention cost per liter (RC). In figure 7 shows that RC must only apply to stocktaking.

There are an infinite ways to evaluate the opportunity cost, one of them is the estimation in terms how the system translates changes in resources into changes in benefits, (Sculpher, Claxton, & Pearson, 2017), and so opportunity cost can be estimated from inefficiency, for the case of study, it is the deficit of supply to reach the demand. Obviously, in this model, opportunity cost should not affect cash flow, so the opportunity cost is estimated only for analyzing the increase of plant size. The opportunity cost is 1.5 cents by liter of deficit. So this rate with matrix of inflation is applied to estimate the opportunity cost.

3.4. Net Income

The equation (2) defines the operating profit by Gross profit minus operating expenses. Net Income is defined by Operating profit minus taxes, in this case, the real taxes in Equator are 33,7%, first 22% for income tax, which is applied to operating profit (SRI, Servicio de Rentas Internas, 2017), and second, 15% to pay utilities to workers, which is applied after the income tax, so equation (3) shows how estimate the Net Income. Figure 9 shows monthly Net income from scenarios.

$$\text{Operating profit} = GP - OE \qquad (2)$$
$$\text{Net Income} = \text{Operating profit} \cdot (1\text{-}33{,}7\%) \qquad (3)$$

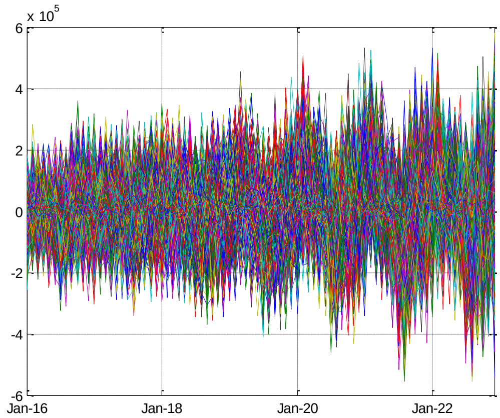


Figure 9: Net Income of L-scenarios

Even though the figure 9 shows a fuzzy information about the Net Income, a statistics analysis will help to understand and interpret this information.

3.5. Investment and depreciation

The table 4 shows a detail the investment for the company of Collection Center. Furthermore, the years for depreciation are showed.

Table 4: Detail of Investment and Years for depreciation

Description	Investment	Useful life in Years
Civil Works	80.000,00	20
Ice Bank construction	28.000,00	20
Generators	12.000,00	10
Stainless Steel Tanks	30.500,00	20
Design	4.500,00	3
Working capital	25.000,00	3
Unexpected	18.000,00	3
TOTAL Investment	198.000,00	

Straight Line Depreciation is the default method used, where the annual depreciation is calculated by equation four (Meigs, Williams, Bettner, & Haka, 2000), where the annual depreciation is equal to initial asset cost divided by the useful life of asset.

$$Annual\ Depreciation = \frac{Asset\ cost}{Useful\ life} \qquad (4)$$

3.6. Cash Flow

The Cash Flow is a financial report that presents a detail of the cash and income flows in analysis period. In this case, the Income is only made up from Gross Sales. The sales cost is the sum of cost of goods sold (CM), retention cost (R_{CT}) and opportunity cost (O_T). The operative expenses include basic services, salaries, social benefits, transportation y depreciation.

Table 5: Equations to estimate Net Flow

Description	Equation
Income (Gross Sales-GS)	P*Sales
Sales Cost (SC)	CM+ R_{CT}+ O_T
Gross Profit (GP)	GS-SC
Operative Expenses (OE)	OE
Operating Profit (OP)	GP-OE
Financial expenses (FE)	
Profit before taxes (PBT)	OP-FE
Income Tax (IT)	0.22*PBT
Net Profit (NP)	PBT-IR
Depreciation (D)	D
Payment of Capital (PC)	PC
Net Flow	NP+D-PC

3.7. Net Present Value NPV

The Net Present Value NPV is a measurement of profitability of investment project, the equation (5) shows how to calculate it from yearly Net Flow (NF), initial investment and rate of minimum return "i".

$$NPV = \sum_{t=0}^{n} \frac{NF_i}{(1+i)^t} \qquad (5)$$

When NPV is larger than zero, it means that investment would add value to investors, then the project may be accepted, if NPV is negative means investment would subtract value from the investors. Once the NPV=0 means that investment would neither gain nor lose value to investors, so NPV is indifferent to accept or reject the project.
The rate of minimum return must be related at least with the following factors: the risk-free rate of return, the "market price of (dollar) risk", and the variance in the project's own present value return (Lintner, 1965), but new approaches raise the Expected return, which depend on Risk-free rate, equity risk premium, small cap risk premium and book to market premium (Damodaran, 2015). The process to estimate the Expected return is quite complex and it is out the aim of this paper, so for the collection center, where risk country rate is about 667 points (Banco Central del Ecuador, 2017), and this can apply to estimate a required throughput rate (Gnecco, 2011), in addition, the maximum inflation rate in last decade was 8.83%, passive interest rate 4,82%. If 5% is applied like as risk-free rate, it means, the Expect return should be over 19%. In this work, the rate of minimum return applied is 20%to estimate the NPV.
Figure 10 shows the NPV from 1600 scenarios, even though, there is a big variation, the most scenarios have a positive NPV.

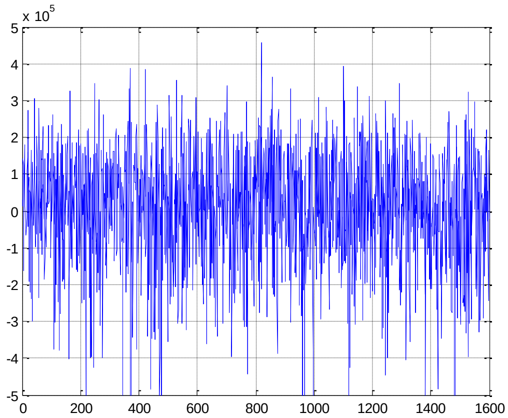


Figure 10: Net Present Value of L-scenarios

3.8. Internal Rate of Return IRR

The internal rate of return (IRR) provide the profitability of project, when their funds remain in the project during whole their life. When this rate is applied to estimate the NPV then the result of NPV is zero, in the context of savings and loans the IRR is also called the effective interest rate. There are a lot of methods to calculate the IRR, the equation (6) are used to estimate de IRR value. The IRR must be over to the sum between rate of inflation and risk premium.

$$NPV = \sum_{t=0}^{n} \frac{NI_i}{(1+IRR)^t} = 0 \qquad (6)$$

For L-scenarios, the calculated IRR shows a big variation, the figure 11 shows the IRR by scenario, where there is a positive tendency.

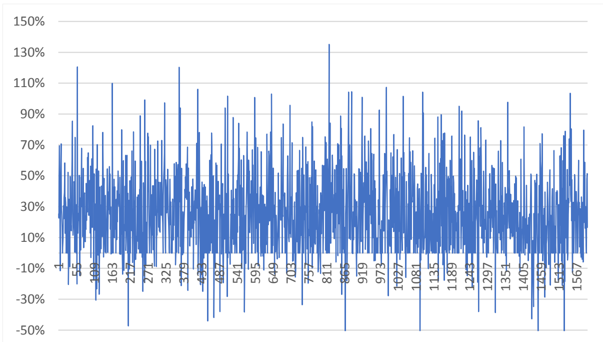


Figure 11: IRR of L-scenarios

3.9. Modified Internal Rate of Return MIRR

The modified internal rate of return (MIRR) is more accurate than IRR, because MIRR assumes that positive cash flows are reinvested, and negatives cash flows have a financial cost. On the other hand, IRR assumes that all cash flows are reinvested at the same IRR. Equation (7) shows how to evaluate MIRR, where PCFi is Positive Cash Flow in i-th year, NCFi is the negative cash flow in i-th year (when i=0 is the initial investment), r is cost of capital of company and fc is the financial cost rate. For this case, the financial cost rate is 11.83% (Banco Central del Ecuador, 2017), which is equal to average interest for loans to SMEs (small and medium-sized enterprises) from banks. The r applied is the 8.83% taking into consideration the maximum inflation, because in the worst case, the reinvestment in the company should be at least higher than inflation.

$$MIRR = \sqrt[n]{\frac{\sum_{i=1}^{n} PCF_i \cdot (1+r)^{n-i}}{\sum_{i=0}^{n} NCF_i/(1+fc)^i}} - 1 \qquad (7)$$

For L-scenarios, the figure 12 displays a comparative between IRR and MIRR, where the calculated MIRR has a slower variation than IRR.

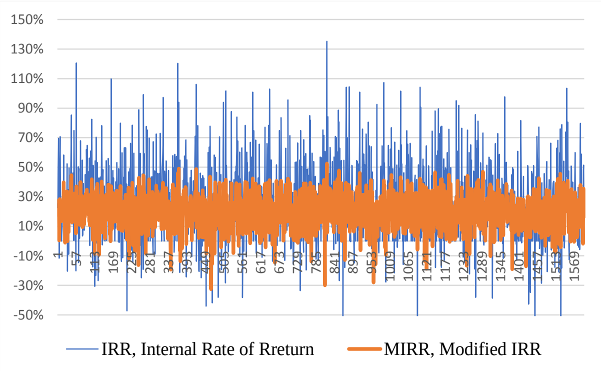


Figure 12: MIRR and IRR of L-scenarios

4. STATISTIC ANALYSIS FROM SIMULATION

Information in figure 9 is confused, but when the average is calculated, like as in the figure 13, these is friendlier to analyze, for example, the average of amplitude will increase over time, and Net Income maintains a steadily increasing form but within five years of prediction, it is still stable, so the information could be used to calculate the NPV, IRR and MIRR with enough accurate.

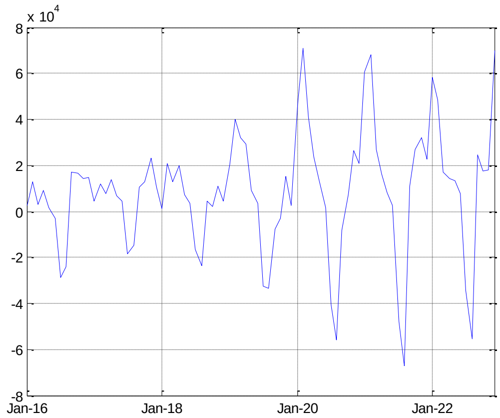


Figure 13: Average monthly Net Income of L-scenarios

The figure 14 shows the histogram of NPV, where this had a bell form, and the Chi-square goodness of fit test (Table 4), point out this is similar to Gauss distribution, and additionally maximum is near to the average, which is shown in table 5.

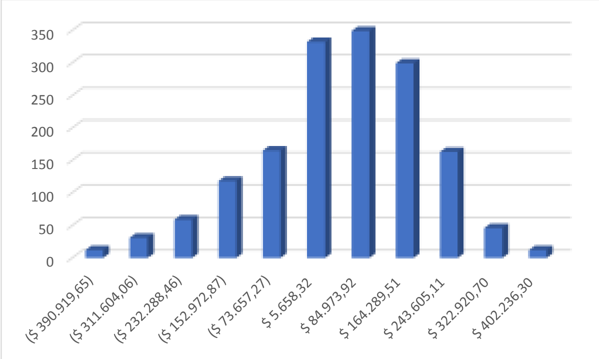


Figure 14: NPV histogram

Histogram of IRR, showed in the figure 15, and Chi-square goodness of fit test (Table 4) indicates that there is a Gauss distribution, furthermore the average is near to the value showed in Table 5.

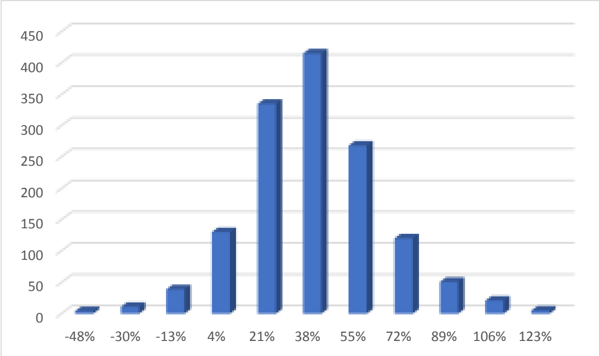


Figure 15: IRR histogram

Figure 16 shows the histograms of IRR and MIRR, where MIRR has a smaller standard deviation than IRR, it means that MIRR values are less spread than IRR values.

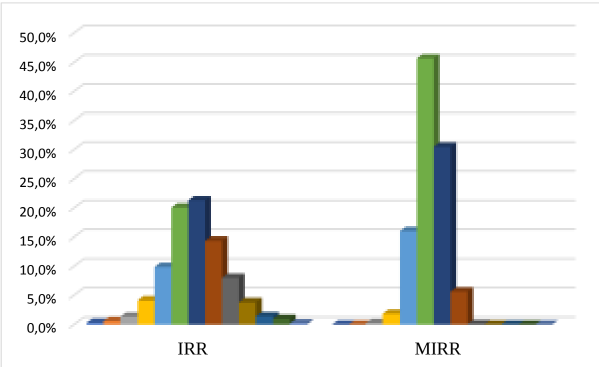


Figure 16: IRR and MIRR

The results of Chi-square goodness of fit test for NPV, IRR and MIRR point out that all distributions are Gaussians, it means, the normal distribution can be applied for the calculus of confidence interval.

Table 6: Chi-square goodness of fit test

Net Present Value NPV			
Degrees of Freedom	χ Reference	χ Calculated	Ho
1599	1.717,28	611,09	Accepted
Internal Rate of Return IRR			

Degrees of Freedom	χ Reference	χ Calculated	Ho
1384	1.493,15	79,19	accepted
Modified Internal Rate of Return MIRR			
Degrees of Freedom	χ Reference	χ Calculated	Ho
1599	1.717,28	53,22	accepted

Average and standard deviation from L-scenarios are calculated and showed in table 7, the averages of NPV, IRR and MIRR indicate that the project will have a good profitability.

Table 7: Statistics of NPV, IRR and MIRR

N = 1600 scenarios		
Factor	Average	Standard Deviation
Net Present Value NPV	$ 5 658	$ 158 631
Int. Rate of Return IRR	29 %	26%
Modified IRR, MIRR	19 %	11 %

Even though the indicators point out that there is a great profitability, it is necessary to use the confidence interval (equation (7)) to obtain a better analysis (Walpole, Myers, Myers, & Ye, 2012). Equations (7) and (8) indicate the confidence intervals of 95% ($Z_{\frac{\alpha}{2}}$=1.96) for the NPV and IRR respectively.

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \qquad (7)$$

$$P((\$2\ 114) \leq \mu_{VPN} \leq \$\ 13\ 431) = 0{,}95 \qquad (8)$$

$$P(27{,}9\% \leq \mu_{IRR} \leq 30{,}5\%) = 0{,}95 \qquad (9)$$

$$P(18{,}8\% \leq \mu_{MIRR} \leq 20{,}0\%) = 0{,}95 \qquad (10)$$

Equation (8) indicates that 95% of all scenarios will have a VPN between $2 114 and $ 13 431, so the investors will earn money with the project. In the same way, equation (9) point out that IRR are between 27,9% and 30,5%, considering an inflation rate about 8,83%, it implies that Collection Center will have a premium between 19,0% and 21,7% into the 95% of L-scenarios. Even though, the opportunity cost doesn't affect the cash flow, the analysis cost could be useful to determinate the optimal plant size, figure 17 sketches the opportunity cost for every scenario, and table 8 shows the statistics, where the average is over two hundred thousand, it implies that company is losing this value into income, by their constraint in size plant.

Table 8: Statistics of Opportunity Cost

N = 1600 scenarios		
Factor	Average	Standard Deviation
Opportunity Cost	$ 212 554	$ 48 089

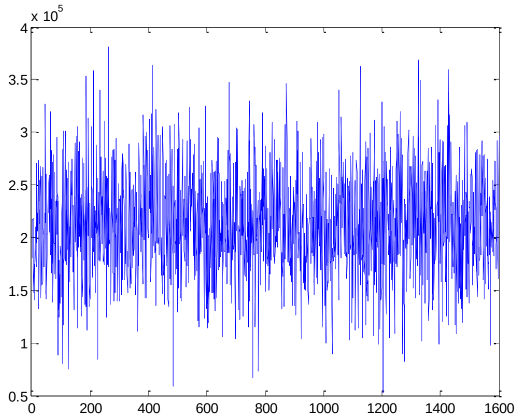


Figure 17: Opportunity Cost from L-scenarios

Table 8 point out the opportunity cost have a high average enough to make a change, it means, plant size should be increased to avoid deficit every month.

5. CONCLUSIONS

The hybrid method and Monte Carlo Simulation could generate L-scenarios to make a measurement of profitability with a big confidence, as can be seen in equation (8) and (9).

In this work, the generation of L-scenarios was applied to get a financial factor, however the hybrid method between ANN and MCS could be used in other applications too, for example to generate L-scenarios of Load on electric transmission lines in order to estimate life time of transformers.

For the application of hybrid method and MCS, is always necessary to have historical information for two reasons: (i) training of ANN, (ii) estimate statistics information to generate Inputs by Monte Carlo Simulation, how it is showed in the diagram of figure 2.

The large of historical information should be as large as at least the half of prediction, in order to be in a stable zone how is shown in figure 11, because sometimes when the prediction is larger than twice of the training period, then the system become unstable.

The Chi-square goodness of fit test is a great tool to define what kind of probability density function could be used, in this work, the test shows that variables (inputs and output) follow a Gauss distribution function.

The confidence interval of NPV, in equation (8), shows that the project has a great likelihood to get successful, so the project must be accepted, however the big standard deviation shows that system is very sensitive to whatever change, so the management of the collection center is an important issue to be considered.

Equation (10), shows that the project has a positive premium for MIRR, added to fact that it has a high probability of occurrence, around of 95%. Moreover, with positive NPV, it implies the project should be executed.

With the generation of L-scenarios, it is possible to calculate other financial factors like Gross margin (see figure 16), operating margin, profit margin, ROE, etc. to make a statistics and financial analysis, so the model used could be applied in future works of budget simulations o supply chain management.

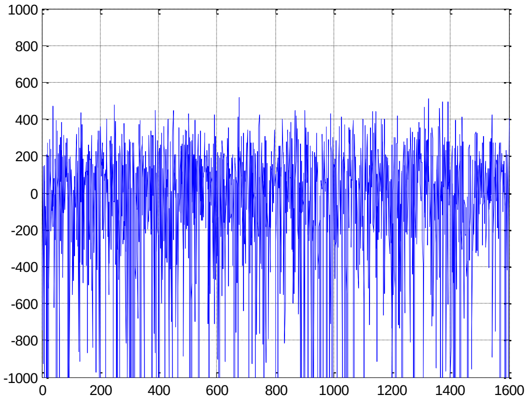


Figure 16: Gross Margin for L-scenarios

The increase of plant size must be realized when the benefit exceeds the cost, the table 6 shows that the average is over two thousands in five years, so if the collection center increase the plant size then they will have enough income to afford that, in addition an extra benefit.

6. REFERENCES

Alexander, A., & Jothivenkateswaran, C. (2016). Analysing genetic algorithm techniques on data mining for constructing Effectual Business Intelligent System (EBIS). *International Journal of Control Theory and Applications, 9*(10), 4479-4484.

Balakrishnan, N., Voinov, V., & Nikulin, M. (2013). *Chi-Squared Goodness of Fit Tests with Applications* (1st ed.). Ontario: Academic Press.

Banco Central del Ecuador. (2017, 3 1). Retrieved from https://contenido.bce.fin.ec/resumen_ticker.php?ticker_value=riesgo_pais

Banco Central del Ecuador. (2017, 03 1). *BCE, Banco Central del Ecuador.* Retrieved from https://contenido.bce.fin.ec/docs.php?path=/documentos/Estadisticas/SectorMonFin/TasasInteres/Indice.htm

Bermeo, J., Castillo, H., Armijos, X., Jara, J., Sanchez, F., & Bermeo, H. (2017). Artificial Neural Network and Monte Carlo Simulation in a Hybrid Method for Time Series Forecasting with Generation of L-Scenarios. *13th IEEE International Conference on Ubiquitous Intelligence and Computing, 13th IEEE International Conference on Advanced and Trusted Computing, 16th IEEE International Conference on Scalable Computing and Communications, IEEE International Conference o* (pp. 665-670). Toulouse; France: Institute of Electrical and Electronics Engineers Inc. doi:10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0110

Brealy, Myers, & Allen. (2013). *PRINCIPIOS DE FINANZAS CORPORATIVAS* (11 ed.). Mexico: Universidad Autónoma de México.

Central Intelligence Agency, CIA. (2017, april 2). *Central Intelligence Agency, the work of nation*. Retrieved from The center of intelligence: https://www.cia.gov/library/publications/the-world-factbook/geos/ec.html

Damodaran, A. (2015). *Applied Corporate Finance.* New Jersey: John Wiley & Sons.

Gnecco, M. (2011). *Riesgo País y Tasa de Rendimiento Exigida.* Mar del Plata: Editorial Académica Española.

INEC, Instituto Nacional de estadística y censos. (2016, 12 1). *Ecuador en cifras*. Retrieved from http://www.ecuadorencifras.gob.ec/inflacion-diciembre-2016/

Kwon, O., Wu, Z., & Zhang, L. (2016). Study of the forecasting performance of China stocks' prices using business intelligence (BI): Comparison between normalized and denormalized data. *Academy of Accounting and Financial Studies Journal, 20*(1), 53-69.

Lebovka, N., Vygornitskii, N., Gigiberiya, V., & Tarasevich, Y. (2016, December 30). Monte Carlo simulation of evaporation-driven self-assembly in suspensions of colloidal rods. *Statistical, Nonlinear, and Soft Matter Physics, 94*(6), Article number 062803. doi:10.1103/PhysRevE.94.062803

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets . *The Review of Economics and Statistics*, 13-37.

Meigs, R., Williams, J., Bettner, M., & Haka, S. (2000). *Contabilidad: La Base para Decisiones Gerenciales.* Mishawaka, USA: Mcgraw-Hill.

Ross, S., Jordan, B., & Westerfield, R. (2014). *Fundamentos de Finanzas Corporativas* (9 ed.). Bogota: MC GRAW HILL.

Sculpher, M., Claxton, K., & Pearson, S. (2017, February). Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions. (Elsevier, Ed.) *Value in Health*, 234-239.

SRI, Servicio de Rentas Internas. (2017, 1 1). Retrieved from SRI, Servicio de Rentas Internas: http://www.sri.gob.ec/de/167

Walpole, R., Myers, R., Myers, S., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias.* Mexico: Pearson Educación.

Zhang, Z., Wang, R., Zheng, W., Lan, S., Liang, D., & Jin, H. (2015). Profit Maximization Analysis Based on Data Mining and the Exponential Retention Model Assumption with Respect to Customer Churn Problems. *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on* (p. 5). Atlantic City, NJ, USA: IEEE.

UTILIZATION OF RAILWAY NETWORK MODEL FOR DYNAMIC CALCULATION OF TRAIN DELAYS

Jan Fikejz[(a)], **Jan Merta**[(b)]

[(a,b)] Department of Software Technologies, FEI, University of Pardubice, Pardubice, Czech Republic

[(a)] Jan.Fikejz@upce.cz ,[(b)] Jan.Merta@student.upce.cz,

ABSTRACT
This article deals with the proposal for dynamic calculation of train delays using the rail network model and satellite navigation. Attention is focused on the description of the location of trains in the designed model of railway network. Further attention is aimed on the design of the dynamic calculation of train delays with utilization of reduced track profile and using computer simulation for experiments.

Keywords: Railway infrastructure models, train positioning, web services

1. INTRODUCTION

Precise dynamic calculation of train delay is not exactly trivial and includes a wide variety of aspects affecting the resulting journey time. Such are both, technical parameters of (i) a train set (e.g. train acceleration or overall train set weight) and (ii) railway infrastructure (character of the line, velocity limits), and (ii) external influences like weather (e.g. temperature, weather conditions, visibility). Delay calculation is always dependant on the current train location in the railway network in given time, thus current information about location is an inseparable part of dynamic calculation of delay. In this case, it is not necessary to demand such accuracy as for safety systems to determine location and thus it is possible to use satellite navigation (GNSS – Global Navigation Satellite System) for finding the current location of a train.

2. POSSIBLE TYPES OF LOCALIZATION

Generally, localisation is prone to a wide range of approaches on how to identify the position of trains on a track. Put simply, localisation may be divided into the following three groups:

- Localization without the use of GNSS,
- GNSS using localization,
- GNSS-based, involving further support systems.

2.1. Trains localization without the use of GNSS

This type of trains localization often requires complementing the rail network infrastructure with additional construction elements, which entails higher costs of the actual implementation. On the other hand, this type of localization shows a high accuracy and reliability and is often used in the railway signalling technology. Essentially, it relates to the system of:

- ETCS (Ghazel 2104; Lieskovský and Myslivec 2010),
- Automatic train control (Chudacek and Lochman 1998; Lieskovský 2004),
- Track circuits (Dorazil 2008),
- RFID.

2.2. Trains localization using GNSS

When using GNSS for various application levels, it is necessary to take an indicated position error into consideration. Indicated position error is generally based on the nature of the satellite navigation. If we use systems that operate with the position information on an informative level only, we can tolerate a certain error; however, such inaccuracy is unacceptable in the railway signalling technology. However, various additional systems can be implemented to eliminate the error (completely or at least partially), thus making the position of the tracked object more accurate. The following systems can be listed in this group:

- EGNOS (Senesi 2012),
- Differential GPS (O'Connor 1997).

2.3. GNSS based localization involving additional support systems

As mentioned above, precise localization of trains using GNSS, especially for the needs of signalling technology, is a priori impossible. Nevertheless, the position of a rail vehicle can be determined significantly more precisely with the use of additional systems. This concern especially solutions using inertial systems (Standlmann 2006), but also less known systems such as those based on GNSS and contactless eddy current measurement (Becker and Poliak 2008).

3. RAILWAY NETWORK MODEL

Undirected graph, as defined graph theory, is a natural candidate for a railway network model. Based on an analysis of data provided by the company SŽDC-TUDC (consisting of service regulations, passports, and codebooks), sets of algorithms were subsequently created, with which it was possible to generate a three-layer model of the rail network (Fikejz and Kavička, 2011). Roughly speaking, the track can be divided into individual so called supertracks, which consist of definition supra-sections (TDNU), where each supra-section contains track definition sections (TUDU) with mileposts (in hectometres). Basic aspects of the description of the rail network are collectively shown in Figure 1.
Mileposts (in hectometres) are shown in Figure with the distance in kilometres and are graphically represented using gray points. TUDU is recorded using a six-digit code (163105, 163106, 16307, 173202) and are graphically represented using solid lines (red, black, orange, brown). Individual supra-sections (CLS 007, CLS008, REG023) are shown in light blue and supertracks (B421021 1 and B421021 1A) are shown in dashed lines. A place significant in terms of transportation (branch line) is symbolized by a green square.

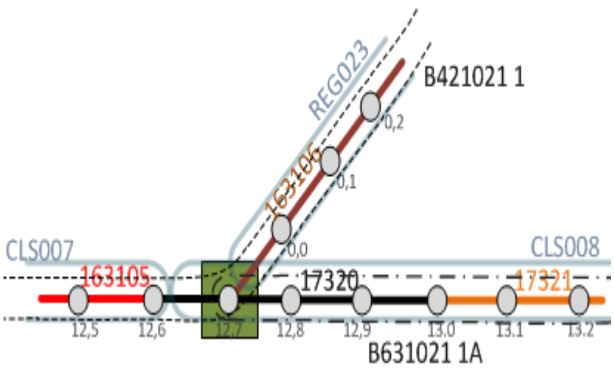


Figure 1: Basic aspects of the description of the rail network

The algorithm of railway network model (Fikejz and Kavička, 2011; Fikejz and Řezanina 2014.) was implemented directly on the database level using PL/SQL language. However, the algorithm had to be adjusted and generalized several times since there are various nonstandard conditions in the data, such as jumps in the mileposts (nonlinear growth of the kilometre succession between the mileposts) or change of an increasing kilometre sequence into a decreasing one and vice versa. The final model includes three data layers:

- **Data-Micro**, consisting of vertices and edges,
- **Data-Mezo**, include mezo-vertices and mezo-edges
- **Data-Macro**, containing super-vertices and super-edges.

Figure 2 presents the overall concept of a complete three-layer railway network model.

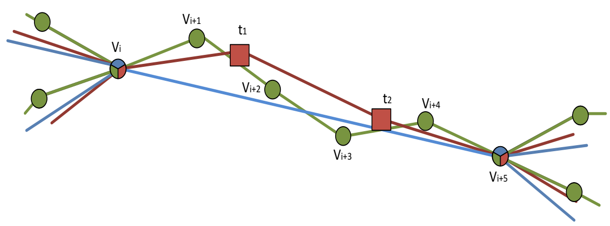


Figure 2: Illustration overall concept of a three-layer module

The data structure non-oriented graph was finally implemented directly in the ORCLE database using the ORACLE Spatial Network Data Model (Kothuri et al. 2007) technology. This technology enables the user to build a various network representation, involving also the object scheme and the communication interface API.

The objects scheme includes metadata and network tables. The interface contains on the server side PL/SQL API (an SDO_NET packet) for the creation, control and analysis of the database network, and a middle layer Java API (on client´s side) for the network analysis. The actual network is then defined by means of two compulsory tables:

- Node table,
- Link table.

For the work with spatial data, ORACLE with Spatial technology defines a special object data type SDO_GEOMETRY, which enables its user to store a number of spatial information and geometric types, such as various points, arcs, linear chains or polygons.

4. LOCALIZATION

The idea of trains localisation access to tracks is based on the correct pairing up of GPS information on position, provided by communication terminals, with the nearest vertex or edge of the graph. The discovered vertex/hectometre post disposes not only of a multi-dimensional key in the form of a GPS coordinate, it is also linked, through definition sections, to further information concerning the railway network infrastructure.
View of the situation that the model of railway infrastructure is stored in the database Oracle we can use the native database functions and operators. The `SDO_NN` (*nearest neighbor*) operator was selected in view of realising this unique trains localisation approach. The aforementioned operator searches for a geometric object that is closest to the object entered (like a point,

for example). In other words, it is possible to find the nearest vertex, or more precisely edge in a model, from the current position trains, Figure 3.

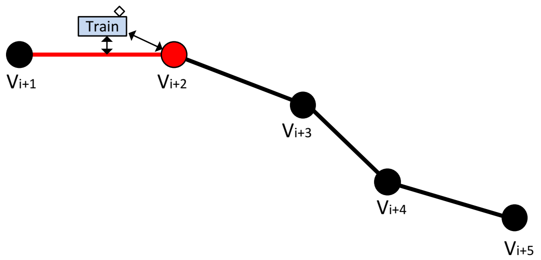


Figure 3: Main concept of localization

The actual detection of the current position of the trains can be divided into the following steps:

1. **Finding the nearest vertex and edge of the graph** – from the current position of the trains given the three-layer railway network model
2. **Assessment of the relevancy of incoming GPS information from the communication terminal** – verification whether the current position is not burdened by a disproportionate error (like, for example, that the distance of the trains from the nearest vertex/edge is a mere few meters or tens of metres, or that the trains is still assigned to the same super-edge, provided that it should still be located on it)
3. **Calculation of the exact position of the trains on the edge of the model** – using perpendicular projection of the point (current trains position) onto the line

The trains position data are collected from the communication terminals. These communication terminals sent position information to the central database every 30 second, Figure 4

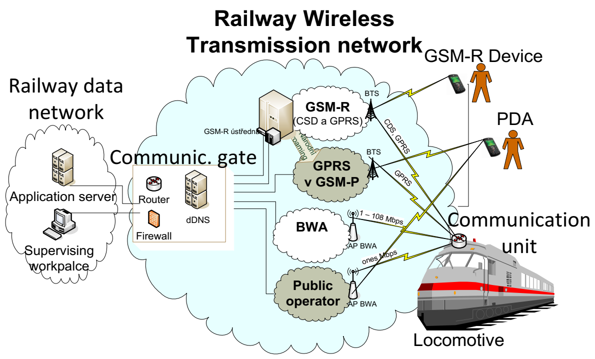


Figure 4 : Communication between the rail vehicle and dispatching centre

4.1. Search algorithm of previous or next railway station

Searching for the previous and next railway station utilises of iterations algorithm to Data-Micro layer. This algorithm allows finding the nearest railway station despite the fact that within the one TUDU there exists more stations, Figure 5.

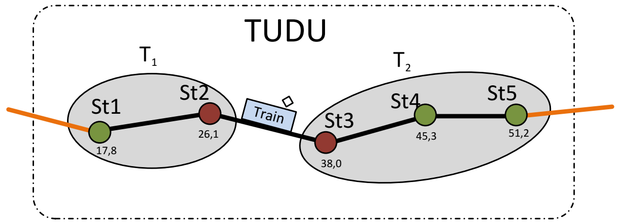


Figure 5: Finding nearest station

Concept of search algorithm consists of the following steps:

1. Sort the railway stations within the same TUDU considering the kilometric values
2. Divide the sorted railway stations to two separated subsets T1 and T2 by current position of rail vehicle
3. Select the first railway station from the subset T1 with the highest value of kilometre and the second railway station from the subset T2 with the lowest value of kilometre

As a next step, there is used the shortest path algorithm for finding the real distance from current train position to the both already found railway stations and according to the civil timetable we are able to calculate the actual time of arrival to the next railway station and actual train delay.

5. REDUCED TRACK PROFILE

Railway infrastructure is rather varied and contains many areas affecting the train dynamic (Bandžuch 2006). For common traction calculations, it is possible to substitute the real track profile by a reduced set of substitute gradients, so called line resistance, which includes:

- gradient resistance
- curve resistance
- tunnel resistance

Curve resistance $\boldsymbol{sl_c}$ is substituted by fictive incline, for which the curve radius R<300 m for secondary and regional lines is defined by:

$$sl_c = \frac{500}{R - 30} \; [‰] \tag{1}$$

Reduced incline $\boldsymbol{sl_r}$ is then defined by:

$$sl_r = \frac{sl_1 l_1 + sl_2 l_2 + sl_3 l_3 + \cdots + sl_k l_k + sl_{c1} l_{c1} + \cdots + sl_{cm} l_{cm}}{l_1 + l_2 + \cdots + l_k} \; [‰] \tag{2}$$

where:

- $\boldsymbol{sl_1}$ **to** $\boldsymbol{sl_k}$ -actual gradient in per mille (incline „+“, decline „-“)
- $\boldsymbol{sl_{c1}}$ **to** $\boldsymbol{sl_{cm}}$ -fictive gradient, substituting set of curves
- $\boldsymbol{l_1}$ **to** $\boldsymbol{l_k}$ -length of gradients s_1 to s_k in meters
- $\boldsymbol{l_{c1}}$ **to** $\boldsymbol{l_{cm}}$ -curve lengths 1 to m

While condition $(l_1 + l_2 + \cdots + l_k) \leq 2{,}5\ (l_{c1} + l_{c2} + \cdots + l_{cm})$ must be valid

An example of a reduced track profile is illustrated on the following Figure 6, in which arrows show the intended direction of movement of the train. The data in tables then depict:

- **kilometric position** of a change in profile $[km]$
- **the direction** where:
  - „+“ expresses change in direction
  - „−“ expresses change opposite the direction
- **track resistance** [‰] where:
  - positive value expresses track incline
  - negative value expresses track decline

6. TRAIN ACCELERATION

Train acceleration is influenced mainly by

- technical parameters of a locomotive (acceleration)
- overall weight of the train set

Using a special software tool for simulation train dynamics developed at University of Pardubice (Diviš and Kavička 2015), a set of measuring simulations focused on acceleration and breaking deceleration of trains was conducted for selected trains for the following track resistances:

- 0.5 %
- 1.0 %
- 1.5 %
- 2.0 %

Assessment of data from individual measurings has shown that data can be approximated by a linear equation for other calculations.
If we consider a train set with the following parameters:

- Acceleration: 0,6497725 m/s^2
- Length: 96 m
- Weight: 234 t

then individual measured values of acceleration can be reflected in a graph on Figure 7.

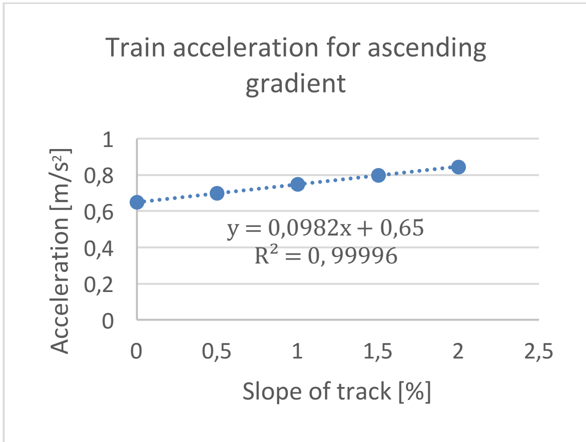
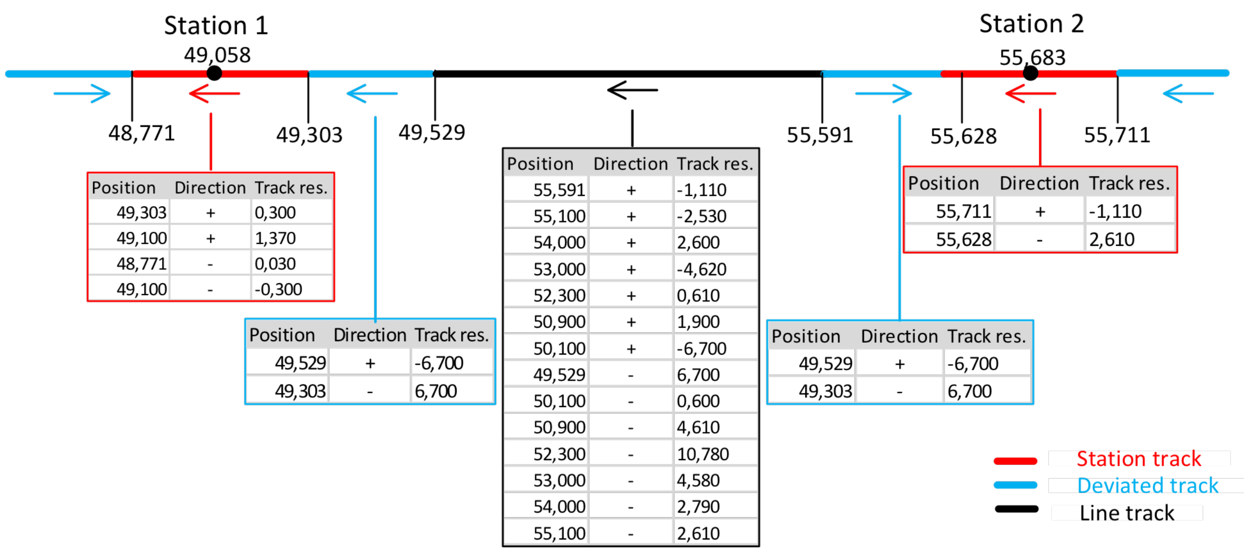


Figure 7: Example of train acceleration

From the graph above, it is clear that measured acceleration for incline can be approximated by a linear equation



Position	Direction	Track res.
49,303	+	0,300
49,100	+	1,370
48,771	-	0,030
49,100	-	-0,300

Position	Direction	Track res.
49,529	+	-6,700
49,303	-	6,700

Position	Direction	Track res.
55,591	+	-1,110
55,100	+	-2,530
54,000	+	2,600
53,000	+	-4,620
52,300	+	0,610
50,900	+	1,900
50,100	+	-6,700
49,529	-	6,700
50,100	-	0,600
50,900	-	4,610
52,300	-	10,780
53,000	-	4,580
54,000	-	2,790
55,100	-	2,610

Position	Direction	Track res.
49,529	+	-6,700
49,303	-	6,700

Position	Direction	Track res.
55,711	+	-1,110
55,628	-	2,610

Figure 6 : Example of reduced track profile

$$y = 0{,}0982x + 0{,}65 \quad (3)$$

for reliability $R^2 = 0{,}99996$.

When data from measured acceleration for a declining track are applied, then through approximation we can achieve linear equation

$$y = -0{,}1044x + 0{,}6511 \quad (4)$$

for reliability $R^2 = 0{,}99974$.

7. CALCULATING DELAY

For dynamic calculation of delay, it is possible to base the calculation on a simplified model, in which the overall journey time between two stations is given by adding their three parts (Figure 8):

- acceleration period to set velocity
- journey period in set velocity
- breaking period from set velocity to zero velocity

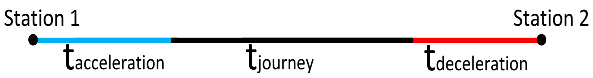


Figure 8: Concept of calculation

$$t_{total} = t_{accel.} + t_{\text{journey}} + t_{decel.} \quad (5)$$

while it is supposed that acceleration and breaking are affected by track profile, however, the train has sufficient performance power to keep required velocity on route between the two track segments.

7.1.1. Calculation algorithm

Delay calculation algorithm uses general formulae for evenly accelerated linear movement, i.e. relations based on:

$$v = v_0 + at \quad \Longrightarrow \quad t = \frac{v - v_0}{a} \quad (6)$$

$$s = v_0 t + \frac{1}{2}at^2 \quad \Longrightarrow \quad v^2 = 2as + {v_0}^2 \quad (7)$$

The algorithm itself is divided into 5 parts.

1. iterative calculation calculates the time $t_{acceleration}$ and path $s_{acceleration}$ needed to reach the required velocity v_{max}
2. iterative calculation calculates the time $t_{deceleration}$ and path $s_{deceleration}$ needed to reach zero velocity from velocity v_{max}
3. path of train movement in velocity v_{max} is calculated based on the difference between the overall path s_{total} and path for acceleration $s_{acceleration}$ and breaking $s_{deceleration}$, i.e.

$$s_{\text{journey}} = s_{total} - s_{accel.} - s_{decel.} \quad (8)$$

4. the time t_{journey} of train movement in set velocity v_{max} is calculated
5. overall time t_{total} is calculated from formula 5

So, if we consider maximal velocity of a train on a regional line v_{max} (for example $65\frac{km}{h}$) and track according to figure 9, then the first part of the algorithm (for calculating $t_{acceleration}$ and $s_{acceleration}$) will conduct individual iterative calculations in points of gradient change on the track sl_i [‰] defined by kilometric location d_i [km]. If the current velocity of the train v_i is higher or equal to the required velocity v_{max}, then the iterative calculation is terminated and subsequently the exact time t_{last} and path s_{last}, when the train reached the required velocity v_{max} are calculated. For starting the calculation, a corresponding railway station $d_{station}$, kilometric location of which is given, is considered.

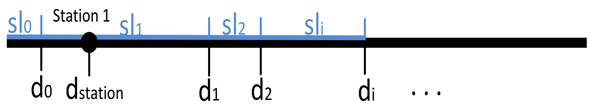


Figure 9: Concept of the acceleration algorithm

Concept of the algorithm for train acceleration to the required velocity v_{max}

1. $i = 1; v_0 = 0$
2. **if** $i = 1$
   then $s_i = d_i - d_{station}$
   else $s_i = d_i - d_{i-1}$
3. **if** $sl_i > 0$
   then $a_i = 0{,}0982 sl_{i-1} + 0{,}65$ (eq. 3)
   else $a_i = -0{,}1044 sl_{i-1} + 0{,}6511$(eq. 4)
4. $v_i = \sqrt{v_{i-1} + 2a_i + s_i}$
5. **if** $v_i < v_{max}$
   then $i = i + 1$ and go to step 2
6. $s_{last} = \frac{{v_{max}}^2 - {v_i}^2}{2a_i}$
7. $t_{last} = \frac{v_{max} - v_i}{a}$
8. $t_{toVmax} = \sum_{k=1}^{i-1} t_k + t_{last}$
9. $s_{toVmax} = \sum_{k=1}^{i-1} s_k + s_{last}$

Analogously, the breaking time $t_{deceleration}$ and the path $s_{deceleration}$ needed to stop the train at the station are calculated accordingly. The relation 8 determines the length of the line on which the train moves at a constant velocity and consequently from the relation

$$t = \frac{v}{s} \quad (9)$$

the journey length t_{journey} is calculated. Overall journey time t_{total} is then given by the sum of the partial times per the relation 5.

From the times between the individual railway stations t_{total}. and the time needed for the train to be serviced in the station $t_{service}$, it is then possible to calculate the total time of the journey to the required station.

For dynamic calculation of the delay of a moving train on the track, it is possible to calculate the time of journey/delay to the next railway station or to the selected station on the track from the knowledge of its current position (from the railway network model).

Complete calculations for the dynamic calculation of the delay were subsequently implemented in the *infraRail* software tool, and further variants of delayed trains were subsequently checked using the discrete simulation, both based on historical data and data generated. The running application captures the current position of the train on the track with a set of information related to its position including the calculation of the current arrival delay to the next railway station, Figure 10.

CONCLUSION

The focus was on the proposal for dynamic calculation of train delays using the rail network model and satellite navigation. A multi-layered model of the railway network was designed reflecting the non-oriented graph. In addition, the algorithm was used to identify the position of trains in the railway network. This algorithm includes the search of the previous or next railway station. The article was also focused on description of the reduced track profile which was used for designs of the algorithm for dynamic calculation of train delays. Proposed algorithms have been implemented in the InfraRail software tool. Discrete simulation was used to test other variants of delayed trains, both based on historical and generated data.

ACKNOWLEDGMENTS

This work has been supported by the project "SGS_2017 Models of infrastructure and operation of land transport systems" (financed by the University of Pardubice).

REFERENCES

Becker, U. and J. Poliak. DemoOrt repositions trains with satellite. In: EURAILmag Business & Technology. 18. BLUE LINE & Bro, France, 2008,s. 216-219.

Chudaček, V. and L. Lochman. Vlakový zabezpečovací systém ERTMS/ETCS. In: Vědeckotechnicky sborník ČD, č. 5/1998

Dorazil, P. Základní vlastnosti kolejových obvodů bez izolovaných styků. Pardubice, 2008. Bachelor thesis. University of Pardubice. Supervisor: Milan Kunhart.

Fikejz, J. and A. Kavička. Modelling and simulation of train positioning within the railway network. In: KLUMPP, Matthias. ESM'2012. The European simulation and modelling conference. Ostende: EUROSIS - ETI, 2012, s. 366 -376. ISBN 978-9077381-73-1.

Fikejz, J. and A. Kavička. Utilisation of computer simulation for testing additional support for dispatching rail traffic. In: European Simulation and Modelling Conference, 2011. Ostende: EUROSIS - ETI, 2011. p. 225-231. ISBN 978-90-77381-66-3.

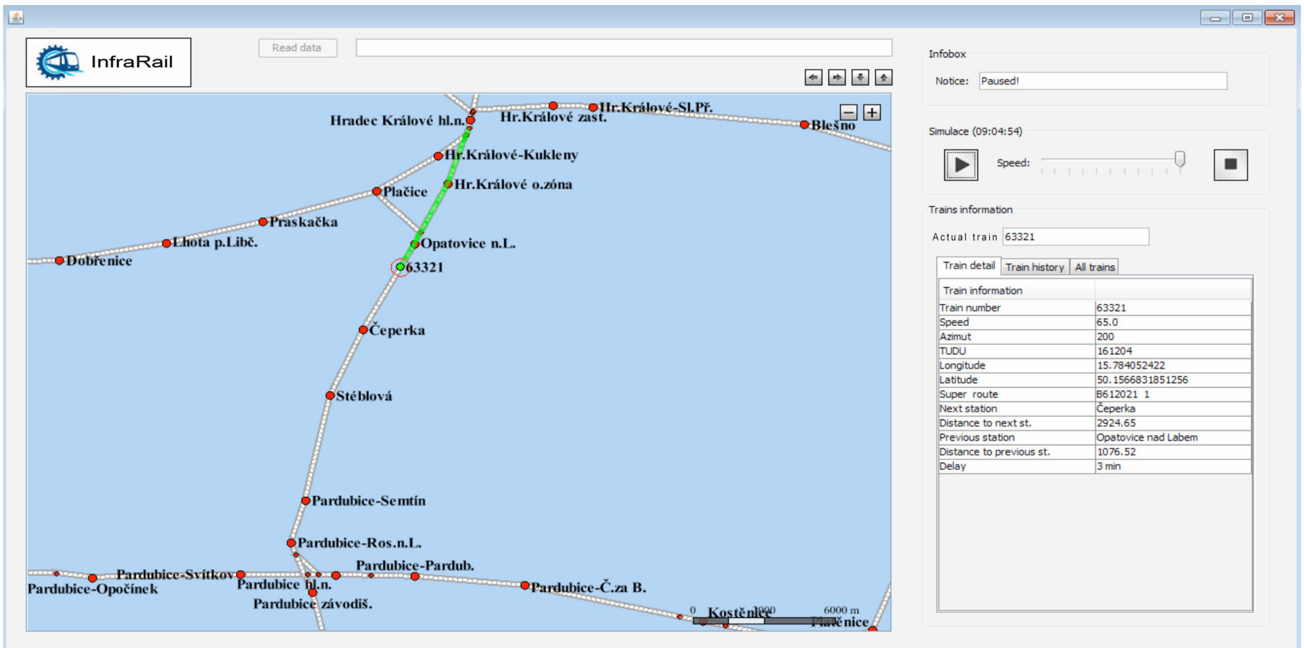


Figure 10: Running application

Fikejz, J. and E. Řezanina, Utilization of computer simulation for detection non-standard situations within the new data layer of railway network model. In: The 26th European Modeling & Simulation Symposium. Bordeaux, 2014 s. 371-377, ISBN 978-88-97999-32-4

Ghazel, M. Formalizing a subset of ERTMS/ETCS specifications for verification purposes. In:Transportation Research Part C: Emerging Technologies. Elsevier Limited, 2014, pp. 60-75 ISSN: 0968-090X

Kothuri, R. et al. Pro Oracle Spatial for Oracle database 11g. New York, NY: Distributed to the book trade worldwideby Springer-Verlag New York, c2007, xxxiv, 787 p. ISBN 15-905-9899-7.

Lieskovský, A. and I. Myslivec. ETCS a AVV - poprvé společně. In: EuroŽel, Žilina, 2010

Lieskovský, A. Automatické vedení vlaků Českých drah. In: Automatizace. Praha: Automatizace, 2004, roč. 10. ISSN 0005-125x.

O'connor, M. L. Carrier-phase differential gps for automatic control of land vehicles, In: Dissertation Abstracts International, Volume: 59-06, Section: B, page: 2876.; 158 p. 1997, Stanford University, ISBN: 9780591909272

Senesi, F. Satellite application for train control systems, In: The Test Site in Sardinia, Journal of Rail Transport Planning and Managemt. Elsevier BV, 2012, s. 73-78, ISSN:2210-9706

Bandžuch, Ľubomír. Modernizácia elektrifikovanej trate v rámci V. koridoru v úseku Košice – Poprad. Žilina, 2006. Diplomová práce. ŽILINSKÁ UNIVERZITA V ŽILINE. Vedoucí práce Doc. Ing. Gabriela Lanáková, PhD.

Diviš, Roman a Antonín Kavička. Design and development of a mesoscopic simulator specialized in investigating capacities of railway nodes. In: The 27th European Modeling and Simulation Symposium (EMSS 2015): 12th International Multidisciplinary Modeling and Simulation Multiconference (I3M 2015). 1. Rende, Italy: CAL-TEK S.r.l, 2015, s. 52-57. ISBN 9781510813762

MODELING PHOTONICS-BASED MICROWAVE FREQUENCY CONVERTER

M.E. Belkin, A.S. Sigov

Moscow State Technological University (MIREA), Scientific and Technological Center "Integrated Microwave Photonics", Moscow, Russian Federation

belkin@mirea.ru

ABSTRACT
Developing previously proposed by us a simplified version of microwave-photonics frequency converter based on period-doubling effect in nonlinear optoelectronic element such as cost- and power-efficient long wavelength wafer-fused vertical cavity surface emitting laser we have detailed and updated the simulation results mainly from a device designer's point of view by numerical modeling using a well-known versatile off-the-shelf soft tool such as MathCAD.

Keywords: microwave photonics, optoelectronic microwave frequency converter, computer aided design

1. INTRODUCTION
Microwave photonics (MWP) is a relatively new branch of photonics science and technology gradually penetrating into telecom (ultra-high-speed optical fiber systems, fiber-to-wireless (FiWi) access networks) and defense (phased-array antenna radars, electronic warfare systems) industries (Urick, McKinney, and Williams 2015). Thanks to capability of MWP devices to accomplish super wideband signal transference and distribution via low loss, lightweight fibers that also have weak dispersion distortion and ultimate electromagnetic immunity, up to this date a plenty of various photonics processing units for transmission, switching, distribution, filtration, and frequency conversion of microwave signals in optical domain has been viewed (Ng 2015). An exciting confirmation is multi-band software-defined coherent radar based on a photonic transceiver that was first in the world constructed by the researchers from CNIT–National Laboratory of Photonic Networks, Italy (Scotti et al. 2015) to demonstrate the practical importance of MWP technology for extending the sensing and signal processing performances of microwave systems.
Previously, we proposed and investigated a simplified version of MWP microwave frequency converter based on period-doubling effect in a semiconductor laser (Belkin, Iakovlev 2015). As a nonlinear element a cost- and power-efficient long wavelength wafer-fused vertical cavity surface emitting laser (VCSEL) was used. This paper details and updates the simulation results mainly from a device designer's point of view by numerical modeling using a well-known off-the-shelf versatile soft tool such as MathCAD.

2. BUILDING PRINCIPLES
The earlier proposed version of cost-effective MWP microwave frequency converter was based on well-known effect of period doubling (Hemery, Chusseau, and Lourtioz 1990) under modulation of a semiconductor laser by a strong radio frequency sinusoidal signal. For period doubling observation, it is necessary to modulate the laser in super large-signal mode with an injection current cutoff similar to a class C regime in electronic amplifiers. To achieve this, the bias current has to be selected in a near-threshold (but higher) or in the initial part of the quasi-linear area of the laser light-current characteristic. This regime results in generation inside the laser cavity of an optical emission spectrum including a fundamental frequency, its higher harmonic tones, their sub-harmonics, and the products of their mixing with higher tones in addition.
Figure 1 depicts the block diagram of the device under modeling entitled optoelectronic frequency multiplicator (OE-FMP). As one can see, the key part of this OE-FMP is an optoelectronic pair including laser (VCSEL) and working in linear mode photodiode (PD) of p-i-n structure. The electronic band-pass filter in the end of the chain post-processes the converted microwave signal. So, VCSEL is the single nonlinear element of the OE-FMP model.

Figure 1: Block diagram of OE-FMX

3. SIMULATION AND RESULTS
With the goal to define the optimal working area of our VCSEL in the period doubling regime, we performed a preliminary analysis based on quantum well semiconductor laser's rate equations describing photon and carrier densities nonlinear interaction (Piprek and Bowers 2002):

$$\begin{cases} \frac{dN_b}{dt} = \frac{\eta_a I(t)}{eV_0} - \frac{N_b}{\tau_t} + \frac{V_a}{V_b}\frac{N}{\tau_e} \\ \frac{dN}{dt} = \frac{V_b}{V_a}\frac{N_b}{\tau_t} - \frac{N}{\tau_t} - \left(AN + BN^2 + CN^3\right) - \upsilon_g g(N,S)S \\ \frac{dS}{dt} = \Gamma\upsilon_g g(N,S)S - \upsilon_g(\alpha_i + \alpha_m)S + \beta_{sp}\Gamma BN^2 \end{cases} \quad (1),$$

where S is average photon density, N_b *is* average carrier density, η_a is current injection efficiency, e is electron charge, $I(t)=I_b+I_m\cdot cos(\omega_m\cdot t)$ is laser current modulated at angular frequency ω_m with a modulation index $m= I_m/I_b$, α_i and α_m are internal and mirror losses. Gain function is given as:

$$g(N,S)=\frac{g_0}{1+\varepsilon S}\ln\left[\frac{N+N_s}{N_{tr}+N_s}\right] \quad (2)$$

The rest of symbols and their values typical for investigated VCSELs are given in Table I.

Table I: VCSEL model parameters

Parameter	Symbol	Value	Dimension
Gain constant	a	$4.8\cdot10^{-20}$	m^2
Linear recombination	A	$1\cdot10^{8}$	m^{-1}
Bimolecular recombination	B	10^{-16}	m^3/s
Auger recombination	C	$3.6\cdot10^{-41}$	m^6/s
Active region thickness	D	0.031	um
Material gain	g	$300\cdot10^{3}$	1/m
Threshold current	I_{th}	2	mA
Resonator length	L	1	um
Initial carrier density*	N_0	$1.2\cdot10^{24}$	m^{-3}
Negative carrier density	N_s	$-0.8\cdot10^{24}$	m^{-3}
Threshold currier density	N_{th}	$5\cdot10^{24}$	m^{-3}
Transparency carrier density	N_{tr}	$3.22\cdot10^{24}$	m^{-3}
Average mirror reflectivity	R	0.995	
Active region volume	V_a	0.891	um^3
Current blocking volume	V_b	2.779	um^3
Group velocity	v_g	$7.7\cdot10^{7}$	m/s
Active region diameter	w	6	um
Internal cavity loss	α_i	1000	1/m
Spontaneous emission factor	β_{sp}	$7\cdot10^{-3}$	
Confinement factor	Γ	0.03	
Nonlinear gain coefficient	ε	$1\cdot10^{-24}$	m^3
Internal quantum efficiency	η_i	0.8	
Emission wavelength	λ	1.31	um
Carrier lifetime at threshold	τ_e	2	ns
Electron relaxation time	τ_{in}	0.1	ps
Photon lifetime	τ_p	6.4	ps
* in formula: $g(N)=a(N-N_0)$			

Solving numerically the analytical model (1) in MathCAD soft tool using the embedded BDF-method we obtained for $m>1$ the values of carrier and photon densities versus time, then applying FFT techniques calculated the optical emission spectra. One example of such optical spectrum (upper sideband only) under strong RF modulation is presented in Figure 2. As seen, we succeeded in identification of such dynamic regime of VCSEL under modeling where the signal levels of a fundamental modulation frequency f_m, a sub-harmonic tone $0.5f_m$, and a component at $1.5f_m$ would be near equal and enough to secure the signal-to-noise ratio needed for typical wireless communication systems.

Also, Figure 3 shows the so-called phase-plane portrait of the VCSEL modulated in super large-signal mode. The loop form on the phase trajectory is an attribute for period doubling regime origination. The results presented on Figures 2 and 3 were obtained for bias current $I_b=1,8I_{th}$, and modulation index $m=2$.

In according with the experimental data (Belkin, Iakovlev 2015), a good fit of the simulated and measured OE-FMP output spectra was obtained at input RF signal power of 0 dBm, frequency f_m=3 GHz, and DC bias current 3.2 mA, which validated the proposed OE-FMP model. Using them we have also succeeded (see Figure 4) to identify clearly the area of VCSEL bias and modulation indexes at which the signal levels of the fundamental modulation frequency f_m, the sub-harmonic tone $0.5f_m$, and the component at $1.5f_m$ would be comparable to each other in the range inside 10 dB.

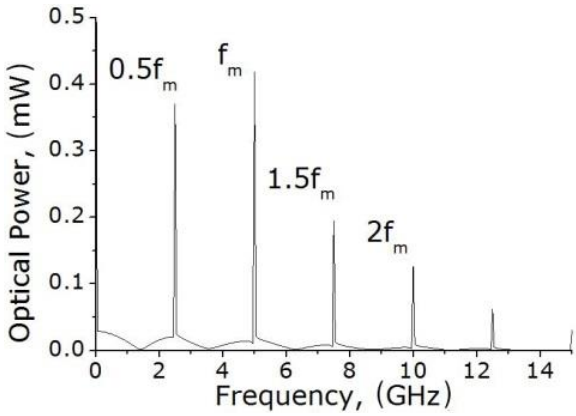


Figure 2: Large signal modulation spectrum.

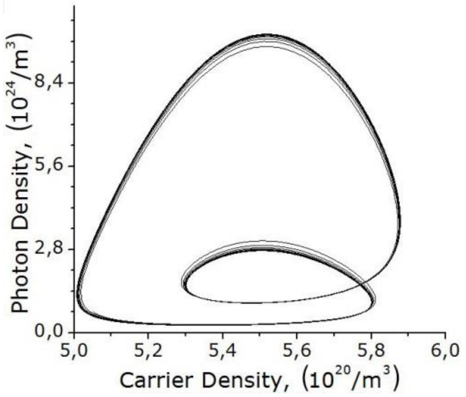


Figure 3: Phase-plane portrait of the VCSEL under simulation modulated in super large-signal mode

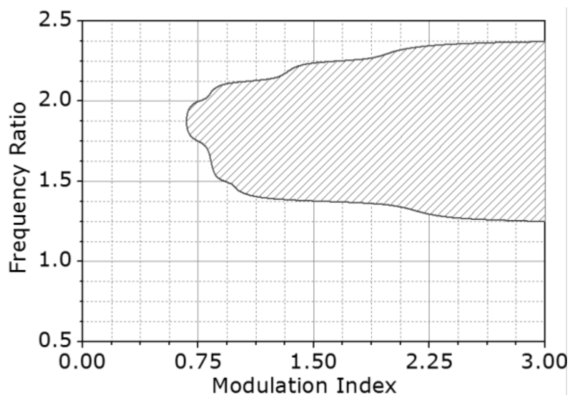


Figure 4: Diagram for VCSEL in sub-harmonic multiplication regime.

Comparing these refined modeling data with the experimental values (Belkin, Iakovlev 2015), we received 1.5-fold reduction for the error of the simulated and measured OE-FMP characteristics.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

4. CONCLUSION

Summarizing the modeling results, the main advantage of using VCSELs in optoelectronic sub-harmonic frequency multiplicator are:

(i) large bandwidth (limited solely by the bandwidths of the laser and photodiode);

(ii) application versatility (up-converter and down-converter have the same block diagram);

(iii) conversion loss independence of the position of microwave frequencies within the operation band (absence of the inherent effect of increasing conversion loss with frequency for standard microwave mixers);

(iv) design simplicity;

(v) lower power consumption of the VCSELs (5-10-fold lower than that of edge-emitting lasers);

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of the Russian Federation under Project RFMEFI60715X0138.

REFERENCES

Belkin M.E., Belkin L., Loparev A., Sigov A.S., Iakovlev V., 2015. Long Wavelength VCSELs and VCSEL-Based Processing of Microwave Signals. In: Pyshkin S., Ballato J., eds. Optoelectronics – Advanced Materials and Devices. Croatia: InTech, Chapter 6, pp. 231-250. Online: http://www.intechopen.com/books/optoelectronics-materials-and-devices/long-wavelength-vcsels-and-vcsel-based-processing-of-microwave-signals

Hemery E., Chusseau L., and Lourtioz J.-M., 1990. Dynamic Behaviors of Semiconductor Lasers under Strong Sinusoidal Current Modulation: Modeling and Experiments at 1.3 μm. IEEE Journal of Quantum Electronics. Vol. 26: 633-641.

Ng W., 2015. Photonics for microwave systems and ultra-wideband signal processing, Optics Communications. http://dx.doi.org/10.1016/j.optcom.2015.09.073i

Piprek J. and Bowers J. E., 2002. Analog Modulation of Semiconductor Lasers. In: Chang W., eds. RF Photonic Technology in Optical Fiber Links. Cambridge University Press. Chapter 3

Scotti F, Laghezza F, Ghelfi P., and Bogoni A, 2015. Multi-Band Software-Defined Coherent Radar Based on a Single Photonic Transceiver. IEEE Transactions on Microwave Theory and Techniques, Vol. 63, No. 2: pp. 546-552.

Urick V.J., McKinney J.D., Williams K.J., 2015. Fundamentals of Microwave Photonics. Hoboken, New Jersey, USA: John Wiley & Sons, Inc.

AN AGENT-BASED MODELLING AND SIMULATION FRAMEWORK TO ASSESS SMALL BUSINESSES' RECOVERY FROM FLOODING

Meshal Alharbi, Graham Coates

School of Engineering and Computing Sciences, Durham University, United Kingdom

meshal.g.alharbi@durham.ac.uk, graham.coates@durham.ac.uk

ABSTRACT

Small and Medium Enterprises (SMEs) form a major part of the economy in many developed countries. For example, in the United Kingdom, SMEs make up 99.9% of all businesses, and thus any disruptions to their operations, such as those caused by flooding, can have severe economic consequences. Given the importance of SMEs, research needs to be undertaken in modelling their behaviour when faced with operational disruptions due to flooding. Agent-Based Modelling and Simulation (ABMS) has been recognized as one approach that allows the study of complex problems in business. This paper outlines initial research carried out towards the development of an ABMS framework aimed at facilitating assessments of the effects of SMEs' behaviours on recovery from flooding.

Keywords: Small and Medium Enterprises; agent-based modelling and simulation; flooding; business recovery.

1. INTRODUCTION

Flooding can result in significant financial losses to a nation's economy due to the disruption caused to individuals, organizations and communities (Coates et al. 2014). In the United Kingdom (UK), Small and Medium Enterprises (SMEs) are essential to the economy as they account for 5.4 million businesses, which represents 99.9% of all businesses with the remaining 0.1% being large businesses (Rhodes 2015). However, the definition of what constitutes an SME varies depending on the country of origin (Hallberg 1999). For example, in the UK, the term SME refers to any business that has fewer than 250 employees (Lukács 2005), whereas in the USA and Australia it is fewer than 500 and 100 respectively (Ayyagari et a. 2007). In relation to the UK definition, SMEs are categorized as micro, small or medium according to their number of employees. That is, a micro-business has 1 to 9 employees, a small business has 10 to 49, and a medium business has 50 to 249 (White 2016). In recognizing the economic importance of SMEs, it has been seen in recent years that these businesses can be vulnerable to the effects of extreme weather events (Crichton 2006, Ingirige et al. 2010). This vulnerability has been reported as being due to their limited resources (Sullivan-Taylor and Branicki 2011, Van Gils 2005) and their tendency to lack disaster mitigation plans (Jones and Ingirige 2008, Li et al. 2015). Flooding, in particular, has caused significant financial losses to SMEs in the UK and remains a serious threat (Wedawatta 2013). According to the Environment Agency (EA), UK businesses' losses due to flooding in 2007 were estimated to be in the region of £740 million (Chatterton et al. 2010). In England, it is projected that the annual loss suffered by residential and business properties due to flooding is over £1 billion (Leinster 2009). Thus, mitigating the potential risk and disruption caused to SMEs by flooding has become an issue of key significance (Wedawatta and Ingirige 2012).

SMEs' preparation for, response to and recovery from disruptive events, such as flooding, depend on a myriad of complex and interdependent processes (Wedawatta 2013). ABMS is recognized as an effective tool for solving complex problems in business and social sciences (Prasad and Chartier 1999; Gilbert and Terna 2000; North and Macal 2007; Dignum and Tick 2008). Agent-based models (ABMs) are computational constructs used to simulate the actions and interactions of autonomous agents in order to evaluate their impacts on the system as a whole (Axelrod 1997). Indeed, it has been demonstrated that ABMs are predictive and analytical research tools that can provide insights into complex social behaviour and generate new theory (Johnson 2011). Chappin et al. (2007) pointed out that, while ABMs originate from research domains such as artificial intelligence and social science, they are flexible enough to be used in a wide range of applications. Clarke (2014) indicated that ABMs are preferred when the goal of the model is the simulation of a behavioural unit, such as a person or household, and when the model represents interactions among one or more types of agent. The versatility of agent-based modelling and simulation (ABMS) has been demonstrated in a variety of domains such as: disaster response (Kwan and Lee 2005; Mysore et al. 2006; Saoud et al. 2006); emergency evacuation (Ren et al. 2009; Wagner and Agrawal 2014); investigating business relationships and networks (Blackmore and Nesbitt 2009; Huang and Wilkinson 2013); risk-based flood incident management (Dawson et al. 2011). While ABMS is proposed as a suitable approach to investigate SMEs' response and recovery from a flood, it is recognized that other approaches exist such as simulation-based serious gaming which can be used to understand business concepts, explore different situations and support decision making in operations management (De Gloria et al. 2014; Van Der Zee et al. 2012).

This paper presents progress made in the initial stages of the development of an ABMS framework to enable assessments of the effects of a range of SMEs' preparatory and responsive behaviours in relation to their recovery from flooding.

2. OVERVIEW OF ABMS FRAMEWORK

The research carried out to date and reported in this paper builds on earlier work (Coates et al. 2014; Li and Coates 2016). The aim of this research is to develop and use ABMS to assess the recovery of flood-affected SMEs' from a variety of industrial sectors. Further, a more comprehensive range of preparatory and responsive SME behaviours will be modelled than seen previously, including those related to promoting greater co-operation between SMEs. Such co-operative behaviours can contribute to flood-affected businesses maintaining some level of operations during (via mutual aid) and in the immediate aftermath of a flood event. To achieve this aim, the research has been divided into two main parts:

1. Enabling the identification of SMEs affected by a simulated flood event in a specified geographical area.
2. Modelling SMEs as agents (and other related organisations), with behaviours and attributes, identified in the specified geographical area to enable the simulation of these businesses in the post-flood recovery stage.

As indicated in Figure 1, within the ABMS framework being developed, a geographical environment is modelled that takes dynamic inundation data as its input from the simulation of a flood event. The modelled geographical environment (MGE) uses Ordnance Survey (OS) data, combined with the inundation data, to enable the identification of each individual business flooded over the timeline of the flood event simulated. Subsequently, flooded-affected businesses, in addition to those unaffected by flooding and other related organisations, are to be modelled as autonomous agents and simulated before, during and after a flood event.

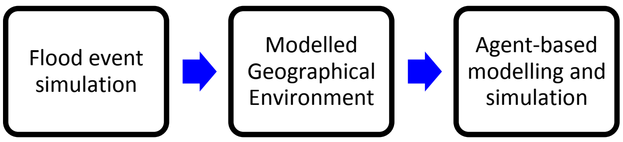


Figure 1: ABMS Framework Overview

SMEs from various industrial sectors, including manufacturing, will be modelled according to behaviours specific to that sector of business. These behaviours of SMEs from a range of industrial sectors have been identified from semi-structured interviews with businesses that have experience of flooding, along with academic and government agency advisory literature. Currently, at this early stage of the research, only manufacturing SMEs have been modelled as agents, with each exhibiting only several pre-flood and post-flood SME behaviours.

3. FLOOD SIMULATION AND GEOGRAPHICAL ENVIRONMENT MODELLING

Prior to the research reported in this paper, the severe flood event of 2007 in Sheffield's Lower Don Valley in the UK was simulated. This geographical area was selected due of its high concentration of SMEs from a range of industrial sectors, some of which have experience of flooding and/or are prone to flooding. Flood event simulation generated a series of inundation data files, each of which represented the water depth throughout the area modelled at 30 minute intervals over a 45.5 hour period, i.e. the duration of the actual 2007 flood event. The flood footprint based on a single inundation data file, corresponding to 25 hours after the flood event commenced, is shown in Figure 2(a).

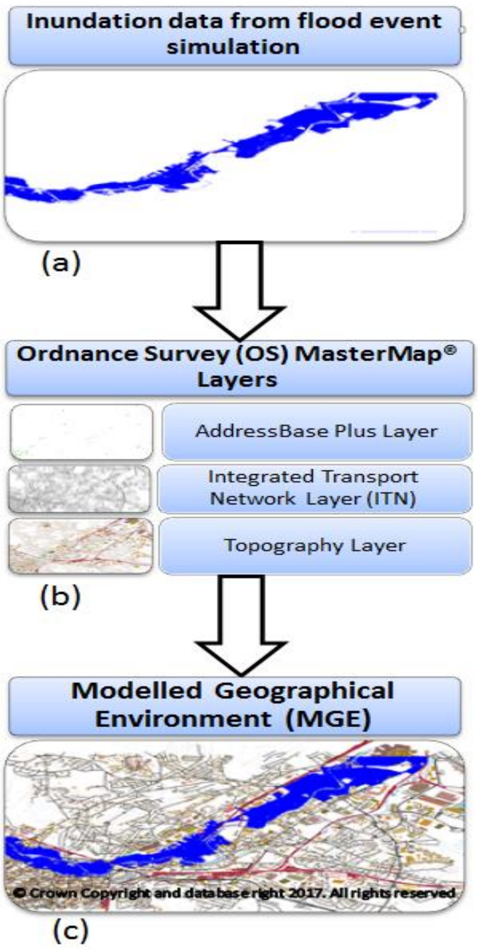


Figure 2: Process of Identifying Flooded SMEs

The inundation data forms the input to the MGE, which is made up from three layers of OS MasterMap®

(Ordnance Survey, 2017) as indicated in Figure 2(b). Specifically, the Topography layer provides data on individual buildings and the Integrated Transport Network (ITN) layer contains data on the road network. In addition, the AddressBase Plus layer provides data related to commercial properties including business name, precise location and classification, as well as other attributes such as number of floors in the premises. Superimposing the inundation data with the MGE's three layers of OS MasterMap® forms a visual representation of Sheffield's Lower Don Valley as shown in Figure 2(c). As stated earlier, inundation data, i.e. the depth of flood water at locations in the geographical area modelled, varies dynamically. Figure 3 shows inundation data, coupled with the OS layer data, at a series of time intervals. That is, at the time the flood commenced (t=0 hours), then 5, 14, and 23 hours after the flood commenced. In Figure 3, two SMEs are highlighted, 'A' and 'B', as these will be referred to in Section 4 in the context of preliminary simulations.

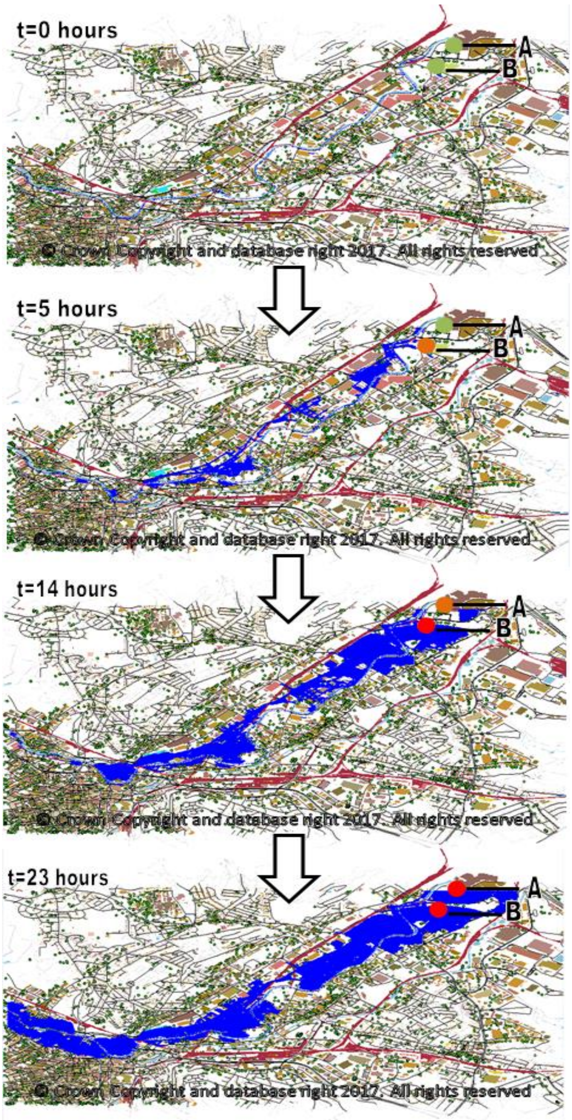


Figure 3: Flood Footprints Based on Inundation Data

In Figure 3, the colour used to represent SMEs 'A' and 'B' signifies the status of the premises of the respective business at the specified time: green for not-flooded; amber for flooding commenced; red for flooded. Throughout the geographical area modelled, the flooded and non-flooded SMEs can be seen as illustrated in Figure 4.

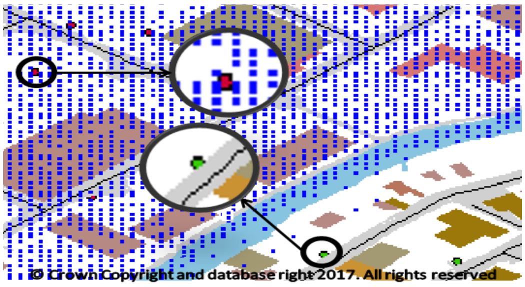


Figure 4: Identified Flooded and Non-flooded SMEs

Based on the MGE and inundation data, in Sheffield's Lower Don Valley, 6224 organizations were identified with 906 flooded at some point during the 45.5 hour event modelled. Table 1 presents a breakdown of these organizations according to the International Standard Industrial Classification of all Economic Activities (ISIC) (United Nations 2008).

Table 1: Organizations in Sheffield's Lower Don Valley

ISIC	**Organizations**	
	Flooded	**Not flooded**
Administrative services	116	802
Accommodation services	41	406
Agriculture/forestry/fishing	2	1
Arts and entertainment	13	75
Education	2	76
Electricity/gas/steam	0	7
Finance	3	68
Households as employers	1	36
Human health / social work	1	127
Information/communication	2	17
Manufacturing	256	1207
Mining and quarrying	0	10
Other service activities	4	25
Professional services	47	234
Public administration	7	78
Real estate activities	56	769
Retail	13	253
Transportation and storage	337	1071
Water	5	56
Total	**906**	**5318**

For the purposes of this research, businesses were mapped from the OS business classifications to the ISIC given this is a more recognizable sector categorization. In relation to Table 1, it is noted that the Environment Agency (2007) indicated that more than 1000 businesses were impacted by the 2007 flood event. The disparity between the EA's indication and the 906 businesses identified in this research may be accounted for by the ten year difference between the actual flood event in 2007 and the OS's AddressBase Plus layer dataset from 2017.

As stated in Section 2, at this stage of the research, only manufacturing SMEs have been modelled as agents; albeit at a fundamental level so far. The manufacturing sector was chosen due to SMEs in this sector suffering significant economic losses and damage to their premises, equipment and machinery due to the flooding in 2007 (Environment Agency 2007). Also, according to the OS AddressBase Plus layer dataset used in this research, the manufacturing sector accounted for the largest number of businesses within the geographical area modelled (as presented in Table 1) including both flooded and non-flooded businesses (n=1463).

4. PRELIMINARY AGENT-BASED SIMULATIONS OF MANUFACTURING SMES

In this research, the Recursive Porous Agent Simulation Toolkit (Repast) is being used to carry out ABMS (Collier and North 2011). Repast has been reported as an appropriate framework for applied modelling of social interventions based on current theories and data (Tobias and Hofmann 2004, Balbi and Giupponi 2009). Also, Repast is viewed as amongst the most effective simulation toolkits in terms of its capability of modelling complex systems (Arunachalam et al. 2008). Robertson (2005), along with Railsback et al. (2006), have indicated that one of Repast's strengths is that it was created and developed by social scientists, which facilitates that it supports projects which entail social networks, generic algorithms, systems dynamics and geographical information systems (GIS).

4.1. Agent Modelling of Manufacturing SMEs

Initial simulation work has focused on modelling manufacturing SMEs as agents based on their size (i.e. micro, small or medium) which, as referred to in Section 1, is associated with number of employees. According to statistical data related to businesses in the UK, 99.9% of all businesses are SMEs with 96%, 3.3% and 0.6% being categorized as micro-, small- and medium-sized respectively (White 2016). Thus, the 1463 manufacturing SMEs (see Table 1) have been modelled in these proportions, then randomly allocated a number of employees corresponding to their respective size, as indicated in Figure 5. For example, 246 of the 256 flood-affected manufacturing SMEs, i.e. 96%, are modelled as micro-businesses, which have between 1 and 9 employees (inclusive).

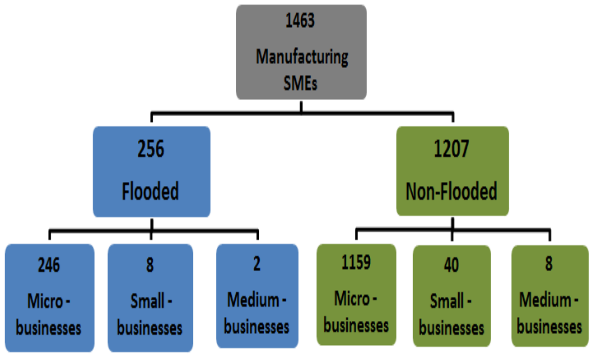


Figure 5: Allocation of Manufacturing SMEs

Flooded manufacturing SMEs are modelled as agents with static attributes including business name, classification code, classification name, size, number of employees, and coordinates of its premises location. Also, from inundation data generated by the flood event simulation, for each manufacturing SME the depth of flood water at 30 minute intervals over the 45.5 hour event period is known. As indicated in Figure 6, the duration of a typical simulation is 21 days which has been divided into three stages: pre-flood; during flood; post-flood. In each day simulated, the 24 hour period is represented by 48 simulation ticks, i.e. one tick per 30 minute period. Furthermore, in each day simulated, working hours start at clock time t_c=8:00 and finish at t_c=18:00, which corresponds to 20 simulation ticks.

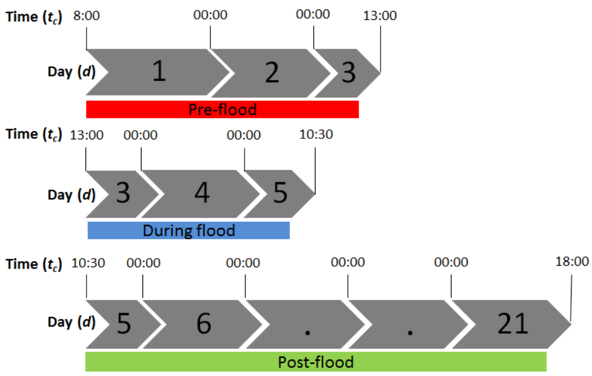


Figure 6: A Typical Simulation Duration (clock hours)

In a typical simulation duration as indicated in Figure 6, the pre-flood stage starts at t_c=8:00 on day 1 until the flood event occurs at t_c=13:00 on day 3. However, the time at which the flood water reaches a particular SME's premises will vary slightly depending on the actual geographical location of the property in the flood affected area. The 'during' flood stage begins at t_c=13:00 on day 3 until the flood water recedes at t_c=10:30 on day 5 given the event modelled has a duration of 45.5 hours. Again, the time at which the flood water recedes from an SME's premises will vary on its location. Post-flood begins from the time the water recedes from an SME's premises until t_c=18:00 on day 21. The period of disruption for an SME begins at the point at which the flood water enters its premises, or at the point the business begins preparing for the flood event (pre-flood), thus reducing or halting production, and ends when it has returned to normal production (post-flood). The term 'normal production' refers to when an SME has resumed its pre-flood level of production. During this period of disruption, manufacturing SMEs' production levels will fall. Business that are flooded will experience their production level fall to zero for a period of time, then gradually recover based on preparatory and response behaviours. In contrast, SMEs threatened by flooding rather than actually being flooded, will see a reduction in production level while preparatory behaviours are enacted, then a resumption of production depending on the degree to which the flood event has disrupted distribution, supply, and other external factors.

In addition to static attributes such as those mentioned earlier, SME agents have dynamic attributes including raw materials (*RM*), machines *(M)*, employees *(E)* and power *(P)*. The relationships between these dynamic attributes are represented in Figure 7, demonstrating that the availability of each is necessary for production.

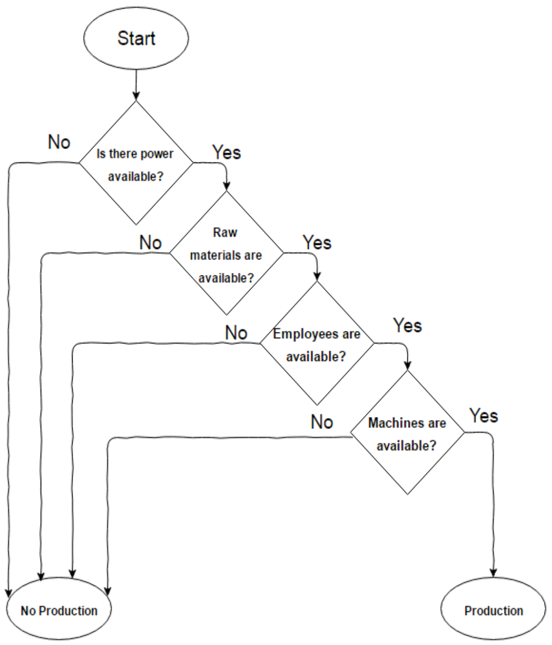


Figure 7: Dynamic Attribute Relationships

These attributes change during the course of the simulation and contribute directly to the SMEs' performance, which is measured as production level *(PL)* in this research. Given the flood event is simulated at 30 minute intervals, each of which corresponds to 1 simulation time step or tick (t_s), then the production level is determined at each half hour interval (or tick) throughout the 21 day period simulated. Post-flood, a flood-affected manufacturing SME's production level can be resumed if:

(a) the power supply is restored to the premises, i.e. $P = 1$ ($P = 0$ indicates that an SME's premises has no power and thus cannot resume the manufacture of products);

(b) raw materials are available, *RM* > 0, where *RM* is the sum of raw material units stored on the ground, RM_{og}, and above ground, RM_{ag};

(c) machines are available, *M* > 0, where *M* is the ratio of an SME's available machines, M_a, to the total number of machines M_t;

(d) production employees are available, $E_a^p > 0$, and $E > 0$ which is the ratio of the number of available production employees working on production, E_p^p, to the total number of production employees, E_t^p.

In relation to (d),

$$E_t^p = 0.81 \times E_t \quad (1)$$

where the total number of employees, E_t, is the sum of the total number of production employees, E_t^p, and the total number of non-production employees, E_t^{np}. Correspondingly,

$$E_t^{np} = 0.19 \times E_t \quad (2)$$

The value 0.81 has been used in equation (1) since, according to the Department of Business Innovation and Skills (BIS 2010), within a *typical* manufacturing business 48.7% of employees work directly in production and 32.1% work in production-related roles. The remaining 19.2% of employees work in roles that are not related to production, such as sales and marketing. Thus, each of the manufacturing SME agents has been assigned 81% of the total employees as being related to production and 19% as being related to non-production.

In terms of the consumption of raw materials during production, if $P = 1, RM \geq 1, M = 1\ and\ E = 1$, then in each 30 minute period a single unit of *RM* is used and, consequently, production level *PL* = 100%. That is, in that half hour period, the premises has power, the raw materials available exceeds or is equal to the unit to be used, all machines are available ($M_a = M_t$), and all production employees are available and work on production ($E_p^p = E_a^p = E_t^p$). However, still with power to the premises and sufficient raw materials available, if not all machines and/or production employees are available, i.e. $M < 1$ and/or $E < 1$, then only a fraction of a unit of raw materials will be used for production (with *PL* < 100%) during the associated 30 minute period. Note that, while in all SMEs a maximum of a single unit of raw materials can be consumed in a 30 minute period (as described above), the size of a unit of raw material varies according to the total number of production employees in that SME. For example, for an SME with $E_t^p = n$, each employee will use $1/n$ units of raw material in a 30 minute period. The code to determine the production level in a 30 minute period, i.e. a simulation tick, is presented in Algorithm 1.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Algorithm 1: Determining Production Level (*PL*)

```
1.  if P = 1 then
2.  {
3.      if raw materials at time t, RM_t ≥ 1 then
4.      {
5.          PL = min {E, M} × 100
6.          RM_{t+1} = RM_t − min {E, M}
7.      }
8.      else if 0 < RM_t < 1
9.      {
10.         if RM_t ≥ min {E, M} then
11.         {
12.             PL = min {E, M} × 100
13.             RM_{t+1} = RM_t − min {E, M}
14.         }
15.         else if RM_t < min {E, M}
16.         {
17.             PL = RM_t × 100
18.             RM_{t+1} = 0
19.         }
20.     }
21.     else if RM_t = 0
22.     {
23.         PL = 0
24.         RM_{t+1} = 0
25.     }
26. }
27. else if P = 0
28. {
29.     PL = 0
30.     RM_{t+1} = RM_t
31. }
```

In Algorithm 1, for each simulation tick, (a) production level and (b) raw materials available at the next tick, are determined depending on whether power is available at the SME's premises (lines 1 to 26) or not (lines 27 to 31). For situations in which power is available, and depending on the amount of raw materials available at the current tick, both (a) and (b) are determined according to the relationship between ratios of production employees working on production to total number of employees, and machines available to total number of machines respectively. Although not included in the algorithm, an SME re-orders raw materials at regular intervals except from the time flooding commences until the flood water recedes from its premises. For example, in relation to Figure 8, at the beginning of the pre-flood period simulated (t_s=0), the SME is modelled as having 100 units of raw materials available (point (a)), which corresponds to the maximum amount required to operate at 'normal' production (i.e. *PL* = 100%) for one working week providing all production employees work on production. The SME consumes raw materials each working day (i.e. 10 clock hours or 20 simulation ticks) until production is stopped due to flooding (at t_s=144).

During the flood (t_s=144 to t_s=197), raw materials may be damaged depending on how much is stored at ground level and above ground level, and the flood water depth in the SME's premises. During the first day after the flood water has receded from the SME's premises, raw materials are re-ordered, which are delivered at the start of the following day as indicated by point (b) in Figure 8. Subsequently, the SME consumes raw materials and re-orders at 5 day intervals. In Figure 8, post-flood, it is noted that raw materials do not always reduce to zero at the end of each working week (points (c) and (d)) due to only a proportion of production employees returning to work, thus working on production, and/or only a percentage of machines being available.

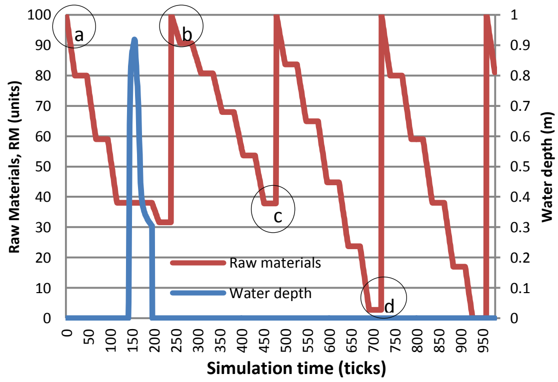


Figure 8: Demonstration of Re-ordering Raw Materials

4.2. Preliminary Agent-based Simulations

As mentioned in Section 4.1, for each SME, the *PL* is determined at each 30 minute interval throughout the three periods (pre, during, post) of the flood event modelled. Pre-flood, which can last up to between 2½ or 3 days (t_s=0 to t_s=120 or t_s=144 ticks in a simulation), *PL* is 100% before any disruption to an SME's operations. However, *PL* drops to zero if the flood water enters an SME's premises and remains at this level during the flood. Post-flood, at some point, production resumes and thus *PL* starts to recover depending on the depth the flood water reached in the premises during the flood and its impact on *P*, *RM*, *M* and *E*. That is, the time at which production is resumed and rate at which it resumes is dependent on the level of damage to machines and raw materials caused by flooding, the distribution of production employees returning to work in the immediate aftermath of the event, and whether or not the premises has power given that it may have been lost during the flood event. At the start of all simulations, each SME has 100 units of raw materials with one unit being consumed per simulation tick providing all production employees are working on production. Given there are 20 simulation ticks per working day (i.e. from t_c=8:00 to t_c=18:00), then, again providing all production employees are working on production, an SME will consume 20 units of raw materials per working day or 100 units per working week. Initially, an SME stores 75 units of raw materials on the ground, RM_{og}, and 25 units above ground at a height of 2.5 metres (m), RM_{ag}, where the above ground storage capacity is 50 units ($RM_{ag}^{c} = 50$). During simulations, on the delivery of raw materials to an SME's premises which occurs every 5 working days, 25 units are stored above ground if storage capacity allows, and 75 units at ground level.

In the following sub-sections, two micro-sized manufacturing SMEs, referred to as 'A' and 'B', which experience different severities of flooding, are considered in terms of how production level was affected by the flood event simulated given these businesses enacted pre- and/or post-flood behaviours:

(a) Pre-flood: on receiving a flood alert from the EA agent, each production employee either lifts-up raw materials from ground level to store them above ground level (see equation (3)) or works on production (see equation (4));

(b) Post-flood: after the flood water has receded from the premises, repair machines at a rate 0.5% of machines per simulation tick either:
(i) by employing a single production employee, if available, while all others work on production, or;
(ii) dividing available production employees such that a number work on repairing machines (equations (5)) while the remainder work on production (equation (7)).

In relation to behaviour (a), the EA issues flood alerts approximately 3 hours before the flood commences, which, if production employees are available, leads to raw materials being lifted-up from the ground to above ground level. Note that (b) is enacted during working hours (i.e. from t_c=8:00 to t_c=18:00) providing the EA alert is received during this time or less than 3 hours from the start of the working day (i.e. after t_c=5:30). Outside these hours, no employees are available to save raw material stored at ground level should the water enter the premises. For behaviour (a), production employees are divided such that they either lift-up raw materials, thus avoiding damage, or continue with production. The number of production employees allocated to lifting-up raw materials, E_l^p, at any simulation tick, t_s, as

$$E_l^p = min\left(E_a^p, \left\lceil \frac{RM_{ag}^c - RM_{ag}}{n_{rm}} \right\rceil, \left\lceil \frac{RM_{og}}{n_{rm}} \right\rceil\right) \quad (3)$$

where n_{rm} units of raw materials can be lifted-up per production employee per simulation tick (in this paper, n_{rm} = 6). Thus, the number of available production employees working on production, E_p^p, is defined as

$$E_p^p = E_a^p - E_l^p \quad (4)$$

In relation to behaviour (b)(ii), in a simulation, the percentage of machines damaged is based on the depth that the flood water reaches in an SME's premises. Production employees available, $E_{\mathrm{a}}^{\mathrm{p}}$, will work on both repairing machines, should any need repairing (i.e. $M_a < 100$) and production. This division of available production employees can be defined according to different strategies. In this paper, the strategy employed determines the number of production employees selected to work on repairing machines, E_r^p, at any simulation tick, t_s, as

$$E_r^p = max\left(1, E_a^p - \left(\left\lceil M \times E_t^p \right\rceil\right)\right) \qquad \text{if} \quad E > M \quad (5)$$

$$E_r^p = 0 \qquad \text{if} \quad E \leq M \quad (6)$$

Thus, the number of available production employees working on production, E_p^p, is defined as

$$E_p^p = E_a^p - E_r^p \quad (7)$$

This strategy ensures that all production employees are assigned to production if there are sufficient machines available for them all to work on. However, if more production employees are available than there are machines available, then the surplus of these employees is assigned to repair damaged machines.

4.2.1. Manufacturing SME 'A'

For SME 'A', the flood commenced at t_s=137 (t_c=4:00 with $RM_{og} = 55$, $RM_{ag} = 25$, M_a =100%) and receded at t_s=174 (t_c=22:30 with $RM_{og} = 0$, $RM_{ag} = 25$, and $M_a = 40\%$) reaching a depth of 1.1m in the premises, as indicated in Figure 9. The flood duration of 37 ticks (i.e. 18.5 hours), coupled with the depth of water reached, led to 60% of machines being damaged. Furthermore, this SME had a total of six production employees ($E_t^p = 6$) but only five available ($E_a^p = 5$) in the first and second day after the flood water receded, then all six available thereafter.

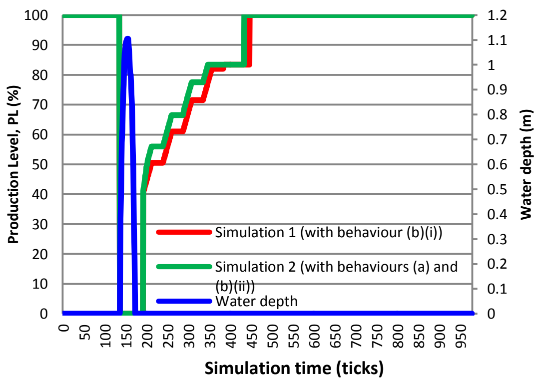


Figure 9: SME 'A' - Production Level and Flood Water Depth in Premises

In Figure 9, two simulations are shown. The red line represents simulation 1 in which SME 'A' has behaviour (b)(i) and the green line signifies simulation 2 in which it has behaviours (a) and (b)(ii). Given the flood commenced outside working hours, (t_c=4:00), the SME 'A' was unable to respond to the flood alert

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

received from the EA agent at t_s=131 (t_c=1:00) in simulation 2. Thus, in both simulations 1 and 2, the SME did not lift-up any raw materials stored at ground level leading to all 75 units being destroyed. However, the 25 units stored above ground level at 2.5m were unaffected by the flood water which reached 1.1m. Once the flood water had receded, employees returned to work at the start of the next working day, t_s=193. At this time, in both simulations, production resumed immediately reaching a level of 40% as governed by the ratio of machines available (0.4), i.e. 40% of machines were available for production.

Subsequently, in simulation 1, a single production employee repaired machines at a rate of 0.5% per tick such that in each working day (20 ticks) 10% of those damaged were repaired and made available for production. At the same time, all remaining employees worked on production from t_s=193 to 386, during which time production level increased steadily from 40% to 83%. During the first day of this period of production, raw materials stored above ground were consumed since those stored on the ground were destroyed during the flood. A delivery of raw materials at the start of the second day after the flood replenished those stored on the ground to 75 units and above ground to 25 units. Subsequently, from t_s=387 to 447, production level remained constant at 83% since during this period the number of production employees was less than the machines available. Furthermore, during this period, 27.6 units of raw materials were consumed. At t_s=447 (i.e. 6 days after the flood), when all machines had been repaired and all production employees were able to work on production, PL=100%.

In contrast, for simulation 2, after reaching 40% at t_s=193 (with $RM_{og} = 0$, $RM_{ag} = 25$, and $M_a = 40\%$), production level increases at a faster rate than in simulation 1 due to two of the five available production employees repairing machines ($E_r^p = 2$) while three worked on production ($E_p^p = 3$) in accordance with equation (5). After 5 hours in the first working day post-flood, at t_s=204 (with $RM_{og} = 0$, $RM_{ag} = 19.4$ and $M_a = 51.5\%$), one of the two production employees allocated to repairing machines was reallocated to production. From this time until a point during the fourth day post-flood (t_s=347), production level increased to 83% with only one production employee repairing machines while the other five worked on production. From t_s=348 to 435, during which time a delivery of raw materials occurred, the production level was governed by the number of employees available, remaining at 83% since $E_p^p = 5$ and $E_a^p = 6$, rather than the number of machines available. At t_s=435, the available machines reached 100%, and thus all production employees worked on production leading to PL=100%. It is observed that the behaviours associated with each of simulation 1 and 2 led to a difference in the time to resume a production level of 100% of 6 hours (447-435=12 ticks).

4.2.2. Manufacturing SME 'B'

As shown in Figure 10, SME 'B' operates at 100% production pre-flood until flood water enters its premises at t_s=119 (t_c=19:00). During the flood, from t_s=119 to t_s=198, the water in the premises reached a depth of 2.2m. This depth of water reached, along with the SME's premises being flooded for 79 ticks (i.e. 39.5 hours), caused all machines to be damaged ($M_a = 0\%$) and thus be in need of repair. SME 'B' has a total of five production employees ($E_t^p = 5$) of which only four return to work in the first day after the flood water receded ($E_a^p = 4$); in the following day the fifth production employee returned to work. In Figure 10, two simulations are shown, as referred to in relation to Figure 9 in Section 4.2.1, with the red and green lines representing simulation 1 and 2 respectively.

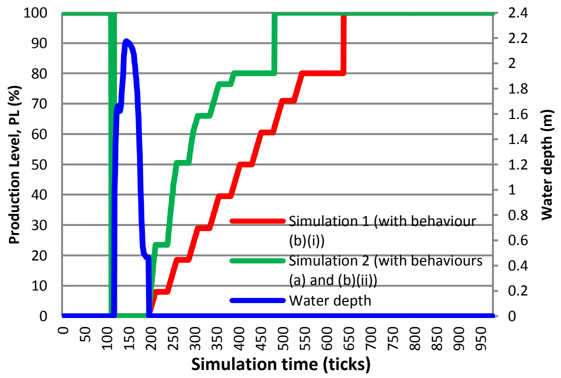


Figure 10: SME 'B' - Production Level and Flood Water Depth in Premises

In simulation 1, from t_s=198 to 640, with raw materials stored above ground being immediately available, along with subsequent deliveries to the SME's premises, production level increases in a gradual and almost linear fashion from 0 to 80%. During this period, production level was constrained by the ratio of the number of available production employees working on production to the total number of production employees (E=0.8 since $E_p^p = 4$ and $E_t^p = 5$; $E_r^p = 1$) as this is less than the ratio of an SME's available machine to the total number of machines ($0.8 < M \leq 1$). Later, at t_s=641, when all machines were repaired and all production employees available to work on production, PL=100%.

In relation to simulation 2, at t_s=113 (t_c=16:00) a flood alert was received from the EA agent and thus all the available production employees were allocated to lift-up raw materials from the ground to above ground level in accordance with equation (3) where $E_a^p = 5$, $RM_{ag} =$ 25 and $RM_{og} = 32$. As a result of this, at t_s=113, the raw materials lifted-up to above ground level reached capacity, i.e. $RM_{ag} = 50$. Subsequently, production was resumed until the end of the working day; the flood water entered the premises afterwards at t_s=119 (t_c=19:00). On the first working day after the flood, at t_s=198 (with $RM_{og} = 0$, $RM_{ag} = 50$, and M_a =0%), all available production employees ($E_a^p = 4$) were allocated to repair machines. Thus, by the end of tick

t_s=198, 2% of the machines were repaired enabling one production employee to be allocated to production at the following tick. Next, at t_s=212, with sufficient raw materials available and M_a =21.5%, one of the production employees repairing machines was reallocated to work on production. Following this, from t_s=212 to 240, two production employees worked on repairing machines while two worked on production. At t_s=241, which is the beginning of the second day after the flood, the fifth production employee returned to work and was allocated to repairing machines leading to PL=23.5%. At t_s=253, with $M_a = 41.5\%$ due to machines being repaired, one of the production employees repairing machines was reallocated to production leading to PL=41.5%. With this allocation of production employees, from t_s=253 to 391, production increased from 41.5% to 80%. From t_s=391 to 483, production level was governed by the ratio of production employees working on production (E=0.8). At t_s=484, with all machines repaired and thus all production employees working on production, PL=100%. Comparing both simulations, it can be seen that in simulation 2, SME 'B' returned to 100% production 78.5 hours (641-484=157 ticks) earlier than in simulation 1 which can be attributed to applying behaviours (a) and (b)(ii) rather than (b)(i).

5. CONCLUSION AND FUTURE WORK

This paper has provided an overview of the initial development of an ABMS framework to assess SMEs' recovery from flooding. For a particular case study area, namely the Lower Don Valley region of Sheffield in the UK, in which the severe 2007 flood event has been simulated, flood-affected businesses have been (a) identified using a modelled geographical environment then (b) represented as agents in preliminary simulations. More specifically, manufacturing SMEs were selected to be modelled as agents given this type of business experienced severe damage to their premises and disruption to their operations as a result of the 2007 flood event. Preliminary simulations carried out showed that a micro-sized SME's recovery from flooding, in terms of production level, depends not only on the depth of water reached in their premises and the number of production employees available, but also on the behaviours of the business pre- and post-flood.

Future research will involve developing a comprehensive set of pre- and post-flood behaviours for SMEs from the manufacturing and other industrial sectors based on data gathered from semi-structured interviews with small businesses having experienced flooding. For example, pre-flood behaviours will include erecting physical flood defences and raising electrical points. Post-flooding behaviours will include cooperating with other businesses unaffected by flooding to facilitate 'mutual-aid' operation across different locations. In addition to modelling SMEs, agents will be developed to model organisations such as customers, suppliers and service companies. Also, simulation experiments will be defined and performed to evaluate the effect of different behaviour combinations on SMEs' recovery from flooding. Finally, once the agent-based model has been fully developed and SME behaviours investigated via simulations, outcomes will be discussed with SMEs.

ACKNOWLEDGEMENTS

The authors thank colleagues from the School of Civil Engineering at the University of Leeds for providing inundation data from flood event simulation.

REFERENCES

Arunachalam, S., Zalila-Wenkstern, R., & Steiner, R. (2008). Environment mediated multi agent simulation tools – a comparison. Proc. of 2nd Int Conf on Self-Adaptive and Self-Organizing Systems Workshops, pp. 43-48, 2008.

Axelrod, R. M. (1997). The complexity of cooperation: Agent-based models of competition and collaboration. Princeton University Press.

Ayyagari, M., Beck, T., & Demirguc-Kunt, A. (2007). Small and medium enterprises across the globe. Small Business Economics, 29(4), 415-434

Balbi, S., & Giupponi, C. (2009). Reviewing agent-based modelling of socio-ecosystems: a methodology for the analysis of climate change adaptation and sustainability. University Ca'Foscari of Venice, Dept. of Economics Research Paper Series, (15_09).

BIS. (2010) Department of Business Innovation and Skills. Manufacturing in the UK: An economic analysis of the sector.

Blackmore, K. L., & Nesbitt, K. V. (2009). Defending against turbulent conditions: Results from an agent-based simulation. International Journal of Business Studies, 17(1), 127.

Chappin, E. J., Dijkema, G. P., van Dam, K. H., & Lukszo, Z. (2007). Modeling strategic and operational decision-making—an agent-based model of electricity producers. In Proc of the 21st annual European Simulation and Modelling Conference (ESM2007), St. Julian's, Malta (pp. 22-24).

Chatterton, J., Viavattene, C., Morris, J., Penning-Rowsell, E. C., & Tapsell, S. M. (2010). The costs of the summer 2007 floods in England.

Clarke, K. C. (2014). Cellular automata and agent-based models. InHandbook of regional science (pp. 1217-1233). Springer Berlin Heidelberg

Coates, G., Hawe, G. I., Wright, N. G., & Ahilan, S. (2014). Agent-based modelling and inundation prediction to enable the identification of businesses affected by flooding. WIT Transactions on Ecology and the Environment, 184, 13-22.

Collier, N., & North, M. (2011). Repast HPC: A platform for large-scale agentbased modeling (pp. 81-110). Wiley.

Crichton, D. (2006). Climate Change and its Effects on Small Businesses in the UK. Axa Insurance.

Dawson, R. J., Peppe, R., & Wang, M. (2011). An agent-based model for risk-based flood incident

management. Natural Hazards, 59(1), 167-189.
De Gloria, A., Bellotti, F., & Berta, R. (2014). Serious Games for education and training. *International Journal of Serious Games*, 1(1).
Dignum, M. V., & Tick, C. (2008). Agent-based analysis of organizations: Formalization and simulation.
Environment Agency (2007) Review of 2007 summer floods,https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/292924/geho1107bnmi-e-e.pdf.
Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. Mind & Society, 1(1), 57-72.
Hallberg, K. (1999). Small and medium scale enterprises: A framework for intervention. The World Bank.
Huang, Y., & Wilkinson, I. F. (2013). The dynamics and evolution of trust in business relationships. Industrial Marketing Management, 42(3), 455-465.
Ingirige, B., Wedawatta, G., & Amaratunga, D. (2010). Building up resilience of construction sector SMEs and their supply chains to extreme weather events. International Journal of Strategic Property Management, (4), 362-375.
Johnson, L. (2011). Complexity modelling and application to policy research. In Complex Systems Institute, University of North Carolina Charlotte
Jones, K., & Ingirige, B. (2008, June). Extreme weather events and business continuity planning. In CIB W70 International Conference in Facilities Management.
Kwan, M. P., & Lee, J. (2005). Emergency response after 9/11: the potential of real-time 3D GIS for quick emergency response in micro-spatial environments. Computers, Environment and Urban Systems, 29(2), 93-113.
Leinster, P. (2009). Flooding in England: a national assessment of flood risk. Environment Agency, Almondsbury.
Li, C., & Coates, G. (2016). Design and development of an agent-based model for business operations faced with flood disruption. Complex Systems: Fundamentals & Applications, 90, 275.
Li, C., Coates, G., Johnson, N., & McGuinness, M. (2015). Designing an Agent-Based Model of SMEs to Assess Flood Response Strategies and Resilience. International Journal of Social, Education, Economics and Management Engineering, 9(1), 7-12.
Lukács, E. (2005). The economic role of SMEs in world economy, especially in Europe. European Integration Studies, 4(1), 3-12.
Mysore, V., Narzisi, G., Nelson, L., Rekow, D., Triola, M., Shapiro, A., ... & Mishra, B. (2006, May). Agent modeling of a sarin attack in manhattan. In Proc of the First International Workshop on Agent Technology for Disaster Management, ATDM (pp. 108-115).
North, M. J., & Macal, C. M. (2007). Managing business complexity: discovering strategic solutions with agent-based modeling and simulation. Oxford University Press.
Ordnance Survey: Britain's Mapping Agency". Ordnancesurvey.co.uk. N.p., 2017. Web. 18 Apr. 2017.
Prasad, M. N., & Chartier, D. A. (1999). Modeling Organizations Using Agent-Based Simulations. In Workshop onAgent Simulation: Applications, Models, and Tools (p. 54).
Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations.Simulation, 82(9), 609-623.
Ren, C., Yang, C., & Jin, S. (2009). Agent-based modeling and simulation on emergency evacuation. Complex Sciences, 1451-1461.
Rhodes, C. (2015). Business statistics Briefing Paper, Number 06152.
Robertson, A. (2005). Agent-based modeling toolkits NetLogo, RePast, and Swarm.
Saoud, N. B. B., Mena, T. B., Dugdale, J., Pavard, B., & Ahmed, M. B. (2006). Assessing large scale emergency rescue plans: an agent based approach. Int J Intell Control Syst, 11(4), 260-271.
Sullivan-Taylor, B., & Branicki, L. (2011). Creating resilient SMEs: why one size might not fit all. International Journal of Production Research, 49(18), 5565-5579.
Tobias, R., & Hofmann, C. (2004). Evaluation of free Java-libraries for social-scientific agent based simulation. Journal of Artificial Societies and Social Simulation, 7(1).
United Nations. Statistical Division (2008) *International Standard Industrial Classification of All Economic Activities (ISIC)* (No. 4). United Nations Publications.
Van Der Zee, D. J., Holkenborg, B., & Robinson, S. (2012). Conceptual modeling for simulation-based serious gaming. Decision Support Systems, 54(1), 33-45.
Van Gils, A. (2005). Management and governance in Dutch SMEs. European Management Journal, 23(5), 583-589.
Wagner, N., & Agrawal, V. (2014). An agent-based simulation system for concert venue crowd evacuation modeling in the presence of a fire disaster. Expert Systems with Applications, 41(6), 2807-2815.
Wedawatta, G., & Ingirige, B. (2012). Resilience and adaptation of small and medium-sized enterprises to flood risk. Disaster Prevention and Management: An International Journal, 21(4), 474-488.
Wedawatta, H. G. S. D. (2013). Resilience of construction SMEs to extreme weather events (Doctoral dissertation, University of Salford).
White, S. (2016). Business population estimates for the UK and regions 2016. B. a. I. Skills(Ed.). Sheffield.

IMPLEMENTATION OF A BUSINESS PLAN FOR A SMALL SOFTWARE DEVELOPER ENTERPRISE USING DISCRETE EVENTS SIMULATION

Joali Evelyn Pérez Bonilla[(a)], Aida Huerta-Barrientos[(b)], Jazmin Dillarza Andrade [(c)], Héctor Bautista Martínez [(d)]

[(a),(b),(c), (d)]Department of Systems Engineering
National Autonomous Mexico University

[(a)]joali.perez@gmail.com, [(b)] aida.huerta@comunidad.unam.mx, [(c)]dillarza_andrade@hotmail.com
[(d)]omarbautista07@gmail.com

ABSTRACT
Nowadays, Mexican small enterprises whose activity is the development of software and related services, face the problem of having a life expectancy of less than two years, mainly due to the lack of a business plan. The objective of this study is to design and implement a business plan for small software developer enterprises that let to improve the economic and operational indicators. In the literature review, several business plans have been formulated from the last ten years, however the problem of implementation persists. In this study, during one year and a half, we worked together with the human resources of a small enterprise for the implementation of the business plan. Additionally, a simulation model based on discrete events simulation methodology was also implemented for the analysis of internal process in terms total costs. We consider that this study and the simulation model implemented could support the decision making process of small enterprises whose main activity are services.

Keywords: business plan, simulation of business process, small enterprises, software developers

1. INTRODUCTION

In Mexico there are around 5 million economic units that generate 29 million jobs, according to INEGI (2015) (National Institute of Statistics and Geography). The economic units are classified into six main activities: manufacturing, commerce, private non-financial services, other economic activities, religious activities and public services. Following INEGI (2015), the 99.8% of the Mexican economy is based on micro, small and medium enterprises (MSME). The life expectancy of MSME in the sector of professional technical and scientific services is approximately two years (see Fig. 1). This situation is unsustainable for the Mexican economy. The solution for this problem has been approached from different perspectives, for instance, through the analysis of financing, which allows a direct confrontation to organizational difficulties for entrepreneurship. However, this approach is mainly economic and elements that have tested to be essential to the viability of firms are usually neglected. Economic resources are provided, however, not only the viability is not guaranteed, but also the internal shortcomings of the company will compromise it. On the other hand, entrepreneurs see as a possibility, to pursue personal advice to improve their managerial and leadership skills, which allows them to visualize in a comprehensive way aspects that should be taken into account when undertaking a venture. However, it is very common for the entrepreneurs to be obligated to understand and intervene in all the areas that make up the company. Despite these solutions, enterprises in this sector continue to disappear affecting the national economy in Mexico. The objective of this study is to design and implement a business plan for a small software developer enterprise that let to improve the economic and operational indicators. The implementation is carried out both in the reality and based on a discrete-event simulation model.

This paper is prepared as follows: the literature review of business plans for technology-based enterprises is presented in Section 2. A business plan for small software developer enterprises is proposed in Section 3. A simulation model for the analysis of internal process and key performance indicators is implemented using Anylogic™ software and validated via face validation technique in Section 4. Concluding remarks are drawn in Section 5.

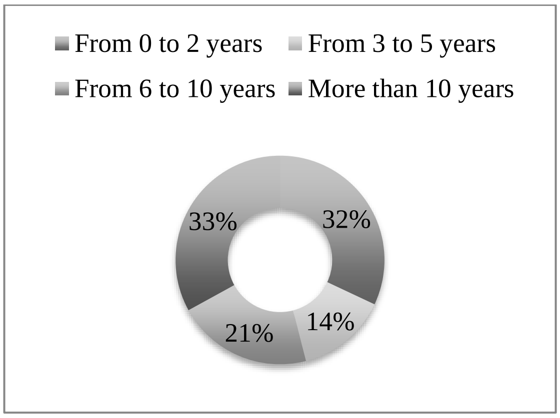


Figure 1: Life expectancy for enterprises in Mexico.

2. LITERATURE REVIEW OF BUSINESS PLANS FOR TECHNOLOGY – BASED ENTERPRISES

As Salazar (2000) explains: a business plan is a document that in an orderly and systematic way details the operational and financial aspects of a company. It is the anticipated and written description, which allows determining where you want to go, where you are and how far to reach the goal set. It is a document that can take different forms depending on each situation. The business plan is based on the idea that all the information necessary to evaluate an investment alternative can be presented in a standardized and systematic way. There are three fundamental concepts that must be contained in the business plan (Salazar, 2000):

1. Business definition. You must understand the clear idea of the product and the business.
2. The environment. It includes market analysis, competition and general context conditions.
3. Financial factors. They are the financial elements that describe the financial conditions of the business.

On the other hand, Diaz et al. (2007) point out that the business plan is a document that identifies, describes and analyzes the business opportunity and, in turn, contains an analysis of economic, market, organizational and economic viability. In order to develop a good business plan, it is necessary to collect a large amount of information and the appropriate sequence of steps, that is, you cannot know the economic viability without knowing the market or the strategies.

The Incubator Program for the Creation of Companies PICE (2016) of INNOVAUNAM, is an environment conducive to the creation of new companies from the early stages. It is an internal organization that offers a range of business development services as well as a space in flexible terms to meet the needs of start-up companies. From the canvas business model of Alexander Osterwalder, the elements that are fundamental to the creation of the organization are identified, starting from the Business Model, the value for an organization, and a guide to generate the business plan. In this case, the business model is defined as a promising idea based on the business plan, which shows that the business idea really constitutes a good investment alternative.

A business plan must contain:

1. Executive Summary
2. Value Proposition: Obtained from identifying a solution to a real need, based on some elements such as: innovation.
3. Market Research: Discover, acknowledge and meet the needs of customers.
4. Technical study: It demonstrates the viability of the project, justifying that it has selected the best alternative to supply the market, in accordance with the constraints of location resources and affordable technologies. It encompasses the means of production, as well as the organization of productive activity.
5. Legal Study. Financial Study. Attachments.

3. A BUSINESS PLAN FOR SMALL SOFTWARE DEVELOPERS ENTERPRISES

The proposal of the plan is based on the structure proposed by Dıaz et al. (2010) including in the section of definition of the company the business model and the proposal that generates value for the small company defined by Osterwalder & Pigneur (2010). This proposal is defined in this way due to the needs of the company and the sector to which it is already directed.

3.1. Executive summary

It must capture the attention and facilitate the understanding of the information. It includes some factors such as:

- Business presentation.
- The product or service.
- The value to the customer.
- The potential market.
- Conformation of team.
- Operations

3.2. Enterprise definition

It must show from the business model all the functional activities of the company. What it must contain is:

- Mission, vision and values.
- Business model.

3.3. Potential market

This element defines whether if the product or service being offered in the market meets the needs of the market. Considering the macro-environment that contemplates the socio-economic context, the legislation and the tendencies; On the other hand, the microenvironment in which the activity is carried out, the clients, the suppliers, the competences and the possible collaborators must be defined.

3.4. Marketing

Once the market and the situation of the sector are known, the strategies to reach the client are defined.

3.5. Internal organization

It should detail the production process or service provision, identify the processes and of these are key, to detail the equipment needs and the quality policy. In this way the functions within the organization chart are identified, the profiles necessary to carry out each activity and create the work team. The financial analysis shows if the company is profitable and can be funded.

The elements that must contain a financial analysis are:

- Cash flow.
- Balance sheet of financial needs.

4. THE BUSINESS PLAN IMPLEMENTATION

Due to confidentiality issues we will name as ENTERPRISE the small software developer enterprise whose data are used in this study.

4.1. Executive summary

- Business presentation. The ENTERPRISE offers customized software development and consulting services based on knowledge of the Brokerage Firms operating processes in the markets operated in Mexico, as well as the regulations of the financial entities and all solutions of information and accompaniment to the user, oriented to support the business processes of the companies that provide financial services, to small, medium and large Brokerage houses. The ENTERPRISE has the mission of providing comprehensive custom software solutions and providing consulting services in information technology for the financial sector. It complements the offer with the specialized consultancy service of the sector. The strategic objectives of the company are mainly positioning to achieve the visibility of being the leading Mexican company in development and software implementation to the extent that it allows the automation of processes. As well as sales and profitability with the marketing plan to get more clients.

- The services. The services offered by ENTERPRISE are based on the own technologies developed internally, a system for automating the operational processes of a brokerage house, which is supported by a proprietary database search, registration and administration technology, security, reliability and traceability of information. The characteristics of this service are installation, configuration and optimization of its developments, execution in update, migration and configuration of new modules.

- The potential market. As Guel and Araiza (2015) states, in the last decade the sector of product development and software services has become a key driver of ma s increasingly growing information economy. Following Guel and Araiza (2015) the software development is capable of generating the most qualified jobs and obtaining foreign exchange for the exports of products and services generated at a distance, such opportunities are possible from the technological advances in the areas of communication, information, architecture and industry. The Mexican financial system is made up of several sectors, most of which operate under the supervision of the National Banking and Securities Commission (CNBV). The market of the financial sector that requires the automation of market operating processes, can be identified that the segment for the product and service we offer represents 5% of the market, in this segment are the Multiple Banking Institutions, Societies Cooperatives of Savings and Loans and the Stockbrokers.

- Conformation of team. In the organization chart, there is a general manager to which are integrated the coordination of design, planning and management of projects and in the third level the directions of the main activities. In addition to managing, there is an outsourcing area for activities that are not carry out by the personal of the ENTERPRISE.

- Operations. The operation of the company is based on the management of key operational processes of said processes are: processes of new developments, maintenance processes and service process.

4.2. Enterprise definition

- Mission. "We are a Mexican company dedicated to providing comprehensive custom software solutions. We develop, implement and provide consulting services in information technology for the financial sector. To adapt the system to the operations of our customers. We serve for our customers to serve".
- Vision. "To be the Mexican company leader in the development and implementation of software made to measure, that allows to automate the critical processes of our clients in the world. To be the main technological partner for the automation of processes and information management in the sectors in which we develop, optimizing our communication in the commercial area and after-sales service".
- Values. Reliability, integrity, service, collaboration, innovation, robustness, honesty, transparency, quality, human capital, and teamwork.
- Business model. In key activities, the management, attention and follow-up to clients is lacking and marketing is scarce. Therefore, in key resources there is no manager, with customer service staff and a salesperson. The clients currently targeted are the financial sector specifically Brokerage Houses, however, the proposed model can be adapted

with its elements to any industry that requires the automation of its processes.

4.3. Potential market
Nowadays, in Mexico 47 banks are in operation. It was identified that the market of the financial sector that requires the automation of processes of operation of the market, it can be identified that the segment for the product and service offered by the ENTERPRISE represents 5% of this market. In this segment are the Multiple Banking Institutions, Cooperative Societies of Savings and Loans and Brokerage Houses.

4.4. Marketing
Digital marketing and social networks. The places to present the ENTERPRISE are congresses, breakfasts and meetings. By the characteristics of the product and services the channel that is handled is direct sale.

4.5. Internal organization
Customer service is provided via telephone and e-mail. The business of the company focuses on three activities that generate value that are the New Development process, customer service process, and maintenance process. The implementation of the new developments process was done based on a simulation model. The development of the simulation model follows the Discrete-events simulation methodology (Banks 1998).

4.6. The conceptual model
Figure 2 shows the flow chart of the activities that are carry out in the New Development process.

4.7. The simulation model
The objective of the simulation model is to measure the total process time and to estimate the resources needed to carry out the New Development process. Data available for each activity were provided by the company's operational staff based on their experience in an approximate way (see Table 1). To verify the data, the company was visited for a week. The human resources for this process are: 2 assistant, 3 programmers and 1 analyst, the working day schedule, and the meal times of each one. We implemented the simulation model using AnyLogic^{TM} software release 7.0.2. We used the Process Modeling Library whose blocks allow users to use combinations of resources and processes to create process-centric models of real world systems (Grigoryev 2015). The simulation model was implemented with three interfaces for users: animation, logic diagram and graphics. In the animation interface, the areas of the company were located as well as the activities carried out (see Fig. 3).

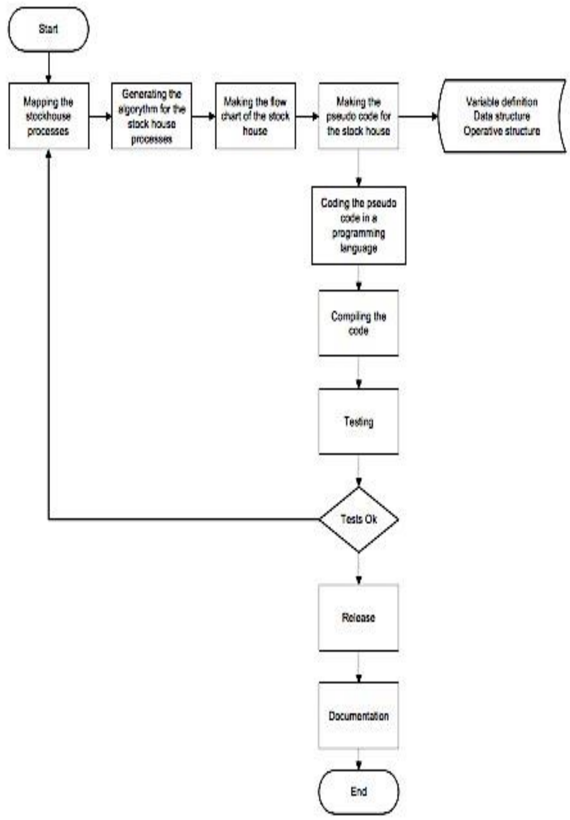


Figure 2: The conceptual model of the New Development process

Table 1: Duration for the main activities of the New Development process

Activities duration			
	Low	Medium	High
Total	-	4 months	-
Direct sales	-	2 years	-
Customer service call	1 minute	5 minutes	10 minutes
Sales time	1 hour	1.5 hours	2 hours
Analysis of the new development	1 hour	5 hours	8 hours
New development	1 hour	20 hours	35 hours
Test	0.5 hour	2 hours	3 hours

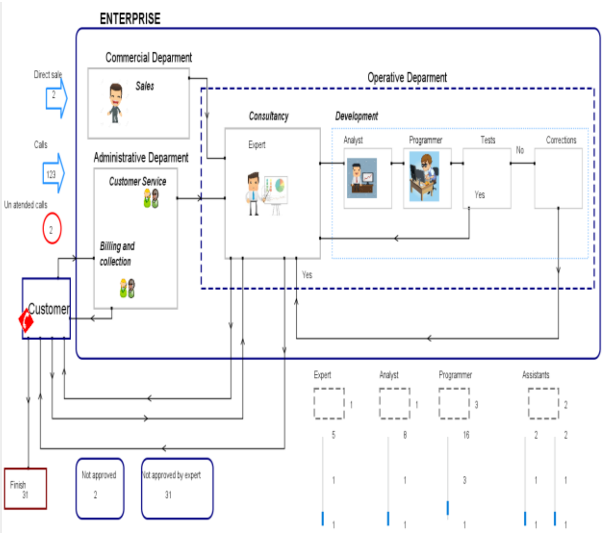


Figure 3: User interface

4.8. The validation of simulation model

As Banks (1998) explains: the validation of simulation model is the determination of whether the conceptual model can be substituted for the real system for the purposes of experimentation. For the validation of the simulation model implemented we used one of the validation techniques proposed by Banks (1998) called face validation. This technique consists on the validation by the user area. The simulation model was presented in a meeting to the team of the ENTERPRISE, and the workers said that the simulation model reflected the behavior of the real system. Therefore, the simulation model of the New Developments process of the ENTERPRISE was validated.

4.9. The verification of simulation model

Following Banks (1998), the verification of simulation model is the determination of whether the computer implementation of the conceptual model is correct. In order to verify the simulation model implemented, the model errors were checked using the compiler included in the simulation software AnyLogic^{TM}. One the compilation process was error free, the simulation model was considered verified.

4.10. Simulation experiments design

The design of experiments has been a very powerful tool in the study of complex systems. On the one hand, it has improved the efficiency and economy of the experimental processes. On the other hand, the use of statistical methods in the analysis of the results obtained has given the scientific objectivity to the derived conclusions (Montgomery & Runger 2003). The deterministic simulation experiments (Kleijnen 2008b) have been applied mainly in the area of engineering and chemistry, in the simulation of airplanes, automobiles and chemical processes. In the case of deterministic simulation experiments, the same simulation model response is always obtained at a given input value. Contrary to what happens in random simulation experiments (Kleijnen 2008a), in which the inputs vary according to a random number generator, thus obtaining an output value of the different simulation model with each value of an input. We conducted simulation experiments considering three factors with two extreme values (see Table 2). The key performance indicator that was analyzed was the internal total time of the New Development process (see Fig. 3) the impact of the three factors on it.

Table 2: Factors for the simulation experiments design

Factors		
	Low	High
A: Management activities time	31 minutes	90 minutes
B: Consulting activities time	90 minutes	480 minutes
C: Production time	180 minutes	2940 minutes

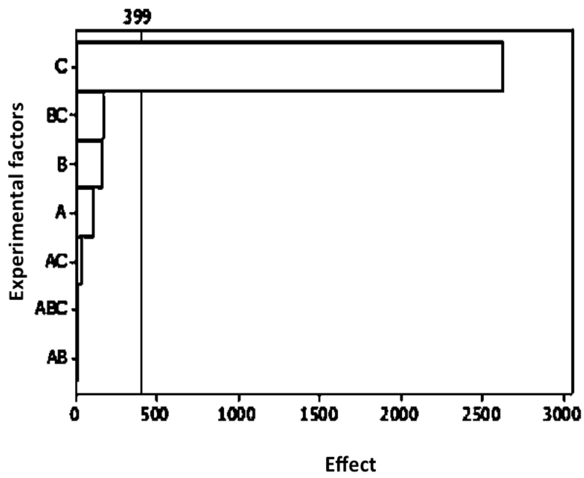


Figure 3: Pareto´s effects

4.11. Service quality analysis

Service quality is measured by the number of complaints, response times, server failures and the number of fines for non-compliance. This is a relevant issue, because it indicates the functionality and compliance of the system. One simulation scenario was proposed to analyze the service quality, in terms of requests no developed. In the scenario analyzed, we considered the number of customer calls and the number of employees as variables. Table 3 shows the results of the simulated scenario.

Table 2: Factors for the simulation experiments design

Simulation experiments			
	Customer calls	Employees	Requests no developed
1	0.2	6	4
2	0.2	18	4
3	5	6	7
4	5	18	7

5. CONCLUSIONS

During the development of this investigation, we were able to conclude that the simulation allowed the partners and collaborators of the ENTERPRISE to recognize the importance of the New Developments process for the improvement of the ENTERPRISE. Likewise, the partners were able to recognize that a further analysis of resources and costs associated to the process can be carried out using the simulation model implemented. Furthermore, we conclude that the simulation of the business plans is a useful tool for micro, small and medium enterprises in Mexico to define, analyze and make decisions regarding the business in their hands. From the statement of different scenarios, beneficial results can be obtained for the problems presented by several enterprises, such as a lack of general vision of the business, with the simulation of the business plan, entrepreneurs can obtain a vision of the enterprise, generating strategies and testing in the model before carrying them out, consequently helping the decision making that allows the survival for more than two years.

ACKNOWLEDGMENTS

The authors appreciate the partial support by National Council for Science and Technology of Mexico (CONACyT) (SNI number 65477).

REFERENCES

Banks, J., 1998. Simulation. New York: John Wiley & Sons.

Díaz, E. et al., 2007. Planes de negocio. In: Navactiva. Available from www.navactiva.com. [accessed 1 February 2017]

Grigoryev, I., 2015. Anylogic 7 in three days. A quick course in simulation modeling.

Guel, C. and Araiza, M., 2015. A Reflection on the Determinants of Development of Software Industry in Mexico. International Journal of Good Conscience. 10(3),71-79.

INEGI, 2015. Instituto Nacional de Estadística y Geografía. Inf. téc. INEGI. Available from:http://www.inegi.org.mx/est/contenidos/Proyectos/ce/ce2014/doc/presentacion/pprdce2014.pdf. [accessed 9 March 2016]

Kleijnen, J.P.C., 2008a. Design and analysis of simulation experiments. New York: Springer.

Kleijnen, J. P. C., 2008b. Design of experiments: overview. Proceedings of the 2008 Winter Simulation Conference, pp. 479-488.

Montgomery, D. C. & Runger, G. C., 2003. Applied Statistics and Probability for Engineers. 3rd edition. Interactive e-Text. John Wiley & Sons.

Osterwalder, A. & Pigneur, Y., 2010. Generación de modelos de negocio. Barcelona: Deusto

Salazar, E., 2000. Guía para la elaboración de un plan de negocios, aplicable a un grupo de desarrollo del IMP Thesis (Master). Universidad Nacional Autónoma de México.

JOALI PEREZ BONILLA

Bachelor in Electric and Electronic Engineering, with major in energetic systems. Work experience as project engineer in solid residues degradation systems. Currently student of the master in system engineering at National Autonomous Mexico University (UNAM). Additionally, works as an external consultant in the enterprise wherein the current investigation took place. Her main interests are on business plans for small enterprises, her work experience and academic formation have led her to study the master in system engineering to obtain the necessary tools of a systemic thinking.

AIDA HUERTA-BARRIENTOS

Prof. Huerta-Barrientos received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the Faculty of Engineering-UNAM. She is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

JAZMIN DILLARZA ANDRADE

Currently is a master student of Systems Engineering, in the disciplinary field Industrial Engineering of the National Autonomous University of Mexico, Faculty of Engineering. She is interested in the optimization of industrial processes that impact on the financial resources and the management of the organizations, besides she is interested in the planning of policies and organizational actions that promote the development and growth of the same ones.

HÉCTOR OMAR BAUTISTA MARTÍNEZ

Currently is Master student of the Department of Systems Engineering at the Faculty of Engineering-UNAM. His research interests are simulation of transport systems, logistics, supply chain and quality.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

AVOIDING RULE EXPLOTION AND MAKING APPROXIMATE INVERSE REASONING IN COMPUTING WITH WORDS APPLICATIONS

Oscar G. Duarte[(a)]

[(a)]Universidad Nacional de Colombia

[(a)]ogduartev@unal.edu.co

ABSTRACT
Computing with words applications are mostly built using rule based systems, which have two important lacks: first, it is not possible to deal with high dimension problems because the size of the rule base increases exponentially; and second, there are no way two compute inputs from outputs. In this paper we show an alternative kind of system that remedy those lacks in some applications. It is based on fuzzy arithmetic rather than fuzzy logic. We also show a web application tool for the environmental impact assessment.

Keywords: Computing with words, Fuzzy arithmetic, Inverse Reasoning, environmental impact assessment

1. INTRODUCTION

Fuzzy sets are an useful tool for representing linguistic concepts, as has been recognized from the earliest Zadeh's papers. The concept of linguistic variable was established by Zadeh (1975a, 1975b and 1976), and it has been the keystone of further developments, such as the computing with words (CW) paradigm shown by Zadeh (1999).

In the CW paradigm, a system compute *words from words using words*. Words involved here must be well defined in the context of a Precisiated Natural Language (PNL), that most of times is a set of Linguistic Variables, Modifiers and Semantic Rules.

Figure 1 shows the structure of a Rule Based CW system. The inputs and outputs are words (it computes words from words); the main block is the Approximate Reasoning block, a typical Mamdani inference engine whose rule base is a linguistic one (it uses words in the computation).

The Linguistic Interpretation block translates words into fuzzy sets, the Inference Engine calculates fuzzy sets, and the Linguistic Approximation block translates fuzzy sets into words.

A simple Linguistic Interpretation block just can process labels of the PNL; its output is a the fuzzy set that is related with the label in a Linguistic Variable.

A simple Linguistic Approximation block compares the output of the Inference Engine with the labels of the Linguistic Output Variable, and selects the most similar. Comparison is made with any kind of similarity measure, for example the consistency:

The consistency between two fuzzy sets x, y over de same Universe of Discourse U and with membership functions $\mu_x(u)$ and $\mu_y(u)$ respectively is:

$$cons(x,y)=\sup_u (\min(\mu_x(u),\mu_y(u))$$

A Rule Based CW system has two major lacks:

1. It is not possible to use a Rule Based CW systems in high dimension applications. The rule base has a combinatorial complexity, and as a consequence we just can manipulate a low number of variables and labels. For a simple 7 inputs and 5 labels in each input, we should define up to 5^7 = 78,125 rules. Even if we would be able to do it, the linguistic meaning of that amount of rules is unintelligible. A single rule whose antecedent has seven atomic expressions is also not easy to understand.
2. It is not possible to make inverse reasoning. By Inverse Reasoning we mean the process of computing inputs from outputs. With a Mamdani Inference Engine we cannot make it.

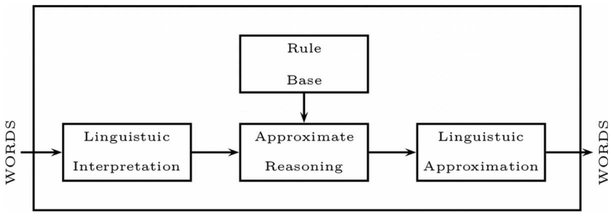


Figure 1: Rule Based CW system

2. ARITHMETIC BASED CW SYSTEMS

In this paper we propose an Arithmetic Based CW system: we propose to compute *words from words using fuzzy numbers* (In fact inputs and outputs are not restricted to words, as we will show later). Figure 2 shows the system structure: The main block is the Approximate Reasoning block that calculates fuzzy numbers from fuzzy numbers using an *approximate reasoning function* (*fra*) instead of a rule base.

In the following, we assume:

- The system has n inputs $x_1, x_2, \cdots, x_n$ and one output y. Multiple output systems are not presented here for simplicity reasons, but can be easily built.

- Every input variable x_i is defined over U_i=[0,1] and the output variable y is defined over V=[0,1]. The sets of all fuzzy sets and fuzzy numbers over U_i are F_{Ui} and N_{Ui} respectively, and the sets of all fuzzy sets and fuzzy numbers over V are F_V and N_V respectively. We define $F_U=F_{U1}$ x FU2 x ... x F_{Un} and $N_U=N_{U1}$ x N_{U2} x ... x $N_{un.}$
- For every input variable x_i there is a linguistic variable X_i with a set of q_i linguistic labels $L_{xi\text{-}1}$, $L_{xi\text{-}2}$, ···, $L_{xi\text{-}qi}$. Every linguistic label $L_{xi\text{-}j}$ has a related fuzzy set $f_{xi\text{-}j}$.
- Analogously, the output variable y has a linguistic variable Y with a set of r linguistic labels $L_{y\text{-}1}$, $L_{y\text{-}2}$,···, $L_{y\text{-}r}$ and every linguistic label $L_{y\text{-}i}$ has a related fuzzy set $f_{y\text{-}i}$.
- Every fuzzy set $f_{xi\text{-}j}$ and $f_{y\text{-}i}$ has a trapezoidal shape. We will use the notation $f = T(a, b, c, d)$ in order to indicate that fuzzy set f has the membership function shown in figure 3.

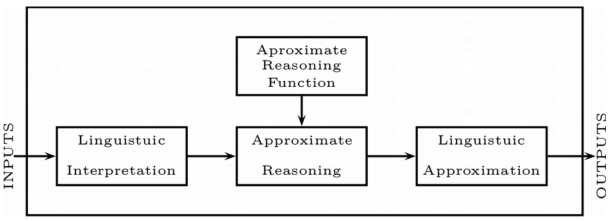


Figure 2: Arithmetic Based CW system

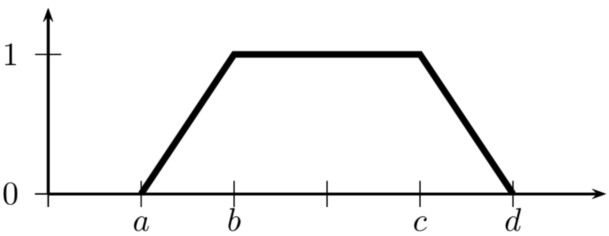


Figure 3: Trapezoidal fuzzy number $f = T(a, b, c, d)$

We will use the usual fuzzy number definition (a normal fuzzy set over R with the upper semicontinuous property). Note that $f = T(a, b, c, d)$ implies that f is a fuzzy number. Now we are ready to explain the blocks shown in figure 2.

2.1. Linguistic Interpretation

The objective of the Linguistic Interpretation block is to translate the inputs into fuzzy numbers. The inputs may be of different types and the output will be always a trapezoidal fuzzy number fn_i for every input. The following are valid types of inputs:

Crisp numbers: if the actual input x_i is the crisp number a, the output will be the singleton fuzzy number $fn_i=T(a,a,a,a)$.

Intervals: if the actual input x_i is the interval [a,b], the output will be the rectangular fuzzy number $fn_i=T(a,a,b,b)$.

Trapezoidal fuzzy numbers: if the actual input x_i is the trapezoidal fuzzy number $T(a,b,c,d)$ the output will be the same input, $fn_i=T(a,b,c,d)$

Linguistic Labels: if the actual input x_i is the linguistic label $L_{xi\text{-}j}$ the output will be its related fuzzy set $f_{xi\text{-}j}$.

Modified Linguistic Labels: valid modifiers are "at least" and "at most" Table 1 shows the corresponding output. In Table 1 we assume the linguistic label $L_{xi\text{-}j}$ has a related fuzzy set $f_{xi\text{-}j}=T(a,b,c,d)$

Simple words: simple valid words are "nothing" and "anything" Table 1 shows the corresponding output.

Table 1: Valid inputs of the types "Modified Linguistic Labels" and "Simple Words"

Input	Output
"at least" $L_{xi\text{-}j}$	$fn_i=T(a,b,1,1)$
"at most" $L_{xi\text{-}j}$	$fn_i=T(0,0,c,d)$
"nothing"	$fn_i=T(0,0,0,0)$
"anything"	$fn_i=T(0,0,1,1)$

2.2. Approximate Reasoning

A typical Mamdani Inference Engine may be viewed as an application $AR: F_U \rightarrow F_V$. However, as typical inputs of this block come from a Linguistic Interpretation block (or from a fuzzyfier block in a fuzzy controller) they are really fuzzy numbers. As a result, we may view the typical Mamdani Inference Engine as an application $AR: N_U \rightarrow F_V$ (note that $N_U \subset F_U$).

We propose a different kind of Inference Engine that may be viewed as an application $ARF: N_U \rightarrow N_V$. In other words, the input of the approximate reasoning block will be the fuzzy numbers $fn_1,fn_2,...,fn_n$ and the output will be the fuzzy number O:

$$O=ARF(fn_1,fn_2,...,fn_n) \quad (1)$$

ARF is an *Approximate Reasoning Function* whose objective is analogous to that of the rule base in a rule based system: it must capture the knowledge about the system.

We argue that in some applications the knowledge about the system is too poor and it has no sense to make a rule base with it. There are some situations in which the only knowledge available is something like:

- "Every time input i increases the output increases (or decreases)"'.
- "Input i is more important than input j "

Of course, we could build a rule base with that knowledge, but a simple crisp arithmetic function can also capture that knowledge. As an example, consider the weighted average function

$$y=\sum_{i=1}^{n} w_i x_i \quad \sum_{i=1}^{n} w_i=1 \; w_i \geq 0 \qquad (2)$$

Equation (2) tell us that output *y* increases as well as any of the inputs x_i increases. If $w_i > w_j$ we also know that the effect of varying x_i is greater than that of varying x_j, in other words, we know that input *i* is more important than input *j*. Suppose (2) refers to a system with 7 inputs every one with 5 labels. Instead of a rule base of 78.125 rules we just need a single equation.

We propose to build (1) using simple crisp arithmetic functions that we will note as *arf* , and the extension principle.

We list some good *arf* of general application. In all cases w_i is a weight variable that let us manipulate the relative importance of every input; it is restricted by $0 \leq w_i \leq 1$ and $\sum_{i=1}^{n} w_i=1$.

We also define s_i as an auxiliary variable that define the sense of the effect of input *i* over the output:

- $s_i=1$ if *y* increases as x_i increases
- $s_i=0$ if *y* decreases as x_i increases

Option 1: A weighted average that includes the sense of the effect of every input:

$$arf_1=\sum_{i=1}^{n} s_i w_i x_i+\sum_{i=1}^{n} (1-s_i) w_i (1-x_i)$$

Option 2: A modified weighted average in which the importance of every input may be varying:

$$arf_2=\sum_{i=1}^{n} s_i w_i g_i(x_i)+\sum_{i=1}^{n} (1-s_i) w_i g_i(1-x_i)$$

where $g_i:[0,1] \rightarrow [0,1]$ is a monotone increasing function such that $g(0)=0$ and $g(1)=1$. As an example, suppose $g_i(x_i)=x_i^r$; if $r>1$ then the lowest values of x_i will be undervalued, and the highest will be overvalued.

Option 3: A weighted average with an offset:

$$arf_3=0.5+[1+\sum_{i=1}^{n} (-1)^{s_i+1} w_i x_i]$$

2.3. Linguistic Approximation

The objective of the Linguistic Approximation block is to translate the output of the Approximate Reasoning block (the fuzzy number *O*) into words. However, we propose different types of outputs for different applications:

A single word: We compute the consistence between *O* and every fuzzy set f_i of the linguistic variable *Y*

$$c_i=cons(f_i,O) \quad i=1,2,\cdots,r$$

Then we select the label with maximum consistency as the output of the system.

A descriptive sentence: We also compute c_i for $i=1,2,...,r$ and then construct a sentence such as

"Output is P_1 L_{y1}, is P_2 L_{y2}, … ,and is P_r L_{yr}"

where P_i is one of the following modifiers:

- "very possibly" if $0 \leq c_i < 1/3$
- "possibly" if $1/3 \leq c_i < 2/3$
- "low possibly" if $2/3 \leq c_i < 1$

A fuzzy number: A valid output is the fuzzy number *O* without any change.

A crisp number: Another valid output is a crisp number representing the central value of the fuzzy number *O*. We use the value of fuzzy numbers defined by Delgado (1988)

A pair of crisp numbers: Another valid output is a pair of crisp numbers representing the central value and the fuzzyness of the fuzzy number *O*. We use the value and ambiguity of fuzzy numbers defined by Delgado (1988)

2.4. Remarks

We want to remark some important aspects of the Arithmetic Based CW systems:

1. Inputs and Outputs are not restricted to words. We can compute words from a set of heterogeneous variables (words, numbers, intervals, fuzzy numbers).
2. As fuzzy numbers are valid input and output variables, it is easy to concatenate two or more Arithmetic Based CW systems *without loosing information about uncertainty*.
3. Using Arithmetic Based CW systems we can design systems of high dimensionality because there are no rule base explosion. But we pay for it:. we must find a crisp function that captures the input-output relationships.
4. There are some applications in which knowledge is enough to construct the *arf* with simple weighted averages; This could be more simple than finding a Rule Base.
5. In low dimensionality problems we can design either Rule or Arithmetic Based CW systems; in section 3 we make a numerical comparison.
6. In section 4 we will show how to compute inputs from outputs using fuzzy arithmetic.

3. RULE BASED VS. ARITHMETIC BASED CW SYSTEMS

Our aim in this section is to make a numerical comparison between Rule and Arithmetic Based CW systems. As we can not design RBCW systems of high dimensionality we limit our experiments to 2 simple cases: a) 1 input - 1 output system and b) 2 inputs - 1 output system.

Our numerical comparison must be done (obviously) between numbers, but RBCW systems can not compute numbers, just words. As both, ABCW and RBCW systems, use the consistency to perform the Linguistic Interpretation we have selected it as the comparison variable. Two types of inputs have been proven: words and numbers. As RBCW systems can not operate with numbers, we have simulated them with words whose associated fuzzy sets are singletons.

Input and Output Linguistic Variables have been always equal: three labels (LOW, MEDIUM, HIGH) with trapezoidal fuzzy sets $T_L(0.0,0.0,0.2,0.4)$, $T_M(0.2,0.4,0.6,0.8)$ and $T_H(0.6,0.8,1.0,1.0)$

3.1. 1 input - 1 output

In this case we have an input x_1 and an output y, whose relationship is increasing. The Rule Base for the RLCW system is:

- R1: IF x_1 is LOW THEN y is LOW
- R2: IF x_1 is MEDIUM THEN y is MEDIUM
- R3: IF x_1 is HIGH THEN y is HIGH

A suitable Approximate Reasoning Function for the ABCW system is

$$fra:\ y=x_1$$

Tables 2 and 3 show the consistency of the output of the systems when inputs are words. There is no difference between them, and we can conclude that the RBCW system is as good as the ABCW system.

Table 2: Consistency of the output of a RBCW system when inputs are words

Input	*cons(O,L)*	*cons(O,M)*	*cons(O,H)*
LOW	1.0	0.5	0.0
MEDIUM	0.5	1.0	0.5
HIGH	0.0	0.5	1.0

Table 3: Consistency of the output of a ABCW system when inputs are words

Input	*cons(O,L)*	*cons(O,M)*	*cons(O,H)*
LOW	1.0	0.5	0.0
MEDIUM	0.5	1.0	0.5
HIGH	0.0	0.5	1.0

Figures 4 and 5 show the consistency of the output of the systems when inputs are numbers. Note the consistency of the output when input is $x_1=0.3$: In the RBCW system the output has the same consistency (0.5) with the three labels, meaning that it can not distinguish if it is LOW, MEDIUM or HIGH; in the other hand, the ABCW system the output has consistency 0.0 with the third, meaning that it is not HIGH. We must conclude that the ABCW performs a better discrimination of the output.

3.2. 2 input2 - 1 output

In this case we have two inputs x_1, x_2 and an output y. The relationship between y and x_1 is increasing, but between y and x_2 is decreasing. The Rule Base for the RLCW system is described in table 4, where as a suitable Approximate Reasoning Function for the ABCW system we design

$$fra:\ y=0.5(1+x_1-x_2)$$

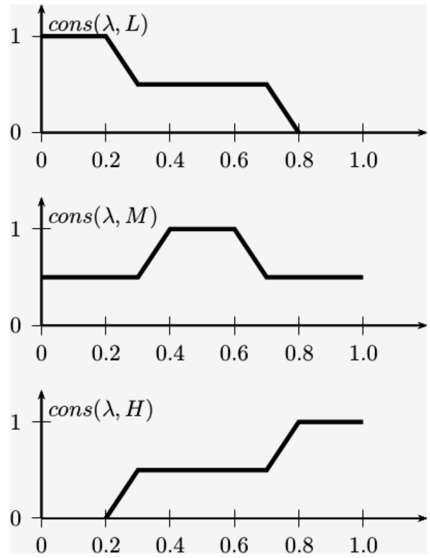


Figure 4: Consistency of the output of a RBCW system when inputs are numbers

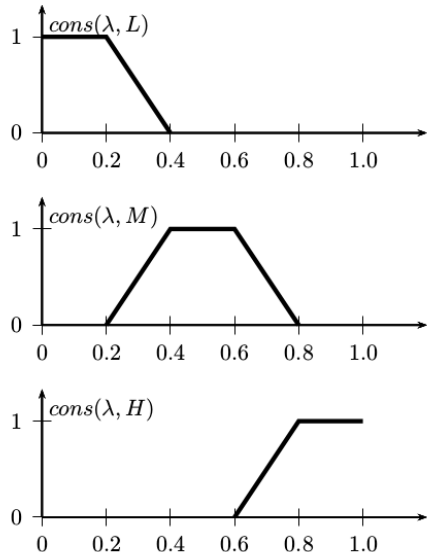


Figure 5: Consistency of the output of a ABCW system when inputs are numbers

Table 4: Rule Base for a RBCW system with 2 inputs

	x_2		
x_1	LOW	MEDIUM	HIGH
LOW	MEDIUM	HIGH	HIGH
MEDIUM	LOW	MEDIUM	HIGH
HIGH	LOW	LOW	MEDIUM

Tables 5 and 6 show the consistency of the output of the systems when inputs are words. Every cell in those tables has three numbers, the consistency of the output with every one of the three labels LOW, MEDIUM and HIGH. Note that every label that has maximum consistency in the RBCW system also has maximum consistency in the ABCW system.

Table 5: Consistency of the output of a 2 inputs RBCW system when inputs are words

	x_2		
x_1	LOW	MEDIUM	HIGH
LOW	0.5/1.0/0.5	0.5/0.5/1.0	0.0/0.5/1.0
MEDIUM	1.0/0.5/0.5	0.5/1.0/0.5	0.5/0.5/1.0
HIGH	1.0/0.5/0.0	1.0/0.5/0.5	0.5/1.0/0.5

Table 6: Consistency of the output of a 2 inputs ABCW system when inputs are words

	x_2		
x_1	LOW	MEDIUM	HIGH
LOW	0.25/1.0/0.25	0.0/1.0/0.0	0.0/0.5/1.0
MEDIUM	1.0/0.5/0.5	0.5/1.0/0.5	0.0/1.0/1.0
HIGH	1.0/0.5/0.0	1.0/1.0/0.0	0.25/1.0/0.25

Figures 6 and 7 show the consistency of the output of the systems when inputs are numbers. We have assumed a monotone relationship between every input and the output; however, the RBCW system produces non-monotone surfaces of the consistencies while thos of ABCW systems are monotone. We argue that ABCW systems represent better than RBCW system the previous knowledge.

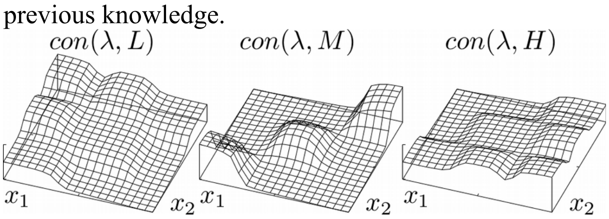


Figure 6: Consistency of the output of a 2-inputs RBCW system when inputs are numbers

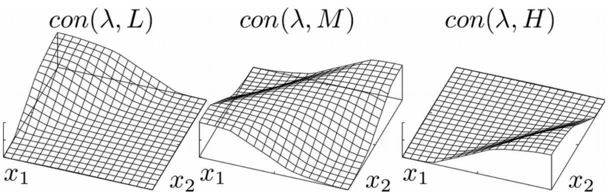


Figure 7: Consistency of the output of a 2-inputs ABCW system when inputs are numbers

4. INVERSE REASONING

CW systems as those of figures 1 and 2 are designed to compute outputs from inputs; however, sometimes we need to perform the inverse calculus: to compute one or more inputs from the remaining inputs and the outputs.

In order to precise the above paragraph, suppose a CW system with three inputs x_1, x_2, x_3 and an output y. Suppose also we know the actual values (words or numbers) of the first two inputs, and we want y to get some desired value; the question is ¿which value must have x_3?

If the CW system is Rule Based, we do not have a procedure to compute x_3. If it is Arithmetic Based, we must deal we the lack of invertibility of fuzzy arithmetic (see Yager (1980) and Bouchon (1997)). We propose to use the Algorithm that computes the Necessary Extension of Inverse Functions (see Duarte (2003) and Duarte (2000)).
In few words, the algorithm verifies if exists a suitable value of x_3 (verifies the invertibility). If it does not exist, the algorithm modifies the desired output y in such a way that the inverse exists. The algorithm is valid for monotone decreasing or increasing functions. *arf* listed in section 2.2 are of this type.

Figure 8 shows the proposed system: We want to compute x_k; Inputs are X_k and y, where X_k is the vector of known x inputs; outputs are x^{nec}_k, the necessary value of x_k, and $\widetilde{y}$, the value of y that we can get with X_k and x^{nec}_k. The inverse reasoning is performed with the Algorithm that computes the Necessary Extension of Inverse Functions. Linguistic Interpretation and Approximation blocks are analogous to those of ABCW systems. We call such system an Arithmetic Based Computing with Words Inverse (ABCWI) system.

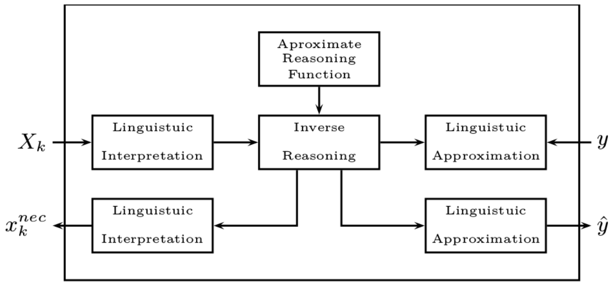


Figure 8: Arithmetic Based Computing with Words Inverse (ABCWI) system

5. AN APPLICATION EXAMPLE

We show in this section an example of risk analysis. It is an ABCW system that computes the risk of damage caused by an atmospheric electrical discharge (lightning) in a specific building (figure 9). A more detailed use of fuzzy techniques for this problem can be found in Gallego (2003); simplification is made here for ilustration proposes. Inputs and outputs are:

- I: it is the *Lightning Intensity*, the peak current of the ligthning.
- D: it is the *Lightning Density*, the number of ligthning per square kilometer that are expected in a year where the building is placed.
- P: it is the *Level of Protection* defined by the type of protection apparatus in the electrical network of the building
- M: it is the *Importance Index* of the facilities in the building.
- R: it is the *Risk of Damage* caused by lightning in the building.

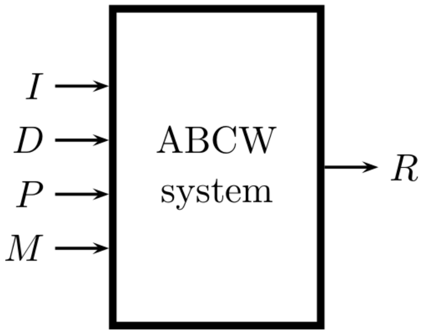


Figure 9: ABCW system of the application example

Note that the relationship between the output R and the inputs *I*, *D* and M is increasing, whereas between *R* and *P* is decreasing (more Level of Protection implies less Risk of Damage). We should study the electrical effects of direct and indirect lightning over the specific electrical network, but this is just an example for illustration, so we omit it. We will use as inputs: numbers for *I* and *D* and words for *P* and *M*. All variables are defined over [0,1] whose linguistic variables are as those of examples in section 3; *I* is the most important variable, so we weight it the double of the others.
In such a situation, we can define *arf* as

$$arf: R= 0.4I+0.2D+0.2(1-P)+0.2M$$

Now suppose the inputs are: *I=T(0.5,0.6,0.7,0.8)*, *D=T(0.2,0.3,0.4,0.5)*, P=MEDIUM, M=LOW. The output of the Approximate Reasoning block is a trapezoidal fuzzy number *O=T(0.28,0.38,0.52,0.66)*.
Consistency between *O* and the three labels are:

$$cons(O,L)=0.40$$
$$cons(O,M)=1.00$$
$$cons(O,H)=0.18$$

Meaning that the risk of damage is MEDIUM. Now we can compute what kind of Level of Protection we need if we want to have a LOW Risk of Damage. The ABCWI system calculates $P^{nec}=T(1.0,1.0,1.0,1.0)$ and the Risk of Damage with that Level of Protection would be

$$\hat{R}=T(0.24,0.30,0.40,50)$$

whose consistency with the three labels is

$$cons(O,L)=0.62$$
$$cons(O,M)=1.00$$
$$cons(O,H)=0.00$$

We must conclude that even with the best Level of Protection P^{nec} we cannot obtain a LOW Risk of Damage.

6. CONCLUSIONS

We can use Fuzzy Arithmetic to design CW systems. Those systems avoid rule explosion for high dimensionality problems; in low dimensionality problems they also have performance as good as the performance of the RBCW systems or even better. Moreover, we can compute inputs from outputs in ABCWI systems that are implemented using the Algorithm that computes the Necessary Extension of Inverse Functions.

ACKNOWLEDGMENTS
If the paper requires an acknowledgements section it can be placed after the main body of the text.

REFERENCES

Bouchon-Meunier B., Kosheleva O., Kreinovich V., and Nguyen H.T.. 1997. Fuzzy numbers are the only fuzzy sets that keep invertible operations invertible. Fuzzy Sets and Systems, pages 155-164, 1997.

Delgado M., Vila M.A., and Voxman W., 1988. On a canonical representation of fuzzy numbers. Fuzzy Sets and Systems, (93):205–216.

Duarte O. G., 2000. Técnicas Difusas En la Evaluación de Impacto Ambiental. PhD thesis, Universidad de Granada.

Duarte O. G., Requena I., and Delgado M., 2003 Algorithms to extend crisp functions and their inverse functions to fuzzy numbers. International Journal of Intelligent Systems, 18: 855–876. doi:10.1002/int.10121

Gallego L. E., Duarte O. G., Torres H., et. al. 2004 Lightning risk assessment using fuzzy logic. Journal of Electrostatics, 60:233–239, March 2004.

Yager R.R., 1980. On the lack of inverses in fuzzy arithmetic. Fuzzy Sets and Systems, pages 73-82.

Zadeh L.A., 1975a. The concept of linguistic variable and its applications to approximate reasoning, part. Information Sciences, 8:199–249.

Zadeh L.A., 1975b. The concept of linguistic variable and its applications to approximate reasoning, part II. Information Sciences, 8:301–357.

Zadeh L.A., 1976. The concept of linguistic variable and its applications to approximate reasoning, part III. Information Sciences, 9:43–80.

Zadeh L.A., 1999. Computing with Words in Information/Intelligent Systems. What is Computing with Words? Physica Verlag.

AUTHORS BIOGRAPHY

Oscar Duarte was born in Bogotá, Colombia. He received the Electrical Engineering degree and the M.Sc. In Industrial Automation from Universidad Nacional de Colombia and the Ph.D. in Computer Science from Universidad de Granada (Spain). He joined the Department of Electrical and Electronics Engineering, Universidad Nacional de Colombia, as a lecturer in 1994. His current research interests include soft computing, modeling and simulation of dynamic systems, control, and engineering education. He is the author of 3 books and more than 40 papers of conference and journals. He was also the Academic Vicedean of the School of Engineering since 2012 to 2016.

COMPLEX NETWORKS OF THE AIR PASSENGER TRAFFIC IN MEXICAN AIRPORTS

Olivia Sashiko Shirai Reyna[(a)], Idalia Flores De La Mota[(b)]

[(a),(b)]Facultad de Ingeniería, UNAM
[(a)]sashikosr@outlook.com, [(b)] idalia@unam.mx

ABSTRACT
Nowadays, the air passenger traffic has been increasing, becoming an excellent, viable and reachable option for many people. This causes that airports may require an efficient organization to serve both, the companies that use the facilities and the passengers. In addition, it is important to consider that the amount of information that is generated may not be easy to analyze, sometimes because the managers don´t know all the information that they have or they don´t know how much this information can help the business.Therefore, in this work, we perform an analysis of the information obtained from some national and international airports of the Mexican Republic, using the methodology of Complex Networks. Also, with the results obtained, we will seek to put forward improvements in the service of this type of facilities, and the infrastructure.

Keywords: complex networks, visibility, time series, airpassengers.

1. INTRODUCTION
Since 2000, a record of all the passengers that arrive and depart at Mexican airports from domestic and international flights have been kept, and have not been analyzed because some of these data are not open data. Some of the data that may concern will be the number of airlines that operate in Mexican airports, the number of passengers, the number of routes, the number of available and taken seats, delays on the flights and more information all of these at domestic and international flights.

The main problem is that the strengths and weaknesses of Mexican airports as a whole systemare not known or identified, causing, among other things, losses in business opportunities and system saturation. For this reason, it is important to perform an initial diagnosis. Then according with the results of this analysis we model the system.Then we built mathematical models of the real data using complex networks, in order toput forward improvements for market strategies.

In order, to improve overall business strategies, it is necessary to apply statistical models and mathematical models to analyze, find patterns and predict data behavior.In this work, we will use different mathematical techniques such as network theory, complex networks, statistics, simulation and time series.

2. STATE OF THE ART
Some papers about complex networks and transportation are the next ones: "A dynamical model for air transportation network" written by Zanin M. et. al. (2009), this article talks about air transportation networks a how we can model this networks. Another article, from the same authors (written by Zanin M. and Lillo F.) is "Modelling the Air Transport with Complex Networks: a short review", this article talks about how to model air transportation as complex networks. Finally, Boccaletti S. et.al. (2014)wrote a summary of many applications of complex networks, but the one´s that we were interested in was the applications of complex networks in transportation systems, they talk about transportation systems and specifically air transportation.These articles were the basement for our research since they give insights on the use of complex networks for air transport problems.

3. METHODOLOGY
For all this analysis, we used the R software, which is an open source programming language and software environment for statistical computing and graphics.We used specific packages for this work, such as,igraph, networks, tkrplot, sand, sna and others, this software allow us to generate graphs/networks, compute different network metrics like clustering or transitivity, different centrality metrics, plot networks and more functions.

We have the monthly information of the air passenger traffic in Mexican airports (domestic, international and total passengers) from January 2000 to March 2017.First, we analyzed one airport, just to prove that all our hypothesis are true. We used the information of Acapulco´s airport, so, with this information we will analyze the data with statistical techniques like time series analysis, then with the obtained results we will do some statistical models we will do some predictions of the air passenger traffic for the next years to analyses the strengths and the weaknesses of all the system in the future. Finally, simulation is used for the validation of the statistical results as well to analyze scenarios.

First, we use all the data that we have, so, we plot this information as a time series. The next figures show the different time series for domestic, international and total passengers of Acapulco´s Airport (respectively).

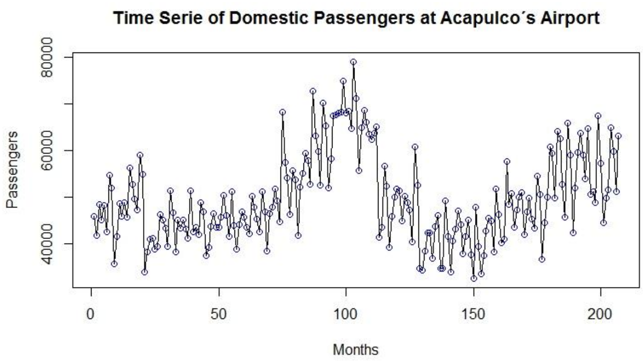


Figure 1: Time Serie Domestic Passengers Acapulco´s Airport

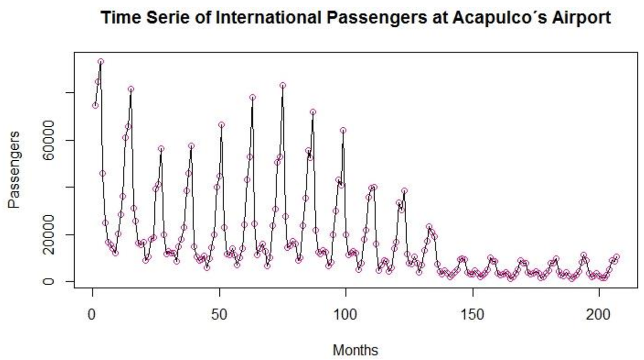


Figure 2: Time Serie International Passengers Acapulco´s Airport

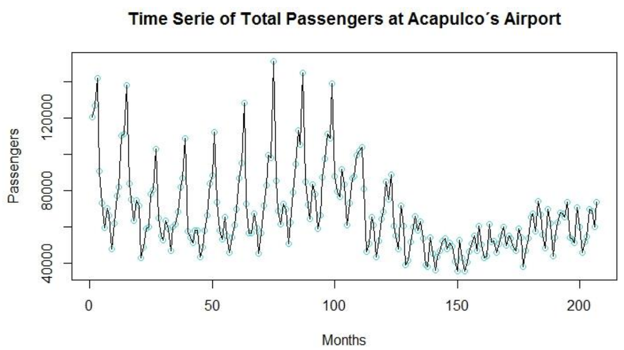


Figure 3: Time Serie Total Passengers Acapulco´s Airport

Now, once we have the plots of the time series we use a technique to transform this time series into complex networks.

3.1. Visibility Algorithm

As we described a little bit, we will transform the time series into complex networks. We use the visibility algorithm to do this transformation (Lacasa L. et al 2008).

The main goal of this algorithm is to map a time series into a network, so, we want to study to which extent the techniques and focus of network theory are useful as a way to characterize time series.

Another important thing is that this network inherits several properties of the time series.

The criterion of this algorithm is to establish two arbitrary data values ($t_a y_a$) and (t_b, y_b)that will havevisibility, and consequently will become two connected nodes ofthe associated graph, if any other data (t_c, y_c) placed between them fulfills:

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a} \quad (1)$$

So, we apply this algorithm to our time series and we obtain the next graphs.

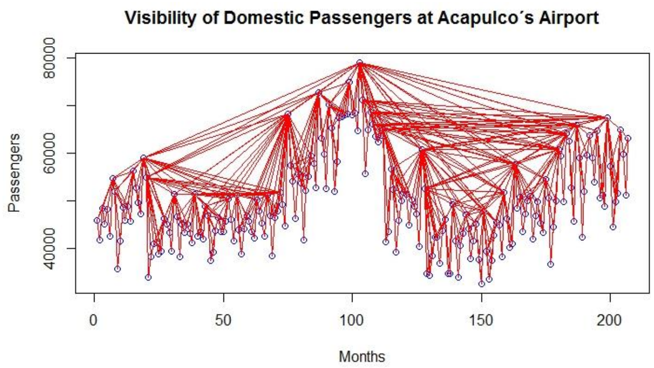


Figure 4:Visibility Domestic Passengers Acapulco´s Airport

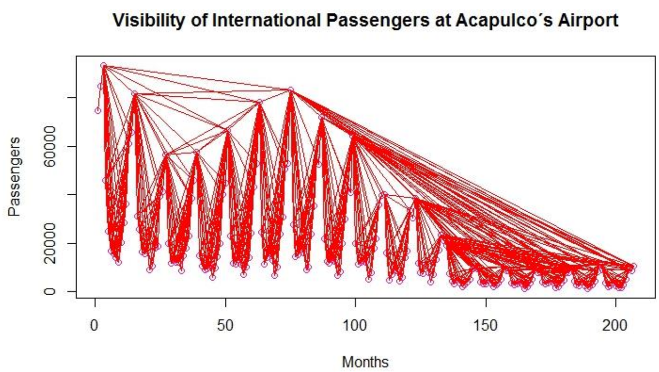


Figure 5:Visibility International Passengers Acapulco´s Airport

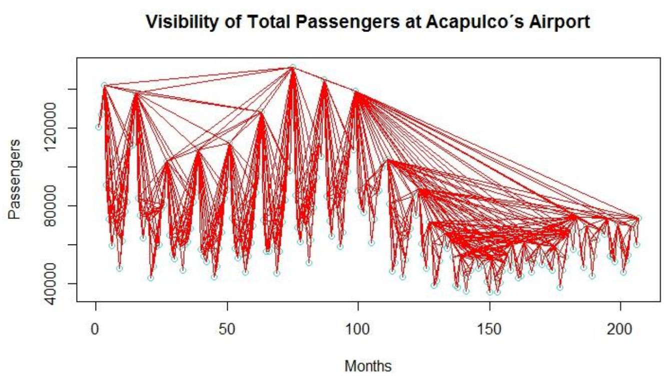


Figure 6:Visibility Total Passengers Acapulco´s Airport

We can easily check that by means of the present algorithm, theassociated graph extracted from a time series is always:

1. Connected: each node sees at least its nearest neighbors (left and right), the first and the last one at least see one.
2. Undirected: the way the algorithm is built up, there is nodirection defined in the links.
3. Invariant under affine transformations of the series data: the visibility criterion is invariant under rescaling of both horizontaland vertical axes, and under horizontal and verticaltranslations.

After, we applied the visibility algorithm, we can obtain the visibility graph, so now, we plot the time series as complex networks, that´s what we will see in the next figures.

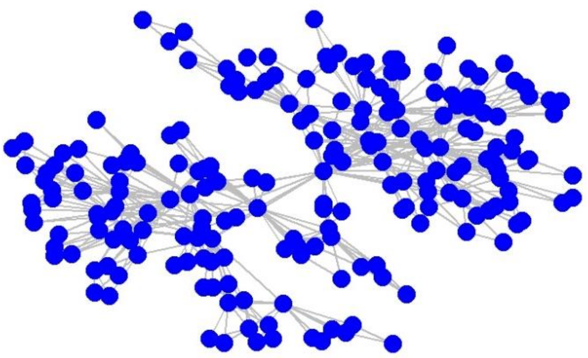


Figure 7:Network Domestic Passengers Acapulco´s Airport

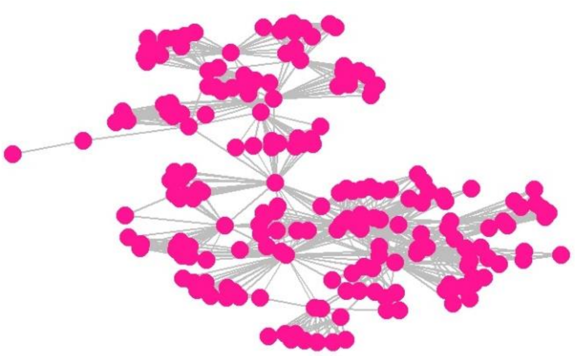


Figure 8:Network International Passengers Acapulco´s Airport

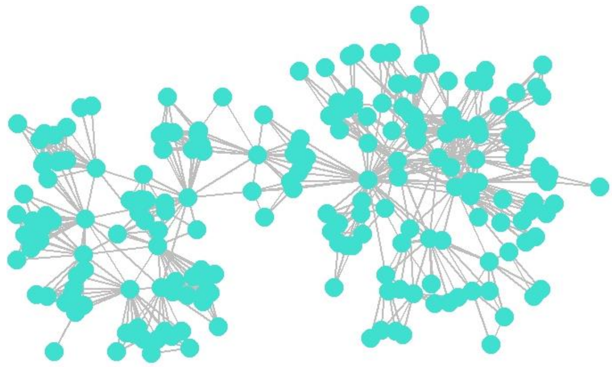


Figure 9:Network Total Passengers Acapulco´s Airport

3.2. Complex Networks

Acomplex networkis graph ornetwork with non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. Such features include a heavy tail in the degree distribution, a high clustering coefficient, assortativityor disassortativity among vertices, community structure, and hierarchical structure. that do not occur in simple networks such as lattices or random graphsbut often occur in graphs modelling of real systems (Newman, 2010).

We will focusthis paper to study the different network metrics, such as the network connectivity. Network connectivity is also a kind of metric to discuss how well parts of the network connect to another. Related terms include network topology, which refers to the structure and makeup of the network as a whole structure.

4. ANALYSIS AND RESULTS

After we have our data as networks we can compute the different network´s metrics.

We compute and compare metrics of our three networks, such as the number of edges, minimum, maximum and mean degree, diameter, mean distance, the number of cliques, the density of the network, if its assortativity or disassortativity among nodes, the global clustering of the network, the mean local clustering, closeness centrality, degree centrality and betweenness centrality. We compute this metrics to study the topology of the networks and to understand and identify different patterns and to classify the networks in the different complex networks models that we have (Random networks, Scale-free networks and Small World networks).

Table 1: Metric´s Results

Results	Domestic	International	Total
Nodes	207	207	207
Edges	638	1013	801
Max. Degree	32	43	44
Min. Degree	2	1	2
Mean Degree	6.164251	9.78744	7.73913
Diameter	8	7	7
Mean Distance	3.663337	3.415224	3.597908
Cliques	8	9	9
Density	0.0299236	0.04751184	0.03756859
Assortativity	0.1387072	0.01292899	0.03675553
Global Clustering	0.3931777	0.504932	0.4548834
Mean Local Clustering	0.7447704	0.7810761	0.7675662
Closeness Centrality	0.3800197	0.3303979	0.3635828
Degree Centrality	0.1254163	0.161226	0.1760236
Betweenness Centrality	0.5545298	0.503284	0.5915958

First, we need to understand the numbers that we have in the Table 1. We notice that there´s a lot of difference between the minimum, maximum and the mean degree, which is because there are just few nodes with a lot of links and there are many nodes with just a few links. The diameter is similar in the three networks. The mean distance is small that means that we can go from any node from the network to another node in a few steps, in this case approximately 3 steps. The number of cliques in the network tells us the number of complete subgraphs in our network. The network density describes the portion of potential connections in a network that are actual connected, so that means that our networks don´t have a lot of actual connections.

To continue with our analysis, our three networks are assortative so that means that there´s a preference for a network´s nodes to attach to others that are similar in some way, soassortativity is often operationalized as a correlation between nodes. The global clustering coefficientis based on triplets of nodes and the local clustering coefficient of a node in a networkquantifies how close itsneighbors are to being a clique (Bollobás, 1998). In our networks the global clustering and the mean local clustering are high.

We also computed the centrality metrics, the closeness centrality of a node is a measure ofcentralityin anetwork, calculated as the sum of the length of theshortest pathsbetween the node and allother nodes in the network. Thus, the morecentral a node is, the closer it is to all other nodes, so, in our networks the closeness centrality is not so high (Caldarelli et. al. 2012).The degree centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has), in our case, the networks have a low degree centrality (Newman, 2010). The betweenness centrality is a measure ofcentralityin a networkbased onshortest paths. For every pair of vertices ina connected network, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through so, the betweenness centrality for each node is the number of these shortest paths that pass through the node; in our networks, we have a high betweenness centrality(Caldarelli et. al. 2012).

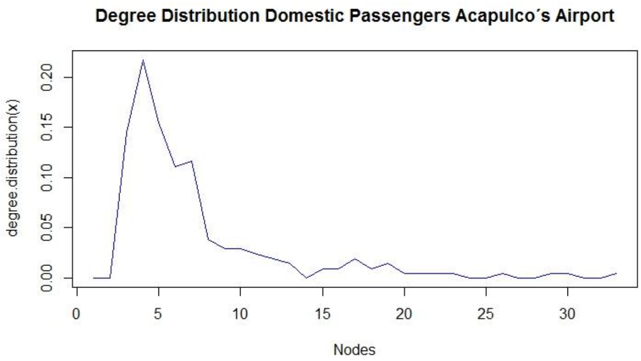


Figure 10:Degree Distribution Domestic Passengers Acapulco´s Airport

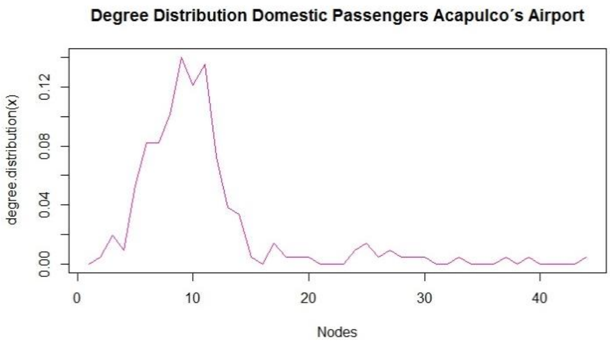


Figure 11:Degree Distribution International Passengers Acapulco´s Airport

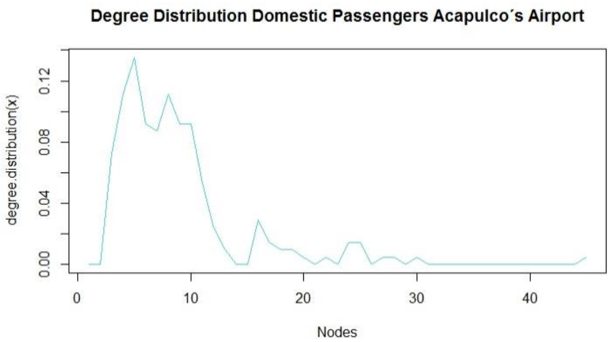


Figure 12:Degree Distribution International Passengers Acapulco´s Airport

With the degree distribution graphs above we notice that it seems that our networks have a power law distribution. Maybe it´s not so clear but there are a few nodes with a high degree and it decays fast, so, there are a lot of nodes with a low degree.

With all these results, we can conclude that our networks follow a scale-free model. The most notable characteristic in a scale-free network is the relative commonness of nodes with a degree that greatly exceeds the average. The scale-free property strongly correlates with the network's robustness to failure. It turns out that the major hubs are closely followed by smaller ones.This implies that the low-degree nodes belong to very dense sub-networks and those sub-networks are connected to each other through hubs (Barabási et. al. 2003).

ACKNOWLEDGMENTS
To CONACYT scholarship for the first author, and DGAPA-PAPIIT Project ITl 02117 Accesibilidad y movilidad del transporte público urbano en la Ciudad de México, el caso de la delegaci6n Tlalpan.

REFERENCES

[1] Barabási A.-L., Albert R., 1999. Emergence of scaling in random networks.Science 286 (5439): 509-512.

[2] Barabási A.-L., Bonabeau E., 2003. Scale-Free Networks. Scientific American. 50-59.

[3] Boccaletti S., Latora V., Moreno Y., Chavez., Hwang D.-U., 2006. Complex networks: Structure and dynamics. Physics Reports. 424. 175-308.

[4] Bollobás B., 1998. Modern Graph Theory.Springer.

[5] Buchanan M., 2002. Nexus: Small Worlds and the Groundbreaking Science of Networks. New York, N.Y.: W.W. Norton.

[6] Caldarelli G., Catanzaro M., 2012. Networks: A Very Short Introduction. Oxford University Press, Oxford.

[7] Caldarelli G., Vespignani A.,2007. Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science. World Scientific Publishing Co., Inc., USA.

[8] Dalgaard P., 2008. Introductory Statistics with R. Springer.

[9] Dorogovtsev S.N., 2010. Lectures on Complex Networks. Oxford University Press, Inc. New York, NY, USA.

[10] Kolaczyk E.D., Csárdi G., 2014. Statistical Analysis of Network Data with R. Springer, New York, USA.

[11] Lacasa L., Luque B., Ballesteros F., Luque J., Nuño J.C., 2008. From time series to complex networks: The visibility graph. PNAS. Vol. 5. No, 13. 4972-4975.

[12] Lacasa, L. et al., 2015. Network structure of multivariate time series. Sci. Rep. 5,15508; doi: 10.1038/srep15508.

[13] Lütkepohl H., 2005. New Introduction to Multiple Time Series Analysis. Springer.

[14] Newman M.E.J., 2010. Networks: An Introduction. Oxford University Press, Oxford.

[15] Newman M.E.J., 2003. The Structure and Function of Complex Networks. SIAM Review 45,167-256.

[16] Newman M.E.J., Barabási A.-L., Watts D.J., 2006. The Structure and Dynamics of Networks. Princeton University Press, Princeton, NJ, USA.

[17] Newman M.E.J., Peixoto T.P., 2015. Generalized Communities in Networks. Physical Review Letters 115, 08870. American Physical Society.

[18] Wei W. W. S., 2006. Time Series Analysis Univariate and Multivariate Analysis. Pearson.

[19] Zanin M., Lacasa L., Cea M., Crístobal S., 2009. A dynamical model for the air transportation network. Proceedings 23rd European Conference on Modelling and Simulation.

[20] Zanin M., Lillo F., 2013. Eur. Phys. J. Spec. Top. 215:5. doi:10.1140/epist/e2013-01711-9

AUTHORS BIOGRAPHY

Sashiko Shirai obtained her bachelor in Actuarial Science in the Faculty of Science of the UNAM. She is studying her Master in Operations Research in the Faculty of Engineering of the UNAM. She is an assistant professor at the Faculty of Engineering and the Faculty of Science. Her research interests lie in complex networks, statistics and operational research.

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 Postgraduate Program at UNAM and her research interests lie in simulation and optimization of production and service systems, based on this research she has been editor of two books edited by Springer.

MODELLING DEPENDENCE OF ARRIVAL SEQUENCING AND METERING AREA TRANSIT TIME ON AIRPORT METEOROLOGICAL CONDITIONS

Margarita Bagamanova[(a)], Juan José Ramos González[(b)], Miquel Àngel Piera Eroles[(c)], Jose Manuel Cordero Garcia[(d)], Alvaro Rodríguez Sanz[(e)]

[(a),(b),(c)] Department of Telecommunications and Systems Engineering
Autonomous University of Barcelona
Carrer de Emprius 2, 08202 Sabadell, Spain
[(d)] CRIDA A.I.E. (Reference Center for Research, Development and Innovation in ATM)
Edificio Allende, Avda. de Aragón, 402, 28022 Madrid, Spain
[(e)] Department of Aerospace Systems, Air transport and Airports
Technical University of Madrid
Plaza Cardenal Cisneros, 3, 28040 Madrid, Spain

[(a)]Margarita.Bagamanova@uab.cat, [(b)]JuanJose.Ramos@uab.cat, [(c)]MiquelAngel.Piera@uab.cat, [(d)]jmcordero@e-crida.enaire.es, [(e)]alvaro.rodriguez.sanz@upm.es

ABSTRACT
Airports are considered complex system in which the coexistence of different actors competing and collaborating for the same resources under operational time uncertainties can cause a poor performance on the overall ATM (Air Traffic Management) system. In order to facilitate the process of decision making to mitigate the propagation of perturbations through the different airport processes a causal model relying on machine learning, using data mining algorithms has been implemented to predict feasible states. This paper introduces a new approach for modelling causal relationships, which can be used for further analysing of feasible scenarios by means of simulation techniques. The state space analysis of reachable airport states is a relevant approach to validate the causal model using a huge amount of historical data for predictive purposes.

Keywords: airport management, Coloured Petri nets, Bayesian networks, decision support tool.

1. INTRODUCTION

The airport is a complex transportation hub serving aircraft, passengers, cargo, and surface vehicles (Office of Technology Assessment 1984). It has three major components: airside, landside and the terminal building, which performs connection between them. Airside is an airport area, where aircraft operate: take off and land, move between the different runways and the terminal. Landside consists of roadways and parking facilities. Terminal complex mainly consists of buildings, serving passengers and air cargo. All these areas are strongly interconnected to each other through different procedures and operations, often fully or partially operated and controlled with the use of IT systems. These operational activities of airports with modernized IT systems are generating an immense amount of data, which can be used for better understanding of hidden dynamics both at the airside and at the landside. However, raw data is quite difficult to be analysed at a glance due to its large volume: for instance, Madrid-Barajas airport airside operations data for one hour of operation with maximum 46 aircraft landed and departed, would make a table of at least 25 different columns with aircraft identification information and data stamps of its movements (landing, taxi in, engine start, taxi out, take off, etc.) and services it went through. The data table of such size can be quite demanding to analyse manually. Therefore for the analysis commodity, these data can be expressed in the form of so called Key Performance Indicators. These Key Performance Indicators (KPIs) are quantitative expressions of effectiveness in achieving performance objectives (European Organisation for the Safety of Air Navigation 2014). As various areas of airport due to their nature can have various KPIs, they are usually merged into Key Performance Areas (KPAs), representing different areas of management interest. An instance of airport KPAs and KPIs is presented in Table 1. The list of KPAs and KPIs can be enlarged according to what targets management team desires to monitor and analyse.

Table 1: Example of KPAs and KPIs (Tabernier 2015)

KPAs	KPIs
Environment / Fuel Efficiency	Average fuel burn per flight.
Airspace Capacity	En-Route and Terminal Manoeuvring Area throughput (average movement per hour).
Airport Capacity	Runway throughput (average movement per hour).
Predictability	Variance of difference in actual & Flight Plan
Punctuality	% Departures < +/- 3 mins vs. schedule due to ATM causes.

Unfortunately, due to tight interdependencies between apparently isolated airport sub processes, airport performance is very sensible to any change in the programmed activities which increase drastically the complexity of airport performance analysis (European Organisation for the Safety of Air Navigation 2017a).
The understanding of the sources of occurred operational issues remains one of the main directions of air transport management scope. Note for instance that European Organisation for the Safety of Air Navigation (EUROCONTROL) aggregates the performance data obtained from European airports and in the form of publicly open documents reveals main European air transport performance problems.
According to one of such reports 2016 was a year with increased volume of flights delay, and furthermore the contribution of reactionary delay has increased up to 45% of total delay minutes (Walker 2017). A reactionary delay is a delay caused by late arrival of aircraft or crew from previous flights (European Organisation for the Safety of Air Navigation 2005). In such manner any delay occurred in the departure airport could lead to severe delays in the following successive flights and their airports of destination. Nevertheless this kind of delay is not the only reason of on-time performance decrease in 2016, as it could be seen on Figure 1.

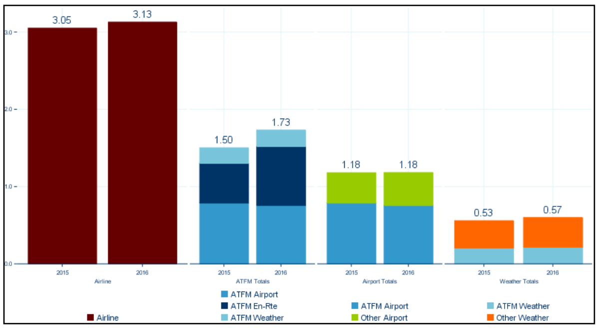


Figure 1: Primary Delay Causes in Europe 2015 vs. 2016, Minutes per Flight (Walker 2016)

Flight delays occurred due to weather conditions also constitute a considerable part of the common delay reasons structure. The fact that weather changes could not be controlled but could be predicted, motivates to obtain the way to efficiently prepare the airport system to any possible impact of weather conditions in order to reduce any negative consequence on its operational activities.
In this paper it is described an approach to model the possible dependence of one of the main airport performance indicators - Arrival Sequencing and Metering Area (further referred to as ASMA) transit time on the weather conditions of arrival airport. Section 2 describes mathematical tools that could be used for the modelling. Section 3 explains the use of Coloured Petri Nets formalism for modelling and simulation of ASMA transit time changes. Section 4 discusses some generated results, directions for further research and some concluding marks are given in Section 5.

1.1. Forecasting in Air Traffic Management
Various organisations perform forecasts for enplanements, airport operations, tracon operations and others. For instance, Federal Aviation Administration (USA) makes its forecast based on demand for aviation services. Econometric and time series modelling are typically used for this purpose. Beside of high potential powerfulness, econometric modelling includes many complex factors and parameters from internal and external infrastructure, which make its application quite difficult and skills demanding. On the other hand, time series modelling seems simpler as it consists of extrapolating knowledge from historical data into the future state. Nevertheless, such extrapolating requires solid statistical analysis and accurate historical data (Federal Aviation Administration 2016).
European Organisation for the Safety of Air Navigation (EUROCONTROL) provides customised analysis and modelling for any airport stakeholders with a use of calculations of performance indicators and different statistical metrics (European Organisation for the Safety of Air Navigation 2017b). International Civil Aviation Organization (ICAO) supports airports planning with medium and long-term forecasts of air traffic for global, regional and route-group levels (International Civil Aviation Organization 2017). These organisations provide open to public global and regional forecasts, however when it comes to the level of particular airport, these organisations could provide only an assistance in analysis and modelling, acting as an external consultant.

1.2. Causal Analysis
Many researchers offer different approaches for understanding and forecasting perturbations of various airport activities. For instance, Quadratic Response Surface (QRS) linear regression models and ensemble Bagging Decision Tree regression (BDT) models have been used to assess weather impact on maximum number of movements per time interval in few USA airports (Wang 2012). Queueing and integer programming models have been used to model the taxi-in process (Idris, Anagnostakis, Delcaire, Hansman, Clarke, Feron and Odoni 1999; Andersson, Carr, Feron and Hall 2000; Roling and Visser 2007). According to the conclusions of these works, the methods used have appeared to be quite helpful, but still not giving perfect approach for airport stakeholders. So the search needs to be continued.
Current modernisation initiatives Single European Sky ATM Research Programme (SESAR) in Europe and NextGen in USA impulse implementation of new operational concepts and technologies, aiming to transform current aviation network into highly efficient, robust and cost optimised system. In order to reach such efficiency it is necessary to understand and fully control any performance area of airport system. For measuring level of success in these tasks airport management can use performance indicators, which permit to compare actual and planned functionality of airport.
It is important to remember, that airports are not operating in isolated conditions, instead, airport

operational disruptions could generate severe reactionary delays through the full aviation network. Thus, it is important the research on new efficient tools for the causal analysis of operational deviations and its prediction, considering the operational conditions that affects each particular airport for the design of mitigation mechanism in the own airport but also at network level.

2. DATA RELATIONSHIP DISCOVERY

We have been provided with Key Performance Indicators data for year 2015, used by analysists of CRIDA (Reference Center for Research, Development and Innovation in Air Traffic Management) and the data from the METAR report, consisting of recorded meteorological conditions on the territory of Madrid-Barajas airport. Some of them are listed in Table 2.

Table 2: KPAs and KPIs

KPA	KPI
TMA	Percentage of flights with holding
	Separations - en NM
	Additional time in ASMA
AIRPORT	Real turnaround time compared to planned
	Additional taxi-out time
	Time between consecutive operations on a runway
	Regulated departures adjustment to CTOT
Capacity	Difference between capacity and demand
	Available capacity
Predictability	Punctual arrivals
	Punctual departures
	Arrivals' standard deviation
	Departures' standard deviation
Meteorology	Wind direction Variable wind direction
	Wind intensity Gusts of wind
	CAVOK
	Predominant visibility Minimal visibility
	Temperature Dew Point Atmospheric pressure
	Phenomenon Cloudiness

For various KPIs' data has been provided in a different form. Some values have been measured for one hour interval, others for 20 minutes interval. Meteorological data consisted of observations for every 30 minutes. Furthermore, we have been commented by CRIDA analysts on the particular interest of discovering hidden causes of perturbations of time in ASMA of radius of 60 nautical miles (NM), expressed as additional ASMA transit time (current performance reports are performed for ASMA with radius of 40 NM).

It has been noted (Klein, Kavoussi and Lee 2009; European Organisation for the Safety of Air Navigation 2015) that weather impact on airport performance is quite significant, but yet not studied well enough. Therefore it has been chosen to study weather impact on one of the KPIs of Madrid-Barajas airport. For the scope of this paper study of weather conditions impact on airport functionality, the following available data has been considered of the first study interest:

- Additional ASMA transit time – a difference between actual time spent by aircraft in ASMA area with radius of 60 NM and average time, statistically measured for particular type of aircraft (for the modelling purpose shortly referred to as ASMA).
- Number of flights with holding patterns – number of flights, which have to take a special route around aerodrome in order to wait for an appropriate moment for landing. (H)
- Wind direction (Wind) and wind intensity (WI).
- Predominant visibility on the aerodrome territory (Vis).
- Dew point (DP).
- Atmospheric pressure (Pres).
- Weather phenomenon type – if fog or any other similar phenomenon occur (Fen).
- Cloudiness (Cloud).

Among the different analytical tools (Marsland 2015; Song 2007) to discover relationship structure between observed variables, the construction of Bayesian networks seems to provide a promising approach to better understanding of complex systems, such as airport, thanks to its capability to cope with high-dimensional problems of different data types (Marsland 2015; Song 2007; Xu, Laskey, Chen, Williams and Sherry 2007) and many powerful computer programs, that made any related computations easy and rather fast.

2.1. Bayesian Networks

Bayesian networks are commonly used for representation of a knowledge about an uncertain area (Song 2007). A Bayesian network is a graphical representation of relationships between different variables, where given variables are represented as nodes and their probabilistic dependencies of each other are represented as directed arcs connecting the nodes. In such manner the absence of direct arc between some two nodes means that these two nodes are conditionally independent of each other (Marsland 2015). When a node has an outgoing arc, it is called *parent*, the nodes with incoming arcs are called *children*. The joint probability distribution P_X of the chosen variables X is represented as a product of conditional probability distributions of each variable X_i (Nagarajan, Scutari, and Lèbre 2013):

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

$$P_X(X) = \prod_{i=1}^{p} P_{X_i}\left(X_i \middle| \Pi_{X_i}\right) \quad (1)$$

Through the conditional probability distribution, calculated for every variable of the studied data, it is also possible to conclude about posterior or future data values. This conclusion is expressed as likelihood function and could serve as the base for prediction model (Gelman, Carlin, Stern, Dunson, Vehtari and Rubin 2014).

The task of discovering a Bayesian network fitting the data consists of two phases: structure and parameter learning. Various algorithms have been developed for the first phase execution. However among all of them only two algorithms have been chosen for the purposes of this paper – Silander - Myllymäki (SM) (Silander and Myllymäki 2006) and Max-Min Hill-Climbing (MMHC) (Tsamardinos, Brown, and Aliferis 2006) algorithms. These algorithms combine constraint-based and score-based algorithms strong sides and are claimed to be highly effective in various situations (Nagarajan, Scutari, and Lèbre 2013). However the approaches, used by these algorithms, are quite different.

2.1.1. Silander – Myllymäki Algorithm

This algorithm was developed for discovering the globally optimal Bayesian network without any structural constraints (Silander and Myllymäki 2006). In order to find the optimal network structure for the specific data, the algorithm has to perform several steps:

1. Find the best parents for all $n2^{n-1}$ pairs of variables, taking the calculated scores for $n2^{n-1}$ as a choice criteria (the higher the score values, the better is the fitness of a candidate variable as a parent).
2. Find the best children node, which cannot be a parent to any other variable.
3. Based on the results of Step 2, find the best arrangement of the variables.
4. Find a best network, taking into account the results of Step 1 and 3 (Silander and Myllymäki 2006).

Despite of quite high quality of the possible SM algorithm results, it has some computational complications. Thus according to the experiments performed by the authors of SM algorithm, the memory requirement for discovering a network of 32 variables is about 16 GB, although distribution of the computation process among few computers could help to overcome this restriction (Silander and Myllymäki 2006). Still, as finding a globally optimal network is NP-hard (Chickering, Meek, and Heckerman 2004), the computational time for SM algorithm is rather long and could easily take 50 hours for a dataset of 30 variables (Silander and Myllymäki 2006). Therefore in order to speed up the discovering of Bayesian network, the use of faster performing algorithm has to be considered as well. One of the most popular algorithms (Nagarajan, Scutari, and Lèbre 2013) with this characteristic is Max-Min Hill Climbing (MMHC) algorithm.

2.1.2. Max-Min Hill-Climbing Algorithm

This algorithm combines principles from local learning and both constraint-based and search-and-score techniques. First, it reconstructs the skeleton of a Bayesian network, and then orients the arcs by performing a Bayesian-scoring greedy hill-climbing search (Tsamardinos, Brown, and Aliferis 2006).

This algorithm has many similarities with the Sparse Candidate (SC) algorithm, which was one of the first successfully performing approaches, applied to large datasets with several hundred variables (Friedman, Linial, and Nachman 2000). Both SC and MMHC perform stepwise reduction of a candidate parents set for each variable and then search for a network that maximise a chosen scoring function. However they do have one important difference. The SC algorithm performs the reduction and network search steps iteratively until there is no improvement in the scoring function value, MMHC performs the candidate parent estimation only once (Nagarajan, Scutari, and Lèbre 2013), therefore fastening the computational process by several times without significant loss in correctness (Tsamardinos, Brown, and Aliferis 2006).

2.1.3. Data Preparation and Learning the Network Structure

As the dataset, provided for analysis, consisted of data for different time intervals, first it has been necessary to transfer all KPIs to the same time interval for facilitation of analysis. It was considered to perform the analysis of data for the time interval of the size of one hour (most common interval of observation that have been seen in the KPIs´ dataset). All chosen for analysis KPIs´ with smaller time interval of observations have been aggregated till the level of one hour.

Additionally, it has been noticed, that provided KPIs values do not all have the same character of values. Some KPIs are observed as *continuous* variables, others – as *discrete*:

- Continuous variable - variable, that can take on any real value within certain interval (Joshi 1989); for instance, additional ASMA time is expressed in minutes.
- Discrete variable - can take on only certain values (Joshi 1989); for instance wind intensity.

Presence of such mixed data can potentially cause a problem in the step of defining a probabilistic model, fitting the data (Nagarajan, Scutari, and Lèbre 2013). Therefore it has been decided to perform a common used solution to avoid the mentioned problem – perform *discretization* or *binning* of the data. Discretization means assigning some particular integer value to the certain intervals of continuous data. There are different ways to define the intervals for data discretization: using expert knowledge on data, using heuristics, performing discretization and structure learning iteratively, etc. (Nagarajan, Scutari, and Lèbre 2013). Taking into account common practice of KPIs´ analysis by CRIDA experts, it has been decided to discretise continuous data as shown in Table 3.

Table 3: Intervals of Discretization

Category	Additional ASMA time, % of unimpeded ASMA time	Visibility, m	Wind direction, °	Actual temperature minus Dew point temperature	Cloudiness	Phenomenon, type	Pressure, QNH
1	(∞;-15)	<50	(22,5; 67,5]	0	FEW	BR	<1013
2	[-15; -10)	[50; 400)	(67,5; 112,5]	-	SCT	DZ	1013
3	[-10; -5)	[400; 8000)	(112,5; 157;5]	-	BKN	FG	>1013
4	[-5; 5)	≥8000	(157,5; 202,5]	-	OVC	RA	-
5	[5; 10)	-	(202,5; 247;5]	-	-	SN	-
6	[10; 30)	-	(247,5; 336,5]	-	-	-	-
7	≥30	-	(337,5; 22,5]	-	-	-	-
0	0	-	VRB	<>0	0	0	-

In Table 3 the following abbreviation have been used:

- VRB – variable wind direction.
- FEW – few clouds.
- SCT – scattered.
- BKN – broken clouds.
- OVC – overcast.
- BR – mist.
- DZ – drizzle.
- FG – fog.
- RA – rain.
- SN – snow.

After data preparation both SM and MMHC algorithms have been executed subsequently in the framework of R software.

As soon as both algorithms have performed their Bayesian network learning for the chosen airport performance data, the best network can be chosen based on the best value of the network scoring functions. Both algorithms have a possibility to evaluate the learnt network with three popular statistical scoring functions: BDeu (Bayesian-Dirichlet equivalent uniform), AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). These scoring functions are common tools for selection between different statistical models and represent goodness of fit of a model to observed data (Brockwell and Davis 1991).

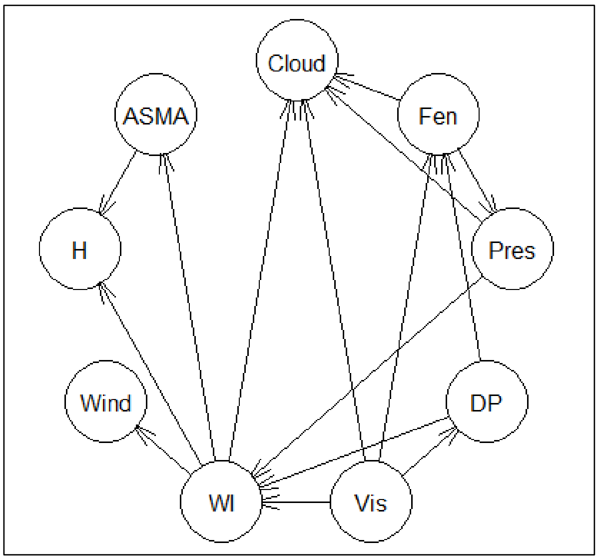


Figure 2: Bayesian network obtained with MMHC and SM Algorithms

In the case of chosen for this paper variables, both SM and MMHC algorithms have come to the same network structure, shown on Figure 2, therefore it was decided not to compare their score functions. Every arc of obtained network had probability of being true of not less than 95% and as MMHC algorithm has come to its results in a shorter computational time (less than one minute for Intel (R) i5-4300M CPU 2.60 GHz, 8 GB RAM), it has been considered to use its results for the further study.

2.2. Bayesian Inference

The knowledge obtained from Bayesian Networks about the data structure and its parameters is used for reasoning on further possible parameters of the chosen airport performance indicators. There are two main approaches for updating the posterior probabilities of data distribution: exact and approximate inference.

Variable elimination and Junction Tree are the two best-known approaches for exact inference task. First approach uses the network structure directly, taking into account the local distributions of the data variables. On contrary, the second algorithm transforms the network by clustering its nodes into a tree. However the feasibility of exact approach is restricted to small networks. Approximate inference algorithms create samples from the local distributions by the use of Monte Carlo simulations and then evaluate them. The sampling can be performed in different ways, implemented in several approximate algorithms (Nagarajan, Scutari, and Lèbre 2013).

The parameters learnt in this step take the form of regression coefficients, belonging to regression functions, describing the conditional dependence between studied variables. For this research it is considered to use the logic sampling approximate inference algorithm, already included in functionality of one of the R software packages for Bayesian Networks. The inferenced parameters of a network have been used for mathematical expression of relationships between observed variables in arc expressions of CPN model in order to perform simulation run and state space analysis.

3. MODELLING WITH CPN FORMALISM

A Coloured Petri Net (CPN) is a formalism, aimed to design, visualise and explore the behaviour of various systems. In order to model the system with CPN formalism it is necessary to define a set of parameters as (Jensen and Kristensen 2009):

- Set of colours – to represent the model entities (key performance indicators).
- Set of places nodes – to represent combinations of the model entities.
- Set of transition nodes – to represent systems' activities (weather changes, arriving aircraft, etc.).
- Set of Arcs – to relate transition and places nodes.
- Guard functions, which are associated to the transition nodes in order to insure their enabling only in case of satisfaction of conditions, described in the corresponding guard function.

For the net elements inscriptions CPN ML, a functional programming language, is included to the modelling framework. It provides the way to make different declarations and perform modelling of data manipulation (Piera and Musič 2010). This language is used in construction of arc functions and in declarations of intervals of possible values for model parameters.
For modelling the chosen KPIs of Madrid-Barajas airport, the colours, representing weather indicators, average additional ASMA time, and number of flights with holding pattern have been chosen. Furthermore it has been considered to introduce the colour, representing system time counter, for having a tool to track system dynamics in time without increasing model complexity.
Design of the developed CPN model is shown on Figure 3.

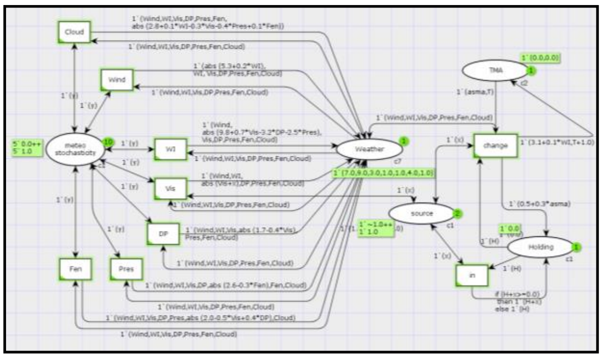


Figure 3: Weather Indicators CPN Model Design

The studied KPIs are distributed among three places as follows:

- Place Holding – number of flights with holding pattern (H).
- Place TMA – additional ASMA time and time counter.
- Place Weather – Wind, WI, Vis, DP, Pres, Fen and Cloud.

Furthermore two supporting places, ensuring the element of stochasticity, have been also added to the model. They are:

- Place Meteo stochasticity – provides tokens for stochastic weather changes.
- Place Source – provides tokens for stochastic changes in number of arriving flights with holding pattern.

In order to formulate the observed ASMA system behaviour in CPN Tools, it is required to define functions for the expressions of arc, connecting elements of the model. The arc functions have the following aspect, based on the maximum likelihood estimation parameters, obtained on the step of Bayesian inference:

$$C_i = \beta + k * C_j \quad (2)$$

Where
C_i = represents CPN colour i, a studied metric.
β = represents intercept value.
k = represents regression coefficient.
C_j = represent CPN colour j, on which CPN colour i is conditionally dependent. When there are more metrics, on which colour i is conditionally dependent, they are included with the corresponding regression coefficients.
After introducing all necessary system parameters, series of simulation runs can be executed in order to verify and validate the model.

4. SIMULATION AND RESULTS

Default tool of CPN Tools v. 4.0.1, which can be used for model verification, is the state space analysis. This analysis consists of generating all states and state changes of a model, that can be reached from the initial state (Jensen and Kristensen 2009). CPN Tools v. 4.0.1 allows to graphically represent all possible system states through *reachability tree* (RT) – a directed graph, where root node represents initial marking of the system, and the successive nodes represent the new states, that can be reached from the initial state, if the corresponding transitions have been fired (Jensen and Kristensen 2009). Few series of state space construction (reachability tree generation) have been performed with a use of CPN Tools v. 4.0.1 software in order to explore how parameters of the system – colours, change their values. The initial markings of the model, used for state space analysis are shown in Table 4. These values have been chosen from the available historical data for the same time period as for Bayesian inference, in order to compare the system dynamics, observed in the historical data and the changes, discovered through RT construction.

Table 4: Simulation Scenarios Initial Markings

Model parameters	Scenario 1	Scenario 2	Scenario 3
	Parameter value		
ASMA time	0	3	0
Flights with holding	0	2	0
Wind direction	7	0	7
Wind intensity	0	2	9
Visibility	4	2	3
Dew point	0	1	1
Pressure	3	1	1
Phenomenon	0	4	4
Cloudiness	0	2	1

In the RT generated for all three chosen scenarios, in every tree a branch with the same weather indicators changes has been found. This has allowed to compare how ASMA transit time has developed in these RT branches and in the historical data. Figure 4, 5 and 6 represent this comparison for each of three simulation scenarios respectively for the time period of 24 hours.

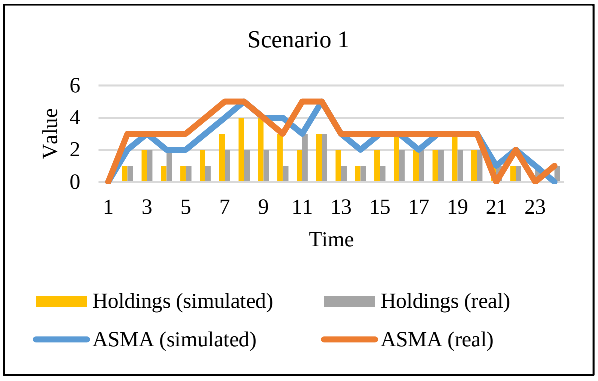


Figure 4: CPN Simulated ASMA Transit Time, Real ASMA Transit Time, CPN Simulated Holdings and Real Holdings Comparison for Scenario 1

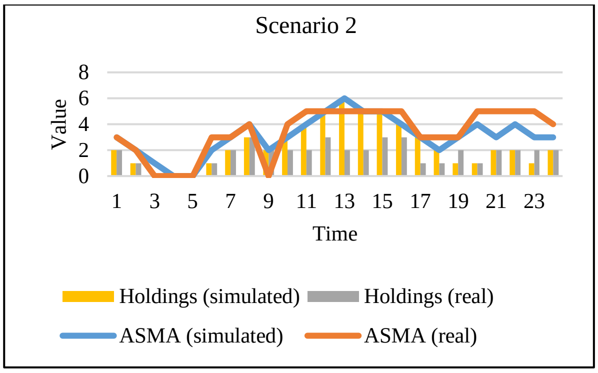


Figure 5: CPN Simulated ASMA Transit Time, Real ASMA Transit Time, CPN Simulated Holdings and Real Holdings Comparison for Scenario 2

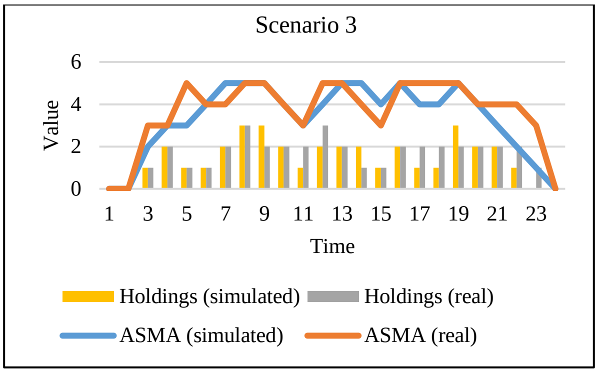


Figure 6: CPN Simulated ASMA Transit Time, Real ASMA Transit Time, CPN Simulated Holdings and Real Holdings Comparison for Scenario 3

All three simulation scenarios have demonstrated that additional ASMA time increases with the delay with the increase of number of flights with holding pattern, and also increases with the development of serious weather conditions (for instance, increasing wind intensity). This is illustrated on Figure 7. Although this correlation becomes not significant in the hours of low number of arriving aircraft (night time). The same behaviour was noted in Scenario 2 and 3 as well.

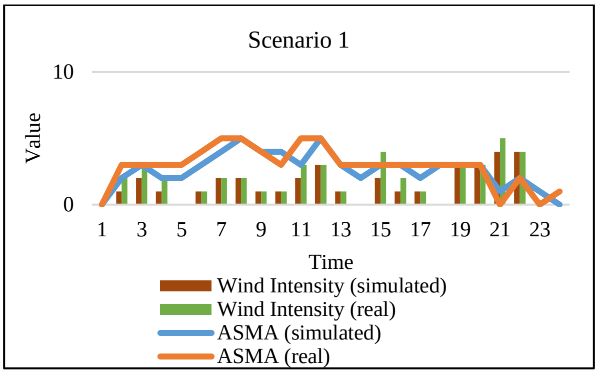


Figure 7: CPN Simulated ASMA Time and Real ASMA Time Comparison to CPN Simulated Wind Intensity and Real Wind Intensity for Scenario 1

Potentially, a set of variables, representing events, preceding the entering of the aircraft into the ASMA, can be added into the model in order to take into account influence of en-route regulations on number of flights with holding pattern.

Furthermore, it has been noticed that both number of flights with holdings and values of additional ASMA time do not increase infinitely. This phenomenon is considered to be probably related to the aerodrome capacity limit: an aerodrome can accept only finite number of aircraft per time interval (due to limited throughput of its runways). Nevertheless it is considered to perform more experiments in the future to better explore this phenomenon.

The explored through RT system dynamics raises the question of adding more metrics to the model, potentially representing en-route events for different flights and also other KPIs, not listed in Table 2, but available in the

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

databases of Madrid-Barajas airport. After adding the new metrics to the model, Bayesian inference and new series of simulation with CPN framework should be performed with various realistic initial markings.

5. CONCLUSIONS AND FURTHER RESEARCH

This paper describes an approach to explore relationships between ASMA transit time deviations, number of flights with holding pattern and weather indicators with the use of Bayesian Network. Mathematical expressions of the discovered relationship have been used in order to build a model, capable to show possible states of the system for different scenarios of ASMA transit time changes. These scenarios are considered to be used by airport decision makers in order to design other scenarios and be prepared for any deviation that could occur in the terminal maneuvering area and its surroundings in the future and be able to explore the possible causes of any deviations of ASMA transit times occurred in the past.

It is considered also that the model could be extended and more airport performance metrics could be added to it in order to perform more wide and complex analysis, considering bigger area of airport operational activities. The noise, representing stochasticity of weather conditions for aircraft on en-route phase, preceding arrival to the studied airport, could be also added and its influence could be observed during the further research. However the computational restrictions of the used software have to be taken into account, as if the model becomes more complex, it would take more time and computational resources in order to explore all possible state spaces and perform the analysis.

ACKNOWLEDGMENTS

This research is partially supported by the Ministry of Economy and Competitiveness. Project "Fire Guided Unmanned Aircrafts and Resources Distribution" (TIN2014-56919-C3-1-R) and the project AIRPORTS (CDTI). Opinions expressed in this paper reflect the authors' views only.

REFERENCES

Andersson K., Carr F., Feron E., and Hall W. D., 2000. Analysis and Modeling of Ground Operations at Hub Airports. 3rd USA/Europe Air Traffic Management R&D Seminar, June 13-16, Napoli.

Brockwell P. J. and Davis R. A., 1991. Time Series: Theory and Methods (2nd ed.). Springer.

Chickering D., Geiger D. and Heckerman D., 1995. Learning Bayesian networks: Search methods and experimental results. Fifth International Workshop on Artificial Intelligence and Statistics, pp. 112-128. January 4-7, Ft. Lauderdale (Florida, USA).

Chickering D., Meek C., and Heckerman D., 2004. Large-Sample Learning of Bayesian Networks. The Journal of Machine Learning Research, 5 (1999): 1287–1330.

European Organisation for the Safety of Air Navigation (EUROCONTROL), 2005. Performance Review Commission, Performance Review Report 2004 (PRR 8).

European Organisation for the Safety of Air Navigation (EUROCONTROL), 2014. SESAR Guidance on KPIs and Data Collection Version 1, D85, Ed. 00.01.01.

European Organisation for the Safety of Air Navigation (EUROCONTROL), 2015. How to proactively manage weather's impact on performance? Available from: https://www.eurocontrol.int/ [accessed 24 May 2017].

European Organisation for the Safety of Air Navigation (EUROCONTROL), 2017a. Managaing complexity in a fast-changing world. Available from: https://www.eurocontrol.int/ [accessed 24 May 2017].

European Organisation for the Safety of Air Navigation (EUROCONTROL), 2017b. Airport capacity and performance (ACAP) activities. Available from: https://www.eurocontrol.int/ [accessed 24 May 2017].

Federal Aviation Administration, 2016. Terminal Area Forecast Summary. Fiscal Year 2016 - 2045. Available from: https://www.faa.gov/ [accessed 10 June 2017].

Friedman N., Linial M. and Nachman I., 2000. Using Bayesian Networks to Analyze Expression Data. Journal of Computational Biology, 7: 601–620.

Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A. and Rubin D.B., 2014. Bayesian data analysis. FL, USA:Taylor & Francis Group, 3-8.

Idris H. R., Anagnostakis I., Delcaire B., Hansman R. J., Clarke J.-P., Feron E., and Odoni A. R., 1999. Observations of Departure Processes At Logan Airport to Support the Development of Departure Planning Tools. Air Traffic Control Quarterly, 7: 229–257.

International Civil Aviation Organization (ICAO), 2017. Forecasting. Available from: https://www.icao.int/ [accessed 17 June 2017).

Jensen K. and Kristensen. L.M., 2009. Coloured Petri Nets: Modelling and Validation of Concurrent Systems. Springer Berlin Heidelberg.

Joshi K.D., 1989. Foundations of discrete mathematics. New Age International Limited, 7.

Klein A., Kavousssi S. and Lee R.S., 2009. Weather Forecast Accuracy: Study of Impact on Airport Capacity and Estimation of Avoidable Costs. Eighth USA/Europe Air Traffic Management Research and Development Seminar. Available from: http://www.atmseminar.org/ [accessed 10 June 2017].

Marsland S., 2015. Machine Learning: An Algorithmic Perspective, Second edition. CRC Press.

Office of Technology Assessment, 1984. Airport System Development. OTA-STI-231, Washington, D. C.: U.S. Congress. Available from: https://www.princeton.edu [accessed 24 June 2017].

Piera M.A. and Music̆ G., 2010. Coloured Petri net scheduling models: Timed state space exploration shortages. Mathematics and computers in simulation (82): 428-441.

Radhakrishnan N., Scutari M. and Lèbre S., 2013. Bayesian Networks in R with Applications in Systems Biology. Springer.

Roling P. C. and Visser, H. G., 2007. Optimal Airport Surface Traffic Planning Using Mixed-Integer Linear Programming. International Journal of Aerospace Engineering, 2008.

Silander T. and Myllymäki P., 2006. A Simple Approach for Finding the Globally Optimal Bayesian Network Structure. Networks, 445–452.

Song P. X.-K. 2007. Correlated Data Analysis: Modeling, Analytics, and Applications. Springer.

Tabernier L., 2015. SESAR Key Performance AREA, Performance targets and assessment. Available from: https://www.eurocontrol.int/ [accessed 15 June 2017].

Tsamardinos I., Brown L.E. and Aliferis C.F., 2006. The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm. Machine Learning, 65 (1): 31–78.

Walker C., 2016. CODA Digest 2016. All-Causes and Cancellations to Air Transport in Europe – 2016. The European Organisation for the Safety of Air Navigation - EUROCONTROL. Available from: http://www.eurocontrol.int/ [accessed 24 May 2017].

Wang Y., 2012. Prediction of weather impacted airport capacity using RUC-2 forecast. Digital Avionics Systems Conference (DASC), 2012 IEEE/AIAA 31st. October 14-18. Williamsburg (Virginia, USA).

Xu N., Laskey K.B., Chen C., Williams S.C., Sherry L., 2007. Bayesian Network Analysis of Flight Delays. Available from: https://pdfs.semanticscholar.org/ [accessed 10 June 2017].

BOOTSTRAPPING THE PAYSIM FINANCIAL SIMULATOR FOR OPEN SOURCE

Edgar Alonso Lopez-Rojas[(a)] and **Katrin Franke**[(b)]

[(a) ,(b)] The Norwegian University of Science and Technology (NTNU in Gjovik)
[(a)] edgar.lopez@ntnu.no [(b)] katrin.franke@ntnu.no

ABSTRACT
PaySim Simulator is an approach to the the lack of legitimate datasets on mobile money transactions to perform and experiment with fraud detection techniques. In this paper we used a technique called bootstrapping which uses the synthetic dataset generated by PaySim to build the parameter files that were initially extracted as aggregated information from the original data which is sensitive to share to the public. This way, researchers will have the possibility to generate diverse dataset with sufficient quality to perform fraud detection research on it. This paper explains the bootstrapping approach and in addition a method to properly reuse the PaySim simulator in the same or similar other domains. This is a solution to ultimately yield the possibility to simulate and generate financial transactions in such a way that they become similar to the original dataset without any direct connection to the original source.

Keywords: Bootstrapping, Multi-Agent Based Simulation, Financial data, Fraud Detection,Mobile Money , Synthetic Data.

Available at:
http://edgarlopez.net/simulation-tools/paysim/

1. INTRODUCTION

PaySim Simulator is an approach to the the lack of legitimate datasets on mobile money transactions to perform and experiment with fraud detection techniques. Obtaining access to data sets of mobile transactions for research is a very hard task due to the intrinsic private nature of such transactions (Lopez-Rojas and Axelsson, 2014). Scientists and researchers must today spend time and effort in obtaining clearance and access to relevant data sets before they can work on them. This is time consuming and distracts researchers from from focusing on the main problem, which is developing and improving their methods, performing experiments on the data, and finding novel ways to solve problems; such as the problem that inspired this paper, which is the fraud detection in financial data. Fraud inspectors on the other hand, are drowning in real fraud data. They are losing the opportunity that qualified people from the research community contribute to their task due to the impossibility to share private datasets.

The work shown in this paper aims to share with the research community our simulator. This is therefore the continuation of our work in this field and presents the development of a tool and a method to generate synthetic data that we previously named *PaySim* (Lopez-Rojas, 2016). PaySim generates synthetic datasets similar to real datasets from mobile money transactions. This is done by the means of computer simulation, in particular, agent based simulation. Agent based simulation is of great benefit in this particular context because the models created represent to some extent the human behaviour during transactions and are flexible enough to easily be adapted to new constraints.

In this paper we overcame the last barrier to share this simulator with the research community which is the sensitivity of the parameter file extracted from the original sample dataset. We used a technique called bootstrapping which uses the synthetic dataset generated by PaySim to build the parameter files that were initially extracted as aggregated information from the original data which is sensitive to share to the public.

PaySim simulates mobile money transactions based on a sample of real transactions extracted from the logs of a mobile money service implemented in an African country. The logs were provided by the multinational company Ericsson (ericsson.com), who is the provider of the mobile financial service which is currently running in more than 14 countries all around the world.

With the help of a statistic analysis and a social network analysis PaySim is able to generate a realistic synthetic dataset similar to the original dataset. PaySim models not only the customers behaviour but the fraudulent behaviour using malicious agents that follow known criminal patterns. By doing this, the resulting dataset is a rich source of data for researchers to perform different sort of test and evaluate not only the performance of fraud detection algorithms, but to measure the cost of fraud, which is otherwise an estimation on the real dataset.

The scope of this paper covers a background of the simulators that lead to PaySim as well as the technique called bootstrap of the parameter files. We also present a method that consist on 7 steps to easy the adoption and use of the PaySim simulator.

2. BACKGROUND AND RELATED WORK

The use of Mobile Money Transfers have grown substantially in the last few years and have attracted greater attention from users, specifically in areas in which banking solutions may not be as procurable as in developed countries. Many providers of mobile money services have been working in several and similar solutions over the past years. There are existing mobile money services in more than 10 African countries which coverage of 14% of all mobile subscribers (Rieke et al., 2013).

The ever growing usage of mobile money has increased the chances and likelihood of criminals to perform fraudulent activities in an attempt to circumvent the security measures of mobile money transfers services for personal financial gain. There is therefore a great amount of pressure on researching the potential security pitfalls that can be exploited with the ultimate goal to develop counter-solutions for the attacks.

Due to the large amount of transactions and the ever changing characteristics on fraud. The most of the measures against fraud start when the customer issue a complain. Many current system still base their detection mechanism on simple thresholds assigned arbitrarily. Therefore there is a need to push forward and investigate the effect of fraud and stop the wrongdoers from profiting from their fraud.

With *PaySim*, we aim to address this problem by providing a simulation tool and a method to generate synthetic datasets of mobile transactions. The benefits of using a simulator to address fraud detection was first presented during our previews work in (Lopez-Rojas and Axelsson, 2012b; Lopez-Rojas et al., 2016). This research states the problem of obtaining access to financial datasets and propose using synthetic datasets based on simulations. The method proposed is based on the concept of MABS (Multi Agent Based Simulation). MABS has the benefits that allows the agents to incorporate similar financial behaviour to the one present in domains such as bank transactions and mobile payments.

Our first implementation of a simulator for financial transaction was introduced in 2012 with a mobile money transactions simulator (Lopez-Rojas and Axelsson, 2012a). This simulator was implemented due to the difficulties to implement a proper fraud detection control on a mobile money system that was under development and that did not produce at the moment any real data sets to use for this research. This was the first paper to present an alternative to the lack of real data problem. The synthetic dataset generated by the simulator was used to test the performance of different machine learning algorithms in finding patterns of money laundering. In this paper we continue with this work. After we obtained access to a real data set of transactions we built a better model and calibrated the model to evaluate the results against the original data set.

The work by (Gaber et al., 2013) introduced another similar technique to generate synthetic logs for fraud detection. The main difference here was that this time there was available real data to calibrate the results and compare the quality of the result of the simulator. The purpose of this study was to generate testing data that researchers can use to evaluate different approaches. This works differers significantly from our work because we present a different method for analysing the data place special attention on evaluating the quality of the resultant synthetic data set.

The work on fraud detection in mobile payments by (Rieke et al., 2013; Zhdanova et al., 2014) is done in a similar domain as the work by (Lopez-Rojas and Axelsson, 2012b; Gaber et al., 2013).

(Rieke et al., 2013) uses a tool named Predictive Security Analyzer (PSA) with the purpose of identifying cases of fraud in a stream of events from a mobile money transfer service (Rieke et al., 2013). PSA is based on a dataset of 4.5 million logs from a mobile money service over a period of 9 months. They use simulation due to the limitation and knowledge of existing fraud in the current logs. The main focus on PSA is to detect money laundering cases that are cause by the interaction of several users of the system in an attempt to disguise the fraud among the normal behaviour of the clients. As a result the paper shows that PSA is able to efficiently detect suspicious cases of money launder with the aim of automatically block the fraudulent transactions.

(Zhdanova et al., 2014) is a continuation of the work done by (Rieke et al., 2013) and uses the simulator developed by (Gaber et al., 2013) to evaluate the results. Semi-supervised and unsupervised detection methods are applied to a mobile money dataset due to the advantage over supervised methods in this type of data where there is a difficulty in having a training data with known cases of fraud.

There is a previews work done about simulations in the domain of financial transactions for retail stores with the purpose of fraud detection (Lopez-Rojas et al., 2013). The work done in that paper is very similar to the work done in this paper. A large collection of data was gathered from one of the Sweden's biggest shoe-retailer. This data was used to produce a simulator called RetSim. RetSim was later used to model fraudulent behaviour from the staff and develop fraud detection techniques. There has been subsequent work on RetSim that produced among other results social network analysis (SNA) which described the relationship between the clients and the staff for each store, measuring the cost of fraud with the purpose of minimize the risk and properly estimate a security budget (Lopez-Rojas, 2015), threshold detection and methods to optimize the setup of thresholds (Lopez-Rojas and Axelsson, 2015a) and finally using this thresholds to properly setup a triage model that prioritize fraud suspiciousness (Lopez-Rojas and Axelsson, 2015b).

Public databases of financial transactions are almost non existent. However our previous work during the implementation of a simulator called BankSim presents a MABS of financial payments (Lopez-Rojas and Axelsson, 2014). BankSim is implemented in a similar way as the RetSim simulator and our simulator using in addition to statistical analysis a social network analysis. BankSim is based on the aggregated financial information of payments during 6 months of the two main cities of Spain that was provided by a bank in Spain with the purpose of developing applications of different kinds that benefit from this sort of data. Our work differs from this work because the source of the data and the characteristics of bank payments and mobile transactions are different as presented later in the following sections.

The key common aspect on previous work is the use of the paradigm of "Multi Agent Based Simulation" approach which incorporates into the behaviour of the agents the main customer logic to reach similar results as the real world. It is important to recognize that a simulation is not an actual "replication" of the original data set. Rather, a simulation will with the aid of statistical methods generate a very similar data set of the original data set. The degree in variance will largely be dependent on how the data on the original data set is structured, hence, different simulations based on different seeds will generate different output data sets but consistent with the real world.

3. PAYSIM BOOTSTRAPPING

We have finally passed the last barrier to share this simulator with the research community which is the sensitivity of the parameter file extracted from the original sample dataset. Aggregated information was originally calculated based on the measure of the STEP of the simulation. As we can see in fig. 1, this information was entirely based on the original dataset, therefore our NDA (Non Disclose Agreement) prevent us from publishing these files. For this reason, it was not possible for us to share the simulator. Without the parameter files the simulator is an empty shell that does not provide any insight from a mobile money domain.

We have previously shared generated synthetic datasets of PaySim in public website for the research community such as Kaggle. But these dataset represent just few instances of all possible scenarios that can be generated with PaySim. Researchers need to be able to answer several different questions regarding several different "What If" possibilities in a future scenario.

We used a technique called bootstrapping which uses the synthetic dataset generated by PaySim to build the parameter files that were initially extracted as aggregated information from the original data which is sensitive to share to the public.

The process is rather simple. We used the original dataset to generate the needed parameters. Once the parameters are available to be used by the PaySim simulator, we generate a dataset that aims to reproduce a scenario similar to the original dataset. This synthetic dataset is used again to calculate the aggregated parameters that were initially used by the PaySim simulator. Once we have obtained these parameters from the synthetic files we are breaking the dependency of the simulator from the original data and therefore we are not violating any NDA.

Next time we run the PaySim instead of using the parameter files previously obtained from the real data, we used the calculated parameter files obtained from the generated synthetic dataset.

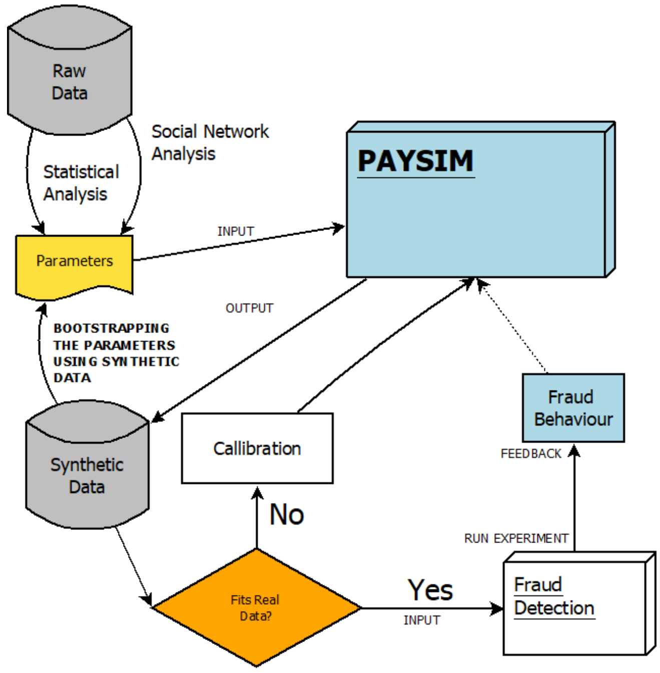


Figure 1: Bootstrapping the PaySim parameters from the Synthetic Dataset

4. MODEL

In this section we describe the parts that are required to understand the build and the use of the PaySim simulator, such as the inputs, the initialization, the execution, and the outputs.

4.1. Inputs

There are multiple inputs required in order for the simulator to function smoothly. As initial input, the number of client neighbours for each agent is assigned. The profile for each agent is then assigned based on probability. Their location in space in relation to their neighbours is also randomly initialized. Some of the inputs used by PaySim are listed here:

Parameter File: This is the file that contains all of the needed parameters that the simulator needs to initiate. Among these parameters we find the seed, and perhaps the most relevant of which, is the paths for where the input files and the output files are placed on the current machine.

Aggregated Transaction File: This file contains the distribution of the transactions from the original data set. More precisely, it contains the number of transactions were made at any given day/hour combination (step), what the average price was, what type of transaction it was etc. This is of paramount importance for the accurate results of the simulator since the synthetic data is generated from the information gathered from this file.

Repetitions File: This file contains the frequency of transactions that the original clients had per type of transaction. This means that some of the agents are schedule more than others based on a social network analysis of the indegree and outdegree of the customers.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Fraud Parameters: These parameters specify the number of fraudulent agents as well as the different probabilities to perform fraud and max/min amount of money for attempting to perform a fraud.

Since the simulator is using MASON as the framework for performing the simulation, it is important to define how each step is going to map real world time. For this simulation we defined that each day/hour combination represents one step. At each step, a Client that represents the agent for the simulator is generated. The client will be placed in an environment in which it is to make decisions based on the information it perceives. The Client is created with the statistical distribution of the possibilities to perform each transaction type for a specific day/hour combination. The client then randomly perform (based on the distribution initiated) different transaction types with the other clients on the simulator. Also, for each client, there is a probability **P** for the client to make future transactions at later steps. This probability is gathered from the database of the original data set.

4.2. Initiation Stage

In this stage, the PaySim simulator must load the necessary input data described in section 4.1.. The first and most important step is to load the values for each parameter in the parameter file. These will among other things contain the file paths for the source data inputs that the simulator needs to load.

Apart from the statistical distribution for each transaction type input to the client, there is another important input, which is the initial balance of the clients. Upon the generation of each client in the simulation, there must be an initial balance attached to that client. Besides the clients, the merchants and the fraudsters are also initialized based on the parameters.

4.3. Execution Stage

After the Initiation Stage, when all the parameters are successfully loaded, the simulator can now proceed to the execution stage. It is at this stage that the simulator will perform the actual simulation that lead to the simulated transaction results.

The agents are the founding blocks of the "Agent Based Simulator". The agent in this context, resembles the clients, the merchants and the fraudsters. Upon each step of the simulation, the PaySim simulator will convert each step to a "Day/hour" combination. This will then be used as an input to extract the statistical distributions from the original data set. Based on the *Aggregated Transaction File*, PaySim harness the probability **P** of performing each each transaction in the simulator and save it into the model of the client. With this information, the client now has gained more knowledge and will know the following important things:

Number Of Transactions: This is the total number of transactions that this generated client will perform.

Make Future Steps: This is the information of whether the client is to participate in future steps. Which means scheduling the tasks of performing more transactions during further steps.

Statistical Distribution: This is the different probabilities that the client will have loaded into it which entails the probability **P** of performing each action.

Initial Balance: This will be the initial balance that the client will have once generated.

After each client is generated, the client will make the decision of what type of transaction it will ultimately make, again this is completely derived from the distribution loaded. The client is in an environment which allows it to freely interact with other clients in the simulation. There are some types of transaction types that are based on that, like "TRANSFER" for instance. The "TRANSFER" type is exchange of money from one client to another; hence, the client will have to interact with other clients to simulate the actual exchange of funds.

The merchants play a passive role during the simulation and the only functions they have is to serve the clients during cash in and cash out transactions and the fraudsters during the cash out operations to fraudulent profit from their victims.

A fraudster will sense nearby clients and perform attempts to take control of their accounts. Upon succeeding, a fraudster will start to empty their accounts either by using a merchant to directly cash out or transferring money to mule accounts which in a short period of time will be also emptied through merchants and the cash out operations.

4.4. Finalization Stage

After each of the agents have completed their role in the simulation and performed all of the actions the results must be saved. There are 4 outputs generated after each simulation. All of which serve a specific purpose which allow a researcher to further test the quality of the generated data and save the configuration of the simulation with the in order to be able to repeat the simulation with the exact initial properties and results.

Logfile: Each transaction that is made will contain a record with the meta-data for that transaction. Data such as what client performed which action, to which other client, the sum of the transaction, and the difference in balance for all clients involved. Each such record will be saved in a logfile unique for the specific simulation.

Database: Apart from the logfile, the record for each transaction will also be saved into a database. The purpose of which is to allow for easier queries when the analysis of the resulsts is to be made.

Aggregated Dump An aggregated dump that is similar to the original aggregated dump from the original data set will also be generated. It is these two files that will be used to generate the plots and graphs resembling the results of the transactions.

Parameter File History This file will contain the exact properties needed for the simulation to be able to reproduce the exact same results again. This is important because each simulator must be able to be reproduced again, and without the original ”seed” used, it will not be possible.

5. METHODOLOGY FOR SIMULATING FINANCIAL TRANSACTIONS

It is important to follow our methodology in order to obtain the best benefit from the PaySim simulator. This section aims to explain how researchers can reuse our simulator to work in the same domain (mobile money) or similar domains that contain financial transactions. As a summary our method follows these steps in order to simulate financial transactions and perform experiments:

1. Obtain a sample of real data.
2. Perform data analysis to extract aggregated information for input into the simulator.
3. Add parametrization for expected fraud scenarios.
4. Run the simulator several times, using different random number seeds and/or different fraud configurations.
5. Apply the fraud detection methods on the generated synthetic dataset.
6. Summarize the results and performance from the experiments.
7. Repeat from step 3 on, for various fraud scenarios.

Obtaining a sample of real data is probably the hardest part of the whole method. Real data is important, because without it the simulation results can not be trusted. Fraud detection results are highly dependent on the dataset. For better results, a sample dataset that represents the financial service is required. This means that it covers enough (interesting) periods of time to learn more during the data analysis. If there is fraud, it should be properly labelled and identified with respect to which class of fraud it belongs in.

The real data sample can be obtained in several different ways: Full dump of a database (100% access), all of the data over a period of time, partial attributes of the data for some period of time, all data anonymised with respect to customer information, anonymisation by adding noise corruption (lowers the data quality) or simply aggregating information over a period of time.

The second step is to perform a data analysis to extract aggregated information for input to the simulator. Depending on the way the real data is provided, we need to perform several operations to convert the data into the format required. The simulator uses aggregations of information over a period of time, as input. The time granularity of the aggregation is specified on the simulation as a STEP. To accurately mimic the data distribution, we must extract aggregated information from the original data that matches each step in the simulation. There are also initial values and other input values extracted from the real data.

The information extracted is represented in terms of probabilities to ease the decision processes of the agents. Social Network Analysis (SNA) helps to recreate the topology of the customers' relations inside the simulation. The agent interacts with other agents within the environment and this interaction is specified by the information extracted during the SNA. The data analysis can also be done by employees of the financial institution that have access to the sample. Researchers only need the output of this step, to continue the process, this allows financial institutions to preserve the privacy of the customers.

In this paper we used the synthetic dataset generated by the PaySim simulator and applied the same data analysis that we performed over the original data. This prevents the problem of sharing the parameter files since they are directly connected to the original data and therefore are in most of the cases protected by NDA.

The third step is to add parametrization about expected fraud scenarios. The simulators are usually built to serve a purpose. Our simulators contain agents that, under certain conditions, act contrary to the law. The synthetic dataset has the benefit that can be generated according to the researcher's needs to study how certain fraud might affect a specific scenario. It can be a representation of the original dataset (sample). That is why we extract the aggregated information from the sample. Part of the simulator validation is to show that, given certain parameters, we can reproduce similar datasets.

The fourth step is to run the simulator several times using different random number seeds and/or different fraud configurations. In order to perform research in this field we need to be able to test different configurations. The Financial Simulator can also be used to answer all the ”WHAT IF” questions that are common during research. Researchers can run the simulator several times, using controlled variation on the parameters, to create new scenarios with normal and fraudulent data. This is specifically useful for answering questions such as WHAT IF: There is no fraud, there is little fraud, there is a lot of fraud, double the number of customers, and so on.

The fifth step is to apply the fraud detection methods on the generated synthetic dataset. This is one of the most important steps in the method. By changing the parameters in the previews step, we can generate diverse scenarios. These scenarios produce datasets with data that are labelled as fraudulent or not fraudulent (as appropriate). Once a dataset is generated, different methods for fraud detection can be tested and evaluated using the fraud label. A method for fraud detection can also be tested and evaluated with different scenarios that use the same fraud label. Fraud prevention methods can be also be added to the simulator to test and evaluate against fraud scenarios with the same flagged fraud.

The fifth step is to summarize the results and performance from the experiments. The biggest advantage of using a simulator over a real dataset is that we know with certainty how much fraud is present and where it is located. In a real dataset, it is impossible to guaranty that there isn't any undetected hidden fraud. Since we control our malicious agents, we can flag all fraudulent behaviour, because we have prior knowledge about the level of fraud injected into the dataset. Measuring all the fraud present in a dataset is one of the biggest challenges when using real data, but not with synthetic data.

Finally the final step is to repeat from step three for different fraud scenarios. This methodology is an iterative process that does not end at the seventh step. The idea is that After analysing the results on the previous step there should be new question that aim to be answer. New questions may lead to new scenarios of fraud. Fraud detection methods can also be modified to improve the results. This way the PaySim simulator can be used in a loop to improve results and perform research in fraud detection.

Re-starting at step three is more effective for research because they can tweak the parameters to generate new scenarios. However, some researchers might chose to work on a previously generated dataset (starting at step 5) only to test different detection methods and compare results against previous research.

6. CONCLUSIONS

Fraud detection in financial transactions is affected by the availability of datasets for testing methods. PaySim is a simulation of mobile money transactions with the benefit that financial transactions can be generated at will following distributions obtained from real data.

Our approach presents an solution to the problem of sharing sensitive information in the parameters. We used a method called bootstrapping in order to generate the parameter files required from the original data using a synthetic dataset that resembles the original dataset. By doing this we avoid breaking any NDA and we are able to share the simulator and the parameter files with the research community.

In addition to this we present a method to reuse the PaySim simulator in similar or other domains that cover financial transactions.

We have available at: http://edgarlopez.net/simulation-tools/paysim/ the source code of the PaySim simulator together with a bootstrapped parameter files. We aim to give support to researchers interested in using our method or our simulator.

References

Chrystel Gaber, Baptiste Hemery, Mohammed Achemlal, Marc Pasquet, and Pascal Urien. Synthetic logs generator for fraud detection in mobile transfer services. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 174–179. IEEE, may 2013. ISBN 978-1-4673-6404-1. doi: 10.1109/CTS.2013.6567225.

Edgar 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, 2012a.

Edgar Lopez-Rojas and Stefan Axelsson. Using the RetSim simulator for fraud detection research. *International Journal of Simulation and Process Modelling*, 10(2):144, 2015a.

Edgar Alonso Lopez-Rojas. Extending the RetSim Simulator for Estimating the Cost of fraud in the Retail Store Domain. In *The 27th European Modeling and Simulation Symposium-EMSS, Bergeggi, Italy*, 2015.

Edgar Alonso Lopez-Rojas. PaySim: A financial mobile money simulator for fraud detection. In *The 28th European Modeling and Simulation Symposium-EMSS, Larnaca, Cyprus*, 2016.

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, 2012b. Linköping University Electronic Press.

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Social Simulation of Commercial and Financial Behaviour for Fraud Detection Research. In *Advances in Computational Social Science and Social Simulation*, Barcelona, 2014. ISBN 9789172952782.

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Using the RetSim Fraud Simulation Tool to set Thresholds for Triage of Retail Fraud. In *20th Nordic Conference on Secure IT Systems, NordSec 2015*, pages 156–171, Stockholm, 2015b. Springer.

Edgar Alonso Lopez-Rojas, Stefan Axelsson, and Dan Gorton. RetSim: A Shoe Store Agent-Based Simulation for Fraud Detection. *The 25th European Modeling and Simulation Symposium*, 2013. (Best Paper Award).

Edgar Alonso Lopez-Rojas, Ahmad Elmir, and Stefan Axelsson. A Review of Computer Simulation for Fraud Detection Research in Financial Datasets. In *Future Technologies Conference, San Francisco, USA*, 2016.

Roland Rieke, Maria Zhdanova, Jurgen Repp, Romain Giot, and Chrystel Gaber. Fraud Detection in Mobile Payments Utilizing Process Behavior Analysis. In *2013 International Conference on Availability, Reliability and Security*, pages 662–669. IEEE, sep 2013. ISBN 978-0-7695-5008-4. doi: 10.1109/ARES.2013.87.

Maria Zhdanova, Jurgen Repp, Roland Rieke, Chrystel Gaber, and Baptiste Hemery. No Smurfs: Revealing Fraud Chains in Mobile Money Transfers. In *2014 Ninth International Conference on Availability, Reliability and Security*, pages 11–20. IEEE, sep 2014. ISBN 9781479942237. doi: 10.1109/ARES.2014.10.

AUTHORS BIOGRAPHY

MSc. Edgar A. Lopez-Rojas

Edgar Lopez obtained his PhD in Computer Science at Blekinge Institute of Technology (BTH) in Sweden and his research areas are 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. He obtained a Masters degree in Computer Science from Linköping University in Sweden in 2011 and a licentiate degree in computer science (a degree halfway between a Master's degree and a PhD) in 2014 from BTH.

Prof. Katrin Franke

Katrin Franke is a professor in computer science within the information security environment at NTNU in Gjovik. In 2007 she joined the Norwegian Information Security Lab (NISlab) with the mission to establish research and education in digital and computational forensics. In this context she was instrumental in setting up the partnership with the Norwegian police organisations as part of the Center for Cyber and information Security (CCIS). Dr. Franke is now leading the NTNU Digital Forensics group. Prof. Franke has 20+ years experiences in basic and applied research for financial services & law enforcement agencies (LEAs) working closely with banks and LEAs in Europe, North America and Asia.

MODELING KEY PHOTONICS COMPONENTS BASED ON OFF-THE-SHELF MICROWAVE-ELECTRONICS COMPUTER TOOL

M.E. Belkin[(a)], V. Golovin[(b)], Y. Tyschuk[(b)] and A.S. Sigov[(a)]

[(a)] Moscow State Technological University (MIREA), Scientific and Technological Center "Integrated Microwave Photonics", Moscow, Russian Federation,
[(b)] Sevastopol State University (SevSU), Sevastopol, Russian Federation

belkin@mirea.ru

ABSTRACT
We review our known, updated, and newer models and simulation results using power microwave-electronics off-the-shelf computer tool NI AWRDE to pursue advanced performances corresponding to the last generation of key microwave photonics components. As a result, we proposed and validated experimentally a new approach to model a broad class of promising analog microwave radio-electronics systems based on microwave photonics technology.

Keywords: microwave photonics, nonlinear equivalent models, microwave-band optoelectronic circuitry elements, computer aided design

1. INTRODUCTION

The modern optical fiber is the information transfer medium for telecom fiber-optics system (TFOS) and has unique operation features, combining low losses and dispersion of the propagating signal, insensitivity to electromagnetic interference, long life cycle and economic feasibility. Thanks to these features, the TFOS's global backbone networks currently exceed 70% of all existing communication facilities. In addition, growing significantly annual demands to transferred information volumes and reduction of delivering time, force hardware developers to constantly capacity increase of this TFOS class in order to use optimally quartz-based optical fiber bandwidth up to its limit that is tens of terahertz (O'Mahony 2006). However, the above-mentioned advantages of TFOS are less obvious in modern local information and communication networks (ICN) of various functionalities. A typical examples are data center (DC) access networks, as well as rapidly developing multi-service subscriber access networks with relatively short (from several meters to 10-15 km) transmission links and much higher requirements to the cost of building and maintenance. In particular, it is impossible to satisfy by TFOS the above requirement to provide communications with mobile subscribers.

Solution of this problem led to the emergence of hybrid local ICNs of radio-over-fiber (RoF) architecture, which combine the fiber-optic and wireless systems

2. MODELS OF MICROWAVE PHOTONICS-BASED CIRCUITRY ELEMENTS

2.1. Semiconductor Laser

A semiconductor laser is one of the most important circuit elements for both fiber-optics and integrated MWP devices, performing the function of electro-optical conversion. Earlier, for simulation of optoelectronic microwave units, we proposed and described in detail (Belkin, Iakovlev 2015) a non-structural (in the form of an equivalent electric circuit) nonlinear model of the laser (Figure 1), suitable for developers of modern analogue fiber-optics systems with a sub-carrier band in the radio-frequency (RF) range, local telecommunication systems of RoF architecture, devices of microwave optoelectronics, as well as optical interconnects in the integrated circuits.

However, the proposed single-band (combined RF and optical bands) model will be correct only in the case of direct modulation of a laser by a microwave signal, but it will be ineffective for modeling MWP device with external modulation and is completely unsuitable in the case for modeling of MWP-based ARES operating on several optical carriers. To investigate such units, an improved dual-band (separated RF and optical bands) laser model with two input ports is proposed (Figure 2), one of which simulates an optical carrier and the other one - a modulating microwave signal. Such an approach is correct for a computer working at the symbolic level.

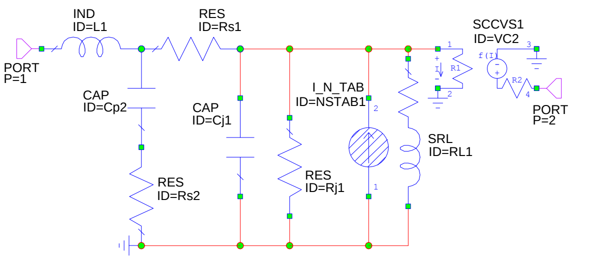


Figure 1: Equivalent electric circuit nonlinear model of the semiconductor laser

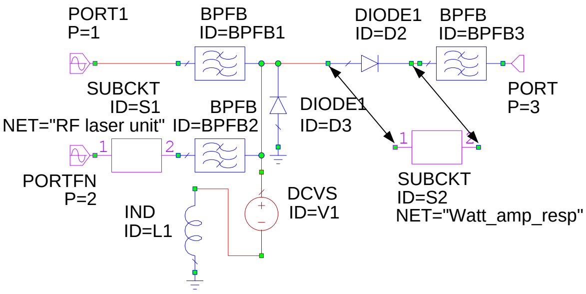
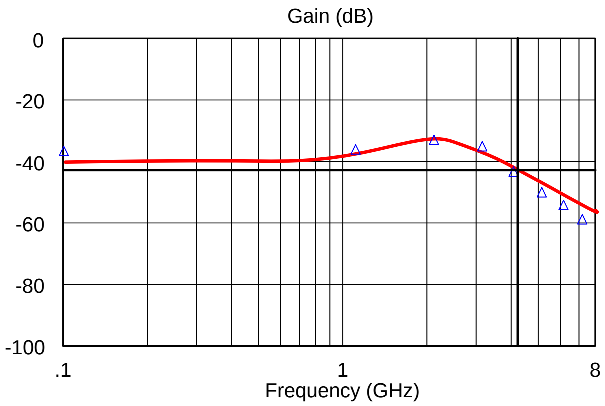


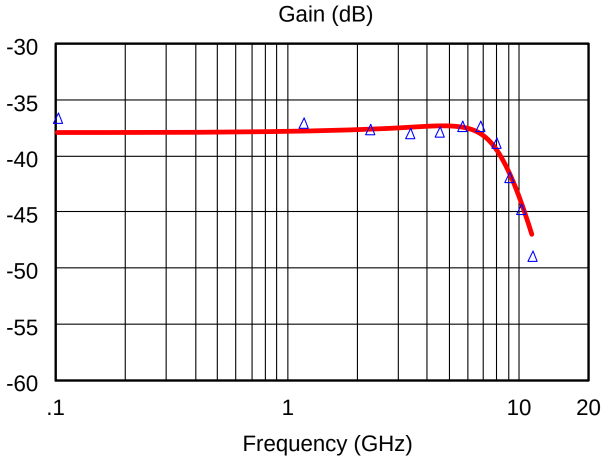
Figure 2: Improved dual-band laser model

The nonlinear models of a laser have been developed make it possible to calculate their key dynamic characteristics with simple and high accuracy, both in the small-signal modulation mode (gain characteristics, standing wave voltage ratio (VSWR), relative intensity noise (RIN)) and in large-signal mode (harmonic and intermodulation distortions, intercept point, dynamic range) that cannot be estimated using optoelectronic CAD tool. As an example, the results of the calculation and experimental verification of small-signal gain characteristics for two bias currents of vertical cavity surface emitting laser (VCSEL) corresponding to a double (a) and five-fold (b) exceeding of its threshold current value are shown in Figure 3. From the Figure one can draw the following conclusions:

- the gain factor is -37 dB, which corresponds to the VCSEL's typical parameters (Piprek and Bowers 2002);
- the -3-dB modulation bandwidth is 4.2 GHz at a bias current 4 mA and increases to almost 8 GHz at a current of 10 mA, which corresponds to the typical tendency (Piprek and Bowers 2002);
- the discrepancy between the calculated and experimental data in both cases does not exceed 2%, which indicates the validity of the proposed model.



a)



b)

Figure 3: Calculated (solid lines) and measured (triangles) gain characteristics of the laser under test

2.2. Photodiode

Another principal component, which should terminate each MWP circuit, is a semiconductor photodiode that performs the function of optical-electrical conversion. The original version of equivalent-circuit nonlinear model of a photodiode with a microwave bandwidth was proposed and described in detail in (Belkin 2012). Recently, an improved model of a photodiode (Figure 4) with updated parameters for a p-i-n photodiode with a passband of 40 GHz was developed (Belkin, Sigov 2015).

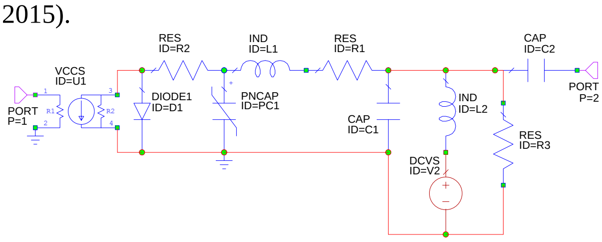


Figure 4: Improved model of a photodiode (Belkin, Sigov 2015)

The nonlinear model of a semiconductor photodiode has been developed allows to calculate its key dynamic characteristics with high accuracy, for example, the frequency characteristic, the linearity of the optical-electrical conversion in the large-signal mode, which cannot be estimated with the help of any available optoelectronic CAD tool. The latter characteristic is especially important for the purposes of MWP, so its study was focused on in our analysis. As is known, the linearity of the optical-electric conversion of a photodiode, which, like for a laser, is determined in terms of intermodulation distortions (IMD) and is affected by its operating mode: in particular, the depth of RF modulation of the optical carrier, the magnitude of the reverse DC bias (Urick, McKinney, and Williams 2015). For example, Figure 5 depicts the results for third-order IMD modeling at two frequencies at 10.0 and 10.1 GHz of a p-i-n photodiode with vertical illumination and a passband of about 40 GHz at a reverse bias of 1 V for three values of the modulation depth of the optical carrier of 10, 25 and 80%.

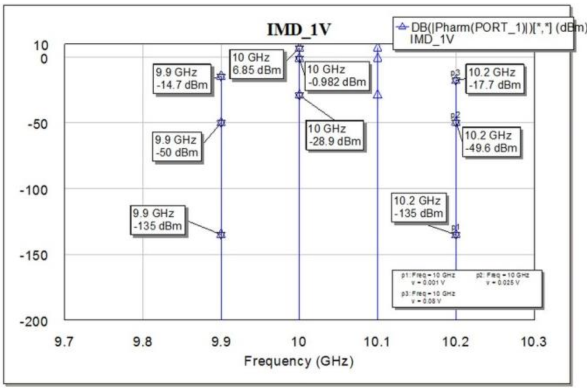


Figure 5: The calculated characteristics of the third-order intermodulation distortions of the photodiode under study at different values of the modulation depth

For convenience of an estimation the received numerical data are listed in Table I, which also includes the results of an experimental study.

Table I: Dependence of intermodulation distortions on the depth of optical modulation

Optical modulation depth (%)	Third-order IMD (dB)	
	Calculation	Experiment
10	< -100	<-80
25	-49	-48
80	-23	-21,5

The following conclusions can be drawn from the Figure and the Table:

- modulation depth up to 25% ensures acceptable linearity of optical-electrical conversion, which corresponds to the photodiode's typical characteristics (Urick, McKinney, and Williams 2015);
- if this value is exceeded to increase the signal-to-noise ratio, the level of intermodulation interference sharply increases, which is correct for any nonlinear circuit;
- the discrepancy between the calculated and experimental data does not exceed 1%, which indicates the validity of the proposed model.

2.3. Electro-Optical Modulator

Despite the simplicity, the direct intensity modulation of a semiconductor laser, which was widely used at the early stages of the TFOS development, has a number of drawbacks: a limited modulation band (up to 18 GHz), parasitic frequency modulation (chirp), necessity of rigid stabilization for the operating point, and so on. In this reason, a remarkable development acquires external intensity modulation, for which a modulator based on a Mach-Zehnder interferometer (MZM) is commonly used. Figure 6 depicts the newer developed MZM nonlinear model based on the equivalent circuit realized in the Schematic environment of AWRDE tool with a carrier at the optical frequency and a modulation signal at the radio frequency.

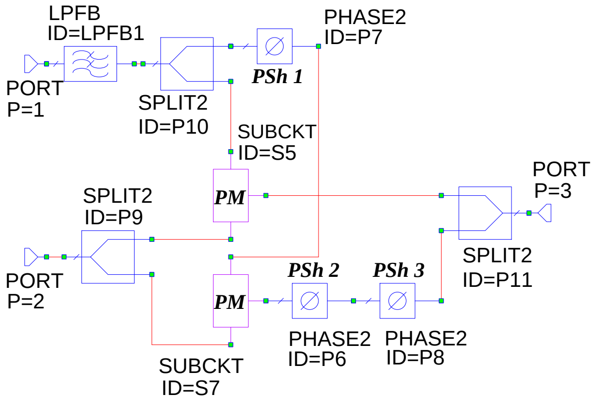


Figure 6: Equivalent circuit-based nonlinear model of Mach-Zehnder optical modulator

The design principle of the model fully reflects the building principle of the MZM (Urick, McKinney, and Williams 2015). Namely, a continuous signal from equivalent quasi-optical port 1 splits in two arms. Each of them includes electrically controlled by antiphase RF signals phase modulator (PM), which model is shown in Figure 7. The core element of both models is the phase-shifting cell (PSC). Figure 8 shows the equivalent circuit of PSC, in which the phase shift is provided by the nonlinear capacitance of the library model of the variable capacitor VRCTR.

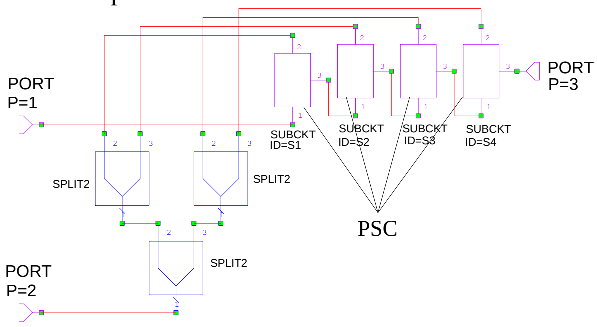


Figure 7: The model of phase modulator based on four sub-models of phase shifting cell

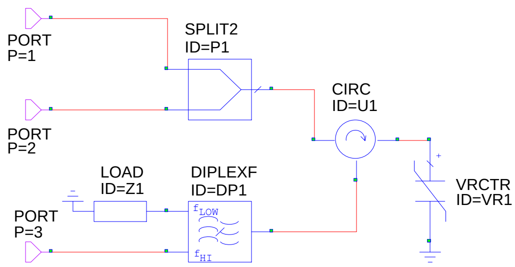
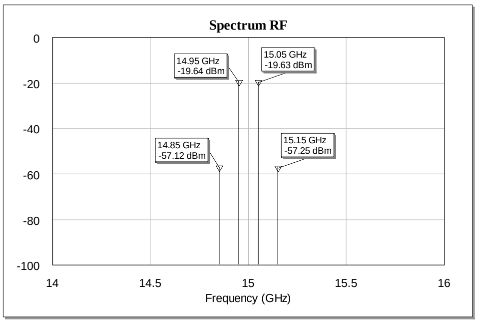


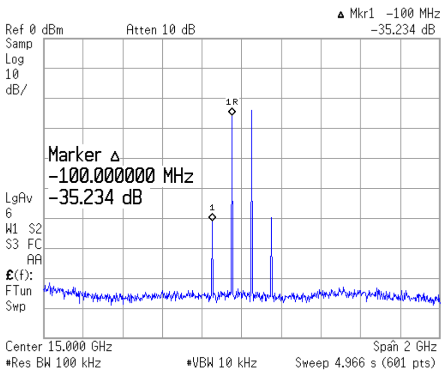
Figure 8: The model of phase shifting cell

The model of the MZM has been developed allows simple and high-precision calculation of its key static and dynamic characteristics. An important feature of the proposed model is its non-linearity that results in possibility to determine a number of large-signal parameters for a microwave MWP unit, such as 1-dB compression power P_{-1dB} or third order intercept point *IIP3*. The latter is determined from the transmission characteristics of the MZM at the fundamental modulating frequency and the frequency of third-order intermodulation distortions (IMD3) under two-tone impact. Note that these key quality parameters cannot be calculated using any optoelectronic CAD tool, but in AWRDE, these model experiments are performed by a one-button operation.

Modeling and real experiments were performed in the same modes when an unmodulated signal with an output of the order of 50 mW at a frequency of 193.3 THz was applied to the optical input of the modulator. Besides, a single sinusoidal RF signal with a frequency of 15 GHz during the P-1 dB test or two sinusoidal RF signals with frequencies of 14.95 and 15.05 GHz during the IIP3 test were applied to its modulating input. In both cases, the power range of the input RF signals corresponded to -15 ... 15 dBm for the calculation and -15 ... 10 dBm for the experiment. The constant bias of the MZM tested was established at the quadrature point of its transfer characteristic. Figure 9 exemplifies the results of calculation (a) and experiment (b) for third-order IMD levels at RF signal power on a modulation input of 10 dBm. The *IIP3* level calculated based on the obtained data was approximately 28 dBm, experimentally determined - 26 dBm. The calculated compression point of the P_{-1dB} is approximately equal to 14 dBm, determined experimentally - about 12 dBm.



a)



b)

Figure 9: Third-order intermodulation characteristics of the optical modulator under test

From the Figure and the subsequent calculations, we can draw the following conclusions:

- levels of *IIP3* and P_{-1dB} correspond to the MZM's typical parameters (Urick, McKinney, and Williams 2015);
- the discrepancy between the calculated and experimental data does not exceed 2 dB, which indicates the validity of the developed model even with modulation in the large signal mode.

3. CONCLUSION

In the paper, we described the results of development for behavior models in the form of physical equivalent circuits, as well as model experiments to calculate the key characteristics of the main microwave photonics components using well-known off-the-shelf microwave-electronics CAD tool AWRDE (National Instruments). The validity of the proposed models was confirmed by the results of experimental studies.

ACKNOWLEDGMENTS

This work is supported by Russian Fund of Basic Researches. The project number is obr-м No 14-29-08141

REFERENCES

Belkin M.E., Belkin L., Loparev A., Sigov A.S., Iakovlev V., 2015. Long Wavelength VCSELs and VCSEL-Based Processing of Microwave Signals. In: Pyshkin S., Ballato J., eds. Optoelectronics – Advanced Materials and Devices. Croatia: InTech Open, Chapter 6, pp. 231-250. Online:

Hemery E., Chusseau L., and Lourtioz J.-M., 1990. Dynamic Behaviors of Semiconductor Lasers under Strong Sinusoidal Current Modulation: Modeling and Experiments at 1.3 µm. IEEE Journal of Quantum Electronics. Vol. 26: 633-641.

Ng W., 2015. Photonics for microwave systems and ultra-wideband signal processing, Optics Communications. http://dx.doi.org/10.1016/j.optcom.2015.09.073i

Piprek J. and Bowers J. E., 2002. Analog Modulation of Semiconductor Lasers. In: Chang W., eds. RF Photonic Technology in Optical Fiber Links. Cambridge University Press. Chapter 3

Scotti F, Laghezza F, Ghelfi P., and Bogoni A, 2015. Multi-Band Software-Defined Coherent Radar Based on a Single Photonic Transceiver. IEEE Transactions on Microwave Theory and Techniques, Vol. 63, No. 2: pp. 546-552.

Urick V.J., McKinney J.D., Williams K.J., 2015. Fundamentals of Microwave Photonics. Hoboken, New Jersey, USA: John Wiley & Sons, Inc.

SIMULATION OF AN EVOLUTIONARY GAME FOR A WEALTH DISTRIBUTION MODEL STRUCTURED IN A SMALL WORLD NETWORK

Javier Lara de Paz(a), Gabriel Policroniades(b), Idalia Flores(c)
(a)EngineeringSchool
Universidad Nacional Autónoma de México
(a)javier_lara_depaz@comunidad.unam.mx
(b) Engineering School
Universidad Nacional Autónoma de México
gpolicroniadesch@gmail.com
(c) Engineering School
Universidad Nacional Autónoma de México
idalia@unam.mx

ABSTRACT
Wealth distribution studies have been reported for almost 200 years using different models to explain the dynamics involved. Also, many kinds of approaches have arisen to fit the registered data. Pareto´s distribution emerged as one of the best empirical model showing a good fitting with real data sets of wealth distribution all over the world and over different time ages. Theoretical models validate their assessments through this distribution. Souma asserted that wealth distribution interaction between agents could be modeled in a small-world network with different rules of wealth exchange. Garlaschelli, found that long-term shape of the empirical distribution strongly depends on the topology of the transaction networks among economic units. In the present work, an evolutionary game theory method was used to establish wealth exchange between economical agents structured in a small world network. The present project constructs a model based on México´s population income data sets dividing it in ten equal sized groups with different range of income stated as agents in the model. Agent based model simulations were performed using NetlLogo software, through different time intervals. Goodness of fit techniques was used to verify simulation results approaches to Pareto and log normal distribution.

Keywords: Complex Networks, Wealth Distribution,Econophysics, Evolutionary Game Theory,Small-World Networks, Pareto´s Law.

1. INTRODUCTION

1.1. Pareto´Law

Wealth distribution has been a controversial affair, that is of general interest. And it is about how the reality affects to everyone who can distinguish emerging economical inequalities. Is a fact which not only concerns to politicians or social science researchers but to many other different knowledge areas and deserves a certain systematic and methodical research. For many years, social sciences have shown more interest on this topic, but the methods employed tried to describe observed inequalities due to wealth distribution, showed inaccurate results unable to estimate different scenarios.

Wealth distribution formal studies (Piketty, 2014) date from about 18th century endings with the famous economist and philosopher Adam Smith, who in 1776 published his outstanding work *"An inquiry into the nature and causes of Wealth of Nations"* in which Smith established the first relations between wealth and means of production. There were also other researchers, who were interested in the behave of wealth among nations and societies. as for example Thomas Malthus, who in 1798 published his essay "Principle of Population" using the little information available, in which he postulated his famous exponential population grow law. Malthus research was based on Arthur's Young research published in 1792 about observations documented of France country life between 1787 -1788, before France Revolution. Another remarkable researcher of the XIX century was David Ricardo (1772-1823). In 1817 published his "Principles of Political Economy and Taxation", which replaced the famous Adam Smith´s "The Wealth of Nations" as a mandatory textbook in economic issues. The study of wealth distribution and economic disparities, were also of interest of Karl Marx and Friedrich Engels, but they focused mainly on how to structure the mentioned means of production in order to close wealth differences gap more than in a wealth distribution data study. At the end of the 19th century the Italian economist Vilfredo Pareto first pointed out a common feature among different wealth and income distributions studies, that emerge even if quantitative difference exists across the data obtained from various economies. Pareto´s investigation was based among household wealth and personal incomes modeling with statistical methods using a power-law distribution and even for the last hundred years data observations, this law seems to be a natural law, irrespective of many differences in culture, history, language, regime and the economic policies followed in different countries, the income distribution is found to follow a universal pattern in the upper tail a power law for the richest 1 - 5% of the population0 (Gaffeo et al., 2008). For the rest of the 95% of the population, wealth distribution fits to a conspicuous log normal decreasing distribution or an exponential fitting is used as well. The mentioned Pareto distribution describes by 'Pareto-tails', which decay as a power law for large wealth

$$\mathcal{P}_{>}(W) \sim (\frac{W_0}{W})^{\mu} \quad \ldots (1)$$

Where $\mathcal{P}_{>}(W)$ is the probability to find an agent with wealth greater than W, and μ is a certain exponent, of order 1 both for individual wealth or company sizes. Pareto estimated this parameter $\mu \approx 1.5$.

Today the Pareto law is usually quoted in terms of the probability density function, $\mathcal{P}(W)$,

$$\mathcal{P}(W) \sim W^{-(1+\mu)} \quad \ldots (2)$$

;for large W

Furthermore, Pareto observed that this distribution is universal, i.e., it "varies very little in space and time; different peoples and different eras yield very similar curves; there is a remarkable stability of form in this curve". This main achievement by Pareto could be seen from the perspective of today's economists and econophysists. (Chatterjee, 2009, 2008, 2005). Both approaches distant themselves from the most traditional one of mainstream economics, which does not consider statistical and heterogeneous aspects of the economy. Nowadays with the fast advances in data mining techniques these fields can analyze economic data with more accuracy, such as personal income, wage income, household wealth and so on, which have become increasingly available with the aids of fast computing and the Internet. More recently, statistical approaches have been used to provide insight on the origin of this law, even more, new methods from both economics and econophysics as well as sociology have emerged. By a much common ground with agent-based modeling and simulation, more recently, econophysists have tried to view this universal, two-class structure of income or wealth distribution as a nature law for a statistical many body-dynamical market system, by analogy with gases, liquids or solids in classical or quantum physical system. Both approximations have tried of all explore the exact shape of the real distribution from economic data, and secondly, to design theoretical models capable to reproduce such distributions.

As mentioned, works done in the field of statistical mechanics have bridged both fields, economic with physics, through the applications of Boltzmann-Gibbs function and the latest k-generalized statistical mechanic (Gaffeo, 2009).

Other paradigm applied is the Agent-Based Modeling (ABM) that uses computational modeling of systems as collections of autonomous interacting entities. This approach, differs from the classic mechanic reductionism considerations presented in economical mainstream models established in the latest 18th and 19th centuries, since ABM allows to represent the systems as a complex adaptive system (Tesfatsion, 2003) an idea introduced first by John von Neumann in the late 1940's. Remarkable ABM advantages are that one doesn´t need to restrict arbitrary assumptions, it is possible to ground the rules chosen for agent´s behaviors against reality, it is able to introduce additional complexity, simulations tools are simplifying any further analysis of the system as well. Using an ABM approach, besides the complex system consideration complex networks topologies are applied to structure agent's interactions. The present project focuses on the above-mentioned ABM approach, modeling the system as a set of economic agents exchanging wealth under an exchange algorithm, towards establishing an evolutionary game rule.

Many different works have been developed including the one´s done by Chakrabarti's group (Chatterjee et al., 2009). They consider in their model an element of proportionality|-saving propensity, thus allowing agents save a fraction of their money ¸m_i, whereas the rest of their money balance (1- λ) m_i is available for random exchanges. It was found the Gamma-like distribution for the case of a fixed saving propensity and a power-law tail in the case of a distributed saving propensity. Drăgulescu and Yakovenko (Drâgulescu, 2003) argued that wealth is equal to money m plus the other property that an agent i has. In the simplest model, it considers only one type of property, denoted as stocks s_i with a price p, and the wealth of an agent i is given by

$$w_i = m_i + p * s_i \quad \ldots (3)$$

It is assumed that the total amount of money and exchanged stocks remains constant. One of the most referenced wealth distribution theoretical modeling works corresponds to Bouchaud and Mézard0. They postulated a time evolution of the wealth w_i of an agent I described by a stochastic differential equation, capturing both exchange between individuals and random speculative trading.

$$\frac{dw_i}{dt} = \eta_i(t)w_i + \sum_{j(\neq i)} J_{ij}w_j - \sum_{j(\neq i)} J_{ji}w_i \quad \ldots (4)$$

The component $\eta_i(t)w_i$ is a Gaussian multiplicative process that simulates the investment dynamics, and the last two terms describe the trade interaction network between the agent i and all other agents in the society. J_{ij} is the exchange rate between agent i and agent j.

Besides the above two types of modeling wealth distribution, there were other wealth exchange models for wealth distribution based on complex networks (Iglesias et al., 2003).

The present work focuses mainly in wealth distribution simulation under an ABM approach and the following papers have been analyzed to establish a background based on economic agents interacting on a small world complex network, under an exchange algorithm as part of a development of an evolutionary game decision rules. Previously among others, Iglesias and his team (Iglesias et al., 2003) presented a simplified model for the exploitation of resources by interacting agents, in an economy with small-world properties having a Gaussian distribution of wealth. However they didn´t show any Pareto´s law distribution for the GNP, they found through network´s mean degree and exchange probability variation that the number of rich agents increases, but poor agents wealth decreases. Garlaschelli and Loffredo (Garlaschelli, 2009) studied Bouchard-Mézard wealth distribution model performing numerical simulations to test the model on more complex network topologies showing that a log-normal or a power-law distribution emerges depending on link's density variation, generally present in real systems. Souma, Fujiwara and Aoyama (Souma et al., 2004) 0 constructed a model of wealth distribution, based on an interactive multiplicative stochastic process on static complex networks. Through numerical simulations they showed that a decrease in the number of links discourages equality in wealth distribution, while the rewiring of links in small-world networks encourages it. Based on the previous results we considered that a small-world network topology can show real patterns on economic agent's interaction. Mao Bin Hu and colleagues (Hua et al., 2008) applied the prisoner's dilemma and snowdrift game cooperation model in a multi-agents system to establish wealth distribution in a population, obtaining good fitting in the simulation results with a log-normal adjustment.

1.2. Evolutionary Game Theory

Adami (Adami et al., 2016) mentions that in game theory, the objective is to find an appropriate strategy to resolve arising conflicts, or alternatively to find the optimal sequence of decisions that leads to the highest payoff. This optimal decision technique takes part through a set of individuals who develop a action, which is called "Movement"; this action, is codified within the genes of each individual granting the possibility of inheriting this characteristic. These actions allow the individual to make decisions in the face of certain events; which are stored in a set of equations; and where each of them preserves those genes that presented a better response to those events. In other words, only those genes most capable of solving a problem are preserved. As well, Nishino (Nishino et al., 2017) considered that one of the great advantages of mixing the agent-based simulation with the theory of evolutionary games is that the different agents obtain the ability to realize cooperation between them if they increase the capacity to obtain a higher performance; a question that otherwise would have been eliminated.

1.3. Data Sources

Present work is based in Mexico´s economic wealth distribution data, obtained from the Statistics and Information Bureau, which systematically present data gathered out from populations profile. Information used is about population average income organized in the so-called *deciles*, that means that population was divided into ten equal population sized groups, each one having the same number of people and ranging the corresponding average income, see table 1.1 below (del Castillo Negrete, 2016).

Table1.1 Mexico´s Population monthly average income distribution in ten equal sized groups (deciles). Simulation color codes for each group are represented in the extreme right column.

Population decile group from a total amount of 127 million	Monthly Average Income (in USD)	Percentage of the total national income	Agents color code
I	147.7	0,3	
II	391.9	0,9	
III	634.4	1,5	
IV	850.3	2,1	
V	1089.4	2,7	
VI	1353.6	3,5	
VII	1678.0	4,7	
VIII	2140.7	6,5	
IX	2956.2	10,8	
X	6930.3	67,0	
Total	18172.5	100	

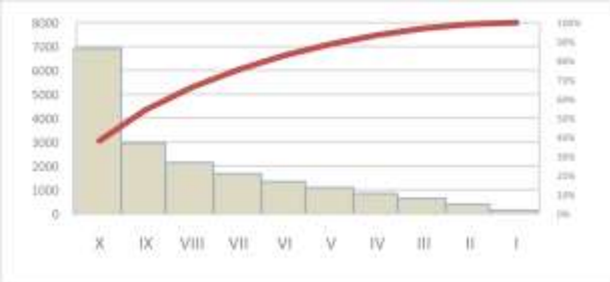


Figure 1.1. Distribution plot for table 1.1. with cumulative tendency line.

2. METHODOLOGY

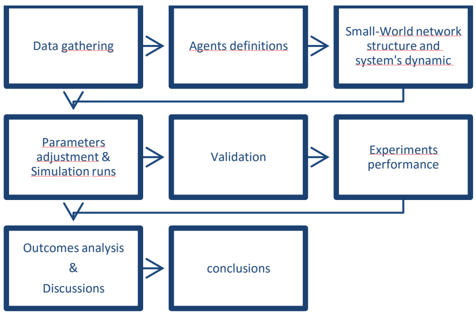


Figure 2.1. Methodology diagram used in the present project. Corresponding sections are described below.

The model is developed using multi-agents based on the income distribution deciles from table 1.1. Agents are set in a small-world network structured in a ring as Strogratz proposed (Watts & Strogratz, 1998). The random network was built upon a topological ring with N vertices and coordination number 2K. Each link connecting a vertex to a neighbor, considering clockwise sense, is then rewired at random, with probability $\boldsymbol{p}$**,** to any vertex of the system. It is assumed any self-connection and multiple connection as prohibited, obtaining a regular lattice at $\boldsymbol{p} = 0$ and progressively random graphs for $\boldsymbol{p}>0$. The system model was programmed in NetLogo6.0.1. a "rewiring link" button was setup for rewiring any link in every step, as (Watss and Strogratz, 1998). Rewiring a link corresponds to long social contacts made shorter by this rewired link, representing what Milgram established (Milgram, 1967), a Small-World. The complex network represents a connected economic system were nodes are agents exchanging an amount of money in every step (tick). Ten different agent sets have been constructed, corresponding to a monthly wealth income *"decil"*, and were spread out randomly over the network as nodes. For an easier visualization, every *decil* agent set was setup as circles with different colors and radius size proportional to its monthly income (figure 2.2). Nodes radius will increase or decrease whether the node gains or loses money due to exchange process. At start up certain amount of wealth based on official data (see table 1.1) is set up for each agent group, and an interaction rule is established following an exchange algorithm (see equation 6).

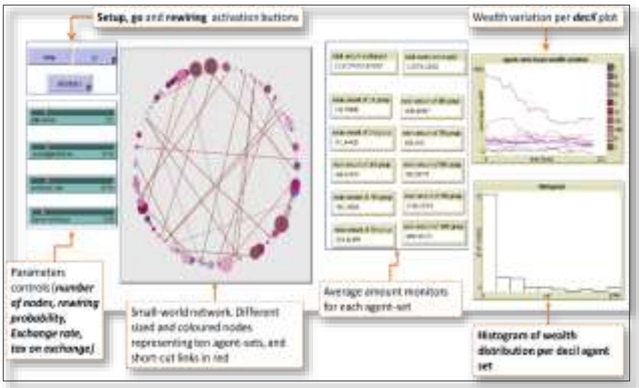


Figure 2.2. Model interface programed in NetLogo 6.1.0. Figure describes elements included.

Once system's topology was structured in a Small-World complex network the go process is constructed as follows, as a Watts & Strogratz Small-World network, system's network has a rewiring probability changing a random number of links wiring them with other different pair of nodes. There are two types of links, active and inactive links, through active links flows exchanged amount of money from a node to recipient, depending on interacting nodes wealth. Number of active or inactive links is randomly defined by "exchange-rate", which corresponds to the opportunity each node has for exchanging money with its corresponding neighbors. This parameter defines the dynamism as in a real economic system. Through every time step, set default as ticks in NetLogo software, a new connection between a pair of nodes is set, new economic interaction starts, based on algorithm equation 6, a node will gain and the other will lose money. Winning node will increase its size proportionally to money gained, as looser decreases its radius. During process running average amount of money each *decil* has, is monitored to determine systems behavior. The process continues with every tick until it is arbitrarily stopped. Data obtained from simulations are collected and analyzed. System's dynamic chart is show below (figure 2.3).

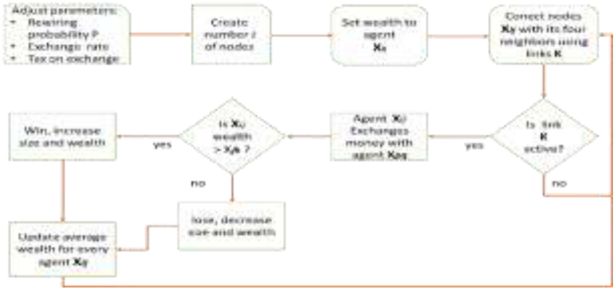


Figure 2.3. System's dynamic process

2.3. Parameters setting

Parameters included represent real economic and social systems behaviors. Rewiring probability, shows heterogeneous agents dynamic in social contact system. In real situation, a contact network is not a full random network nor a regular graph, but intermediate to both. In this system rewiring probability represents long path interactions between two agents theoretically far separated. As Milgram demonstrated, these long-distance relations are possible as we live in a Small-World. For simulation performance, this parameter is set to 30% due to social contacts models perform more realistic around this value (Lara & Flores de la Mota, 2014). "Exchange rate" parameter creates a random number of directed-active links less than or equal to adjusted value connecting agents through directed flow edges enabling an agent financial interaction. It is possible that an agent is connected to its four neighbors, this node will exchange with its connected neighbors winning or losing money with any of them therefore increasing or decreasing its wealth, changing visual agent size as well.

This parameter represents opportunities a person has to exchange money for commodities. If there is a bigger exchange rate, the economic system has a better dynamism. By this project an insight in rate exchange variation scope is being look for and its affectation to wealth distribution. "Tax-on-exchange" parameter establishes tax rate paid in every carried-out transaction. System is not adjusted to reload the initial quantity so that as time steps keep on running agent's money tends to cero owing to tax amount paid. The other parameter present in model is the "number of nodes". By this control, nodes are created considering ten different equal sized *decil* groups.

Figure 2.4. Parameters controls panel

2.4. Money exchange process

If an agent is linked with another by a directed edge exchange action is executed following a trade rule established by an exchange algorithm 0

$$m_i' = m_i - \Delta m$$
$$m_j' = m_j + \Delta m \quad \ldots (5)$$

where m_i is starting money each ***agent$_i$*** has at the beginning of period (see table 1.1). Being Δm the transaction amount of money or wealth.

$$\Delta m = profits - (profits * tax) \ldots (6)$$

Where profits correspond to exchanged money gained in a transaction between agents *i* and *j* .

2.5. Simulation

To perform simulations, once the network was set up with a certain number of agents and wealth exchange interaction algorithm, NetLogo software is used to program the model, with agents representing nodes and the small-world network topology, wealth exchange between agents flows through diffusion links. In each time step, tick, money or wealth is permitted to flow across specified links between agents, having more preferential flow agents with bigger wealth concentration due to connectivity, every link has a preferential attachment with the richer link. Network has a rewiring probability as well, and every step a rewiring is performed. After simulation took part each node could earn more money, or lose it, and the new outcoming wealth distribution scenario is then analyzed. First system is setup with fixed rate-exchange and tax-rates values, as well as rewiring probability, to fit with Pareto´s distribution (equation 2) for validation. Five hundred agents were created, with a rewiring-probability P = 0.30, and an exchange - rate = 0.4, tax-on-exchanged value is set to 0.16 value (Mexico's corresponding VAT). A sample of 150 runs were performed, and each simulation had about 30 ticks of duration, the wage earned and described in table 1.1 is the amount earned monthly and the system doesn´t consider a reload of money. Results and validation analysis are shown in next section.

Figure 2.5 A simulation framework

3. VALIDATION

Since the system is based on Mexico´s income distribution the starting money set in the model based on Table 1.1. outcomes are adjusted to this distribution parameters. For validation, a goodness of fit was performed using Chi-square distribution, using simulation outcomes of final mean value of remaining money each *decil* agent-set has at the end of 30 days (ticks), also maximum and minimum values are included (see table 3.1). Average is obtained to construct *decil* income distribution obtained from simulations. As main objective of present project is to represent an agent based model income distribution with an exchange rule, adjustment to theoretical Pareto Function is also performed with μ = 1.5. Results are shown in the analysis below.

Table 3.1. Data outcomes from 150 simulation runs with 500 nodes. Columns show mean, maximal and minimal value of money after 30 ticks (a month).Extreme right column represent average of the other columns.

Decil group	**mean value**	**max value**	**min value**	**average value (Max-Mean-min)**
I	**418.13**	**518.82**	**0.00402**	**312.32**
II	**502.36**	**757.35**	**0.00447**	**419.91**
III	**654.51**	**859.04**	**0.00361**	**504.52**
IV	**833.22**	**1099.61**	**0.00299**	**644.28**
V	**804.75**	**1025.13**	**0.00389**	**609.96**
VI	**963.83**	**1278.29**	**0.00323**	**747.37**
VII	**826.87**	**1015.21**	**0.00450**	**614.03**
VIII	**972.98**	**1622.86**	**0.00320**	**865.28**
IX	**1332.51**	**1685.83**	**0.00196**	**1006.11**
X	**4158.16**	**4789.32**	**0.00410**	**2982.49**

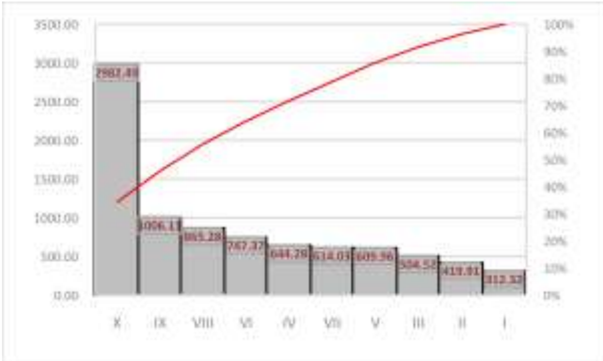


Figure 3.1. Decil-sets distribution plot with cumulative tendency line from table 3.1.

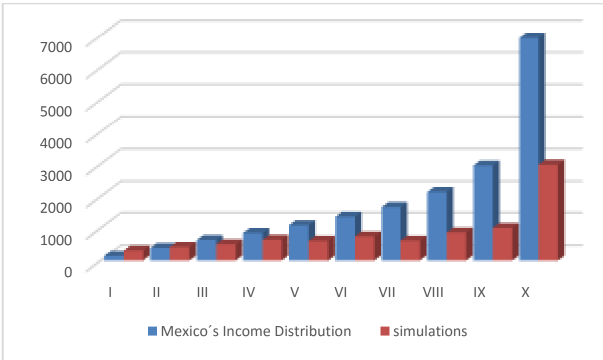


Figure 3.2. Bar-Plot for both data-set. It is shown a decreasing in simulation data outcomes. Decil-sets Distribution remains in simulation data's as shown in previous graphs.

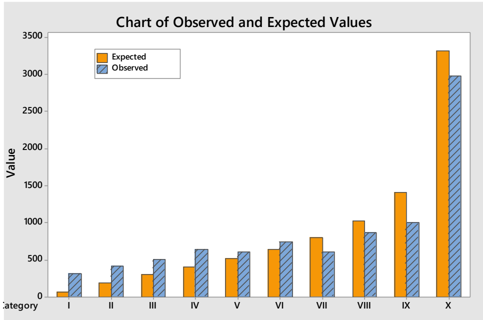


Figure 3.3. Plot for contrasting both distributions, historical Mexico's income distribution and observed simulation results, for a Chi-Squared Goodness of fit validation Analysis.

Table 3.2. Chi-Squared Goodness of fit Analysis, for 150 Simulation Outcomes vs Table 1.1. information. At the bottom of the table adjustments parameters are shown (Chi-Squared Value aand P-Value)

Decil set	Simulation data's	Table 1.1 information	Proportion	Expected	Contribution to Chi-Sq
I	312.32	147.7	0.00813	70.76	824.585
II	419.91	391.9	0.02157	187.76	287.045
III	504.52	634.4	0.03491	303.93	132.374
IV	644.28	850.3	0.04679	407.37	137.776
V	609.96	1089.4	0.05995	521.92	14.852
VI	747.37	1353.6	0.07449	648.5	15.076
VII	614.03	1678	0.09234	803.91	44.852
VIII	865.28	2140.7	0.1178	1025.59	25.058
IX	1006.11	2956.2	0.16267	1416.29	118.791
X	2982.49	6930.3	0.38136	3320.24	34.357

Degrees of Freedom	Chi-Sq	P-Value
9	*1634.77*	*0*

As seen on Analysis shown above, it exists a good fit between Mexicós Income Distribution information used to perform the model and our simulation results. Next a Pareto´s Distribution Adjustment is carried out to verify if our model corresponds to a Pareto Income Distribution. Analysis results are shown below.

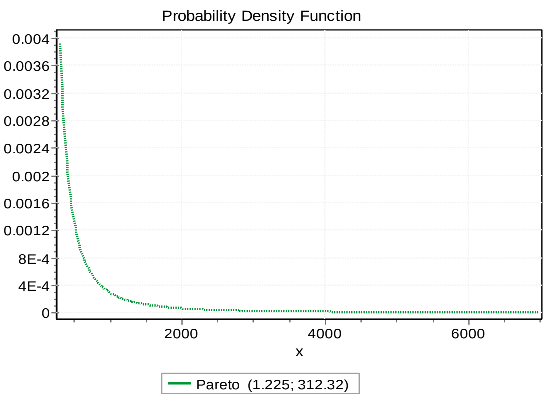


Figure 3.4. Pareto´s Density Function Plot for Simulation outcomes of Table 3.1.

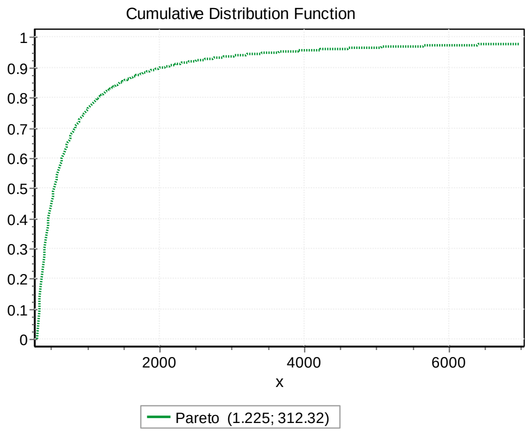


Figure 3.5. Pareto´s Cumulative Distribution Function Plot for Simulation outcomes of Table 3.1.

Tabla 3.3. Summary of Pareto' s Distribution fitting parameters for simulation outcomes

μ = 1.225	℘ = 321.32				
Kolmogorov-Smirnov					
Sample Size	10				
Statistic	0.25957				
P-Value	0.43717				
Rank	25				
α	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**
Critical Value	0.322	0.3687	0.40925	0.45662	0.48893
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	10				
Statistic	2.096				
Rank	35				
α	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	No	No	No

Observed analysis show a good fitting with Pareto Distribution Function (equations 1 and 2). With the shown validation analysis, the model can be considered for simulation process of Mexico´s income distributions analysis. Once the model is significantly accepted, it should be considered repercussions model parameters, (rewiring-probability, taxes and exchange-chance) would have in system´s dynamic. Experiments are designed to get insight on this question. In next section this experiments are described.

4. Experiments Performance

Gini coefficient is used to measure, in general, incomes and wealth distribution. Gini coefficient is a statistical measurement of dispersion and variability of income or wealth distribution of a nation. It is and indicator of inequality (Gini, 1909). Its values range from 0, which represents a perfect distribution, that means that all residents would have the same wealth or income. On the other side if its value is 100 % that means that just one person concentrates total wealth. Parameters modeled in this project represent social dynamics by rewiring probability and exchange rate, and the influence that taxes have in money flow between agents or population. By a design of experiment will get insight on effects of these parameters on income distribution. The DoE performed is a full factorial design with 3 factors and different levels for each parameter. See Table 4.1. for experiment's description.

Table 4.1.Full Factorial Design of Experiment Parameters

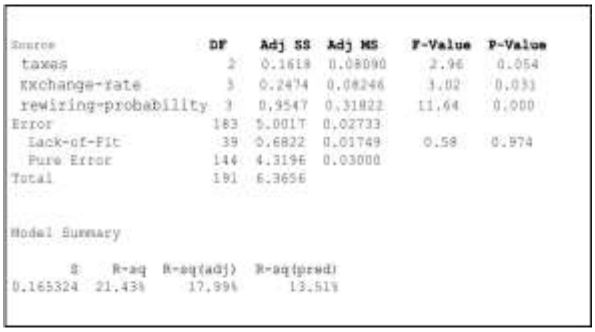
Factors: 3	Replicates: 4
Basic runs: 48	Total runs: 192
Basic Blocks:1	Total Blocks: 4
Rewiring-Probability Factor	4 levels: 20%, 40%, 60% and 80%
Rate-exchange Factor	4 levels: 20%, 40%, 60%, 80%
Tax-on-exchange factor	3 levels: 15%, 25%, 35%
Response value:	Gini coefficient

After running all replicas, results obtained are analyzed by an AnoVa tool. Results are then analyzed and discussed in section 5.

5. RESULTS ANALYSIS AND DISCUSSION

5.1 Results analysis

After runs were carried out, and AnoVa tool is applied to simulation outcomes as described below



Source	DF	Adj SS	Adj MS	F-Value	P-Value
taxes	2	0.1618	0.08090	2.96	0.054
exchange-rate	3	0.2474	0.08246	3.02	0.031
rewiring-probability	3	0.9547	0.31822	11.64	0.000
Error	183	5.0017	0.02733		
Lack-of-Fit	39	0.6822	0.01749	0.58	0.974
Pure Error	144	4.3196	0.03000		
Total	191	6.3656			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.165324	21.43%	17.99%	13.51%

Figure 5.1. ANOVA results from performed experiment

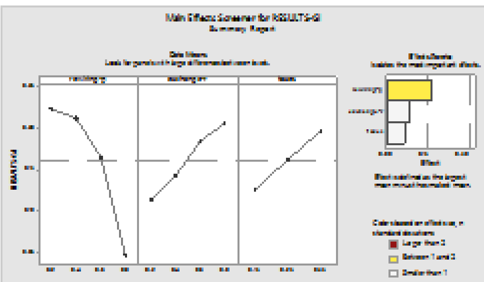


Figure 5.2. Main effect Plot for the three factors. At the upper right corner it is indicate which factor affects more than others

Figure 5.2. indicates the significantly effects a factor has over response. Rewiring-Probability (***P***) affects more than other two factors, but at higher levels of P results on negative Gini-coefficient ***G***, which are considered as outlier data. It is suggested that for higher levels of P and lower levels of exchange-rate (***R***) as well as for the lower levels of taxes ***(T)***.

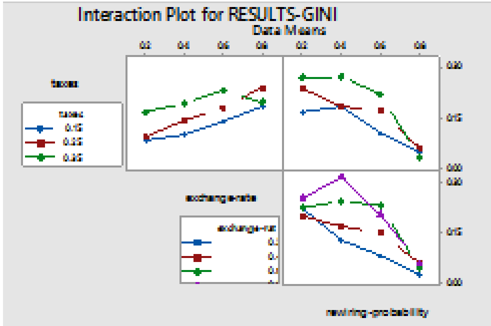


Figure 5.3. Interaction Plots for three factors. At top left interaction between exchange-rate and taxes are compared. Top right interaction between rewiring-probability and taxes are compared. Bottom left, interaction between rewiring probability and exchange rate is observed.

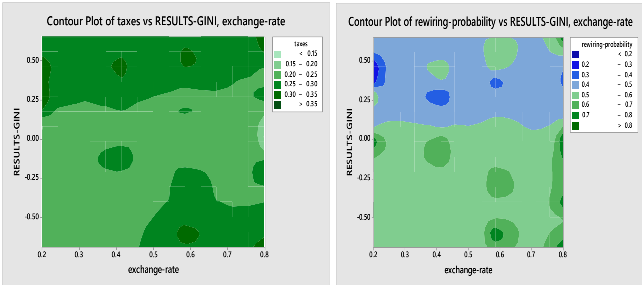


Figure 5.3. Contour Plots, left interaction between T, R and Gini coefficient G response is shown. Right interaction between R, P and G is shown

In figure 5.3 graphs of interaction between factors is explored. There appears to be a moderate interaction between P and R, but there seems to be more significant interaction between T and P factors. With contour plots effects between factors is more distinguishable, what is observable in both graphs is that are outlier data with negative values for G, and are produced mostly by higher values of P. An individual value Plot is used to identify this values.

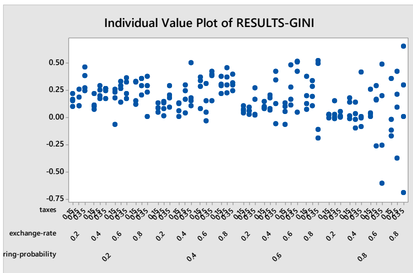


Figure 5.4. Individual Value Plot for Gini Coefficient. Negative data outcomes are observable with this plot.

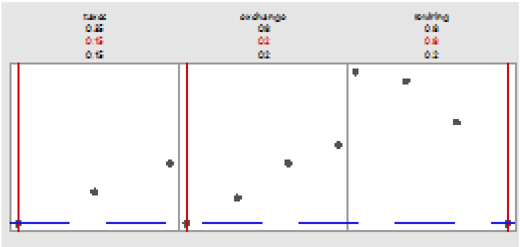


Figure 5.7. Optimal Plot. Red labels indicate optimal values to minimize Gini Coefficient. Blue Dot line indicates lower value of G.

As seen in graph 5.6 for individual values, there are many outlier data that can affect more significantly to general outcomes. Design of experiment applied should be reviewed. Figure 5.7 indicates the parameters value to obtain the lowest Gini coefficient value using this design.

5.2. Discussion

As P parameter represent social contacts in a real system, it is observed for high levels of P with lower level of R , G values tend to negative values, which are not possible, a review on this affection should be done with an one factor experiment design. Also, it is noted that interaction with taxes does not affect significantly to response since tax imposed affects equally to all agents. A remarkable overview is the optimal combination of factor levels that indicates that for the highest level of P and with the lowest levels of R and T optimal is obtained. This results indicates that with social interaction is important to diverse with whom to make trades, and exchange rate if it is low there are less opportunities for agents to lose its money. And more interesting analysis is that low taxes increase social equality.

CONCLUSIONS

This project is part of a doctoral research which main scope is to get insight on more significant factors that affect wealth and income distribution, and as a first insight a network model should be designed, with an evolutionary game interaction rule. But as there is limited space that part is left for a further research. Although with the present work it is possible to gain insight on some important factors and its interactions. And a more interesting result is taxes influence in income equality. Results on this project should be considered by politicians and stakeholders to design more adequate economic policies.

REFERENCES

Adami, C., Schossau, J. &Hintze, A., 2016. The reasonable effectiveness of agent-based simulations in evolutionary game theory: Reply to comments on “Evolutionary game theory using agent-based methods”. Physics of Life Reviews, Volume 19, pp. 38-42.

Anglin P., 2005, “Econophysics of wealth distributions: a comment”, in Eds. Chatterjee, Yarlagadda and Chakrabarti, pp.229-238 (2005).

Atkinson B. and T. Piketty, 2007, “Top Incomes over the Twentieth Century”, Eds.(2007) (Oxford University Press, Oxford).

Bouchaud J. P. and Mézard M., 2000, “Wealth condensation in a simple model of economy”, Physica A 282, (2000) 536-545.

Chatterjee A., 2009, “Kinetic models for wealth exchange on directed networks”, The European Physical Journal B 67, (2009) 593-598.

Chatterjee, B. K. Chakrabarti and R. B. Stinchcombe, 2005, “Master equation for a kinetic model of a trading market and its analytic solution”, Physical Review E 72, (2005) 026126.

Chatterjee, S. Yarlagadda and B. K. Chakrabarti, 2005, “Econophysics of Wealth Distributions”, Eds. (2005) (Springer, Milan).

del Castillo Negrete M., 2016, “La magnitud de la desigualdad en el ingreso y la riqueza en México, una propuesta de cálculo”, CEPAL, Eds. UNO. 215.

Di Matteo T., Aste T. and Hyde S. T., 2004, “Exchanges in complex networks: income and wealth distributions”, in The Physics of Complex Systems (New Advances and Perspectives) edited by F.Mallamace and H. E. Stanley (IOS Press, Amsterdam) (2004) p. 435

Drâgulescu A. and V. M. Yakovenko, 2003, “Statistical mechanics of money, income, and wealth: a short survey”, in Modeling of Complex Systems: Seventh Granada Lectures, edited by P.L. Garrido and J. Marro, American Institute of Physics (AIP) Conference Proceedings 661, (2003) pp. 180-183.

Gaffeo E., DelliGatti D., Desiderio S. and Gallegati M., 2008, “Adaptive microfoundations for emergent macroeconomics”, Eastern Economic Journal 34, (2008) 441-463.

Garlaschelli Diego and Loffred Maria I., 2004, “Wealth Dynamics on Complex Networks”, Physica, feb 2004.

Hua Mao-Bin, Jianga Rui, Wua Qing-Song, Wub Yong-Hong, 2008, “Simulating the wealth distribution with a Richest-Following strategy on scale-free network”, Physica A 387 (2008) 5862–5867.

Ichinomiya Takashi, 2012, “Wealth distribution on complex networks”, Physical Review E, 86, 066115(2012).

Iglesias J. R., Goncalves S., Pianegonda S., Vega J. L. and Abramson G., 2003 “Wealth redistribution in our small world”, Physica A 327, (2003) 12-17.

Nishino, N., Okazaki, M. & Akai, K., 2017. Effects of ability difference and strategy imitation on cooperation network formation: A study with game theoretic modeling and multi-agent simulation. Technological Forecasting and Social Change, pp. 1-12.

Piketty Thomas, “Capital in the Twenty-first Century”, Harvard University Press, 2014.

Souma Wataru y Fujiwara Yoshi, Aoyama Hideaki,2001, “Small-World Effects in Wealth

Distribution", Statistical Mechanics (cond-mat.stat-mech), KUCP0189, 2001.
Tesfatsion Leigh, 2003, "Agent-based computational economics: modeling economies as complex adap8tive systems", Information Sciences 149 (2003) 263–269.
Watts D. J. and Strogatz S. H., 1998"Collective dynamics of 'small-world' networks." Nature, 393 (6684): 440–442,1998.
Gini, C., 1909. "Concentration and dependency ratios" (in Italian). English translation in *Rivista di Politica Economica*, **87** (1997), 769–789.

Acknowledgments
To CONACYT scholarship for the first author, and DGAPA-PAPIIT project ITl 02117 Accesibilidad y movilidad def transport e publico urbano en la ciudad de Mexico, el caso de la delegación Tlalpan

AUTHORS BIOGRAPHY

Javier Lara de Paz studied Physics at the Faculty of Science of National Autonomous University of Mexico (UNAM), a Specialization in Applied Statistics at the UNAM. He studied his Master degree in Operation Research at the Faculty of Engineering of the UNAM. He has been professor of Science at the and Operation Research professor at the Engineering School at the UNAM. Currently he is studying a Ph.D. in Operational Research at UNAM.

Idalia Flores de la Mota 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.

Gabriel Policroniades Chipuli He studied Industrial Engineering at Technological Institute and Superior Studies of Monterrey , Campus Toluca (ITESM, Campus Toluca), he studied an engineering mastery with operation research orientation and ha advanced metrology studies for the calculation of uncertainty, and knowledge of Lean Six Sigma Black Belt and Logistics and supply chain Currently he is studying a Ph.D. in Operation Research at UNAM.

CALCULATION OF THE CAPACITY OF SWITCH AREA WITHIN RAILWAY STATIONS WITH THE USE OF SIMULATION METHODS

Michael Bažant[(a)], Josef Bulíček[(b)], Pavel Krýže[(c)], Petr Veselý[(d)]

[(a),(d)]Faculty of Electrical Engineering and Informatics, University of Pardubice
[(b)]Faculty of Transport Engineering, University of Pardubice
[(c)] SŽDC, s.o., Timetable & Path Capacity Department

[(a)]Michael.Bazant@upce.cz, [(b)]Josef.Bulicek@upce.cz, [(c)]Kryze@szdc.cz, [(d)]Petr.Vesely@upce.cz

ABSTRACT

The paper deals with the research on the capacity of switch area within railway stations using computer simulation. The paper briefly introduces some of the currently used methodologies for the calculation of the capacity of switch area and subsequently introduces the concept of the new method, the basis of which is based on the application of the simulation approach. The simulation approach is particularly useful in view of the fact that it can better reflect the reality of traffic compared to purely analytical approaches. The paper also presents a comparison of results from selected currently used methodologies, results of simulation runs with train delays and results from the newly created methodology.

Keywords: railway station capacity, switch area, simulation,

1. INTRODUCTION

Assessment of railway infrastructure capacity is very important activity that applies both to planning new or reconstructing existing infrastructure, planning and timetabling, and last but not least in the operational management of transport processes.
The capacity of the railway infrastructure is closely related to the economy of operation and to the quality of the realized transport.
The performance and quality of railway transport closely dependents on the capacity of the railway infrastructure on which the transport processes are implemented. Capacity calculations are performed primarily for individual tracks of railway lines and railway stations. For railway stations the calculations are individually performed for station tracks and for individual switch area. This article focuses on the capacity calculations of switch area.

Thoughts about capacity calculations are linked to the entire history of rail transport. First methods were focused mainly on quantitative indicators, such as the number of trains (or technological operations) that can be performed on the given infrastructure in compliance with the rules for operating the transport on the given infrastructure. With increasing traffic density qualitative indicators were also observed in the calculations. While such level of capacity utilization is reached, a gradual reduction in the level of transport quality can be occurred in the case that the range of traffic will be still grown on a particular infrastructure. Individual features of this process can be different for each individual infrastructure element (case) due to different local operation conditions. On the other hand, this principle is valid and a challenge for research in the field of railway capacity estimation.

2. EXISTING APPROACHES TO CALCULATE THE CAPACITY OF THE SWITCH AREA WITHIN RAILWAY STATIONS

Currently, a number of approaches to assess the capacity of switch area are being used. Mostly used are analytical and graphical-analytical methods. In general, within SŽDC (dominant railway infrastructure manager in the Czech Republic) there is primarily used the procedure under the D24 directive (originally the ČSD directive D24) - Assessment of the capacity of railway lines. This directive was approved in 1965 and, at the time of its inception, was a good and progressive methodology for the implementation of capacity calculations. From today's point of view, the directive D24 is morally obsolete and, in many respects, does not reflect the nature of the current operation and does not meet the requirements of the current railway transport. Nevertheless, it provides a large number of output indicators that indicate the capacity of the infrastructure. Currently, efforts are being made to replace it with a new modern methodology.

Such kind of so-called graphical methods can be included among the existing methodologies. Graphical methods are not used too often. This is due in particular to their difficult algorithmization and the consequent increased demands on the experience of the worker who carries out the calculation. This is the way in which the so-called additional trains are inserted into suitable gaps within timetable that is tested. When inserting additional trains, care should be taken to avoid any conflict with the trains already recorded in the timetable. The number and ratio of individual types of trains to be introduced and their characteristics must principally be based on the composition of the tested timetable. Second way of graphical methods utilization is so called compression characterized in following paragraph.

The Codex UIC 406 - Capacity, was established in 2004 under the UIC as a recommendation for a uniform and standardized approach to the calculation of capacity in railway transport, it is dedicated to all members of the UIC. This methodology is based on a compression method. Timeslots of trains in tested timetable are adjusted so as to determine the minimum interval for which the range of traffic in the given timetable can be realized. In this compression the collision-free approach and the original sequence of trains must be maintained. From the value of the minimum possible interval of the occupancy of the infrastructure by the given transport range and the length of the interval in which the calculation took place, a basic indicator of this methodology - occupancy time rate, is evaluated.
The main problem of this method is that the train paths are not able to be moved in reality and existing gaps usually cannot be used with the same effectivity as free gap occurred on the end of compressed timetable. For instance, real gaps between existing train paths can be too short for placement of an additional train, so that the capacity is free, but utilizable.
The individual existing methodologies use a variety of computational procedures and provide a different set of output variables in the area of capacity.

The individual existing methodologies use a variety of computational procedures and provide a different set of output indicators in the area of railway capacity. One of the basic output values which is common to most analytical (analytical-graphical methods) values, is the indicator occupancy time rate.

The proposal for a new methodology for the calculation of switch area capacity was carried out in the framework of the expert study which dealt with the calculation of railway station capacity focusing both – switch areas and station tracks. The contracting authority of this study was SŽDC. During the preparation of the new methodology for station capacity calculation, some existing methodologies have been thoroughly tested. Within the project, a comparison of selected output indicators, which these different methodologies provide, was made. The testing was carried out on selected railway stations and was carried out on a large set of different transport scenarios that represented different characteristics and different intensities of traffic.

Outputs of currently used indicators (e.g. occupancy time rate) mainly provide quantitative view on switch area assessment whether it is still more important to check qualitative aspects nowadays due to increasing volume of traffic and sometimes due to relative invariable extent of infrastructure (long-term investments).

3. APPLYING NEW APPROACHES IN CALCULATING THE CAPACITY OF SWITCH AREA WITHIN RAILWAY STATIONS

The new methodology combines the analytical calculation method with the simulation calculation when the mesoscopic level of detail is applied within the simulation section. Main aim to provide the new approach is to add qualitative aspect of capacity assessment of infrastructure because this aspect is not included in traditional assessment.
The analytical part is the calculation of station intervals, which are a necessary part of the calculation. These intervals represent headway times between 2 subsequent technology operations (e.g. train drives).
New methodology for switch area capacity estimation is supposed to simplify those using abstractions of number of station tracks in terms of facilitating the calculation of these intervals.
Already in 1974 prof. Schwanhäußer came up with a probable approach to calculate secondary train delays, which was innovated in 1998 [Schwanhäußer]. This was also an analytical approach that did not reflect a concrete timetable, leading to inaccurate conclusions in assessing secondary delays and other statistical indicators. However, the approach based on the evaluation of secondary delays is so inspirational and objectivizing that there is an attempt to follow it.
Just replacing this analytical part with the simulation part is intended to eliminate such inaccuracies, even at the cost of having to enter specific timeslots of considered trains. It also includes generating random aspects - train delays with the appropriate probability distribution and parameterization.

Using simulation allows to assess infrastructure with indicator focused on secondary delay (increment of possible primary delay of train at estimated part of railway infrastructure) which significantly reflects on the infrastructure possibilities to handle the amount of traffic and secondary delay also indicates quality of the infrastructure. Other indicator that was suggested and can help with assessment of quality of railway infrastructure is average length of queue (how many trains in average have to wait to be able to use switch area or part of switch area).
Advantages of this approach are: indicators of quality are provided; all needed details (timetable, train priorities, parameters of delays etc.) are taken into account as well. Disadvantages of this approach are: only the switch area that is assessed is evaluated. So in case more switch areas in the station and all of them should be assessed, it must be done separately, but this limitation occurred also in the most of used state-of-art analytical approaches, so

that this disadvantage can be seen as acceptable. Another disadvantage is that tracks in railway station or on connected railway lines are not included in the assessment. Potential conflicts on these tracks are not considered. Switch areas are evaluated in an 'isolated' way also from the operational point of view. On the other hand, operational limitations (possible capacity problems) of neighbouring infrastructure are not related to the switch area capacity itself. There is not united opinion if switch area capacity should be evaluated in isolated or complex point of view.

4. CALCULATIONS AND COMPARSION OF OLDER METHODOLOGIES WITH THE NEW METHODOLOGY

Within the project, comparative calculations were performed using selected existing methodologies (see Chapter 2). Additionally, the disadvantages and limitations of the methodologies used were identified and a new methodology has been proposed (see Chapter 3). Comparative calculations were performed using the newly proposed methodology.

Within this chapter the basic output indicators of the mentioned methodologies will be presented and the results will be presented for selected examples of the switch area capacity calculation.

Three different railway stations were examined within the project. These stations differed in both their size and topology. These were the stations Zdice, Lysá nad Labem and Praha hl.n. Zdice railway station is a simple station on a double-track railway line with one branch single-track line.

Lysá nad Labem is a railway junction of busy lines. There is a higher rate of train crossing from one line to another. Station Praha hl.n. is an example of a large node with high traffic (it is primarily about passenger transport). Stations Lysá nad Labem and Praha hl.n. were chosen for the relatively complicated topological layout of switch areas.

For each of these stations, several traffic scenarios were created. These scenarios are differing both in intensity and type of traffic. The current and future planned (able to be occurred with high probability) scenarios of traffic are also included. Other scenarios are hypothetic, but representing legitimate kinds of traffic situation (e.g. increased volume of suburban transport, situation with freight transport in dominant role etc.).

The corresponding calculations were made for all the created scenarios by the above-mentioned methodologies for all switch areas at all stations (2switch areas at each station).

The disadvantages of the existing methods (especially SŽDC D24 methodology) were considered in designing the new methodology. Existing methods have certain insufficient aspects, such as: the need to divide the switch area into elements (groups of switches) for the purpose of calculating, to determine probably time of inaccessibility of infrastructure due to occupation of collective elements by other trains (e.g. going in other direction), missing incorporation of the fact, that train routes are cancelled stepwise (how the individual switches are becoming unoccupied behind a train), difficult interpretation of the resultant value of occupancy time rate in practice and more.

The new methodology was designed with respect to elimination the above shortcomings.

Each of these methodologies provides a set of output indicators. These indicators determine the quantitative and qualitative relationship between the traffic intensity and the range and capabilities of the given transport infrastructure. Most of the output indicators are not directly comparable to the indicators of other methods.

The D24 methodology provides many output parameters. Basic indicators include the occupancy time rate,

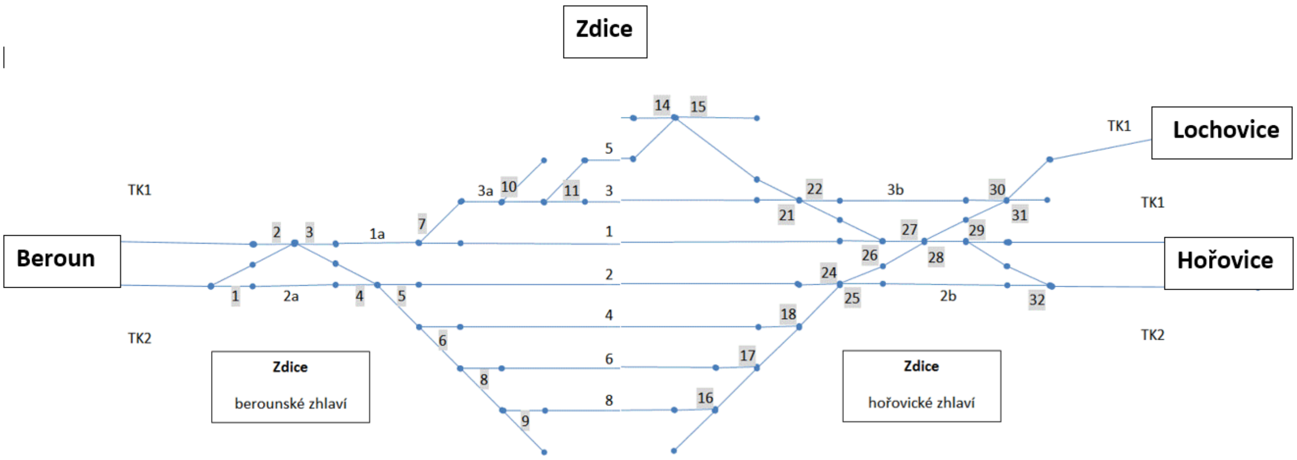


Figure 1 – Schematic layout of Zdice station

capacity, capacity utilization, gap (between operation and average reserve time related to one operation. The UIC406 methodology provides occupancy time rate, number of trains in concatenation, concatenation rate and list of concatenation trains for the given switch area.

Newly proposed methodology, based on the mesoscopic simulation of separated operation of a switch area, provides primarily qualitative indicators secondary delay and average queue length. Secondary delay is defined as increment of delay connected to examined part of infrastructure – switch area (incl. possible waiting in front of this switch area).

Calculation for the chosen transport scenario for the switch area in direction of Hořovice at the railway station Zdice was selected as an example to be presented.

Schematic layout of Zdice railway station and its connection to the adjacent railway stations are shown in Figure 1. The station consists of two switch areas and 8 station tracks, which can be used for train movements or stay (dwelling of trains).

One of the station tracks is dead-end track and is only available from the direction Hořovice and Lochovice. Other main tracks are accessible from all external lines.

The station is connected to three lines. The line Beroun-Zdice includes 2 tracks that are connected to the station tracks via the “Beroun’s switch area”. Both double track line Zdice-Hořovice and single track branch line Zdice-Lochovice are connected to the station via the “Hořovice-Lochovice’s switch area”.

The sample transport scenario is based on the current timetable. Due to the fact that the station is able to realize higher traffic, additional trains (long-distance, regional and freight) have been added to the timetable. These additional trains have been added to examine the capabilities of individual methodologies in a situation approaching capacity limits. In such cases, accurate results are very important. Adding of additional trains was an attempt to create a real expansion of existing traffic. The realistic procedure has been chosen with regard to application in practice.

Within this scenario, in addition to standard train paths, one independent locomotive shunting movement was realized. It is the withdrawal of locomotive from the train number 7803 and its shunting movement on the opposite side of the train. This shunting movement isn’t shown in the timetable in Figure 2 for obvious reasons.

Calculations were performed at the reference interval of 120 minutes between 8:00 and 10:00.

The following trains and shunting movement were included in the calculation for the given period: 60003, 60100, 753, 752, Lc_7803, 60004, 7950, 7802, 60011, 1241, 60010, 7801, 60005, 60002, 7903. Trains 754 and 755 are not used in the calculation because they are trains running on train paths 750 and 751 shifted by 120 minutes. For the same reason, a pair of trains 7903/7904 is omitted from the calculation. Timetable of this transport scenario is shown in Figure 2.

Shifting of solved time period is recommended in the case, when this shifting will allow consider all trains as complete (if some of trains is intersecting a border of solved time period, e.g. 8:00 or 10:00). Naturally, the defined length (120 min) must be preserved. The results can be interpreted in easier and more accurate way than when some train intersecting the border must be considered only by part of its occupation time. On the other hand, in the cases of rush operation this situation can be inevitable. Due to this fact above mentioned modifications (omissions of same trains) are made.

The running time used in the calculations are not determined directly according to regulation (SŽDC V7)

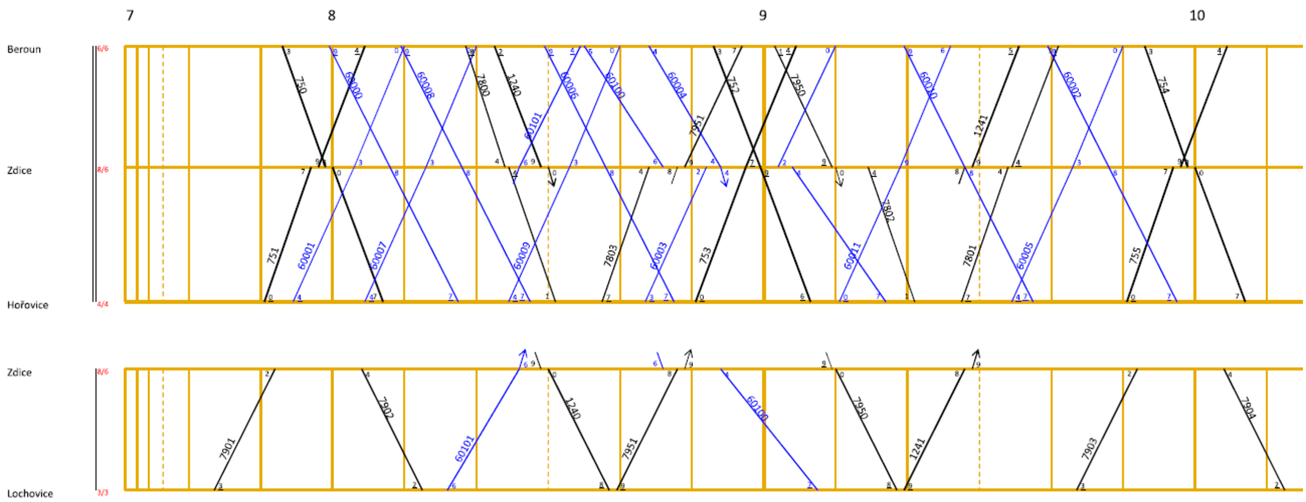


Figure 2 - Timetable

as the sum of theoretical running time and the corresponding layover time (a part add to running time due to reliability reasons). Running times are taken from the simulation model created within the project. The simulation model follows the above principles. Also, other input data, such as time of occupancy, are taken from the outputs of the simulation model, which is determined in accordance with applicable regulations (SŽDC D24, SŽDC 104).

The results of the capacity calculation using D24 directive

The first step in the capacity calculation of the switch area is the assembling of "Overview of trains movement through the switch area". The second step is the assembling "Train path dependency table". The last step is to calculate the values of the capacity indicators for the given switch area. The resulting indicators are shown in the following table.

k_p	Ratio of trains in all operations	0.96
$N_ú$	Capacity [operation]	60.643 (60)
N	Capacity [train]	58.217 (58)
K_{vp}	Capacity utilization	0.487
S_o	Occupancy time rate	0.333
z	Average reserve time related to one operation [min]	2.759
t_{mez}	Gap between operation [min]	0.599

The results of the capacity calculation using UIC406 methodology

An own proprietary software tool was used to calculate using the UIC406 methodology. The input data for the calculation are data on the individual trains and occupation times of the individual elements of the switch area. Based on these input data, an overview of the occupancy of switches is created. Next, using the compression method, the total occupation time of the switch area is determined

Total occupation time after compression [hh:mm:ss]	00:35:34
Occupancy time rate	0.296
Number of trains in concatenation	18
Concatenation rate [%]	62.1
Concatenation trains	751, 60001, 60007, 60101, 1240, 60009, 7803, 7951, 60003, 60100, 753, Lc_7803, 7950, 7802, 60011, 1241, 7801, 60005

The results of the capacity calculation using the Z-SIM method

For the Z-SIM method, the results from the mesoscopic simulation of separated operation of a switch area are presented. The presented results represent an average of 100 simulation replications.

Input delay – sum [min]	240,202
Input delay – average per train [min]	8.283
Secondary delay – sum [min]	9.656
Secondary delay – average per train [min]	0.333
Average queue length [number of trains]	0.080
Ratio of waiting [%]	18.20

5. CONCLUSIONS

Result of investigations it is still recommended to use currently used key indicator - occupancy time rate. This indicator allows to objectively assess the use of transport infrastructure over time due to its formulation as a ratio between a sum of time occupation and total length of calculation time period. The indicator allows you to take into account the different occupation times arising for each operation as well as the expression for only a part of the transport infrastructure.

The project compares two analytical methods SŽDC D24 and UIC 406 that provide similar results.The key indicator of both methodologies is the occupancy time rate, so it is possible to compare its values achieved in individual scenarios. Comparison of the individual scenarios for stations Zdice, Lysá nad Labem and Praha hl.n. was evaluated, after excluding extreme cases, an average deviation of 6.5%. It can be said that the results achieved by both methodologies are comparable.

It is assumed that along with the increasing occupancy time rate, the sum of secondary train delays in the Z-SIM method will increase. The results of the comparison are depicted in Figure 3. It can be seen from this figure that higher-grade switch areas are generally more prone to unfairness in the form of average secondary delays per train.

However, the relationship between the secondary delay and the occupancy time rate is not completely tight (ie. that each specific value of the occupancy time rate would exactly correspond to the specific delay value). Relationship between the two variables is a framework. This confirms that it is necessary to address not only the occupancy time rate as a quantifier but also the above-mentioned indicator, reflecting other aspects of switch area and the operation affecting quality. At the same time, it underlines the necessity of a comprehensive perception of capacity in the context of multiple indicators, not as a single aggregate variable.

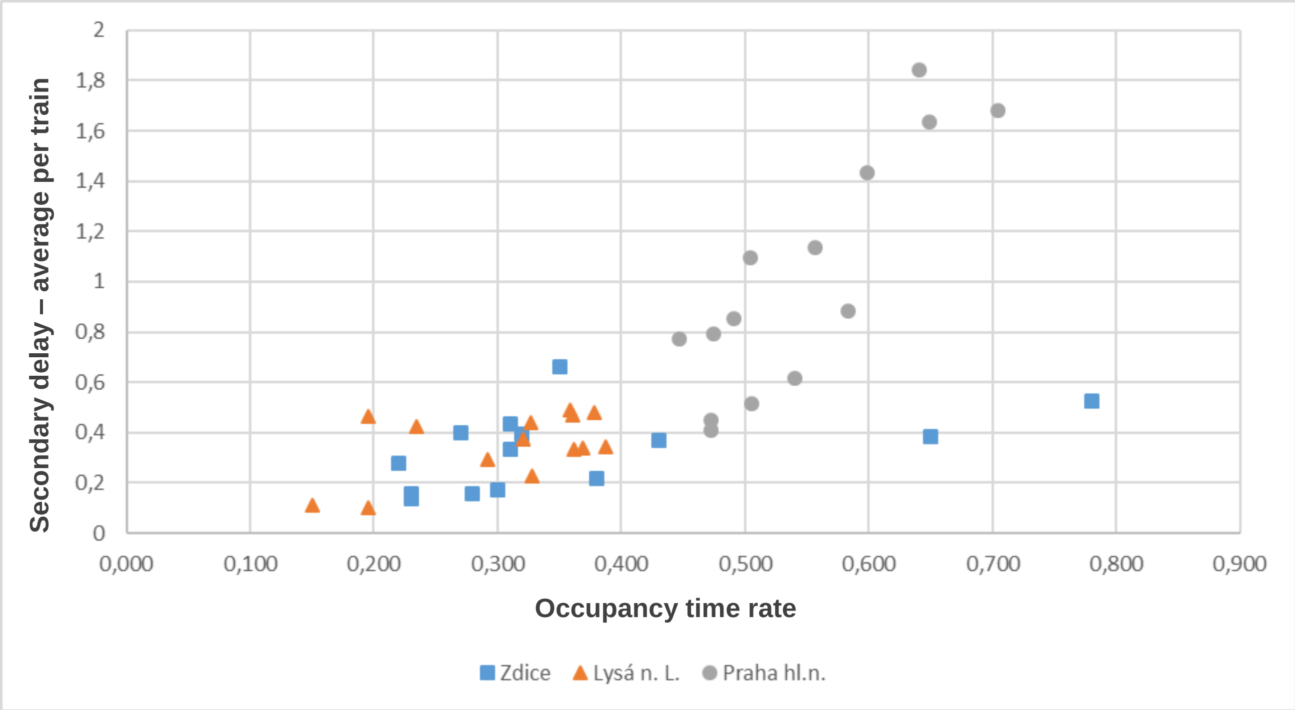


Figure 3 – Dependency between occupancy time rate and average delay time

On the other hand, this also raises the question of whether the average delay per train is indicative. Authors also recommend to take into account values of the secondary delays for individual trains, as the distribution of secondary delays between individual trains can be a key issue. Acceptable average value may be subject to absolutely unsatisfactory values for some of trains, and if time and organizational options allow it, authors recommend to present results in a more detailed breakdown.

From the quality point of view, it is recommended to use mesoscopic simulation of separated operation of a switch area. The value of secondary delay is able to characterize the operation quality of estimated switch area. If this value is too high, it is obvious that switch area is a cause of delays due to a lack of capacity or due to an inappropriate allocation of possible free capacity as well.

On the other hand, it is not necessary to eliminate this delay strictly to value of zero, because such volume of secondary delays (e.g. delay able to be eliminated by other layover times in time schedule) can be accepted. It is necessary to consider all results (values of capacity indicator) in a complex point of view with an effort to discover possible capacity bottlenecks and to prevent their occurrence.

Final conclusion is that mesoscopic simulation can improve the possibilities of switch area capacity estimation. It can also contribute to combined estimation of quantity and quality. This can lead to a complex point of view on capacity estimation of switch area.

Complexity of results also can be improved by ‘additional’ application of microscopic simulation for detail estimation as well. This will be necessary for the case of complicated arguable cases with the results that are close to limits.

REFERENCES

SŽDC (ČSD) D24 – Předpisy pro zjišťování propustnosti železničních tratí, účinnost od 1. 10. 1965.

UIC Code 406 – Capacity. UIC International Union of Railways, Paris, 2004.

DANĚK J., VONKA J.: Dopravní provoz železnic. Bratislava: Alfa, 1988.

Schwanhäußer, W.: Stuttgart 21 – Ergänzende betriebliche Untersuchungen, Teil III“, Leistungsverhalten und Bemessung des geplanten Stuttgarter Hauptbahnhofes und seiner Zulaufstrecken, RWTH Aachen, 1997.

AUTHORS BIOGRAPHY

Michael Bažant is a lecturer in Computer Programming and Simulation in the Faculty of Electrical Engineering and Informatics at the University of Pardubice. He completed his PhD in Technology and Management in Transport and Telecommunications. His research interests lie in the areas of decision-making in the simulation models of transportation systems.

Josef Bulíček is a lecturer in Transport Modelling, System Analysis and Capacity of Railway Transport in the Faculty of Transport Engineering at the University of Pardubice. His research interests lie in interconnection of different modes of public passenger transport, including application of modelling in this field.

Pavel Krýže is employee of the Railway Infrastructure Administration (SŽDC), which is the main manager of railway infrastructure in the Czech Republic. He deals with issues of railway infrastructure capacity.

Petr Veselý is a lecturer in Computer Programming and Computer Graphics in the Faculty of Electrical Engineering and Informatics at the University of Pardubice. His research interests lie in the areas of technological aspects in railway systems.

IDENTIFYING SENTIMENTS ON HOTELS IN TENERIFE PUBLISHED ON SOCIAL MEDIA BY ENGLISH-SPEAKING TOURISTS

C.A. Martín[a], R.M. Aguilar[b], J.M. Torres[c], A. García-Aguilar[d]

[a],[b],[c],[d]Department of Computer and Systems Engineering, University of La Laguna, 38200 La Laguna (Tenerife), Spain.

[a]carlos.martin.galan@iac.es, [b]raguilar@ull.edu.es, [c]jmtorres@ull.edu.es, [d]albertogaragui@gmail.com

ABSTRACT

This paper describes an algorithm for automatically identifying the sentiments expressed by tourists on eWOM (Electronic Word of Mouth) platforms. Based on reviews published by tourists after staying in hotels, a classifier is trained using an SVM (Support Vector Machine) supervised learning algorithm. We present a use case for this method involving a group of hotels located on the island of Tenerife (Canary Islands).

Keywords: sentiment analysis, social net, machine learning, support vector machine, turism

1. INTRODUCTION

The tourism industry accounts for approximately ten percent of the world's GDP (WTTC, 2015), comprising almost the entirety of the economy in some regions. In Spain, its contribution to the GDP is sixteen percent. In the ranking of the importance of tourism to the national economy, Spain is in seventh place in absolute terms, and in forty-seventh if its size is taken into account (WTTC, 2016).

According to the latest figures, in 2014 the Canary Islands received over eleven and a half million foreign tourists for the first time in its history (11,511,108). This was 8.68% more than in 2013, or an additional 919,848 foreign tourists. The tourism industry is the real engine of growth and development in the Canary Islands, accounting for a high percentage of its GDP. This has a knock-on effect on the remaining industries and services on the islands, especially in the development of trade, transportation, food production and industry.

Tourism also comprises a very important component in creating jobs in the services sector of the archipelago, which encompasses direct employment in the sun and beach sector, as well as workers in activities that support tourism, such as restaurants, hotels, travel agencies, passenger transport, car rental and recreational, cultural and sports activities.

In 2016 Tenerife received over 7 million tourists, most of them from the United Kingdom, which accounted for thirty-one percent of the passengers arriving at Tenerife's airports in 2016. These figures indicate that the opinion that English tourists have of Tenerife is particularly relevant.

The Canary Islands are home to 17% of all hotel rooms available in Spain (1,363,934). These are offered primarily in medium- to high-category establishments, with 56% of the rooms in four-star hotels, 28% in three-star hotels and 10% in five-star hotels (Instituto Canario de Estadística, 2010a; Instituto Nacional de Estadística, 2010).

Future forecasts agree on the need to adapt tourist activity to a) the new requirements and demands of clients, as well as to changing conditions in the industry, such as the emergence of new destinations and markets, namely Asia and Central and Eastern Europe, b) shorter stays and shorter travel distances, c) the financial situation, d) the growing autonomy of users in arranging their travel and the democratization of the supply, e) the effect of new information and communication technologies in the marketing, distribution and communication processes f) the limits on the environmental carrying capacity of the destination, g) the demand for greater corporate social responsibility, and so on (European Travel Commission, 2010).

In light of the financial importance of tourist activity to the Canary Islands, and of the constantly changing scenario, the market survival of our destinations and companies will depend to a large extent on their ability to continue offering experiences, products and services to tourists that are sufficiently appealing and that respond to their changing tastes and demands. As a result, the important role of research in tourism is undeniable as a means of identifying the key socio-economic, sectorial and corporate variables, and for making decisions that will lead to successfully managing both destinations and companies.

According to Peppers and Rogers (Peppers, 2011), a company that caters to its customers is one that uses information to gain a competitive edge, grow and become profitable. In its most general form, Customer Relationship Management (CRM) may be considered as a group of practices designed to put a company in much closer contact with its customers so it can learn more about them. The broader goal is to increase the value of its customers, and thus of the company. This is why when a management model is based on customer satisfaction, we have to consider four dimensions: customer identification, customer acquisition, customer retention and customer development. This fourth dimension involves finding the real market for a tourism company and product by discovering and learning from its own customers. Focusing on the customer allows developing a product that the customer actually needs. In short, the goal is to have an in-depth knowledge of the customer.

In this paper, we seek to monitor users' perceptions of hotel establishments in Tenerife through comments made on social media so as to ascertain the level of satisfaction with the tourism destination.

The analysis or tourists comments are part of the sentiment analysis field. According to Liu B. (Liu B, 2012) the purpose of sentiment analysis is to define automatic tools capable of extracting subjective information to create structured knowledge that can be put into operation.

Sentiment analysis is a complex field of research (Sentiment Analysis in Social Networks, 2016), in which great difficulties can appear like context analysis or figurative language devices such as irony and sarcasm. These sort of problems are common when users comments in social networks. Despite this, companies are increasingly mindful of the importance of opinions on their products and services that users publish on social media. Specifically, the appearance of social media has provided us with a new and valuable source of information in the tourism industry for studying the perception of various destinations. This way of learning tourists' opinions offers advantages, since the information provided is spontaneous, natural and not conditioned by how a questionnaire is designed. We can thus learn which factors are truly important at the destination from the tourists' point of view and their opinions of said factors. However, due to the diversity and amount of information, monitoring the social networks in real time is a huge challenge that can only be undertaken through automation. Even so, the problem poses serious difficulties due to the heterogeneity of the information and the added challenge of processing natural language. In this regard, the application of machine learning or network analysis techniques is essential to overcome these barriers.

The goal of our work is to automatically analyze the perception of hotels in Tenerife (Canary Islands) through information provided by English-speaking tourists on social media. To achieve this objective, we propose creating a tool that detects the sentiment expressed in the comments published on the tourism platforms booking.com and tripadvisor.com so as to determine the level of satisfaction with the services received during stays at hotels on the island of Tenerife.

2. METHODS

A text can be automatically classified through machine learning. This entails using a classification algorithm that, based on a set of words, is capable of assigning text to a predefined category. To create this algorithm we need to have a set of data (text) whose category is known (sentiment) that can serve as training data. The next step involves using a classification algorithm to determine the different parameters of the model. In this paper we describe the design for classifying text with a support vector machine algorithm. This process will be divided into the four main steps shown below:

- Definition of the category tree
- Gathering of training data
- Labeling of training data
- Training the classifier

2.1. Definition of the category tree

The first issue to address is what categories or labels we want to assign to the texts. This will depend on the problem to be solved. If the goal is to analyze the sentiment of the text, we could assign the following categories: Positive or Negative. Then, the texts that are entered into the classifier will be labeled as positive ("the hotel was clean") or negative ("the food was bad").

The sentiment analysis can have two possible goals:

- Polarity analysis: to ascertain if a given text has positive, negative or neutral connotations. One of the relevant applications in this area is to analyze user reviews of a product.
- Identification of names: to output a list of the names mentioned in the text, and certain implementations also classify them. For example, one feature of the identification of names is to list the people, organizations or locations mentioned in a text.

Specifically, in this work we will conduct a polarity analysis to determine if, based on their comments, tourists have a good or bad perception of the service received in hotels in Tenerife.

2.2. Gathering of training data

Once the category tree is defined, the next step is to gather training data. This refers to the texts that will be used as training examples and that are representative of the future texts that we want the model to classify automatically. These data are everywhere and can be obtained automatically from the web through APIs and open data sets. If our goal is to learn tourists' perception of the service received at a hotel, we can resort to comments published on eWOM tourism platforms like tripadvisor.com or booking.com.

2.3. Labeling of training data

Once the data are collected, they have to be labeled with the corresponding categories in order to create the set of training data. This is a tedious process because each text has to be labeled using the set of predefined categories. When determining tourist sentiment, each comment used for training has to be identified as positive or negative. The set of training data tells the classifier that for a given input (text), it has to assign a specific output (category). The success of any training algorithm based on supervised learning (it has a data set with its associated classification) depends on the accuracy of how the training data set is initially labeled.
The number of training samples needed depends to a large extent on the specific use case; that is, on the complexity of the problem and on the number of categories to be used. For example, it is not the same to train a sentiment analysis model for comments as it is to train a classifier to identify specific elements in a text. Sentiment analysis is a much harder problem to solve and requires much more training data. Analyzing comments published on eWOM platforms is much more difficult than analyzing well-written and well-structured critiques. In summary, the training data have to be sufficient (at least 3,000 samples for sentiment analysis) and adequately proportioned in terms of the number of positive and negative texts. (Pozzi, 2017)

2.4. Labeling of training data

A support vector machine (SVM) is a supervised learning algorithm; therefore, it has to be trained using a set of properly categorized samples so as to yield a classifier that outputs the category of a new case. The SVM model is supported by the statistical theories proposed by Vapnik (Vapnik,1999).

This type of model can be used for binary or multi-category classifications or, in the most general case, to approximate a regression.
Its performance relies on finding the vectors that define the hyperplanes separating the groups with the maximum margin. To understand this idea, suppose that we want to make predictions about a binary classification. For example, we have the values for a hotel inspection and we want to determine if it increases its star rating or not. We could graphically represent the samples on a Cartesian plane on whose edges we only consider two of the variables included in the inspection (Figure 1). The hotels whose star rating is increased are drawn as circles, and those whose rating is not increased are drawn as squares.

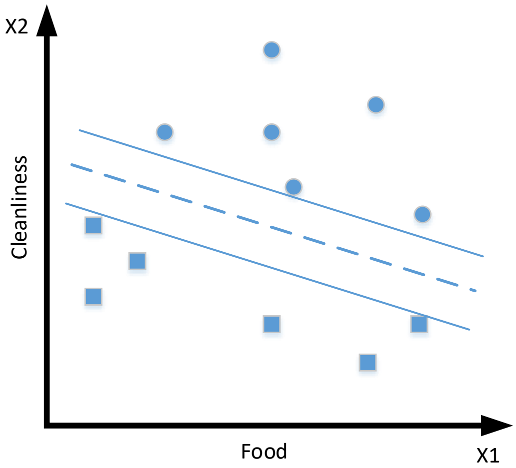


Figure 1: Hotel with increased star rating (circles) and with same star rating (squares)

The SVM model will assign a surface to separate the negative and positive samples by the largest margin possible on either side of the hyperplane (the solid lines in Figure 1). In this case, since it is a two-dimensional problem, the SVM will look for a straight line. In a case with three attributes to classify (e.g. food, cleanliness, staff), it would be a plane. In a case with more than three dimensions, it would be a hyperplane with the appropriate number of variables.
There are numerous methods for determining this hyperplane, such as neural networks and optimization. What sets SVMs apart is that the resulting hyperplane is such that the margin separating the data is as large as possible (dashed line in Figure 1). The further away the hyperplane is from the samples to be classified, the better the result will be because small variations in the variables do not result in being assigned to another group.
Therefore, if we regard each available case (squares and circles) as a vector in space, the SVM algorithm looks for the vectors on which to

support the parallel planes, one toward the positive set and another toward the negative. The hyperplane of maximum separation desired will be the vector drawn exactly in between these vectors.

In the event that a non-linear surface has to be used to separate the sets (Figure 3 left), the SVM transforms the space of attributes into a linear space of higher dimensionality. It does this by mapping each point in the data set by means of a non-linear transformation $\phi(x)$: $R^2 \rightarrow F$, in the so-called higher dimension feature space F, such that each point X_n is projected to a point $\phi(X_n)$. If the transformation is selected correctly, the feature space F will be a higher dimensional space in which there will be a linear relationship between the projected data set and the factors that underlie the variability in the original data set. This process of mapping data in a feature space by means of a non-linear transformation is part of what is known as kernel-based methods. Figure 2 left shows two data sets in R^2. In principle, it would be very complicated to train a classifier that would be able to separate the two sets; however, each point can be mapped using a non-linear transformation $\phi(x)$ into a feature space F with M dimensions. If the non-linear transformation is properly selected, both data sets could be easily separated along a boundary by hyperplane h in F. A schematic illustration of this process is shown in Figure 2.

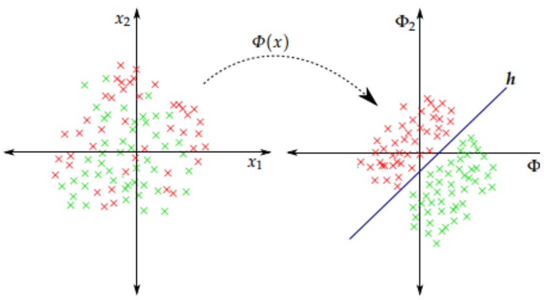


Figure 2: In an SVM, two data sets that are hard to classify in R^2 (left) are mapped via ϕ (x) into the feature space (right), where they can be classified by using hyperplane h as a boundary.

3. RESULTS

The goal of this work is to create an automatic classifier that can determine the sentiment that tourists who visit the island of Tenerife have regarding the hotels where they stayed. This sentiment is taken from the comments left on eWOM platforms like booking.com and tripadvisor.com. (García-Pablos, 2016)

As noted in the above section, the first step in creating a classifier is to obtain a set of previously classified comments to use as a training sample for the classifier that will be used later. In order to train the classifier properly, we need to find 1500 positive comment samples and 1500 negative comment samples. The first part of the work involved finding said samples. We decided to take the training samples from the British version of the booking.com website, where comments on hotels feature the following characteristics:

- They are scored numerically between 0 and 10 (decimals allowed).
- In some case, information is available on the source and date of the comment.
- The person making the comment has the option of giving the comment a title with a positive and negative aspect from his review.

With these characteristics, and after doing a mass data extraction, we can set up the training sample as follows:

- Take the reviews with a score above 6 and label them as “good”, using as training text the concatenation of the title and the positive part of the comment.
- Take the reviews with a score below 5 and label them as “bad”, using as training text the concatenation of the title and the negative part of the comment

From the resulting samples, we select 1500 good ones and 1500 bas ones distributed uniformly along the range of scores. An example of the samples used to train the algorithm is shown in Figure 3.

Once the training samples are obtained, we generate an SVM training algorithm using Monkeylearn (infrastructure and software service for analyzing text) (MonkeyLearn, 2017). The following parameters are specified to define the algorithm:

1. Language: must match that in the samples (English in this case).
2. Normalize weights: we enable the normalization to inform the classifier that it must take into account the number of samples for each category when defining their probability. In our case, since the number of positive and negative samples is balanced, it is not enabled.

3. N-gram: defines whether the features used to characterize the text will be:
    - Unigrams or words (n-gram size = 1)
    - Bigrams or terms consisting of two words (n-gram size = 2)
    - Trigrams or terms consisting of up to three words (n-gram size = 3)

The classifications are based on words (unigrams).

4. Stemming: this setting specifies if words have to be derived. The derivation process transforms words into their root version, meaning that inflections and derivatives are grouped. For example, the words “fishes”, “fishing” and “fisherman” are all transformed into the root word “fish”. This will generally help the classifier to generalize and improve its classification. However, in the case of sentiments based on reviews written in

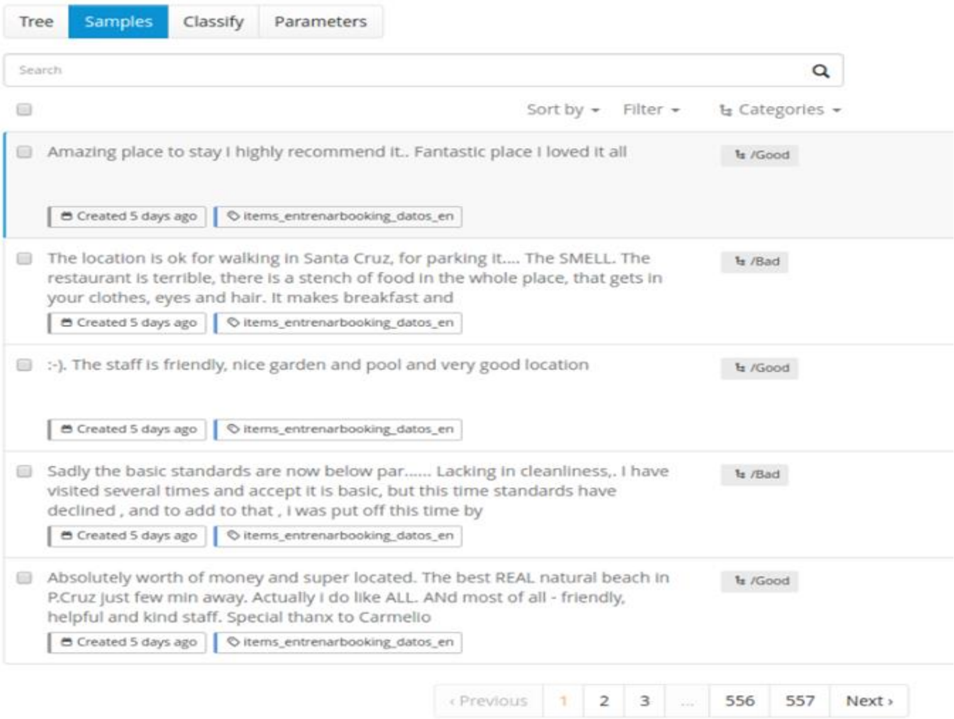


Figure 3: Example of some of the training samples used

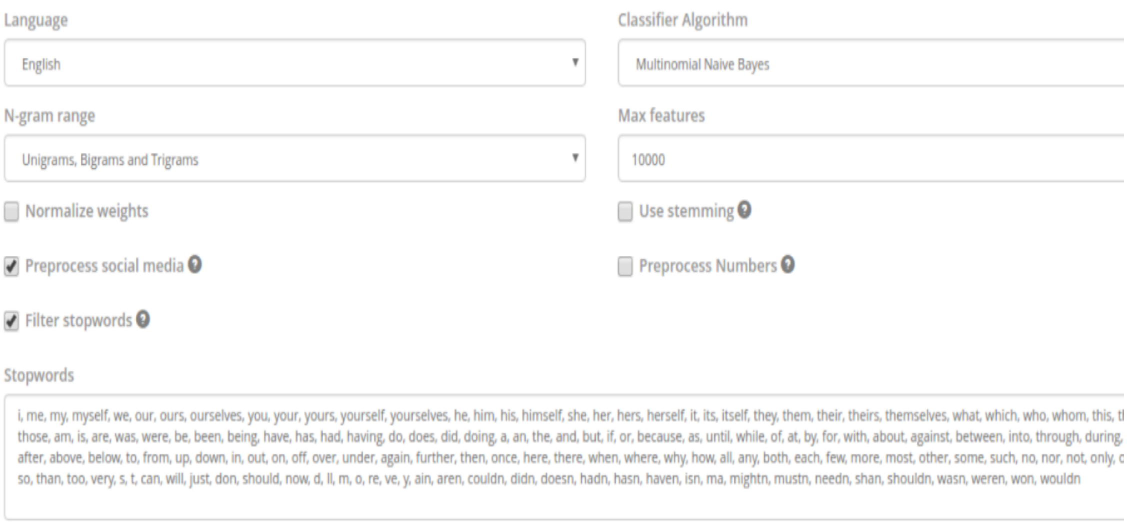


Figure 4 – Configuration parameters for training the SVM

digital platforms, the setting is better without this transformation.

5. Filter inactive words: this active setting filters stopwords. These are high-frequency words that do not normally contribute as classification characteristics, such as articles, conjunctions, etc.
6. Pre-processing of social media: expressions used in conversations on social media are transformed into a word. For example, :-) is transformed into "happy".

The training of the classifier yielded the results shown in Figure 5. The method was calculated to be 95% accurate in identifying good sentiments and 91% for the bad. We can see that the words that are repeated most often in good sentiments are "beautiful" and "amazing", whereas in the bad it is "rude", "dirty" and "cockroaches".

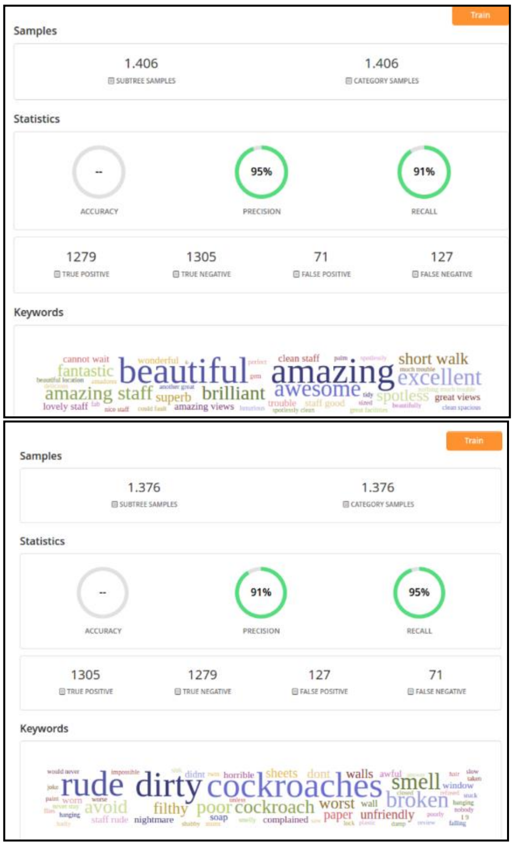


Figure 5 – Performance of the training process for good sentiments and bad sentiments.

We monitored a group of hotels in a city on the island of Tenerife; specifically, we analyzed 13,541 comments, yielding the results shown in Figure 6. The X axis shows the actual rating of the reviewer, while the Y axis shows the predictions of the classifier. Note that better results are obtained by classifying negative comments (1-2), than positive (4-5), and as expected better predictions at the extremes.

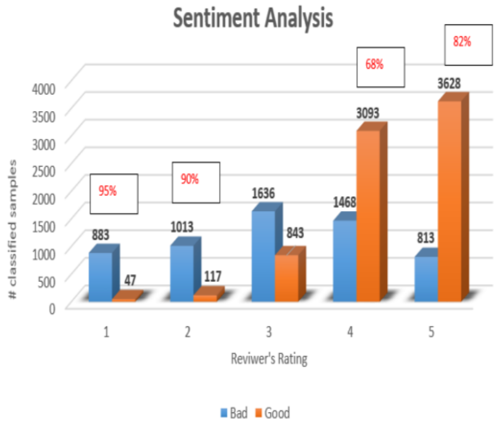


Figure 6 – Results of the classification of 13541 comments on hotels.

4. CONCLUSIONS

This paper clearly shows the need to monitor customer perception in order to maintain a competitive edge. In the case of tourism, the main driver of the economy in the Canary Islands, it is a key component. In this area, however, in which there are millions of users, finding information on their perceptions of the service received is a cumbersome task. Moreover, the use of questionnaires does not provide accurate information on how customers feel about the service they received. As a result, in this paper we propose using the reviews that tourists publish on eWOM platforms like tripadvisor.com and booking.com.

In order to identify the sentiment of a comment, we used an SVM training algorithm that, according to the bibliography, exhibits very good efficiency indices with classification problems, (Xiang'zi, 2017). In this paper we used the online service Monkeylearn to create the classifier.

In the final step, we used this classifier to determine the sentiment expressed for a group of hotels on the island of Tenerife in the Canary Islands. In this use case we can see how the classifier obtains better results by evaluating potentially negative comments and, as expected, shows a better behavior with less neutral comments. In future work we plan to create a classifier that can identify what tourists are discussing so that we can determine what aspects they like and thus what to promote, as well as what elements are associated with negative comments so that they can be improved.

We also plan to conduct a more exhaustive analysis of the results in order to be able to monitor sentiments based on the tourist's place of origin, the length of their stay on the island and other factors.

ACKNOWLEDGMENTS

This work is being supported by the project #2016TUR15 "VITUIN: Vigilancia Turística Inteligente de Tenerife en Redes Sociales " from research funds of the Fundación CajaCanarias

REFERENCES

García-Pablos A., Lo Duca A., Cuadros M., Linaza M.T., and Marchetti A. 2016, "Correlating Languages and Sentiment Analysis on the Basis of Text-based Reviews," in Information and Communication Technologies in Tourism 2016, pp. 565–577.

MonkeyLearn Inc, ©2017. All rights reserved. http://monkeylearn.com [accessed 15 May, 2017]

Peppers D. and Rogers M., 2011, Managing Customer Relationships: A Strategic Framework (Second Edition), Wiley.

Pozzi F.A., Fersini E., Messina E., and Liu B., 2017,, "Chapter 1 - Challenges of Sentiment Analysis in Social Networks: An Overview," in Sentiment Analysis in Social Networks, Eds. Boston: Morgan Kaufmann, pp. 1–11.

Vapnik Vladimir N., 1999, An Overview of Statistical Learning Theory, IEEE Transactions On Neural Networks, Vol. 10, No. 5

WTTC, 2015, Econocmic Impact Analysis 2015, World Travel & Tourism Council. Available from: http://sete.gr/media/2614/150430-economic-impact-2015.pdf [accessed 15 May, 2017]

WTTC, 2016, Econocmic Impact 2016 Spain, World Travel & Tourism Council. Available from: https://www.wttc.org/-/media/files/reports/economic%20impact%20research/countries%202016/spain2016.pdf [accessed 15 May, 2017]

Xiang'zi Leng, Jinhua Wang, Haibo Ji, Qin'geng Wang, Huiming Li, Xin Qian, Fengying Li, Meng Yang, 2017, Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses, Chemosphere, Volume 180, pp: 513-522

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, & Bing Liu. (2016).Sentiment analysis in social networks (1st ed.). US: Morgan Kaufmann Publishers Inc. Retrieved from http://lib.myilibrary.com?ID=961656

NUMERICAL ALGORITHM FOR WAVEGUIDING PROPERTIES RECONSTRUCTION FROM INTENSITY-ONLY MEASUREMENTS IN MULTICORE FIBERS WITH STRONG MODE COUPLING

Alexey Andrianov[(a)], Nikolay Kalinin[(b)], Arkady Kim[(c)]

[(a),(c)]Institute of Applied Physics of RAS, Nizhny Novgorod, Russia
[(b)]Nizhny Novgorod State University, Nizhny Novgorod, Russia

[(a)]alex.v.andrianov@gmail.com
[(b)]nikolay@kalinin.nnov.ru
[(c)]arkady.kim@gmail.com

ABSTRACT

A numerical algorithm for reconstruction of supermodes structure and propagation constants in multicore optical fibers with strong core coupling is developed. The method is based on measurements of intensity distribution patterns at the fiber output under excitation of individual cores at the input for several fiber lengths and subsequent solution of the inverse problem. We used our method for a 7-core fiber and showed that the inverse problem can be solved efficiently by the stochastic hill climbing method and all waveguiding properties of the fiber can be retrieved with high accuracy.

Keywords: multi-core fibers, supermodes, numerical modeling, stochastic hill climbing

1. INTRODUCTION

Multi-core optical fibers (MCF) have recently attracted much attention as a promising medium for high-speed data transmission systems, and also as a component of various photonic devices and laser systems (Saitoh and Shoichiro 2016). Of particular interest are MCFs with sufficiently large coupling between the cores, which leads to efficient energy exchange. Such fibers can be used to manipulate the beams of high-power laser radiation, in particular, for coherent combining of radiation from many lasers to increase the peak intensity (Ramirez 2015). In such fibers, parameters of the individual cores and the coupling between the cores must be controlled with great accuracy to ensure predictable interaction, especially in the presence of strong nonlinearity characteristic of high-power laser systems. To carry out a full-scale numerical simulation with realistic parameters of experimentally produced MCFs, it is necessary to know with high accuracy real parameters of their cores and coupling coefficients between cores, which determine the properties of propagating modes. Direct measurement of these parameters with demanded accuracy is a rather difficult task. In this paper, we present an alternative algorithm based on solving the inverse problem of reconstructing the required parameters from measurements of linear light propagation.

2. RETRIEVING FIBER PARAMETERS FROM LINEAR PROPAGATION EXPERIMENT

The MCF can be seen as a system of several coupled single-mode waveguides. Coupling coefficients between the cores depend on the internal structure of the fiber. Assuming that the fiber structure does not change along its length, it is possible to calculate the supermodes (eigenmodes) of the MCF, each of which propagates with its own phase velocity. If only one supermode is excited in the MCF, the power distribution over the cores remains constant during its propagation. In an ideal case, when the internal structure of the fiber (shape and size of the cores and the distance between them) is known, it is possible to calculate the coupling coefficients between the cores and calculate the propagation constants and the supermode profiles. However, even with slight deviations from the ideal fiber structure, the shape of the supermodes and their propagation constants differ from the calculated ones. If several supermodes are excited (for example, when only one MCF core is illuminated at the input), the pattern of the intensity distribution at the fiber output is very sensitive to the deviations of the propagation constants from the ideal case.

Typically, designed core parameters (sizes, refractive index differences), as well as core spacing, are well known for multi-core fibers. However, due to technological issues in the fabrication of a multi-core fiber, the properties of the cores are subject to random perturbations (deformation of the shape, additional mechanical stress that cause some changes in the refractive index); the distances between the cores are also slightly different from the target value. This leads to slight differences in the propagation constant of the individual modes of the cores and the coupling coefficients between the cores and, consequently, the structure and propagation constants of the supermodes of MCF. These usually small changes significantly affect the propagation of laser radiation due to the accumulation of the phase difference over a sufficient propagation length. In experiments that are promising for observing nonlinear effects and studying the coherent combining of laser beams in MCFs, fiber lengths corresponding to several beat lengths (the length of linear energy transfer between cores) are used, which are sufficient to observe a significant difference of the intensity pattern in real and ideal fiber.

Further we show that it is possible to reconstruct the propagation constants and the coupling coefficients of the cores, as well as the profiles of the supermodes, on the basis of several measurements of intensity patterns at the fiber output and optimization algorithm.

The propagation of light in the MCF can be described by the equation

$$\frac{dA}{dz} = i\hat{C}A, \quad (1)$$

where $A=(a_1..a_N)^T$ is the column vector of slowly varying amplitudes in each of N cores, z is the coordinate along the fiber, and $\hat{C}$ is the matrix of the coupling coefficients and the propagation constants of the cores (Chan, Lau and Tam 2012, Koshiba, Saitoh, Takenaga and Matsuo 2011). The eigenvalues of the matrix are propagation constants of N supermodes, and its eigenvectors are profiles of the supermodes.

Based on several intensity measurements in each core $(I_1(z_j)..I_N(z_j))$ at several different propagation lengths z_j and several known initial conditions, the matrix can be retrieved by solving the minimization problem. The difference between calculated and measured intensities

$$\Delta = \sum_{j=1}^{M}\sum_{i=1}^{N} | I_i(z_j) - | a_i(z_j) |^2 |^2, \quad (2)$$

where M is the number of different points along the fiber at which the measurements were taken, needs to be minimized, which can be done by well-known algorithms.

The main part of the developed method is a modified numerical minimization algorithm for finding the matrix $\hat{C}$, and consequently, the coupling coefficients between the cores and the propagation constants of individual cores. We used modified stochastic hill climbing method (Russel and Norvig 2010, Brownlee 2011), which was tailored for our particular problem.

Since the number of unknown variables in this matrix $\hat{C}$ is sufficiently large (N^2, where N is the number of cores), then the use of direct numerical algorithms to search for minima is not possible for multi-core fibers. It should be noted that not all of the unknowns are independent since we assume the matrix $\hat{C}$ to be Hermitian, which corresponds to the absence of absorption in the fiber. To handle the minimization problem with the above restriction, we built our own implementation of the stochastic minimization algorithm.

As an initial approximation in our method we use the matrix $\hat{C}$, which can be found on the basis of the known geometric structure of the fiber and the refractive indices of the cores and cladding materials by using an appropriate mode-solver. Usually, these data describe the real fiber quite well, except for small deviations, which however lead to significant distortions in the light propagation patterns, and are to be found by our iterative method.

At each iteration, each value in the matrix is changed randomly within certain limits leaving the matrix symmetric, then the value of Δ is computed. At each iteration, several such changes are made to the same matrix, from which the approximation with the smallest value of Δ is selected. If this value is better than the previous one, the approximation is replaced by the matrix that gave this smallest value of Δ. The number of changes at each iteration is of the order of 10-50. It is important to note, that in our implementation of the iterative algorithm the limits of changes decrease with increasing number of iteration, which allows obtaining more accurate results and at the same time reduces computational costs. Namely, at each iteration, which did not lead to an improvement in the approximation, the limit of variation of each matrix element was multiplied by 0.995. In addition, as the iteration number increases, the number of changes at each step increases. For retrieving experimental data for 7-core MCF (see below) we used 10 changes per iteration at the initial stages of the algorithm and 50 changes per iteration at the final stages.

To satisfy the Hermitian condition for the matrix $\hat{C}$, all random variations were chosen so as to preserve its symmetry. It turns out that even in this class of approximations it is possible to significantly improve the initial approximation, calculated from the geometric structure of the fiber. However, the modification of the algorithm, allowing asymmetric, but at the same time Hermitian changes in the matrix $\hat{C}$, did not significantly improved the approximation.

To demonstrate applicability of this method, we experimentally tested it for a seven-core fiber with the structure shown in the inset in Fig. 1.

The laser radiation was alternately launched into each core and the output intensity in each core was measured at several fiber lengths by the cut-back method. Few measured intensity patterns at the fiber output are shown in Fig. 2. The initial approximation for the matrix was calculated on the basis of the fiber structure. With the help of our mode-solver the direct problem of finding propagation constants and coupling coefficients in an idealized fiber was solved and corresponding ideal propagation constants and coupling coefficients were calculated, which were then used as an initial approximation for the matrix $\hat{C}$ at the input of the iterative algorithm. Next, our modified stochastic hill climbing method was used to find the matrix that best approximates the measured intensity dependences. After several iterations we achieved significant reduction in the discrepancy between the measured and calculated dependences compared to the initial approximation (i.e., approximation calculated on the basis of the refractive index profile of the MCF preform), see Fig. 1.

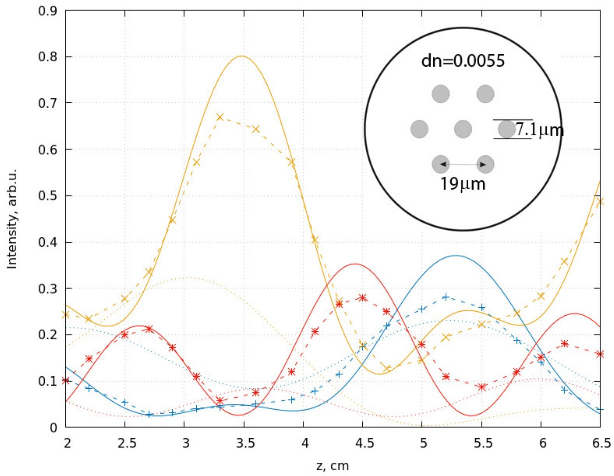


Figure 1: Intensity in some MCF cores measured in experiment (dashed lines with symbols), calculated by our algorithm (solid lines) and calculated based on the refractive index profile (dotted lines).

It can be seen that the overall structure of the reconstructed supermodes is close to the structure of the supermodes, calculated from the ideal refractive index profile of the MCF (Fig. 3). The propagation constants of the supermodes obtained by our algorithm show only a small difference at the relative level of 10^{-5} from the values calculated on the basis of the MCF refractive index profile. Nevertheless, this gave a much better matching of measured and calculated dependences of the intensities in the cores along the fiber. Since the obtained matrix determines all the waveguiding properties of the MCF, the propagation constants in individual cores can be readily obtained. Thus, our algorithm allows retrieving with high accuracy small corrections to the propagation constants (at the level of 10^{-5}), arising from imperfectness of the fiber manufacturing technology.

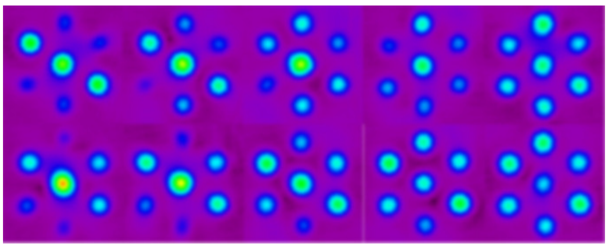


Figure 2: Measured intensity patterns at the output of different lengths of 7-core MCF while the central core was excited at the input.

Compared with the known method for reconstruction of mode properties in multimode and multi-core fibers based on the observation of spectral interference of a broadband signal and subsequent Fourier processing (Nicholson, Yablon, Fini and Mermelstein 2009), our method does not require a broadband light source. Another known method utilizes similar idea that underlies our method and based on the measurements of the output intensity patterns and corresponding iterative algorithm (Mosley, Gris-Sánchez, Stone, Francis-Jones, Ashton and Birks 2014). However, we believe that our method has much better convergence and greater reliability due to utilization of multiple data sets at several fiber lengths.

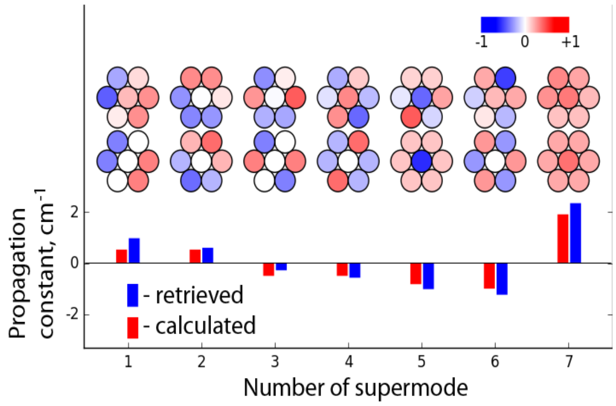


Figure 3: Structure of supermodes and propagation constants (relative to propagation constant of single core) retrieved by our algorithm (upper row) and calculated on the basis of refractive index profile (lower row).

Typically, to reconstruct the parameters of 7-core fiber, about 30000 to 50000 iterations were required, depending on the quality of the initial approximation. Our C++ implementation of the algorithm took about one minute to perform the reconstruction on a typical laptop computer. Thus, our algorithm is significantly faster than previously reported iterative algorithm utilizing intensity-only measurements for reconstructing parameters of multi-core fibers (Mosley, Gris-Sánchez, Stone, Francis-Jones, Ashton and Birks 2014).

3. ACCURACY AND STABILITY OF THE ALGORITHM

According to the results obtained from the processing of the measurement data for the 7-core fiber, it can be seen that our method allowed us to find corrections to the coupling coefficients and propagation constants, which give a much better coincidence of the radiation propagation pattern to the measured one, than the pattern calculated on the basis of the ideal fiber structure.

However, it was not possible to achieve perfect coincidence of the measured intensity and intensity calculated by our algorithm. This can be caused by

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

various factors, for example, additional changes in the structure of the fiber and the coupling coefficients along the fiber, unaccounted dynamics of the polarization and nonmonochromaticity of propagating radiation, and inaccuracies in the experimental measurement of intensities. In addition, this could be caused by the inefficiency of our method. In order to test the efficiency of the method, the following numerical study was carried out.

Our method works under the assumption that the propagation of light in a multi-core fiber is described by the matrix $\hat{C}$, which contains the coupling coefficients between the cores and the propagation constants of individual cores (Eq. 1). It is reasonable to check numerically the efficiency and the limits of our method. To do this, the matrix $\hat{C}$ calculated on the basis of the refractive index profile of a certain fiber, for example, shown in the inset in Fig. 1, was changed randomly. Then, using the solution of the direct propagation problem (Eq. 1), the intensities at several points along the fiber were calculated upon excitation of various cores at the fiber input, i.e. experimental measurements were simulated numerically. After that, our method was applied to solve the inverse problem using simulated intensity values. As a result, the original and reconstructed values of the coupling coefficients and propagation constants were compared.

These calculations were performed for the various amount of random changes of the matrix $\hat{C}$. We introduced parameter *p*, and each matrix element was changed by a random value not exceeding *p*% of its original value while maintaining the symmetry of the matrix. It turned out that correct retrieval of the matrix $\hat{C}$ strongly depends on the value of *p*. For example, for our 7-core fiber, the refractive index profile of which is shown in the inset in Fig. 1, the correct parameters are retrieved for *p* not exceeding 80. For large values of *p*, the initial approximation for the algorithm is not accurate enough, and the reconstructed values differ substantially from the correct ones. To estimate the accuracy, we used the parameter Δ, as well as the average deviation of the elements of the reconstructed matrix from their correct values

$$e = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} | C_{ij} - C_{ij}^* | \quad , \qquad (3)$$

where $\hat{C}^*$ is the retrieved matrix.

The obtained dependencies of the reconstruction quality *e* and Δ on the parameter *p* for a different number of the fiber cores *N* are shown in Figs. 4 and 5, respectively. For values of *p* giving an unstable result, several realizations showing the spread are displayed in the figures. In this case, we consider results with the parameter $\Delta<10^{-2}$ and the parameter $e<0.1$ as sufficiently good reconstruction. It can be seen that for MCFs with a smaller number of cores, the reconstruction quality is improved.

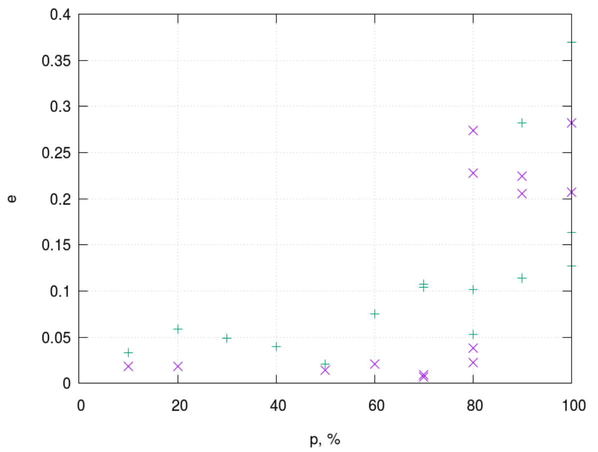


Figure 4: Dependence of the quality of reconstruction *e* on the parameter *p* for 7-core fiber (x) and 4-core fiber (+).

4. MINIMUM DATA SET FOR SUCCESSFUL RECONSTRUCTION

Our method uses data on the intensity of light in different cores at different propagation distances with several variants of excitation at the fiber input.

In the experimental measurement of the intensity and in numerical simulations, we used *N* variants of different input excitation; in each variant one of the *N* cores was illuminated at the input. We used from *2N* to *3N* different lengths of the fiber within the interval corresponding to several beat lengths. Such choice made it possible to obtain good results with a sufficiently small number of initial data and a small amount of computation. We also note, that if only half of the variants of the fiber input excitation was utilized, the accuracy of the reconstruction degrades only slightly. However, the choice of several fiber lengths at which intensity patterns are measured is important for successful reconstruction. If all the lengths are too close (their spread is less than one beat length), then the accuracy of the method degrades significantly.

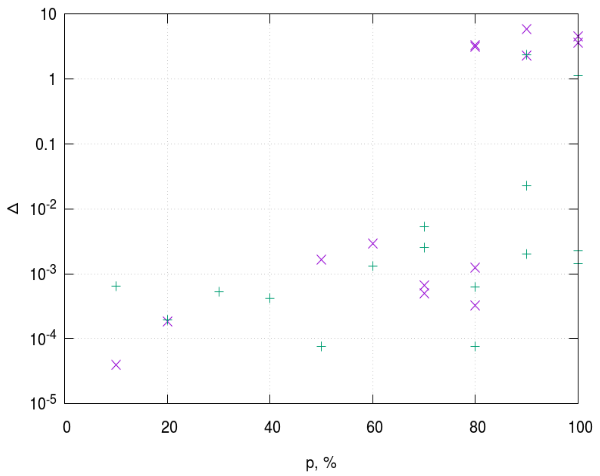


Figure 5: Dependence of the minimization error Δ on the parameter *p* for 7-core fiber (x) and 4-core fiber (+).

It is also worth noting that our method can be successfully applied even if some individual measurements are missing, since the method considers each measurement independently of the others.

5. CONCLUSION

In conclusion, we proposed a new method for reconstruction of supermodes structure and propagation constants in MCF with strong core coupling based on the intensity-only measurements. The method was successfully applied to 7-core MCF and allowed us to retrieve propagation constants of supermodes and modes of individual cores with high accuracy.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation, project #16-12-10472.

REFERENCES

Brownlee J. (2011). Clever Algorithms: Nature-Inspired Programming Recipes (1st ed.). Jason Brownlee.

Chan F.Y., Lau A.P.T., Tam H.Y., 2012. Mode coupling dynamics and communication strategies for multi-core fiber systems. Optics Express, 20, 4548.

Koshiba M., Saitoh K., Takenaga K. and Matsuo S., 2011. Multi-core fiber design and analysis: coupled-mode theory and coupled-power theory, Optics Express, 19, B102.

Mosley P.J., Gris-Sánchez I., Stone J.M., Francis-Jones R.J., Ashton D.J. and Birks T.A., 2014. Characterizing the variation of propagation constants in multicore fiber. Optics Express, *22*, 25689.

Nicholson J.W., Yablon A.D., Fini J.M. and Mermelstein M.D., (2009). Measuring the modal content of large-mode-area fibers. IEEE Journal of Selected Topics in Quantum Electronics, 15, 61.

Ramirez L.P., Hanna M., Bouwmans G., El Hamzaoui H., Bouazaoui M., Labat D., Delplace K., Pouysegur J., Guichard F., Rigaud P., KermèneV., 2015. Coherent beam combining with an ultrafast multicore Yb-doped fiber amplifier. Optics Express, 23, 5406.

Russell S., Norvig P. (2010). Artificial Intelligence: A Modern Approach (3rd ed.). New Jersey: Prentice Hall.

Saitoh K., and Shoichiro M., 2016. Multicore fiber technology. Journal of Lightwave Technology, 34, 55.

ELECTRO-OPTICAL COMPOSITE POLYMER LAYERS FOR HIGH SPEED RADIOPHOTONICS MODULATORS, EXPERIMENT AND MODELING

Maxim M. Nazarov[a,b], Viktor I. Sokolov[a], Eugeny V.Polunin[c], Yulia E. Pogodina[d]

[a] Crystallography and Photonics Federal Research Center, Russian Academy of Sciences
[b] Kurchatov Institute National Research Center, pl. akad. Kurchatova 1, Moscow 123182, Russia
[c],[d] Institute of General Chemistry, Russian Academy of Sciences, Moscow, Russia

[a]Visokol@rambler.ru, [b]Nazarovmax@mail.ru, [c]Polunin-507@yandex.ru, [d]Yulia_jwr@rambler.ru

ABSTRACT
The properties of polymer materials with embedded electro-optical (EO) chromophores capable for refractive index modulation under the applied electric field are studied. Thin film waveguide structure are manufactured based on SU-8, PMMA and new synthesized fluorine polymers and a number of chromophores like DR1, DR13 and specially synthesized ones. To measure the electro-optic response inside modulator prototype we produced metalized sandwich structures from those composites, polled chomophores by corona or contact polling and measure induced refraction changes Δn = 0.002 with a prism-coupling method. Special modeling algorithm to precisely determine optical characteristics of each of six layers from prism-coupler data is developed. Adapted fitting procedure allows to extract electro-optics coefficient values of inner layers. Under the influence of intense short wavelength radiation this material refractive index decreases on Δn = 0.01 - 0.03, that allows one to create a waveguide with a photo mask method or by direct laser writing.

Keywords: electro-optical polymer, refraction, prism-coupler, layer reflection modeling, waveguide.

1. INTRODUCTION

The efforts of the researchers are directed to the synthesis of new chromophores possessing a large EO coefficient in the telecommunication (1.5 μm wavelength) C-band, an increase of the concentration of chromophores in the polymer matrix in the absence of agglomeration, a reduce of the applied voltage, and a better temporal and temperature stability of the EO materials. The latest achievements in this direction are presented in the reviews of Dalton, Sullivan and Bale (2010); Kajzar, Lee and Jen (2003). This work is based on an interdisciplinary approach that includes chemical, optical and modeling directions: 1) Synthesis of new polymers and chromophores. 2) Optical methods for characterization and evaluation of electro-optical response. 3) Modeling and fitting prism-coupler data to obtain linear and nonlinear optical parameters of multilayer planar waveguide structure. The final our goal is the manufacturing of compact high-speed electro-optical modulators for the telecommunications C-band of wavelengths 1530-1565 nm.

New monomers based on halogenated acrylates with a record low absorption coefficient in the telecommunications C-band have been synthesized. Monomers have a high activity in the process of radical photopolymerization and are capable for the formation of polymer waveguides. New chromophores (not described earlier) have been synthesized, possessing a high electro-optical coefficient r_{33}. Composite polymeric materials have been created, both of the guest-host and of the side-chain type. Methods for measuring the EO of the coefficients r_{13} and r_{33} of polymeric composites, including second-harmonic generation and resonant excitation of waveguide modes using a prism-coupler have been developed. Composite films are oriented (poled) using a corona discharge or contact polarization, we obtained the EO coefficient in a planar waveguide r_{33} = 4-20 pm/V. The calculation algorithm has been developed that describes the reflection of TE and TM of polarized electromagnetic waves from a multilayer structure, including a polymeric buffer layer, a composite waveguiding layer, and semitransparent metal electrodes. The code takes into account the optical constants and thicknesses of all layers and allows obtaining exact value of r_{33} when compared with the experimental angular spectra. A new method for the formation of Mach-Zehnder waveguide interferometers is proposed, based on the effect of photodestruction of a composite under the action of intense visible radiation – Sokolov et.all. (2017).

2. METHODS AND RESULTS

With the introduction of EO chromophores in the polymer matrix and polling them it is possible to control the refraction of light in such a composite medium under the external electric field. This effect is used in the manufacturing of high-speed integrated optical modulators based on EO composites. As the first step we produced a nonlinear optical film of polymethylmethacrylate (PMMA) or SU8 with embedded chromophore DR13 into it. By side-chain method we were able to increase chromophores concentration in transparent polymer to 25%.

The obtained absorption characteristics on Figure 1 are consistent with the literature data, EO chromophores like DR1 and DR13 has a strong absorption line with a center near 510 nm and bandwidth of 100 nm. Note that, depending on the type of matrix (solvent, polymer, etc.) into which the chromophores are embedded, their absorption peak can be shifted by 10-20 nm toward larger or shorter wavelengths.

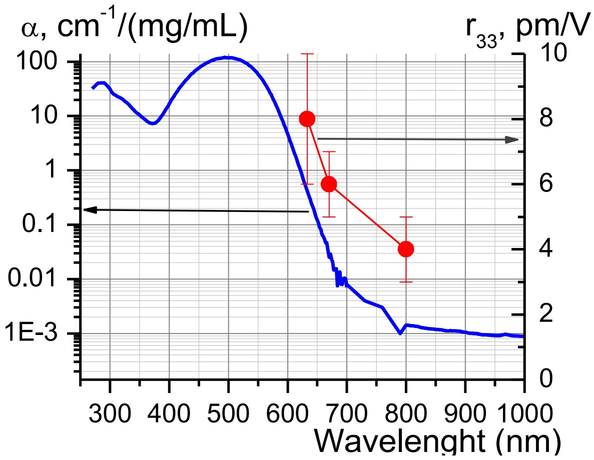


Figure 1. The absorption spectrum of the chromophore DR13 (solution in chlorobenzene) normalized to the concentration of the chromophore is the left axis. The right axis is the measured EO coefficient r_{33} for three wavelengths.

P polymer films were deposited on a substrate by spin-coating method. After forming thin film and drying it at 60° C for several hours, the solvent was evaporated from the polymer material. Films of PMMA/DR13 composite with thickness h_1 = 4-5 μm were deposited on quartz (SiO_2) and silicon (Si) substrates with thermally grown oxide layer of SiO_2. That is sample S1. Sample S2 was formed on a quartz substrate with a copper layer previously deposited on it with a thickness h_0 = 1 μm (lower electrode). For the mechanical and electrical protection of the PMMA/DR13 film, a layer of photoresist SU-8 with thickness h_2 was applied to it, see Fig. 2. Sample S3 was produced on a low-resistance silicon (LR Si) plate with a thermally grown oxide layer (h_1) 3.5 μm thick, onto which the PMMA/DR13 film (h_2) was deposited.

With the method of the prism coupling adapted for a metallized film structure we studied optical and electro-optical properties of the polymer composite. To measure the electro-optic response we polled chomophores by a corona or contact polling method and analyzed the induced refraction changes with original prism-coupling modeling method

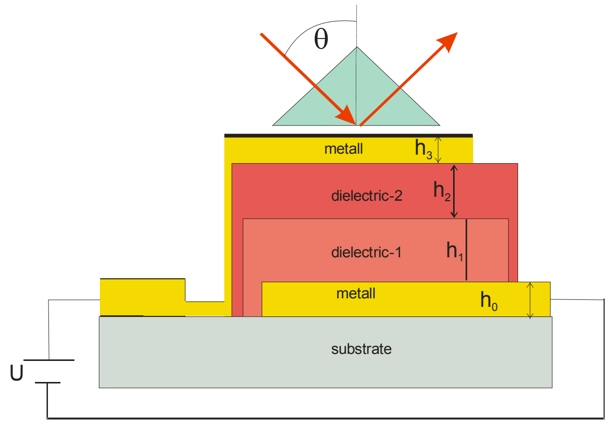


Figure 2: A sketch of a planar multilayer waveguide and coupling prism for measuring reflection angular spectra.

In the reflection angular spectra several quasi-waveguide modes are excited in a stack of thin films. One of the films contains oriented chromospheres, the upper metal film is semitransparent, conducting Cu layer of 30 nm. A sequence of dark resonances (m-lines) - Horsthuis and Krijnen (1989) - is measured in this structure when incidence angle θ is scanned for S or P-polarization.

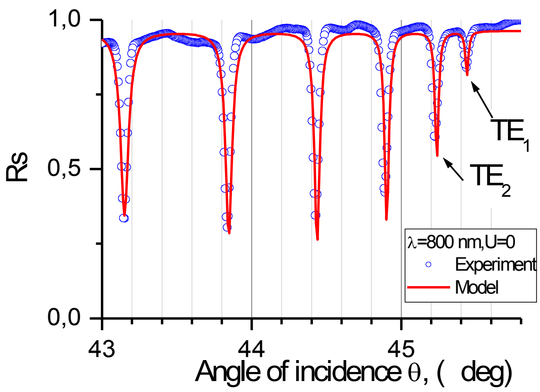


Figure 3: The Angular Reflection Spectrum Rs(θ) of six layers sample in the case of TE (S) polarization of the incident radiation. The solid line is the modeling, the circles are the experiment. The wavelength is λ = 800 nm. Sample S1

The algorithm for calculating the coefficients of reflection of TE(Rs) and TM(Rp) polarized laser beams from multilayered anisotropic structure (see Fig. 2 and 3) is developed. It allows to determine the characteristics of all films in a multilayer structure.

Besides the thicknesses of the layers h_1, h_2, h_3 and their permittivities ε_1, ε_2, ε_3, on Rs and Rp influences the thickness of the air gap H_i between the sample and the prism and the refractive index N_p. The numerical program (written in FORTRAN) takes into account the wavelength λ, complex refraction and thickness h_i of each layer, takes into account the anisotropy of the composite polymer film, characterized by the values of n_o and n_e. Expressions for the reflection coefficients have the form

$$R_s(\theta)=\left|\left(\frac{k_z-i\gamma^{II}}{k_z+i\gamma^{II}}+A_s e^{-2\gamma^{II}H_i}\right)\Big/\left(1+\frac{k_z-i\gamma^{II}}{k_z+i\gamma^{II}}A_s e^{-2\gamma^{II}H_i}\right)\right|^2 \quad (1)$$

$$R_p(\theta)=\left|\left(\frac{k_z-i\varepsilon_p\gamma^{II}}{k_z+i\varepsilon_p\gamma^{II}}+A_p e^{-2\gamma^{II}H_i}\right)\Big/\left(1+\frac{k_z-i\varepsilon_p\gamma^{II}}{k_z+i\varepsilon_p\gamma^{II}}A_p e^{-2\gamma^{II}H_i}\right)\right|^2 \quad (2)$$

Where $k = 2\pi/\lambda$ is the wave vector in vacuum, $k_z = kN_p\cos(\theta)$ is the projection of the wave vector of the electromagnetic wave in the prism (see Fig.2), $\gamma^{II}=k\sqrt{(N_p\sin\theta)^2-1}$, is z- projection of the wave vector in the gap between the measuring prism and the sample, $\varepsilon_p = N_p^2$ the dielectric constant of the prism. Angle dependent complex parameters A_s and A_p determine the reflections of the light beam in the layers of the thin-film structure.

$$A_s=\left(\frac{\gamma^{II}-\gamma^{III}}{\gamma^{II}+\gamma^{III}}+B_s e^{-2\gamma^{III}H_m}\right)\Big/\left(1+\frac{\gamma^{II}-\gamma^{III}}{\gamma^{II}+\gamma^{III}}B_s e^{-2\gamma^{III}H_m}\right),$$

$$A_p=\left(\frac{\varepsilon_m\gamma^{II}-\gamma^{III}}{\varepsilon_m\gamma^{II}+\gamma^{III}}+B_p e^{-2\gamma^{III}H_m}\right)\Big/\left(1+\frac{\varepsilon_m\gamma^{II}-\gamma^{III}}{\varepsilon_m\gamma^{II}+\gamma^{III}}B_p e^{-2\gamma^{III}H_m}\right),$$

where H_m is the thickness of the upper semitransparent metal electrode,

$$B_s=\left(\frac{\gamma^{III}-\gamma^{IV}}{\gamma^{III}+\gamma^{IV}}+C_s e^{-2\gamma^{IV}H_1}\right)\Big/\left(1+\frac{\gamma^{III}-\gamma^{IV}}{\gamma^{III}+\gamma^{IV}}C_s e^{-2\gamma^{IV}H_1}\right),$$

$$B_p=\left(\frac{\varepsilon_{1y}\gamma^{III}-\varepsilon_m\gamma^{IV}}{\varepsilon_{1y}\gamma^{III}+\varepsilon_m\gamma^{IV}}+C_p e^{-2\gamma^{IV}H_1}\right)\Big/\left(1+\frac{\varepsilon_{1y}\gamma^{III}-\varepsilon_m\gamma^{IV}}{\varepsilon_{1y}\gamma^{III}+\varepsilon_m\gamma^{IV}}C_p e^{-2\gamma^{IV}H_1}\right),$$

where H_1 is the thickness of the upper anisotropic polymer layer,

$$C_s=\left(\frac{\gamma^{IV}-\gamma^{V}}{\gamma^{IV}+\gamma^{V}}+\frac{\gamma^{V}-\gamma^{VI}}{\gamma^{V}+\gamma^{VI}}e^{-2\gamma^{V}H_2}\right)\Big/\left(1+\frac{\gamma^{IV}-\gamma^{V}}{\gamma^{IV}+\gamma^{V}}\frac{\gamma^{V}-\gamma^{VI}}{\gamma^{V}+\gamma^{VI}}e^{-2\gamma^{V}H_2}\right),$$

$$C_p=\left(\frac{\varepsilon_{2y}\gamma^{IV}-\varepsilon_{1y}\gamma^{V}}{\varepsilon_{2y}\gamma^{IV}+\varepsilon_{1y}\gamma^{V}}+\frac{\varepsilon_s\gamma^{V}-\varepsilon_{2y}\gamma^{VI}}{\varepsilon_s\gamma^{V}+\varepsilon_{2y}\gamma^{VI}}e^{-2\gamma^{V}H_2}\right)\Big/\left(1+\frac{\varepsilon_{2y}\gamma^{IV}-\varepsilon_{1y}\gamma^{V}}{\varepsilon_{2y}\gamma^{IV}+\varepsilon_{1y}\gamma^{V}}\frac{\varepsilon_s\gamma^{V}-\varepsilon_{2y}\gamma^{VI}}{\varepsilon_s\gamma^{V}+\varepsilon_{2y}\gamma^{VI}}e^{-2\gamma^{V}H_2}\right),$$

where H_2 – is the thickness of the lower anisotropic polymer layer. Here, $\varepsilon_1 = \varepsilon_{1x}x + \varepsilon_{1y}y + \varepsilon_{1z}z$, $\varepsilon_2 = \varepsilon_{2x}x + \varepsilon_{2y}y + \varepsilon_{2z}z$ are the permittivity tensors of polymer films 1 and 2; $\gamma^{III}=k\sqrt{(N_p\sin\theta)^2-\varepsilon_m}$, $\gamma^{VI}=k\sqrt{(N_p\sin\theta)^2-\varepsilon_s}$, ε_s is the permittivity of the substrate. In the case of TE polarization of the incident light wave $\gamma^{IV}=k\sqrt{(N_p\sin\theta)^2-\varepsilon_{1x}}$, $\gamma^{V}=k\sqrt{(N_p\sin\theta)^2-\varepsilon_{2x}}$, and In the case of TM polarization $\gamma^{IV}=k\sqrt{(N_p\sin\theta)^2\varepsilon_{1y}/\varepsilon_{1z}-\varepsilon_{1y}}$, $\gamma^{V}=k\sqrt{(N_p\sin\theta)^2\varepsilon_{2y}/\varepsilon_{2z}-\varepsilon_{2y}}$.

We note that in z-projections of wave vector γ^{IV}, γ^{V} in layers h_1 and h_2 imaginary parts of the corresponding refractive index (extinction coefficient) m_i are introdused $\varepsilon_i=(n_i+im_i)^2$. Namely m_i determines the width of the observed m-lines.

When an external electric field E is applied to a composite film with aligned EO chromophores, the m-lines are shifted due to a change in the refraction of the film. This shift depends on the electro-optical coefficient r_{33} for TE polarization and on r_{13} for TE polarization. The sensitivity of the method to a small refraction changes is limited by the width of the m-lines, and accordingly by absorption. Moving to long wavelengths (like 800 nm), where absorption (and m-lines width) is small allows more resize fitting.

Comparison with original analytical model of six-layer structure attenuated total reflection (see Fig. 2 and 4) allows us to determine precisely all optical characteristics. For the example, for the sample used at Fig. 3 determined optical parameters of EO layer are the following: h_2=4.5±0.02 μm is the thicness, n_2=1.5421±0.0005 is the refraction, m_2=(25±1)E-4 is extinction coefficient determining absorption, all that for λ = 800 nm.

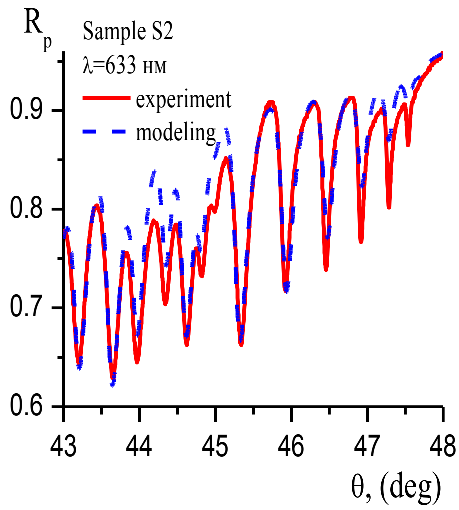


Figure 4. The measured and calculated angular reflection spectrum of Rp(θ) at a wavelength of 633 nm for sample S2.

Fig. 4 shows the measured and calculated dependences of the reflection coefficient Rp(θ) (TM polarization, λ = 633 nm) for sample S2 after alignment of chromophores in the PMMA / DR13 film. The structure parameters were determined from the measured positions of the m-lines after each subsequent layer was deposited. After deposition of all the layers, these parameters were refined by varying them for the best match of the measured and calculated angular spectra Rp(θ).
Sample S2 does not support true waveguide modes in the EO polymer layer h_1, since the refractive index of this layer is lower than that of layer h_2. Nevertheless, the reflection at the boundaries of the layers h_1 and h_2 allows the radiation to propagate in the layer h_1, although with high radiation losses (leaky modes). In Fig. 4 true guided modes excited at high angles of

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

incidence (θ = 48 - 45.5 °) they refer to the layer h_2 of SU-8 with DR13 polymer, and the resulting modes at θ <45° refer to the polled layer h_1. The most sensitive to the change in the refractive index are the low-order m-lines.

When analyzing sample S2 at λ = 633 nm, the following thickness values h_j and refractions n_j of layers j numbered, as in Fig. 2: n_3 = 0.16+i*3.37, h_3 = 40 nm, h_2 = 4.80 μm, n_2 = 1.5965 + i*0.002; h_1 = 4.75 μm, $n_{1,o}$ = 1.522 + i*0.002, $n_{1,e}$ = 1.532 + i*0.002, $n_0 = n_3$.

When an external electric field *E* is applied to a composite film with polled EO chromophores, the m-lines are shifted on *Δθ* due to a change in the refraction of the film *δn* see Fig.5.

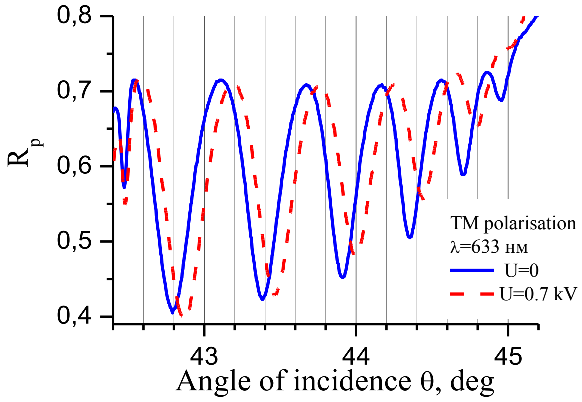


Figure 5: The measured Angular Reflection Spectrum for λ = 633 nm, with and without applied voltage. TM (P) polarization. Sample S3

The largest angular shift of m-lines is observed for TM polarization due to a larger change in the refractive index n_e

$$\Delta\theta_{TM} \propto \delta n_e = -0.5 n_e^3 r_{33} E$$

Using the calculations, we can compare the magnitude of the angular shift of a particular m-line and the change in refraction that leads to such a shift. For example, for λ = 633 nm and for the TM_3 mode in the PMMA / DR13 layer (θ = 44.6 ° in Figures 4), the refraction of change of n_e to 0.0006 results in a shift of the mode angle θ by 0.017°, hence $\Delta n_e = \Delta\theta_{TM}/29$. The resolution of the prism coupler Metricon2010 is Δθ = 0.01°, thus the accuracy of determining the refractive index at low absorption reaches Δn = 0.0001. Thus, the induced changes in the refractive index in the fourth digit can be registered.

The change in the refractive index due to the EO effect is related to r_{ij} and the applied electric field $E_z = U/(h_1 + h_2)$. Let us estimate the value of r_{33}, which leads to the observed variation Δn_e = 0.006 when the control voltage U = 0.3 kV is applied to sample S2. With the total thickness of polymer composite films $h_1 + h_2$ = 9.5 μm, we obtain r_{33} = 8 pm/V. As the distance from the EO resonance decreases, the coefficient decreases (see Figure 1).

At 800 nm, the film is transparent (see Fig. 1), there is already a nonresonance part of the EO coefficient. A further increase in the wavelength does not lead to a significant decrease in r_{33}. To characterize the coefficient EO by the method of prism coupler, the optimum wavelength is the wavelength corresponding to the edge of the absorption band, in the case of DR1 and DR13 dyes this is 670-700 nm. Moreover, the EO coefficient still has a noticeable resonant contribution, and m - the line are already narrow enough.

Using calculation algorithm we relate Δθ and δn and thus we found EO coefficients r_{33} and r_{13} composite PMMA/DR13 in multiple wavelengths (633, 800, 980 nm) near the absorption peak of chromophore (at 520 nm) and far from it. The proposed method allows us to measure the induced external electric field changes in the refractive index in the thin-film light-guiding structure of the real electro-optic modulator with an accuracy of ± 0.0003.

3. CONCLUSSION

Electro-optical response of perspective chromophores embedded into polymer waveguide structure is studied. Modeling is based on analytical solution of Fresnel formulas for six layer system under prism coupling conditions for guided modes excitation. Complex transition and multiple reflections in each layer are taken into account. Adapted fitting procedure to the experimental data allows obtaining optical linear (refraction, absorption, thickness) and nonlinear (EO Pockels effect) properties of each layer.

The sensitivity of the method makes it possible to measure the EO of the coefficients r_{33} and r_{13} for the TE and TM of the polarization of the incident light beam from 2 pm/V, the sensitivity is limited by the breakdown voltage of the metallized structure. The method makes it possible to determine the refractive indices of several dielectric layers (even those covered by a translucent metal film) with an accuracy of ±0.0005, and their thickness with an accuracy of ±0.05 μm.

The advantage of the proposed method in comparison with the known approaches – Nahata and Shan (1993) is that it can be used to measure the characteristics of real polymer EO modulators with a multilayer structure and external electrical contacts for different wavelengths.

The results of the research can be used to create active integrated-optical devices of radio-photonics: waveguide amplifiers, high-speed optical switches and modulators for the telecommunications C-band of the spectrum. The design of integrated - optical modulator for radio-photonics is suggested

ACKNOWLEDGMENTS
The work is supported by Russian foundation of basic research (RFBR) grants № 14-29-08265, 17-07-01478.

REFERENCES

Dalton L.R., Sullivan P.A., Bale D.H., 2010. Electric Field Poled Organic Electro-optic Materials: State of the Art and Future Prospects. Chemical Review, 110, 25-55.

Horsthuis W.H.G., Krijnen G.J.M., 1989. Simple measuring method for electro-optic coefficients in poled polymer waveguides. Applied Physics Letters, 55, 616.

Kajzar F., Lee K.-S., Jen A. K-Y., 2003. Polymeric Materials and their Orientation Techniques for Second-Order Nonlinear Optics. Advances in Polymer Science, 161. © Springer-Verlag Berlin Heidelberg.

Nahata A., Shan J., Yardley J.T., Wu Ch., 1993. Electro-optic determination of the nonlinear-optical properties of a covalently functionalized Disperse Red 1 copolymer. Journal of Optical Society of America B, 10, 1553-1564.

Sokolov V.I., Akhmanov A.S., Asharchuk I.M., Goryachuk I.O., Khaydukov K.V., Nazarov M.M. 2017. Formation of Channel Optical Waveguides in Polymethylmethacrylate with Embedded Electro-Optic Chromophore DR13 by the Photoinduced Bleaching Method. Optics and Spectroscopy, 122, 469-74.

AUTHORS BIOGRAPHY

Maxim M. Nazarov received the Ph.D. degree in physics from the M.V. Lomonosov Moscow State University in 2002 on the topic of surface plasmon enhanced optical harmonics on metal grating. He was scientific researcher at MSU till 2012 on the topics of femtosecond lasers and THz spectroscopy. In 2012 he is a scientific researcher in ILIT of Russian Academy of Sciences (RAS) on the topic of polymer waveguides for THz and visible ranges, nanoparticles and integrated optics. In 2017 he is the head of laboratory of ultraintense laser pulses in National Research Center "Kurchatov Institute", Moscow, Russia. He is the author of more then 60 publications in a peer-reviewed scientific journals.

Victor I. Sokolov received the Ph.D. in Physics and Mathematics in 1991 Moscow State University (Moscow, Russia). He was Scientist, Senior Research Scientist, Head of the laboratory at Institute on Laser and Information Technologies of the Russian Academy of Sciences (1991-2016). In 2016 he is Head of the Institute on Photon Technologies of Federal Scientific Research Centre «Crystallography and Photonics» Russian Academy of Sciences. He is the author of more, then 100 publications in the field of interaction of laser radiation with matter, photonics, integrated optics, polymeric materials, scientific instrumentation.

Eugene V. Polunin - in 1973 he graduated from the chemical faculty of Moscow State University, in 1983 he defended his thesis on stereospecific synthesis of linear isoprenoids at the Institute of General Chemistry of the Russian Academy of Sciences. From 1983 to 1995 he worked and taught at the chemistry department of Moscow State University, then returned to IGCh RAS. Author and co-author of about 60 publications in the field of synthesis of biologically active compounds and polymer chemistry. Co-author of 4 inventions.

Yulia E. Pogodina graduated from the Moscow state university of fine chemical technologies in 2001. Now she is engineer at the Institute of General Chemistry, RAS. Specialization: synthesis of spacer groups on the basis of amino acids for porphyrins, development of medicines for phototherapy of cancer.

A METHOD FOR FACTOR SCREENING OF SIMULATION EXPERIMENTS BASED ON ASSOCIATION RULE MINING

Lingyun Lu[a], Wei Li[b], Ping Ma[c], Ming Yang[d]

Control and Simulation Center, Harbin Institute of Technology, Harbin 150080, P.R. China

[a]jyzluhit@gmail.com, [b]frank@hit.edu.cn, [c]pingma@hit.edu.cn, [d]myang@hit.edu.cn

ABSTRACT
Since complex simulation models contain a large number of inputs and parameters, it is meaningful to identify the most important ones for further studies. The traditional procedure of factor screening is to design simulation experiment schemes and generate simulation output data first. Furthermore, due to rigorous hypothesis and fewer levels for each factor, the existing screening methods are usually not appropriate for complex simulation. In this paper, a new method for factor screening of simulation experiments is proposed. It is based on data mining technique and existing simulation data. The relationships between factors and the output are discovered from the sets of simulation data using the association rule mining algorithm firstly. Then the important factors can be identified by synthesizing the association rules. Finally, the proposed method is verified by the test function.

Keywords: simulation experiments, factor screening, association rules, quantitative attributes

1. INTRODUCTION

Complex simulation models consist of a large number of inputs and parameters which are generally referred to as factors in design of experiments (DOE). DOE is essential for doing certain analyses include validation, uncertainty analysis, sensitivity analysis, optimization, etc (Kleijnen 2008, Li 2016). However, it is usually prohibitive or impractical to do these analyses with the large number of factors involved, especially for computer simulation models with high computational cost. An important question is – which factors are really significant when there are potentially a large number factors involved (Alam 2004)? The Pareto principle or 20-80 rule implies that only a few factors are really important (Wright 2010). In order to improve efficiency, it is reasonable to screen a small number of factors before analyzed. Factor screening is performed to eliminate unimportant factors so that the remaining important factors can be more thoroughly studied in later experiments, regardless of the application domain of area, including national defense, logistics, industry, healthcare, etc (Li 2016, Bruzzone 2014, Padhi 2013, Rantanen 2015).

Several types of factor screening methods have been developed to identify important factors with a limited set of simulation experiments. The most common ones are fractional factorial, central composite, and Plackett–Burman designs (Myers 2002). However, these screening designs developed for physical experiments are not appropriate for simulation physical experiments. Procedures developed for stochastic simulation experiments include Morris' randomized one-factor-at-a-time design (Morris 1991), frequency domain method (Morrice 1995), edge designs (Elster 1995), iterated fractional factorial designs (Campolongo 2000) and the Trocine screening procedure (Trocine 2001). Some methods developed for discrete-event simulation experiments include sequential bifurcation (SB), SB with interactions (SB-X), Cheng's method, controlled sequential bifurcation (CSB), CSB with interactions (CSB-X), controlled sequential factorial design (CSFD), fractional factorial controlled sequential bifurcation (FFCSB), etc (Wan 2007, Shen 2009). However, the majority of these methods have many assumptions or limitations, such as monotonicity of the function, the smoothness of the inputs/outputs function, etc. These methods are efficient and effective when their assumptions are satisfied.

From the factor screening methods described above, it can be concluded that these methods easily neglect some important factors due to rigorous hypothesis and few factor levels. It is possibly unsuitable for the complex simulation models. As is known to all, the importance of factors depends on the experimental domain. The effective way is to select levels as many as possible during factor screening which can represent simulation model more precisely and improve the accuracy of the classification. So the data mining algorithm is adopted to solve this problem, which can be based on the existing simulation data and regardless of mathematics assumptions.

In this paper, the association rule mining algorithm is investigated as a potential factor screening method. It should be based on simulation data, including simulation inputs/parameters and the output. The remainder of the paper is organized as follows. In Section 2, the procedure of the Apriori algorithm and the algorithm of partitioning quantitative attributes are described. In Section 3, a method for factor screening of simulation experiments is proposed. Through this method, important factors can be identified by synthesizing the association rules. In Section 4, the proposed method is illustrated and verified by the test function. Finally, the paper is concluded in Section 5.

2. RELATED WORK

The objective of factor screening is to identify the important factors among many factors involved. In this section, the procedure of the Apriori algorithm will be described. However, this algorithm can only cope with boolean attributes instead of quantitative attributes. But quantitative attributes usually exist in simulation models. So the algorithm of partitioning quantitative attributes is adopted.

2.1. Apriori Algorithm

The Apriori algorithm is proposed by Agrawal (1994) which is used to find potential relationships between the items (Data_Attributes). The result of the algorithm will be expressed in the form of association rules. In order to clarify the Apriori algorithm clearly, some notations is illustrated in the following.

- *k*-itemset: An itemset having k items.
- L_k : Set of large *k*-itemsets with minimum support.
- C_k : Set of candidate *k*-itemsets which has potentially large itemsets.
- D : Set of transactions where each transaction T is a set of items, i.e., $T \subseteq D$.

Given a set of transactions D , the use of the Apriori algorithm is to generate all association rules which the support and confidence are larger than the given minimum values. The minimum support and minmum confidence are referred to as *min_sup* and *min_conf* respectively. The procedure of the Apriori algorithm is described as follows:

Apriori algorithm

```
L_1 = {large 1-itemsets};
For ( k = 2 ; L_{k-1} ≠ ∅ ; k++ ) do begin
    C_k = apriori-gen( L_{k-1} );
    For all transactions t ∈ D do begin
        C_t = subset( C_k , t );
        For all candidates c ∈ C_t do
            c.count++ ;
        end
    end
    L_k = { c ∈ C_k | c.count ≥ min_sup };
end
return L = ∪_k L_k .
```

At the beginning, the large 1-itemsets is determined by simply counts item occurrences. Then subsequent pass k consists of two phases. One is the candidate itemsets C_k which is generated by the apriori-gen function; and the other is the support of candidates in C_k which is counted by scanning the database.

The apriori-gen function described above returns a superset of the set of all large *k*-itemsets. And the subset function returns all the candidates contained in a transaction t with some rules. The more details of these two functions can be referred to the literature Agrawal (1994).

2.2. The Algorithm of Partitioning Quantitative Attributes

However, quantitative attributes need to be discretized when using the Apriori algorithm. Two problems may be occurred after discretization (Srikant 1996).

- ***min_sup***. If the number of intervals for a quantitative attribute (or values, if the attribute is not partitioned) is large, the support for any single interval can be low. Hence, without using larger intervals, some rules involving this attribute may not be found because they lack minimum support.
- ***min_conf***. There is some information lost whenever we partition values into intervals. Some rules may have minimum confidence only when an item in the antecedent consists of a single value (or a small interval). This information loss increases as the interval sizes become larger.

In order to solve problems described above, the algorithm of the equi-depth partitioning (Fukuda 1999) is adopted, which can divide N data into M buckets almost evenly. The procedure of the algorithm is described as follows.

- **Step 1**: Make an *S*-sized random sample from N data.
- **Step 2**: Sort the sample in $O(S \log S)$ time.
- **Step 3**: Scan the sorted sample and set $i(S/M)$ -th the smallest sample to p_i for each $i = 1, \ldots, M-1$. Let p_0 be $-\infty$ and p_M be $+\infty$.
- **Step 4**: For each tuple x in the original N data, find i such that $p_{i-1} < x \le p_i$ and assign x to the *i*-th bucket. This check can be done in $O(\log M)$ time by using the binary search tree for the buckets. Thus, for all i , the size u_i of B_i can be computed in $O(N \log M)$ time.

3. THE METHOD FOR FACTOR SCREENING OF SIMULATION EXPERIMENTS

To identify important factors from many factors involved in simulation models, the Apriori algorithm described in Section 2 is adopted. However, in many cases, factors used in simulation are quantitative attributes, which the Apriori algorithm cannot cope with them. So data preprocessing is used to partition the domain of quantitative attributes into several intervals and mapped these intervals into discrete attributes. And the algorithm of the equi-depth partitioning is adopted to pruning the search space. What need to be pointed out that calculation of strong rules is not interesting while using support-confident framework, lift for mining positive non-redundant association rules is adopted here.

It is assume that itemsets A and B are denote the premise and consequent of association rules, respectively, where $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B = \emptyset$. Here $A \Rightarrow B$ is used to represent the relationship between

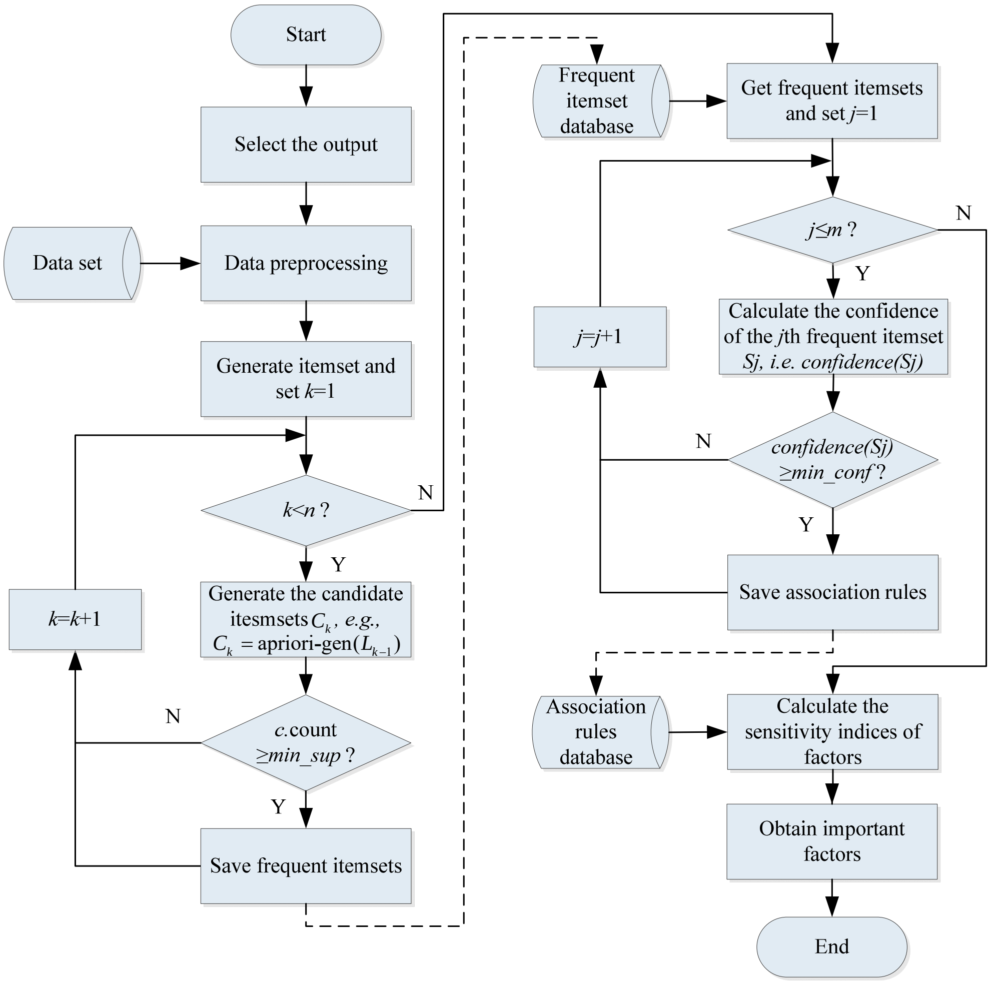


Figure 1: Flow of Factor Screening Based on the Improved Apriori Algorithm

factors and the output. Then the confidence of $A \Rightarrow B$ can be calculated as

$$confidence(A \Rightarrow B) = P(B \mid A) = \frac{support(A \cup B)}{support(A)} \quad (1)$$

where $support(A)$ is the support of A.

The flow of factor screening based on the improved Apriori algorithm is shown in Figure 1. The procedure of the proposed method mainly consists of four steps:

- **Step 1**: Data preprocessing consists of partition quantitative attributes into several intervals and discretization.
- **Step 2**: Obtain the set of all frequent itemsets by finding all itemsets where the support of the corresponding itemsets satisfies *min_sup*.
- **Step 3**: Generate association rules using frequent iemsets where the confidence of the corresponding frequent itemsets satisfies *min_conf*.
- **Step 4**: Calculate the sensitivity indices of factors by Equation (2), and eventually obtain the important factors by comparing the values of sensitivity indices.

The equation of calculating sensitivity indices is given as

$$\begin{aligned} r_i = {} & \max(\text{confidence}(x_i \Rightarrow y)) - \\ & \min(\text{confidence}(x_i \Rightarrow y)) \end{aligned} \quad (2)$$

$$S_i = \frac{r_i}{\sum r_i}$$

where x_i is factor and y is the output.

4. CASE STUDY

The method proposed in this paper can be used for factor screening of simulation experiment. In this section, a test function is used to demonstrate the effectiveness of the proposed method.

Consider the test function

$$y = 11x_1 + 9x_2 + 2x_1x_2 + 15x_3 + 3x_4 - 2x_5 + 16x_6 + x_7 \quad (3)$$

where $x_1, x_2, \ldots, x_7$ is an i.i.d. sample from $\boldsymbol{U}(0,1)$.

The set of data $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_7, \boldsymbol{y}\}$ is obtained first, where the sample size is 5000. Then each factor x_i is partitioned into seven intervals through the algorithm of the equi-depth partitioning. Each interval of x_i and y are discretized in the form of $\{x_i\_1, x_i\_2, \ldots, x_i\_7\}$ and $\{y\_1, y\_2, \ldots, y\_7\}$, respectively. The association

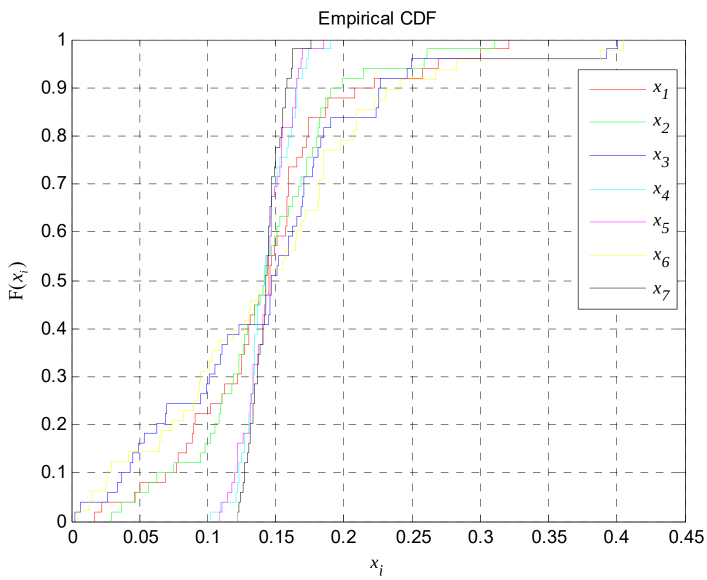


Figure 2: An Ensemble of 7 CDFs of the Factors

rules are generated by the Apriori algorithm, as shown in Table 1. Only 2-itemsets is taken to be considered, i.e., A only contains one factor in each rule. According to the fourth column of Table 1, the cumulative distribution function (CDF) for the confidence of each factor x_i is drawn in Figure 2. The relative importance of the factors is qualitatively expressed by the gradients of the curves, which the smaller one is more important. In other words, the range of confidence for each factor reflects the importance of factors. Furthermore, the sensitivity indices for each factor are calculated and shown in Figure 3. It can be quantitatively conclude that the order of factors is $x_6 > x_3 > x_1 > x_2 > x_4 > x_5 > x_7$ according to the importance. The important factors are x_6, x_3, x_1 and x_2. The experimental results obtained above are in agreement with the theoretical results which are easily concluded from Equation (3).

Table 1: List of the Association Rules

ID	A	B	$confidence(A \Rightarrow B)$
1	x_1_2	y_2	0.1737
2	x_1_2	y_5	0.1275
3	x_1_4	y_6	0.1569
137	x_3_1	y_5	0.0700
249	x_6_7	y_7	0.4050
342	x_7_3	y_4	0.1261

Above all, it is demonstrated that the proposed method in this paper is an effective method which can be used for factor screening. Compared with the traditional screening methods, the proposed method can identify important factors based on the existing simulation data and regardless of mathematics assumptions.

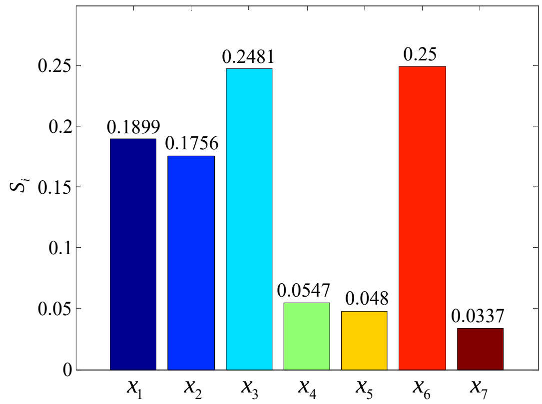


Figure 3: Histogram of Sensitivity Indices for Test Function Using the Proposed Method

5. CONCLUSIONS

Factor screening is very important while simulation models to be analyzed with many factors involved. In this paper, the use of association rule mining algorithm as a kind of factor screening method has been studied. The relationships between factors and the output are considered to be a measure of index to identify important factors from many factors involved. As a classic algorithm of association rule mining, Apriori algorithm can only be used to solve the problem of mining boolean association rules. So the algorithm of equi-depth partitioning is adopted to transform quantitative attributes into boolean attributes. After the association rules are generated, the results of confidence are used to calculate the sensitivity indices of factors. And the important factors are identified according to the sensitivity indices. It is demonstrated from the case study that the proposed method appears to perform well in factor screening.

To improve the efficiency of factor screening and insure its accuracy, two aspects are deserved to be further studies. One is to study how to synthesize the association rules in other ways, and the other is how to process the unbalanced data.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (Grant No. 61403097).

REFERENCES

Kleijnen J.P.C., 2008. Design and analysis of simulation experiments. New York:Springer.

Li W., Lu L.Y., Liu Z.Z. et al., 2016. HIT-SEDAES: An integrated software environment of simulation experiment design, analysis and evaluation. International Journal of Modeling, Simulation, and Scientific Computing, 7 (3), 1650027.

Alam F.M., McNaught K.R., Ringrose T.J., 2004. Using Morris' randomized OAT design as a factor screening method for developing simulation metamodels. Proceedings of the 2004 Winter Simulation Conference, pp. 949-957. December 5-8, Washington DC, United States.

Wright A., Bates D.W., 2010. Distribution of Problems, Medications and Lab Results in Electronic Health Records: The Pareto Principle at work. Applied Clinical Informatics, 1 (1), 32–37.

Sanchez S.M., Wan H., 2015. Work smarter, not harder: a tutorial on designing and conducting simulation experiments. Proceedings of the 2015 Winter Simulation Conference, pp. 1795-1809. December 6-9, Huntington Beach, United States.

Bruzzone, A., Longo, F., 2014. An application methodology for logistics and transportation scenarios analysis and comparison within the retail supply chain. European Journal of Industrial Engineering, 8 (1), 112–142.

Padhi S.S., Wagner S.M., Niranjan T.T. et al., 2013. A simulation-based methodology to analyse production line disruptions. International Journal of Production Research, 51 (6), 1885–1897.

Rantanen J., Khinast J., 2015. The future of pharmaceutical manufacturing sciences. Journal of Pharmaceutical Sciences, 104 (11), 3612–3638.

Myers, R.H., Montgomery, D.C., 2002. Response surface methodology: process and product optimization using designed experiments, second ed. John Wiley & Sons, New York.

Morris M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics, 33: 161–174.

Morrice, D.J., Bardhan, I.R., 1995. A weighted least squares approach to computer simulation factor screening. Operations Research, 43 (5), 792–806.

Elster, C., Neumaier, A., 1995. Screening by conference designs. Biometrika 82 (3), 589–602.

Campolongo, F., Kleijnen, J.P.C., Andres, T., 2000. Screening methods. In: Saltelli, A., Chan, K., Scott, E.M. (Eds.), Sensitivity Analysis. John Wiley & Sons, New York, pp. 65–89.

Trocine L., Malone L.C., 2001. An overview of newer advanced screening methods for the initial phase in an experimental design. Proceedings of the 2001 Winter Simulation Conference, pp. 173-176. December 9-12, Arlington, United States.

Wan H., Ankenman B., 2007. Two-stage Controlled Fractional Factorial Screening for Simulation Experiments. Journal of Quality Technology, 39 (2), 126–139.

Shen H., Wan H., 2009. Controlled sequential factorial design for simulation factor screening. European Journal of Operational Research, 198, 511–519.

Agrawal R., Srikant R., 1994. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487-499. September 12-15, Santiago, Chile.

Srikant R., Agrawal R., 1996. Mining quantitative association rules in large relational tables. Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data pp. 1-12. June 4-6, Montreal, Canada.

Fukuda T., Morimoto Y., et al., 1999. Mining optimized association rules for numeric attributes. Journal of Computer and System Sciences, 58 (1), 1–12.

Rastogi R., Shim K., 2002. Mining optimized association rules with categorical and numeric attributes. IEEE Transactions on Knowledge and Data Engineering, 14 (1), 29–50.

Ke Y.P., Cheng J., Ng W., 2008. An information-theoretic approach to quantitative association rule mining. Knowledge and Information Systems, 16 (2), 213–244.

AUTHORS BIOGRAPHY

LINGYUN LU is a Ph.D. student at Harbin Institute of Technology (HIT), and received the M.E. from HIT in 2013. His research interests include simulation experiment design and simulation data analysis. His email is jyzluhit@gmail.com.

WEI LI is an associate professor at HIT, and received the B.S., M.E. and Ph.D. from HIT in 2003, 2006 and 2009 respectively. His research interests include simulation evaluation, simulation data analysis, and distributed simulation. His email is frank@hit.edu.cn.

PING MA is a professor at HIT, and received Ph.D. from HIT in 2003. Her research interests include distributed simulation technique and VV&A. Her email is pingma@hit.edu.cn.

MING YANG is the corresponding author. He is a professor and the director of control and simulation center at HIT. Also he is the vice editor-in-chief for Journal of System Simulation and editor for International Journal of Modeling Simulation and Science Computing. His research interests include system simulation theory and VV&A. His email is myang@hit.edu.cn.

EVOLUTION STRATEGY - TESTING DIFFERENT TOURNAMENT SELECTION STRATEGIES

Pavel Raska[(a)], Zdenek Ulrych[(b)]

[(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

[(a)]praska@kpv.zcu.cz, [(b)] ulrychz@kpv.zcu.cz,

ABSTRACT

The paper deals with testing Evolution Strategy and different tournament selection strategies on testing functions and discrete event simulation model. We developed a simulation optimizer used for simulation optimization, for testing different settings of optimization algorithm parameters and evaluating the success of finding the optimum by the optimization algorithms and other evaluation criteria.

Keywords: evolution strategy selections, testing functions, discrete event simulation model

1. INTRODUCTION

The basic common problem of simulation optimization algorithms is to quickly find the global/local optimum of the objective function (function maximization can be converted to function minimization):

$$\breve{\mathbf{X}} = \operatorname{argmin}_{\mathbf{X}\in\tilde{X}} F(\mathbf{X}) = \{\breve{\mathbf{X}} \in \tilde{X} : F(\breve{\mathbf{X}}) \leq F(\mathbf{X}) \forall \mathbf{X} \in \tilde{X}\} \quad (1)$$

Where symbols denote:

- $\breve{\mathbf{X}}$ … Global minimum of the objective function
- $F(\mathbf{X})$ … Objective function value of candidate solution – the range includes real numbers, i. e. $F(\mathbf{X}) \subseteq \mathbb{R}$. Objective function represents the aim of simulation optimization
- $\tilde{X}$ … Search space

This optimum has to respect the specified constraints. We use Box constraint – search space is limited:

$$\tilde{X} = \prod_{j=1}^{n} \tilde{X}_j = \prod_{j=1}^{n} [a_j, b_j],\ a_j \leq b_j \quad (2)$$

Where symbols denote:

- j… j-th decision variable of the simulation model
- n … Dimension of the search space
- a_j … Lower bound of the interval of j-th decision variable
- b_j … Upper bound of the interval of j-th decision variable

The candidate solution represents the values of each decision variable of the simulation model. Some optimization algorithms (described as pseudo-pascal algorithms) need to access these values hence the element will be transformed into a list of values of decision variables (vector of point coordinate in the search space). Decision variables represents the axes in the search space. These axes are indexed from zero to $n-1$:

$$\mathbf{X}[j] = x_j \forall j : j = \{0,1,2,\dots,n-1\} \quad (3)$$

We have tested different optimization methods (Pseudo gradient – Hill Climbing; Local Search; Tabu Search; Downhill Simplex; Metaheuristic - Simulated Annealing, Evolution Strategy; Differential Evolution and Self Organizing Migrating Algorithm) to find the global optimum of the objective function of the discrete event simulation models. This paper focuses on using Evolution Strategy which is a very general method which can be used for different types of objective functions.

Evolution Strategy generates more than one candidate solution. Each candidate solution represents the individual in the list of the generated candidate solutions – population – in the context of Evolution algorithms.

These candidate solutions are processed in a different way. To distinguish these candidate solutions from each other, the generated candidate solutions are placed into a list. Each item can be accessed by the index in this list:

$$\mathbf{X}_i = S[i] \forall i : i = \{0,1,2,\dots,m-1\} \quad (4)$$

Where symbols denote:

- S … List of candidate solutions
- m … Length of the list S

Most optimization methods are very sensitive to setting their parameters. Hence we tested different settings of Evolution Strategy to reduce some incorrect settings of the optimization methods parameters. We also tested different selection strategies to find a suitable and effective strategy for the optimization method. Considering the time requirements of testing the optimization method efficiency of finding the optimum of discrete event simulation models we substituted the testing objective function of the simulation models with different testing functions – De Jong´s, Rosenbrock´s, Michalewicz´s and Ackley´s functions.

2. EVOLUTION STRATEGY

The foundations of the first evolution strategy were laid in the 1960s at the Technical University of Berlin by three students, namely Hans-Paul Schwefel (Schwefel, 1995), Ingo Rechenberg and Peter Bienert. Inspired by lectures about biological evolution, they aimed at developing a solution method based on principles of variation and selection. In its first version, a very simple evolution loop without any endogenous parameters was used. (Bäck, Foussette, and Krause, 2013)

We implemented the Evolution Strategy algorithm using Steady State Evolution in the simulation optimizer to test different types of selection. (Marik, Stepankova, and Lazansky, 2001, Miranda, 2008, Hynek, 2008, Tvrdik, 2004)

The optimization algorithm contains (Raška & Ulrych, 2015) the following parameters and functions:

- n … Search space dimension
- φ … Relative frequency of success
- $sum\varphi$ … Sum of relative frequencies of success
- A … List of lower boundaries for each decision variable (axes of the search space)
- B … List of upper boundaries for each decision variable (axes of the search space)
- j… j-th decision variable
- i… Individual´s index (order in population)
- σ … List of standard deviations for each axis (decision variable of the simulation model) of the search space. The standard deviation is affected by Rechenberg 1/5th-rule – line number 25 (Schwefel, 1995)
- ϑ … List of steps for each axis
- Mutate $_{\text{ES\_n}}$ … Process of individual mutation – line number 11 (algorithm of the mutation is shown in Figure 2. Algorithm parameters are described in the following text)
- m … Size of the population
- m_{MP} … Number of offspring
- q … Number of successes (the offspring is better than the parent) to be monitored
- k … Number of other contestants per tournament
- X_{Pop} … Population of individuals
- X_{Arch} … Archive of offspring
- ExtractOptimalSet … Extracting best elements from the population
- TerminationCriterion … Termination criterion of simulation optimization
- Length … Function providing the length of list
- DeleteListItem … Function returning a new list by removing the element at defined index from the list
- AddListItem … Function inserting one item at the end of a list
- AppendList … Function adding all the elements of a list to another list

```
1   begin
2     φ ← ( );
3     σ ← ( );
4     n ← min{Length(A),Length(B)};     //dimension of the search space
5     for j ← 0 to n − 1 do
6         σ ← AddListItem(σ, |(B[j] − A[j])/3|); //initial list of standard deviations
7     X_Pop ← CreatePop(m, A, B);       //initial population
8     while not TerminationCriterion( )  do begin
9         X_Arch ← ( );
10        for i ← 0 to m_MP − 1 do    begin
11            X ← Mutate_ES_n(X_Pop[i mod m], σ, P_1, P_2, A, B, ϑ);
    //mutation using normal distribution – the offspring of parent
12            if Length(φ) ≥ q then
    //if the relative frequency is too long (the list exceed specified length)
13                φ ← DeleteListItem(φ, 0);      //delete first item in the list - FIFO
14            if CF_F(X)(X, X_Pop[i mod m]) < 0 then
    // the offspring is better than the parent
15                φ ← AddListItem(φ, 1);
    //if the offspring is better than parent add 1 to the end of the list φ else 0
16            else φ ← AddListItem(φ, 0);
17            sumφ ← 0;
18            for i ← 0 to Length(φ) − 1 do
19                sumφ ← sumφ + φ[i];
20            if (sumφ/Length(φ)) < 0.2 then
21                for j ← 0 to n − 1 do begin
22                    σ[j] ← σ[j] * 0.82;
23                    if σ[j] < ϑ[j]/100 then σ[j] ← ϑ[j] ;
    //increase the standard deviation – not to copy identical gene
24                end
25                else   if (sumφ/Length(φ)) > 0.2 then
26                           for j ← 0 to n − 1 do begin
27                               σ[j] ← σ[j] * 1.22;
28                               if σ[j] > |B[j]−A[j]|/2 then σ[j] ← ϑ[j] ;
    //reduce the standard deviation
29            X_Arch ← AddListItem(X_Arch, X);
30        end;
31        X_Pop ← AppendList(X_Pop, X_Arch);
    //the combined population of parents and offspring
32        AssignFitnessRank(X_Pop);
    // assign fitness (rank) according to objective function value
33        X_Pop ← TournamentSelect_r(X_Pop, m, k);       //tournament selection
34    end;
35    result ← ExtractOptimalSet(X_Pop);  //extract candidate solution(s)
36  end;
```

Figure 1: Evolution Strategy Algorithm - Steady State Evolution

The mutation uses parent $\mathbf{X}$ to generate new offspring $\mathbf{X}_{\text{Mut}}$ - Figure 1: Evolution Strategy Algorithm - Steady State Evolution– line number 11. Mutation uses normal distribution. The mutation is shown in Figure 2.

```
begin
  Randomize;
  X_Mut ← X;
  j ← 0;
  while j ≤ (Length(X) − 1) do begin
      if (P_1) ≤ Random_u( ) then
          if ((P_2) ≤ Random_u( )) and ( j < Length(X) − 2) then begin
              X_Mut[j] ← X[j + 1];
              X_Mut[j + 1] ← X[j];
              X_Mut[j] ← Perturbation_u(X_Mut[j], A[j], B[j]);
              j ← j + 1;
          end
          else X_Mut[j] ← Random_n(X[j], σ[j]);
      else X_Mut[j] ← Random_n(X[j], σ[j]);
      X_Mut[j] ← Perturbation_u(X_Mut[j], A[j], B[j]);
      j ← j + 1;
  end;
  result ← X_Mut;
end;
```

Figure 2: Mutation of the Parent (using normal distribution) - “Mutate $_{\text{ES\_n}}$”

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The mutation algorithm contains the following parameters:

- P_1 … Probability of mutation
- P_2 … Probability of swapping neighbouring genes
- Random_u … Function returning single uniformly distributed random number in interval $[0, 1)$

Algorithm of mutation uses the "Perturbation" function – see Figure 3: "Perturbation" algorithm - using correction of the individual - mirroring the individual coordinates from the space of unfeasible solution back to search space: (Tvrdik, Evolutionary algorithms - Study Texts (in Czech language Evoluční algoritmy - učební texty), 2004)

```
begin
  X_Pert[j] ← X[j];
  while (X[j] < A[j]) or (X[j] > B[j]) do begin
      if (X[j] < A[j]) then
          X_Pert[j] ← 2 · A[j] − X[j]
//mirror the individual coordinate back to search space
      else if (X[j] > B[j]) then
              X_Pert[j] ← 2 · B[j] − X[j];
      X_Pert[j] ← Step(X_Pert[j], A[j], B[j], ϑ[j]);
//round or truncate the coordinate of the individual in the search space to the nearest neighbour
  end;
  result ← X_Pert[j];
end;
```

Figure 3: "Perturbation" algorithm

The population is sorted according to the objective function values - Rank-Based Fitness Assignment procedure – see Figure 4: Rank-Based Fitness Assignment - "AssignFitnessRank" (in the context of this paper, fitness is subject to minimization). This sorting algorithm contains the following parameters:

- $\mathrm{CF}_{F(\mathbf{X})}$ … Comparing function for comparing individuals using their objective function value
- r … Individual order
- fit … Fitness value

The procedure uses the process of assigning a scalar fitness value to each solution candidate in the population according to their order in the population.

```
begin
  X_Pop ← Sort_a(X_Pop, CF_F(X));
//sorting of the whole population using comparing function
  r ← 1;
  AssignFitnessTo(X_Pop[0], 1) ; //assigning fitness to the first individual
  for i ← Length(X_Pop) − 1 downto 0  do begin
     if CF_F(X)(X_Pop[i], X_Pop[i − 1]) < 0 then r ← r + 1;
  //objective function value of the individual X_Pop[i] is better than X_Pop[i − 1]
     fit ← r;
     AssignFitnessTo(X_Pop[i], fit);
  end;
end;
```

Figure 4: Rank-Based Fitness Assignment - "AssignFitnessRank" procedure

This procedure sorts all individuals in a population using the comparing function in ascending order. The function compares an individual according to its value of the objective function (objective function minimization):

$$\mathrm{CF}_{F(\mathbf{X})}(\mathbf{X}_1, \mathbf{X}_2) = \begin{cases} -1 & \text{if } F(\mathbf{X}_1) < F(\mathbf{X}_2) \\ 1 & \text{if } F(\mathbf{X}_1) > F(\mathbf{X}_2) \\ 0 & \text{else} \end{cases} \quad (5)$$

3. ALGORITHMS OF THE SELECTION

Selection is the process of choosing individuals according to their fitness values from the population and places them into the mating pool. (Bäck, Foussette, and Krause, 2013)

Generally, there are two classes of selection algorithms:

- with replacement (annotated with a subscript $_r$) - each individual from the population is taken into consideration for reproduction at most once and therefore also will occur in the mating pool one time at most
- without replacement (annotated with a subscript $_w$) - the mating pool returned by algorithms can contain the same individual multiple times. Like in nature, one individual may thus have multiple offspring.

Another possible classification of selection algorithms is (μ, λ) notation where μ denotes the number of parents and λ denotes number of offspring – e.g. (μ, λ) selection strategy is applied to λ offspring while their parents are "forgotten". This selection does not use the information about the parent´s fitness according to a new generation. This strategy relies on the excess of offspring - the selection uses Darwinian natural selection where λ> μ. (Beyer and Schwefel, 2002)

Normally, selection algorithms are used in a variant with replacement. One of the reasons therefore is the number of elements to be placed into the mating pool.

The selection algorithms have a major impact on the performance of evolutionary algorithms. (Weise, 2009)

3.1. Tournament Selection

Tournament selection proposed by Wetzel (Wetzel, 1983) is one of the popular and effective selection schemes. This type of selection has been analysed by Blickle and Thiele (Blickle and Thiele, 1995), Miller and Goldberg (Sastry and Goldberg, 1996), etc.

The Evolution Strategy optimization algorithm (shown in Figure 1: Evolution Strategy Algorithm - Steady State Evolution) uses Tournament selection - Figure 6. The final population consists of children and parents with good fitness. This strategy supports elitism – the individual with good fitness survives. (Tvrdik, 2010

The next algorithm uses tournament selection with replacement – an individual can compete against itself (its copy). This situation is impossible in reality, but remember, these individuals have the same fitness value therefore the first individual wins and not its copy according to the use of the comparing function. This comparing function selects the first individual if both fitness values are the same (line number 7).

The algorithm creates an empty mating pool (line number 2). Individuals are selected according to their fitness values in the initial stage (line number 3). Individuals are randomly picked from the population (line number 5) and compete with each other in k tournaments (line number 7). The winner of these competitions enters the mating

Proceedings of the European Modeling and Simulation Symposium, 2017
ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

pool (line number 8). Although it is a simple selection strategy, it is a very powerful selection.
The other tournament selection algorithms contain the same following parameters:

- X_{Pop} … List of the individuals to select from
- X_{MP} … Mating pool
- m_{MP} … Number of individuals to be placed into the mating pool
- $\mathfrak{f}$ … Fitness function
- i … Individual index
- $\text{CF}_{\mathfrak{f}(\mathbf{X})}$ … Comparing function for comparing individuals using their fitness function value
- a … Index of the tournament winner
- k … Number of contestants

```
begin
  X_MP ⟵ ();
  X_Pop ⟵ Sort_a(X_Pop, CF_f(X));
  for i ⟵ 0 to m_MP − 1 do begin
    a ⟵ ⌊Random_u(0,Length(X_Pop))⌋;
    for j ⟵ 1 to k − 1 do
      a ⟵ min{a, ⌊Random_u(0,Length(X_Pop))⌋};
    X_MP ⟵ AddListItem(X_MP, X_Pop[a]);
  end;
  Result ⟵ X_MP;
end;
```

Figure 5: Tournament Selection with replacement - “TournamentSelect $_{\text{r}}$” (Weise, 2009)

The absolute values of the fitness play no role. The only thing that matters is whether or not the fitness of one individual is higher than the fitness of another one, not the fitness difference itself. With rising k, the selection pressure increases - individuals with good fitness values create more and more offspring, whereas the chance of worse solution candidates to reproduce decreases. (Weise, 2009)

The tournament selection without replacement (shown in Figure 6: Tournament Selection - “TournamentSelect $_{\text{w1}}$”) uses the same principle, but the winner of the tournaments does not participate in other tournaments (is deleted from the population – line number 9).
The second variant of the tournament selection without replacement is identical to the previous algorithm. The difference is that all individuals (their indexes) are selected for one overall round (one list) where the individual cannot be included. The main winner with the lowest fitness value is selected from this list. This process is repeated several times until the mating pool is filled.

```
1   begin
2     X_MP ⟵ ();
3     X_Pop ⟵ Sort_a(X_Pop, CF_f(X));
4     for i ⟵ 0 to min{Length(X_Pop), m_MP} − 1 do begin
5       a ⟵ ⌊Random_u(0,Length(X_Pop))⌋;
6       for j ⟵ 1 to min{Length(X_Pop), k} − 1 do
7         a ⟵ min{a, ⌊Random_u(0,Length(X_Pop))⌋};
8       X_MP ⟵ AddListItem(X_MP, X_Pop[a]);
9       X_Pop ⟵ DeleteListItem(X_Pop, a);
10    end;
11    Result ⟵ X_MP;
12  end;
```

Figure 6: Tournament Selection - “TournamentSelect $_{\text{w1}}$” (Weise, 2009)

```
begin
  X_MP ⟵ ();
  X_Pop ⟵ Sort_a(X_Pop, CF_f(X));
  for i ⟵ 0 to m_MP − 1 do begin
    A ⟵ ();
    for j ⟵ 1 to min{k,Length(X_Pop)} do begin
      repeat
        a ⟵ ⌊Random_u(0,Length(X_Pop))⌋;
      until Search_u(a, A) < 0;
      A ⟵ AddListItem(A, a);
    end;
      a ⟵ min{A};
    X_MP ⟵ AddListItem(X_MP, X_Pop[a]);
  end;
  Result ⟵ X_MP;
end;
```

Figure 7: Tournament Selection - “TournamentSelect $_{\text{w2}}$” (Weise, 2009)

The previous tournament selection algorithms can be called deterministic tournament selection algorithms. The winner of the k contestants that take part in each tournament enters the mating pool. In the non-deterministic variant, a probability for the individual selection p is defined. The best individual in the tournament is selected with probability p, the second best with probability $p(1-p)$, the third best with probability $p(1-p)^2$ and so on. The i-th best individual in a tournament enters the mating pool with probability $p(1-p)^i$. (Weise, 2009)

```
begin
  X_MP ⟵ ();
  X_Pop ⟵ Sort_a(X_Pop, CF_f(X));
  for i ⟵ 0 to m_MP − 1 do begin
    A ⟵ ();
    for j ⟵ 0 to k − 1 do
      A ⟵ AddListItem(A, ⌊Random_u(0,Length(X_Pop))⌋);
    A ⟵ Sort_a(A,CF(a_1,a_2)≡(f(a_1) − f(a_2)));
    for j ⟵ 0 to Length(A) − 1 do
      if (Random_u() ≤ p) ∨ (j ≥ Length(A) − 1) then begin
        X_MP ⟵ AddListItem(X_MP, X_Pop[A[j]]);
        j ⟵ ∞;
      end;
  end;
  Result ⟵ X_MP;
end;
```

Figure 8: Tournament Selection - "TournamentSelect $_{r,p}$" (Weise, 2009)

4. SETTINGS OF EVOLUTION STRATEGY AND TOURNAMENT SELECTION PARAMETERS

We tested different settings of the evolution strategy algorithm and tournament selection parameters. We defined a step and lower and upper boundaries for these parameters.

Table 1: Settings of Evolution Strategy And Tournament Selection Parameters

Parameter	Step	Lower Bound	Upper Bound
m … Size of population	1×n	1×n	6×n
m_{MP} … Number of offspring	1×n	1×n	6×n
q … Number of successes (the offspring is better than the parent) to be monitored	1×n	1×n	6×n
k … Number of other contestants per tournament	1×n	1×n	6×n
p … Probability of the individual selection	0.1	0.1	0.6

5. TESTED SIMULATION MODELS

Considering the time required for testing the behaviour of optimization methods, we substitute the testing on the discrete simulation models (and its objective function) by a different testing function to reduce the duration of testing. We tested a different setting of the evolution strategy on four testing functions - De Jong´s, Rosenbrock´s, Michalewicz´s, Ackley´s functions - and discrete event simulation model. (Raska & Ulrych, 2015)

We simulated all feasible solutions of the discrete event simulation model to build a database of simulation experiments to increase the speed of simulation optimization.

5.1. Testing functions

The domain of the function is a defined step for each axis – substitution of the simulation model input parameter (discrete) values of the discrete event simulation model. All testing functions were minimized.

5.2. De Jong´s Function

A convex and unimodal testing function. The function definition: (Pohlheim, 2006)

$$F(\mathbf{X}) = \sum_{j=1}^{n} x_j^2,$$
$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01} \right\rfloor \cdot 0.01 = 0 \right) \wedge \left(-30 \le x_j \le 30 \right),$$
$$j = 1 : n, n \in [2,10,20,30,44] \qquad (6)$$

Where symbols denote:

- $F(\mathbf{X})$… Objective function
- j … Index of control
- n … Dimension of the search space - the dimension of the search space is 10; 20; 30; 40; 50
- x_j … Value of control - testing functions (except Michalewicz function) input parameters values range from -30 (lower boundary) to 30 (higher boundary)

We substitute the testing on the simulation models by testing on the testing function, and we found that the smallest step that can be performed by the optimization methods is 0.01 for each axis in the search space (x_j mod 0.01=0). The input parameters are not continuous.
This resolution represents 8,1504×10^{188} possible solutions (combinations of testing function input parameters) in a fifty dimensional search space. To achieve a better idea of the testing functions landscapes the continuous testing functions are shown in the following four figures. De Jong´s continuous function is shown in Figure 9.

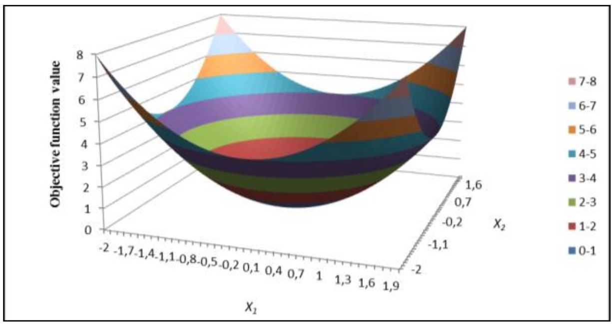


Figure 9: De Jong´s Function

5.3. Rosenbrock´s Function

Rosenbrock´s (Rosenbrock's valley, Rosenbrock's banana) function is a unimodal and non-convex testing function. The function definition: (Pohlheim, 2006)

$$F(\mathbf{X}) = \sum_{j=1}^{n-1} 100 \cdot (x_j^2 - x_{j+1})^2 + (1 - x_j)^2,$$
$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01} \right\rfloor \cdot 0.01 = 0 \right) \wedge \left(-30 \le x_j \le 30 \right),$$
$$j = 1 : n, n \in [2,10,20,30,44] \qquad (7)$$

Rosenbrock´s continuous function is shown in Figure 10.

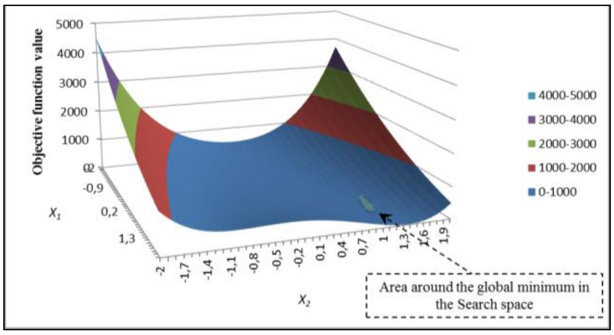


Figure 10: Rosenbrock´s Function

5.4. Michalewicz Function

Michalewicz 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)

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

$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01}\right\rfloor \cdot 0.01 = 0\right) \wedge \left(0 \le x_j \le \pi\right), \qquad (8)$$

$$j = 1:n, n \in [2,10,20,30,44]$$

We selected $m = 5$ in our simulation model. The Michalewicz continuous function is shown in Figure 11.

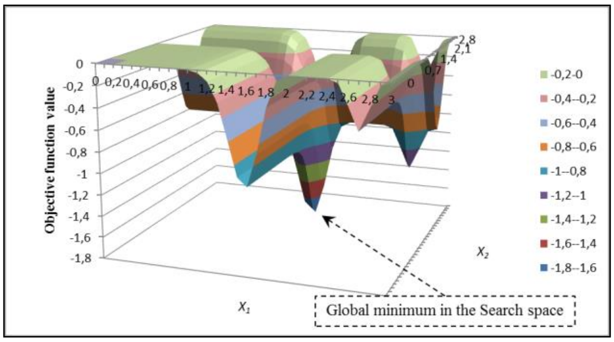


Figure 11: Michalewicz Function

5.5. Ackley´s Functions

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

$$F(\mathbf{X}) = -20\cdot\exp\left(-0.02\cdot\sqrt{\frac{1}{n}\cdot\sum_{j=1}^{n}x_j^2}\right) - \exp\left(\frac{1}{n}\cdot\sum_{j=1}^{n}\cos 2\cdot\pi\cdot x_j\right) + 20 + \exp(1), \qquad (9)$$

$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01}\right\rfloor \cdot 0.01 = 0\right) \wedge \left(-30 \le x_j \le 30\right),$$

$$j = 1:n, n \in [2,10,20,30,44]$$

Ackley´s continuous function is shown in Figure 12.

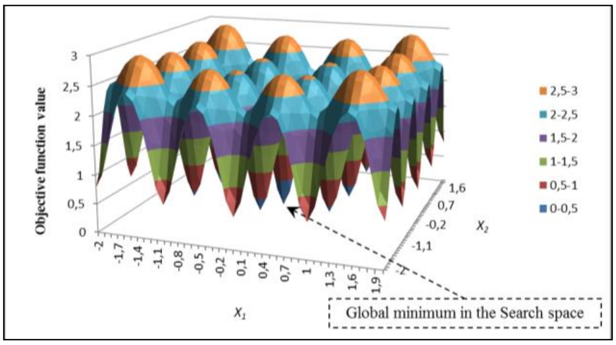


Figure 12: Ackley´s Function

5.6. Discrete event simulation model

The tested discrete event simulation model is focused on a production workshop consisting of the workplaces shown in Table 2: Production workplaces. The simulation model was built in Tecnomatix Plant Simulation software – see Figure 13.

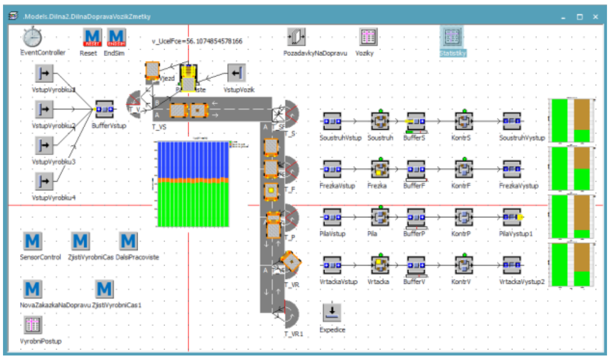


Figure 13: Discrete Event Simulation Model

Table 2: Production workplaces

Workplace Number -WN	Description
1	Sawmill
2	Lathe
3	Milling machine
4	Drill
5	Control station - input
6	Control station - output

The product passes through the following workplaces: $5 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 6$. Transportation between workplaces uses a forklift truck with a speed of 60 $[m/min]$. The distance between the workplace is 20 metres. Four types of products are processed at the workshop. The first product arrives every 13 minutes, the second product arrives every 5 minutes, the third product arrives every 20 minutes, and the fourth product arrives every 18 minutes at the workshop.

The next table contains the sequences of the workplaces which the product passes through. This table also contains the time of processing (and also intervals) at the workplace.

Table 3: Sequence of the Workplaces

Product	Sequence of Workplaces/Time of Processing [min]						
	5	1	2	3	4	5	6
1	5	[1,15]	[3,22]	[2,24]	[4,15]	6	
2	5	[1,23]	[3,15]	[2,25]	[3,12]	[4,23]	6
3	5	[1,18]	[4,19]	[3,17]	[4,30]	6	
4	5	[1,25]	[2,26]	[3,25]	[2,15]	[3,25]	6

The product is placed in the buffer with a maximum capacity of 15 units before the inspection station after the processing. The time of product inspection is from 20 seconds to 30 seconds (mean - 27 seconds). If the product fails the inspection, the worker must immediately rework the product.
The next table contains the probability of a defective product (marked as PoDP) and rework time (RT) in minutes using different random distribution at different workplaces (workplace number - WN).

Table 4: Probability of a Defective Product (PoDP) And Rework Time (RT) – The First and Second Product

WN	Product			
	1		2	
	PoDP	RT [min]	PoDP	RT [min]
1	3%	NORM(20,5)	2%	NORM(25,5)
2	5%	NORM(18,5)	4%	NORM(10,5)
3	5%	NORM(20,5)	5%	TRIA(25,30,35)
4	4%	NORM (20,8)	7%	TRIA(14,18,25)

Table 5: Probability of a Defective Product (PoDP) And Rework Time (RT) – The Third and Fourth Product

WN	Product			
	3		4	
	PoDP	RT [min]	PoDP	RT [min]
1	2%	NORM(22,5)	1%	NORM(20,10)
2	4%	NORM(20,5)	3%	NORM(30,10)
3	2%	TRIA(20,25,30)	7%	NORM(20,10)
4	2%	TRIA(20,25,30)	2%	NORM(25,10)

The main goal is to determine the number of machines and controllers at individual workplaces according to a number of lift trucks, machines and controller utilization (maximizing production processes).

The objective function:

$$F(\mathbf{X}) = \frac{\text{NumberOfProcessedProducts}}{10} + \sum_{WN=1}^{4} (\text{MachineUtilization} + \text{ControllerUtilization}) \qquad (10)$$

Decision variables of the simulation model (decision variables):

- Number of machines at the first, second, third, fourth workplace
- Number of controllers at the first, second, third, fourth workplace
- Number of forklifts

6. OPTIMIZATION EXPERIMENTS

We specified the same conditions which had to be satisfied for each optimization method, e.g. the same termination criteria. The optimization method could perform a maximum of 100,000×*n* (parameter *n* denotes the dimension of the search space) simulation experiments to find the global optimum in the search space of the testing function (Tvrdik, Evolutionary algorithms - Study Texts (in Czech language Evoluční algoritmy - učební texty), 2004).

The termination criterion for the discrete event simulation model 10,000×*n* (n=9). Another termination criterion is VTR (value to reach) – stopping the simulation optimization if the optimum is found. We tested all possible solutions of the simulation model, so we could specify the value to reach. We also created a database of the simulation experiments. If the simulation optimizer wants to perform the simulation run, the optimizer searches for the simulation experiment in this database. If the simulation experiment is found, the optimizer downloads the simulation experiment and its objective function value.

If the optimization method has the same parameters as another optimization method, we set up both parameters with the same boundaries (same step, lower and upper boundaries).

We tested different settings for each evolution strategy selection. The next table (see Table 6) shows the same settings of evolution strategy parameters for each type of tournament selection. Parameter *D* denotes the dimension of the search space. We tested four types of the tournament selection:

- With replacement - "Tournament R" (see Figure 5); "Tournament RP" (see Figure 8)
- Without replacement - "Tournament W1" (see Figure 6); "Tournament W2" (see Figure 7)

If the tournament selection "Tournament R", "Tournament W1" or "Tournament W2" is tested, the parameter - The Number of Contestants - varies from 1×*D* (lower boundary) to 6×*D* (higher boundary) with step 1×*D*. This means that we tested 1,296 different series - different settings - for each selection and also for each testing function or the discrete event simulation model.
If we test the tournament selection "Tournament RP", the parameter - The Probability for The Individual Selection - varies from 0.1 to 0.6 with step 0.1 We have to test 7,776 series for this tournament selection testing for each simulation model.
We tested 233,280 series on the testing functions and 11,664 series on the discrete event simulation model. We replicated these series several times to reduce the random behaviour of the tested optimization method and its selection strategies.

Table 6: Settings of Evolution Strategy Parameters

Method Parameter	Step	Bounds		Number of Series
		Lower Bound	Upper Bound	
Number of Offspring	1×*D*	1×*D*	6×*D*	6*6*6*6 = 1296
Number of Success (the Offspring Is Better Than the Parent) To Be Monitored	1×*D*	1×*D*	6×*D*	
Population Size	1×*D*	1×*D*	6×*D*	
Number of Other Contestants Per Tournament	1×*D*	1×*D*	6×*D*	

We evaluated these optimization experiments with different settings - series. The following charts show the average optimization method success of finding the optimum (suboptimum if the optimum was not found).

The first criterion is the function whose output is the standardized scalar value $\mathfrak{f}_1 \in [0,1]$ of not finding the known VTR (value to reach). This value represents the failure of finding the global optimum by the optimization method in a particular series – value minimization. This value is expressed by pseudopascal code and shown in Figure 14. This algorithm contains the following parameters:

- X^* … List of found optima in each optimization experiment in the series
- $\mathbf{X}^*$ … Global optimum $\mathbf{X}^*$ in the search space
- ε …Tolerated deviation from the value of the objective function value of global optimum
- $F(\mathbf{X})$ … Objective function value
- n_{Succ} …Counter of successful finding optimum
- $\mathfrak{f}_1$ … Standardized scalar value

```
begin
  n_Succ ← 0;
  for i ← 0 to Length(X*) − 1 do
    if |F(X*[i]) − F(X*)| ≤ ε then
      n_Succ ← n_Succ + 1 ;
  (*Optimum or acceptable candidate solution
  was found *)
  result ← (Length(X*) − n_Succ) / Length(X*);
  (*standardization - % share of unsuccessful
  series*)
end;
```

Figure 14: Pseudopascal Algorithm of First criterion – Finding the Global Optimum or Suboptimum

If the failure is 100[%] the first criterion equals 1 therefore we try to minimize this criterion. Average Method Success of Finding Optimum can be formulated as follows:

$$f1\,AVG = \left(1 - \frac{\sum_{i=1}^{s} f_{1_i}}{s}\right) \cdot 100[\%] \qquad (11)$$

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 series were also evaluated regarding specified tolerance between the best optimum (suboptimum) found in the series and the specified parameter ε. We initially specified $\varepsilon = 0.001$. The optimization method had to find the candidate solution whose objective function value is nearly the same as the objective function value of the global optimum in the search space (the tolerance equals 0.001).

The following chart provides us with information about the success of finding the optimum:

- We can assume high average selection strategies success of finding the optimum (suboptimum) if the dimension of the search space of the testing function is lower or the objective function surface is simple
- If the testing function surface is hard – multimodal, planar regions - the optimization failure rate of the optimization selection strategies is high (the high number of absolutely unsuccessful series $f_1 = 1$, i.e. series does not contain any optimization experiment where the optimum was found)
- Tested tournament selection strategy types have little effect on evolution strategy success
- The success of finding the De Jong´s testing function optimum rapidly decreases if the dimension increases. We obtained high success of finding the optimum on testing function Ackley10.
- Tested selection strategies – "Tournament R" and "Tournament W2" - have almost the same success of finding the optimum

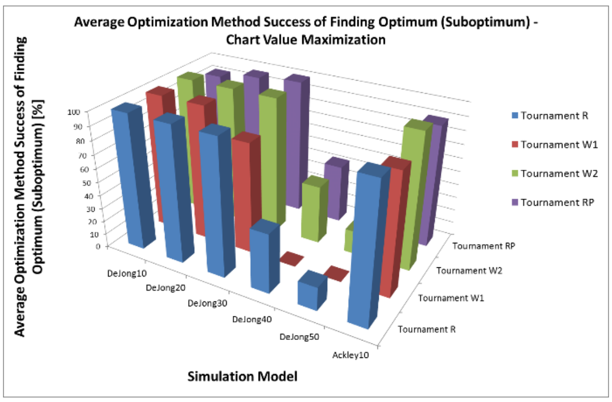


Figure 15: Average Optimization Method Success of Finding Optimum (Suboptimum) – Testing Functions - the Number Denotes the Dimension of the Search Space

- Selection strategies – “Tournament R” and “Tournament W2” – tested on the discrete event simulation model have almost the same success of finding the optimum
- The failure rate (the high number of absolutely unsuccessful series $f_1 = 1$) of finding the optimum of the objective function of the simulation model is almost seventy percent

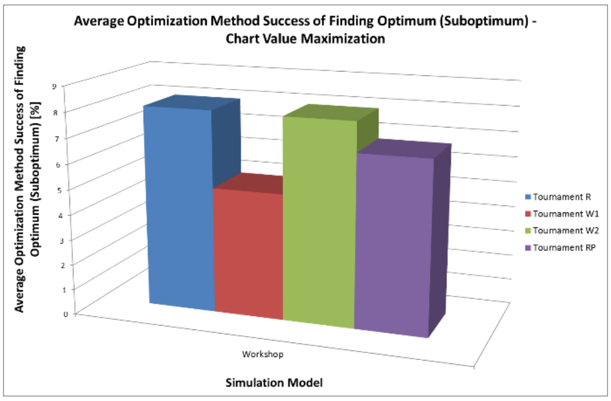


Figure 16: Average Optimization Method Success of Finding Optimum (Suboptimum) – Discrete Event Simulation Model

The second criterion f_2 is useful when there is no series which contains any optimum or a solution whose objective function value is within the tolerance of the optimum objective function value (the first criterion f_1 equals zero in this case). 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 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:

$$f_2 = \left(\frac{F(\mathbf{X}^*) - F(X_{Best})}{F(\mathbf{X}^*) - F(X_{Worst})} \right) \quad (12)$$

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 a concrete series; $F(X_{\text{Worst}})$ denotes objective function value of the worst found solution (element) of the search space. Output of function can take as value $f_2 \in [0,1]$.

The average of the second criterion of absolutely unsuccessful series where $f_1 = 1$ is calculated using the formula:

$$f_{2\,AVG} = \left(1 - \frac{\sum_{i=1}^{s} f_{2_i}}{s} \right) \cdot 100[\%] \quad (13)$$

where i denotes the index of one series, f_{2_i} denotes the value of the second criterion (optimization method success – the best value is zero), s denotes the number of performed series.

The average of the difference between the optimum and the local extreme tested on the testing functions is shown in Figure 17. The charts contain only series where $f_1 = 1$ (no optimum was found in the series).

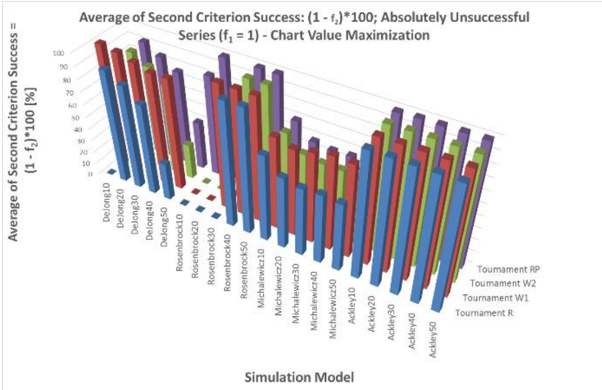


Figure 17: Average of Second Criterion Success: $(1 - f_2) * 100$; Absolutely Unsuccessful Series ($f_1 = 1$) - Chart Value Maximization – Testing Functions

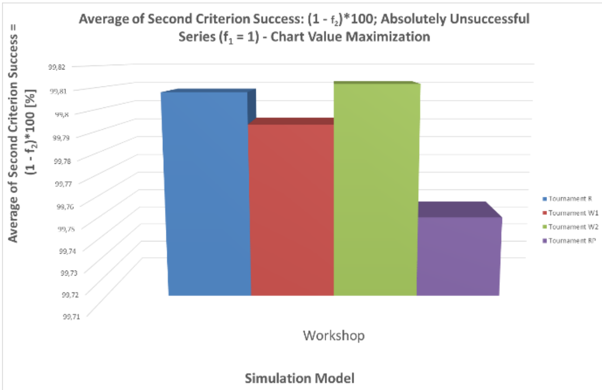


Figure 18: Average of Second Criterion Success: $(1 - f_2) * 100$; Absolutely Unsuccessful Series ($f_1 = 1$) - Chart Value Maximization – Discrete Event Simulation Model

We can see that each tournament selection type has a problem with the complicated objective function landscape – Michalewicz´s function. The optimization method provides local optima far from the global optimum. The charts (see Figure 17 and Figure 18) also contain other information. Evolution strategy selections found the global optima very near to global optimum of Ackley’s function. Tournament selection “Tournament W1” was better (according to the difference between optimum and local extreme) than the other types of selections on average. Averages of the second criterion success are: “Tournament R”: 63.7[%]; “Tournament W1”: 82.7[%]; “Tournament; W2”: 63.5[%]; “Tournament RP”: 77.3[%]. Tournament selections “Tournament R” and “Tournament W2” have almost the same success of the second criterion calculated for the testing functions and discrete event simulation model.

Objective functions values of found global optima were very close to the objective function of the global optimum of the discrete event simulation model.

7. CONCLUSION

The goal of the research is to test different settings (series) of the tournament selection strategies of the Evolution Strategy on testing functions and a discrete event simulation model. We evaluated the success of finding the optimum by these strategies and we also used a function evaluating the difference between the objective function value of the best solution found in the series and the optimum objective function value.

The success of finding the optimum by the Evolution Strategy selection strategies strongly depends on the objective function surface. The problems occurred with the finding of the optimum of the Michalewicz's function – flat areas, multimodal function.

Tested selection strategies – "Tournament R" and "Tournament W2" - have almost the same success of finding the optimum. Tournament selection "Tournament W1" was better (according to the difference between the optimum and local extreme) than the other types of selections on average.

ACKNOWLEDGMENTS

This contribution has been prepared within project LO1502 'Development of the Regional Technological Institute' under the auspices of the National Sustainability Programme I of the Ministry of Education of the Czech Republic aimed at supporting research, experimental development and innovation.

REFERENCES

Bäck, T., Foussette, C., and Krause, P., 2013. Contemporary Evolution Strategies,1 ed., Vol. XIII. Berlin, Germany: Springer-Verlag Berlin Heidelberg.

Beyer, H. G., and Schwefel, H. P., 2002. Evolution Strategies - A Comprehensive Introduction. Natural Computing, 1(1), 3-52.

Blickle, T., & Thiele, L. (1995). A mathematical analysis of tournament selection. Proceedings of the Sixth International Conference on Genetic Algorithms, 9-16. Pittsburgh, PA: Morgan Kaufmann Publishers Inc. Available from: http://citeseer.ist.psu.edu/blickle95mathematical.html [accessed March 10, 2015],

Holland, J. H., 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Ann Arbor: The University of Michigan Press.

Hynek, J., 2008. Genetic Algorithms and Genetic Programming (in Czech language: Genetické algoritmy a genetické programování). Prague: Grada.

Marik, V., Stepankova, O., and Lazansky, J., 2001. Artificial Intelligence (3). Prague: Academia Praha.

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.

Pohlheim, H., 2006. GEATbx: Example Functions. Available from: http://www.geatbx.com/docu/fcnindex-01.html#P204_10395 [accessed 20 November 2011]

Raska, P. & Ulrych, Z., 2013. Simulation Optimizer and Optimization Methods Testing On Discrete Event Simulations Models and Testing Functions. ATHENS, GREECE, DIME Universita, pp. 50-59.

Raska, P., and Ulrych, Z., 2015. Comparison of Optimization Methods Efficiency in Different Dimensions of the Search Space of Simulation Models. The 27th European Modeling & Simulation Symposium 2015, pp. 181-190. Bergeggi, Italy: DIME Universita of Genova.

Raška, P., and Ulrych, Z., 2015. Comparison of optimisation methods tested on testing functions and discrete event simulation models. International Journal of Simulation and Process Modelling, Volume 10(Issue 3): pp. 279-293

Sastry, K., and Goldberg, D. E., 1996. Genetic algorithms, selection schemes, and the varying effects of noise. Evolutionary Computation. IV, pp. 113-131. Massachusetts Institute of Technology.

Schwefel, H. P., 1995. Evolution and Optimum Seeking. New York, USA: John Wiley& Sons.

Sumathi, S., Hamsapriya, T., and Surekha, P., 2008. Evolutionary Intelligence: An Introduction to Theory and Applications with Matlab (Vol. I). Berlin: Springer-Verlag Berlin Hedilberg.

Tvrdik, J., 2010. Stochastic Algorithms for Global Optimization (in Czech language: Stochastické algoritmy pro globální optimalizaci). Available from: http://www1.osu.cz/~tvrdik/wp-content/uploads/STAGO_10.pdf [accessed 5 January 2014]

Tvrdik, J., 2004. Evolutionary algorithms - Study Texts (in Czech language Evoluční algoritmy - učební texty). Available from: http://prf.osu.cz/doktorske_studium/dokumenty/Evolutionary_Algorithms.pdf [accessed February 6, 2012]

Weise, T., 2009. E-Book "Global Optimization Algorithms - Theory and Application" 2nd Edition. Available from: http://www.it-weise.de/projects/book.pdf [accessed 2 February 2011]

Wetzel, A., 1983. Evaluation of the Effectiveness of Genetic Algorithms in Combinatorial. Pittsburgh, Pennsylvania: University of Pittsburgh.

ANALYSING FOR MEASURE CHARACTERISTICS OF NASH-SUTCLIFFE EFFICIENCY TYPE INDICATORS USED FOR MODEL VALIDATION

Kai-Bin Zhao [a], **Ke Fang** [b], **Ming Yang** [c]

Control & Simulation Center, School of Astronautics, Harbin Institute of Technology, Harbin, China

[a]kaibin.zhao.HIT@hotmail.com, [b]hitsim@163.com, [c]myang@hit.edu.cn

ABSTRACT
The validation of simulation model is commonly quantified by various goodness-of-fit indicators. The selection and use of specific goodness-of-fit indicators and the interpretation of their results can be a challenge for even the most experienced evaluators since each indicator may put different emphasis on different types of simulated and observed behaviours. In this paper, first of all, two goodness-of-fit indicators calculated on weighted errors are developed, presented and recommened, which are more suitable for evaluating simulation model compared with high and low magnitude within observed data respectively. Then, the specific measure characteristics (that is, which part of the simulated and observed datasets most influences the various indicators) of several Nash-Sutcliffe efficiency (*NSE*) type indicators is analyzed and compared through two hypothetical examples using a simple observed and simulated datasets. Finally, some misunderstanding and ambiguity is corrected and clarified for measure characteristics of *NSE* type indicators.

Keywords: methods of quantitative validation, Nash-Sutcliffe efficiency, comparison between indicators, measure characteristics

1. INTRODUCTION

Krause, Boyle, and Bäse (2005) put forward three main reasons for model validation requirement: (1) to provide a process of quantitative indicator of model's ability to replicating past and predicting future behavior; (2) to provide a means for evaluating improvements to the model, such as through adjustment of model parameter values and model structural modifications, and the important characteristics expression changes of model behavior in space and time; (3) to compare modeling efforts in model development life cycle process.

Model validation deals with the issue of whether or not a model is a sufficient accurate representation of real system for the intended applications of the model (Martens, Put, and Kerre 2006). From the perspective of practicality and operability, a commonly pragmatic approach is comparison of simulated data (modeled data, predicted data, or model output data) with observed data (reference data, measured data, experimental data, test data, expected data, or real-word data) to determine the degree of behavioral similarity between simulation model and real system. This process of model performance evaluation may appear the following situations (Pushpalatha, Perrin, and Le Moine 2012): (1) data set used for model validation vary span several orders of magnitude that may not be equally important for the modeller who may be interested in agreement between the part of observed and simulated graph; (2) the model may be used for diferent applications, which may require specific criteria (e.g. either high or low magnitude simulation studies); (3) the variance of error varies significantly throughout validation period with space and time, instead of a nearly constant, which often leads to emphasize larger errors within a wider range of the observations while smaller errors within a small range tend to be neglected.

For these situations, a large variety of goodness-of-fit indicators have been proposed and used over the years in the field of model validation, as shown for example by the lists of indicators given by Bennett, Croke, and Guariso (2013); Hauduc, Neumann, and Muschalla (2015); and Crochemore, Perrin, and Andreassian (2015). Among these indicators, the Nash and Sutcliffe (1970) efficiency (*NSE*) type indicators has received considerable attention from many evaluators, especially in the area of environmental science, due to its flexibility to be applied to various types of simulation models (Oudin, Andreassian, and Mathevet 2006; Nicolle, Pushpalatha, and Perrin 2014). Modifications of the *NSE* have been proposed by several authors, including those based on transformed variables, others using relative or absolute instead of squared errors, and those adopting benchmark series different from the mean of observations. Until by now, the focus of discussion work with the existing *NSE* type indicators is primarily on their interrelationship and measure characteristics, which have gained some research achievements that can be found in the literature (i.e. Krause, Boyle, and Bäse 2005; Oudin, Andreassian, and Mathevet 2006; Dawson, Abrahart, and See 2007; Pushpalatha, Perrin, and Le Moine 2012; Hauduc, Neumann, and Muschalla 2015).

Although there are a lot of pre-analysis for *NSE* type indicators, but there are still some limitations. The present work intends to complement previous studies, and this paper has two main objectives: first, in order to effectively select and use different *NSE* type indicator and interpret their results, we wish to discuss further specific measure characteristics of each indicator，that is, which part of the $\{O_i, S_i\}$ most influences the various indicators; Second, we wish to propose two goodness-of-fit indicators suited for the evaluation of high-magnitude and low-magnitude simulations using actually observed data.

The remainder of this paper is arranged as follows: Section 2 describes different *NSE* type indicators and our proposed goodness-of-fit indicators on weighted errors; Section 3 provide a hypothetical example used for analysing measure characteristics of various indicators; Results and discussion are presented in Section 4, while conclusions are drawn in Section 5.

2. GOODNESS-OF-FIT INDICATORS

A number of goodness-of-fit indicators have been proposed in the simulation literature, however, there is no consensus on the best approach. In this section, the goodness-of-fit indicators included in the study are provided. These are the seven indicators: Nash-Sutcliffe efficiency and some of its modified forms: Legates-McCabe efficiency (LME), NSE on squared transformed series, NSE on root-squared transformed series, NSE on logarithmic transformed series, NSE on inverse transformed series, LME on deviations weighted by observations cubed, LME on deviations weighted by observations to the power of minus three.

These indicators above are generally viewed as different normalized forms of different error measures calculated on point-by-point concurrently and are generally used in order to make decisions about the validity of the simulation model.

In the following equations, $\boldsymbol{O}$ represents the sample containing the observations obtained from real system, but could more generally containing the expected values, $\boldsymbol{O}=(O_1, O_2, \ldots, O_n)^{\mathrm{T}} \in \mathrm{R}^n$; $\boldsymbol{S}$ represents the sample containing the model simulated outputs, $\boldsymbol{S}=(S_1, S_2, \ldots, S_n)^{\mathrm{T}} \in \mathrm{R}^n$; O_i and S_i are respectively the ith observed and simulated value; The subscript i is data index (usually referring to time, but could more generally indicate spatial location or some other kind of index); n is is total number of pairwise-matched observed and simulated data used for validation; $\overline{O_i}$ is the mean of the observations.

2.1. Nash-Sutcliffe Efficiency

The *NSE* suggested by Nash and Sutcliffe (1970) is expressed as one minus the mean of the squared errors between the observed and simulated datasets normalized by the variance of the observed values during the validation period, which is the most widely attention goodness-of-fit indicator in the model validation fields due to its flexibility. It is calculated as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O_i})^2}, \quad -\infty < NSE \le 1 \tag{1}$$

The *NSE* ranges from minus infinity to one: an *NSE* value of one implies that observed value are in complete agreement with simulated value for all subscript i, which is generally viewed as model's ability to reproduce and predict system behaviour perfectly; *NSE* values between zero and one implies that the simulation model is superior than the use of the benchmark model (the mean of the observation series) as a accurate representation of real system; an *NSE* value of zero implies that the simulation model, on average, performs as good as the use of the benchmark model; whereas negative *NSE* values implies that the simulation model is poor than the use of the benchmark model, which ususlly indicates unacceptable or unsatisfactory model performance rating.

2.2. Legates-McCabe Efficiency

The *LME* suggested by Legates and McCabe (1999) is expressed as one minus the mean of the absolute errors between the observed and simulated datasets normalized by the mean absolute deviation of the observed values. It is calculated as:

$$LME = 1 - \frac{\sum_{i=1}^{n}|O_i - S_i|}{\sum_{i=1}^{n}|O_i - \overline{O_i}|}, \quad -\infty < LME \le 1 \tag{2}$$

The modified *NSE* is less sensitive to large extreme absolute errors values due to using absolute instead of squared errors within the numerator of the fractional part of *NSE*, but this indicator was not catch much attention and deeply discussed because of its limited application and resulting relative lack of reported information. The range of *LME* lies between minus infinity and one (perfect fit). An *LME* value of lower than zero indicates that the mean value of the observation series series would have been a better predictor than the model.

2.3. NSE Based on Transformed Data Series

Additional modifications of Nash-Sutcliffe efficiency are based on transforming observed outcomes and model outputs, including NSE calculated on a squared transformation of $\boldsymbol{O}$ and $\boldsymbol{S}$ (NSE_{squ}) (Oudin, Andreassian, and Mathevet 2006; Crochemore, Perrin, and Andreassian 2015), NSE calculated on a root-squared transformation of $\boldsymbol{O}$ and $\boldsymbol{S}$ (NSE_{sqrt}) (Perrin, Miche, and Andréassian 2001), NSE calculated on a logarithmic transformation of $\boldsymbol{O}$ and $\boldsymbol{S}$ (NSE_{ln}) (Oudin, Andreassian, and Mathevet 2006), NSE calculated on inverse transformation of $\boldsymbol{O}$ and $\boldsymbol{S}$ (NSE_{inv}) (Le Moine 2008). The NSE_{squ}, NSE_{sqrt}, NSE_{ln} and NSE_{inv} ranges from minus infinity to one (perfect fit). They is calculated as:

$$NSE_{\text{squ}} = 1 - \frac{\sum_{i=1}^{n} (O_i^2 - S_i^2)^2}{\sum_{i=1}^{n} (O_i^2 - \overline{O_i^2})^2}, \; -\infty < NSE_{\text{squ}} \le 1 \qquad (3)$$

$$NSE_{\text{sqrt}} = 1 - \frac{\sum_{i=1}^{n} (O_i^{0.5} - S_i^{0.5})^2}{\sum_{i=1}^{n} (O_i^{0.5} - \overline{O_i^{0.5}})^2}, \; -\infty < NSE_{\text{sqrt}} \le 1 \qquad (4)$$

$$NSE_{\text{ln}} = 1 - \frac{\sum_{i=1}^{n} [\ln(O_i) - \ln(S_i)]^2}{\sum_{i=1}^{n} [\ln(O_i) - \overline{\ln(O_i)}]^2}, \; -\infty < NSE_{\text{ln}} \le 1 \qquad (5)$$

$$NSE_{\text{inv}} = 1 - \frac{\sum_{i=1}^{n} (O_i^{-1} - S_i^{-1})^2}{\sum_{i=1}^{n} (O_i^{-1} - \overline{O_i^{-1}})^2}, \; -\infty < NSE_{\text{inv}} \le 1 \qquad (6)$$

The NSE_{squ} indicator is more sensitive to very high magnitudes and shows nearly no reaction on low magnitudes. The NSE_{sqrt} indicator provides more balanced information as the errors are more equally distributed on high- and low- magnitudes parts. The NSE_{ln} indicator is more sensitive to low magnitudes but still reacts to peak magnitudes. The NSE_{inv} indicator is more sensitive to very low magnitudes and shows nearly no reaction on high magnitudes.

2.4. LME on Deviations Weighted by Observations

Nash-Sutcliffe efficiency type indicators based on data transformations use the model error series computed after arithmetic transformation instead of original error series to evaluate model performance, which result in the fact that the implicit weightings to emphasise specific magnitudes of interest usually involve simulated variable. As a result, this fact is very adverse to the interpretation and choice of the efficiency criteria. In this section, we present a kind of Modified Legates-McCabe efficiency indicators calculated on weighted deviations between the observed and simulated datasets as a alternate solution to address the problem described above. These are the two indicators: LME on deviations weighted by observations cubed (LME_{vh}), LME on deviations weighted by observations to the power of minus three (LME_{vl}). They is described respectively as:

$$LME_{\text{vh}} = 1 - \frac{\sum_{i=1}^{n} \left| O_i^3 \cdot (O_i - S_i) \right|}{\sum_{i=1}^{n} \left| O_i^3 \cdot (O_i - \overline{O_i}) \right|}, \; -\infty < LME_{\text{vh}} \le 1 \qquad (7)$$

$$LME_{\text{vl}} = 1 - \frac{\sum_{i=1}^{n} \left| O_i^{-3} \cdot (O_i - S_i) \right|}{\sum_{i=1}^{n} \left| O_i^{-3} \cdot (O_i - \overline{O_i}) \right|}, \; -\infty < LME_{\text{vl}} \le 1 \qquad (8)$$

The LME_{vh} is used to emphasize errors that occur in very high mignitudes of absolute ***O***. The LME_{vl} is used to emphasize errors that occur in very low mignitudes of absolute ***O***. Through this modifications to the *LME*, the differences between the observed and simulated values are quantified as absolute errors weighted by given observation series, which reduce the adverse effects of factors such as interpretation of indicator measures and the choice of efficiency indicators;

3. APPROACH FOR INDICATORS ANALYSIS

$$\begin{cases} \{x_i;\; i = 1,\ 2,\ \ ,n\} = [-10:0.1:10] \\ O_i = 1.1 + \tanh(x_i) \\ S_i = 1.1 - \tanh(x_i) \end{cases} \qquad (9)$$

In this example, $\{O_i, S_i\}$ generated by equation (9), as shown in Figure 1. The 201 separate synthetic simulated graph were generated by the following processes: for the simulated graph number 1, the first ordinate was the same as the first ordinate of the observed graph and the remaining ordinates of the simulated graph stays the same; for the second simulated graph, the first two ordinate was the same as the first two ordinate of the observed graph and the remaining ordinates of the simulated graph stays the same; for the simulated graph number 3, the first three ordinate was the same as the first three ordinate of the observed graph and the remaining ordinates of the simulated graph stays the same; and in the remaining model simulations (4 to 201), the observed graph values were progressively substituted for the simulated graph until the last model simulation (number 201) was the actual observed graph. Figure 2 shows the observed graph and simulated graph for model simulation number 100 (the first 100 time steps are the same as the observed and the remaining values are the arithmetic mean of entire observed graph).

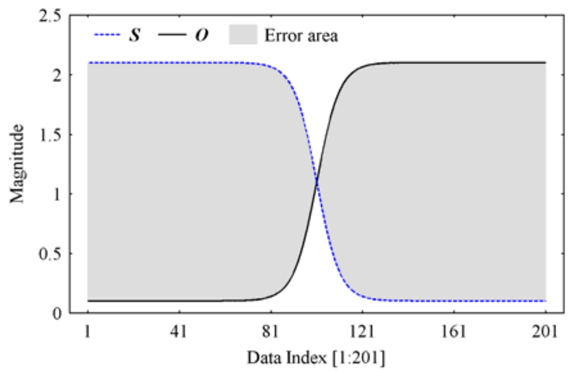


Figure 1: Original Observed (Black Line) and Original Simulated (Blue Dotted Line) Graph and (Gray Area) Error Area

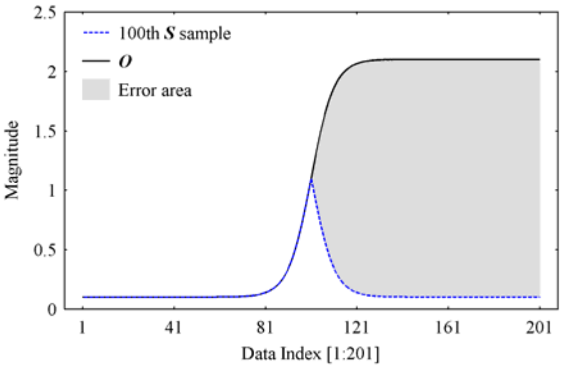


Figure 2: Original Observed (Black Line) and 100th Simulated (Blue Dotted Line) Graph and (Gray Area) Error Area

4. RESULTS AND DISCUSSION

The computed values of the goodness-of-fit indicators described in section 2 for each of 201 separate synthetic simulated graph are shown in Figures 3-4 and Table 1.

The plot in Figures 2 shows the Nash-Sutcliffe efficiency and its modified forms (NSE_{squ}, NSE_{sqrt}, NSE_{ln}, NSE_{inv}), the plot in Figures 4 shows Legates-McCabe efficiency and *LME* on deviations weighted by observations (LME_{vh}, LME_{vl}). Table 1 shows the Sample Index (colums) 1, 41, 81, 121, 161, 201 and the computed values (rows) of the different indicators at Sample Index.

For simulated sample index 1, all goodness-of-fit indicators have negative values, which implies that the simulation model is poor than the use of the mean of the observation series as a accurate representation of real system. The computed values of *NSE*, NSE_{squ}, NSE_{sqrt}, NSE_{ln}, NSE_{inv} adopting squared errors is less than *LME*, LME_{vh}, LME_{vl} adopting absolute errors.

For simulated sample number 1 to 81, there were major increases of of *NSE* (1.76), NSE_{squ} (1.74), NSE_{sqrt} (1.77), NSE_{ln} (1.78), NSE_{inv} (1.74), *LME* (0.85), LME_{vl} (1.94), which indicates that these indicators very sensitive to very low magnitudes conditions. LME_{vh} (0.00) exhibited nearly no reaction on or is not sensitive to very low magnitudes conditions.

For simulated sample number 81 to 121, there were major increases of of *NSE* (0.46), NSE_{squ} (0.45), NSE_{sqrt} (0.41), NSE_{ln} (0.31), NSE_{inv} (0.13), *LME* (0.29) which indicates that these indicators very sensitive to low magnitudes conditions. LME_{vh} (0.23) showed a immediate reaction on or is sensitive to high magnitudes conditions, and slight increases of LME_{vl} (0.03) showed nearly no reaction on or is not sensitive to high magnitudes conditions.

For simulated sample number 121 to 201, there were major increases of of *NSE* (1.76), NSE_{squ} (1.74), NSE_{sqrt} (1.77), NSE_{ln} (1.78), NSE_{inv} (1.75), *LME* (0.85), which indicates that these indicators very sensitive to high magnitudes conditions. LME_{vh} (1.77) exhibited showed a strong reaction on very high magnitudes conditions, and LME_{vl} (0.00) showed nearly no reaction on or is not sensitive to very high magnitudes conditions.

The modified forms LME_{vh} did exhibited nearly no reaction on low magnitudes of observed and simulated series and therefore was mostly sensitive for better model realisation during high magnitudes of observation series, while LME_{vl} did show nearly no reaction on high magnitudes of observed and simulated series and was mostly sensitive for better model realisation during low magnitudes of observation series.

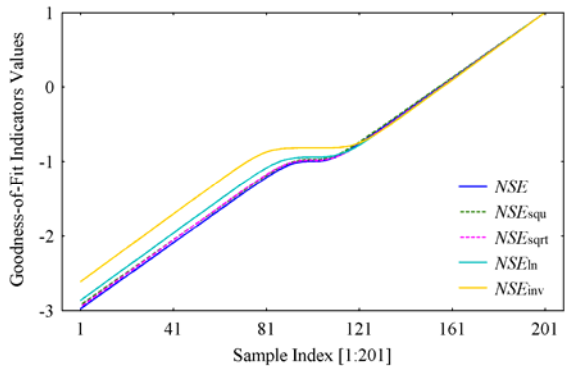


Figure 3: Evolution of *NSE*, NSE_{squ}, NSE_{sqrt}, NSE_{ln}, NSE_{inv} During Example in Sect. 3

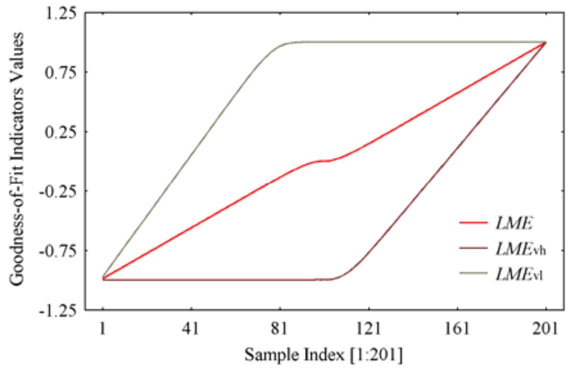


Figure 4: Evolution of *LME*, LME_{vh}, LME_{vl} during Example in Sect. 3

Table 1: Goodness-of-fit Indicators Values for The Six Sample Index of Example

S.S.I.[a]	1	41	81	121	161	201
NSE	-2.98	-2.09	-1.22	-0.76	0.12	1.00
LME	-0.99	-0.56	-0.14	0.15	0.57	1.00
NSE_{squ}	-2.93	-2.05	-1.19	-0.74	0.13	1.00
NSE_{sqrt}	-2.95	-2.06	-1.18	-0.77	0.11	1.00
NSE_{ln}	-2.87	-1.96	-1.09	-0.78	0.10	1.00
NSE_{inv}	-2.62	-1.71	-0.88	-0.75	0.09	1.00
LME_{vh}	-1.00	-1.00	-1.00	-0.77	0.11	1.00
LME_{vl}	-0.97	0.05	0.97	1.00	1.00	1.00

[a] S.I. = Simulation Sample Index.

5. CONCLUDING REMARKS

Eight different goodness-of-fit measures for the evaluation of model performance or validity were investigated with a simple example. Through the computed results of the different goodness-of-fit indicators obtained for a simple example, their measure characteristics in terms of emphasis on different types of errors are analyzed and identified. To increase the sensitivity of indicators measures to very high- and very low- magnitudes conditions and reduce adverse effect on the choice of the goodness-of-fit indicators, LME on deviations weighted by observations cubed (LME_{vh}) and LME on deviations weighted by observations to the power of minus three (LME_{vl}) were proposed. Contrary to what was expected, the results from the a simple example revealed that LME_{vh} did show nearly no reaction on low magnitudes of observed and simulated series and therefore was mostly sensitive for better model realisation during high magnitudes of observation series, while LME_{vl} did show nearly no reaction on high magnitudes of observed and simulated series and was mostly sensitive for better model realisation during low magnitudes of observation series.

ACKNOWLEDGMENTS

The paper was also made possible through the financial support of the National Natural Sciences Foundation of China (Grant No. NNSFC 61374164), for which the authors are most grateful.

REFERENCES

Bennett, N.D., Croke, B.F., Guariso, G., et al., 2013. Characterising performance of environmental models. Environmental Modelling & Software, 40: 1-20.

Crochemore, L., Perrin, C., Andreassian, V., et al., 2015. Comparing expert judgement and numerical criteria for hydrograph evaluation. Hydrological Sciences Journal, 60(3): 402-423.

Dawson, C.W., Abrahart, R.J. and See, L.M., 2007. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modelling & Software, 22(7): 1034-1052

Hauduc, H., Neumann, M.B., Muschalla, D., et al., 2015. Efficiency criteria for environmental model quality assessment: a review and its application to wastewater treatment. Environmental Modelling & Software, 68: 196-204.

Krause, P., Boyle, D.P. and Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5: 89-97.

Legates, D.R. and McCabe, G.J., 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1): 233-241.

Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modéles pluie–débit? PhD Thesis, Université Pierre et Marie Curie, Pairs, France, pp. 324.

Martens, J., Put, F. and Kerre, E., 2006. A fuzzy set theoretic approach to validate simulation models. ACM Transactions on Modeling and Computer Simulation, 16(4): 375-398.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: part I - a discussion of principles. Journal of Hydrology, 10(3): 282-290.

Nicolle, P., Pushpalatha, R., Perrin, C., et al., 2014. Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. Hydrology and Earth System Sciences, 18(8): 2829-2857.

Oudin, L., Andreassian, V., Mathevet, T., Perrin, C. and Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resources Research, 42.

Perrin, C., Michel, C. and Andréassian, V., 2001. Does a large number of parameters enhance model performance? comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology, 242(3): 275-301.

Pushpalatha, R., Perrin, C., Le Moine, N. and Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. Journal of Hydrology, 420: 171-182.

AUTHORS BIOGRAPHY

Kai-Bin Zhao received the M.S. degree in guidance navigation and control from the Harbin Engineering University, China, in 2013. He is currently working toward the Ph.D. degree in control science and engineering at Harbin Institute of Technology, China. His current research focuses on model validation for complex simulation system. His e-mail address is kaibin.zhao.HIT@hotmail.com.

Ke Fang is an Associate Professor of the Control & Simulation Center, School of Astronautics at Harbin Institute of Technology, China. He holds a Ph.D. in control science and engineering from Harbin Institute of Technology, and was a visiting scholar at Arizona State University from year 2014 to 2015 in AZ, USA. His research interests include complex simulation systems, model validation and VV&A. His e-mail address is hitsim@163.com.

Ming Yang is a Professor of the Control & Simulation Center, School of Astronautics at Harbin Institute of Technology, China. He holds a Ph.D. in control science and engineering from Harbin Institute of Technology. He involved in major simulation conference and committees in China. His e-mail address is myang@hit.edu.cn.

MultiMAuS: A Multi-Modal Authentication Simulator for Fraud Detection Research

Luisa M Zintgraf [(a)], Edgar A Lopez-Rojas [(b)], Diederik M Roijers [(c)], Ann Nowé [(d)]

[(a),(c),(d)]Vrije Universiteit Brussel
[(b)]Norwegian University of Science and Technology

[(a)]lmzintgraf@gmail.com, [(b)]edgar.lopez@ntnu.no, [(c)]droijers@ai.vub.ac.be, [(d)]ann.nowe@como.vub.ac.be

ABSTRACT
MultiMAuS is an agent-based simulator for payment transactions, intended for the analysis and development of dynamic on-line fraud detection methods via a multi-modal user authentication system. The multi-modal authentication procedure allows for a flexible number of authentication steps a user has to do before a transaction is processed (or rejected). It can thus adapt to the risk associated with a certain transaction, in the context of a given user. Our simulator is based on real-world credit card transaction data, to realistically model customer behaviour. The simulator can be used to study short and long term consequences of fraud detection algorithms, for different scenarios like varying levels of fraud or authentication steps. The implementation was done in Python, and is publicly available together with aggregated real transaction data (which serves as input to the simulator) and an example simulated transaction log.

Keywords: multi-modal authentication, fraud detection, multi-agent simulation, credit card transactions

Implementation: github.com/lmzintgraf/MultiMAuS

1. INTRODUCTION

Fraud detection research is committed to the development of methods to detect and prevent fraud. However, researchers are frequently hindered by the fact that there is little publicly available data, and it is often difficult to get access to datasets, due to security and privacy concerns. This means that research in this field is either restricted to the few publicly available datasets, synthetic (potentially unrealistic) data, or private datasets which cannot be shared, making it hard to reproduce results or compare different approaches.
One way to circumvent these problems is to use realistic simulations, based on real data that does not have to be shared together with the simulator (Lopez-Rojas and Axelsson 2012a). Such simulators are able to produce transaction logs which are similar enough to real-world data to allow reliable fraud detection research, and can safely be shared publicly. Besides increased reproducibility and comparability of published fraud detection methods, such simulators have additional benefits over real-world data: first, they allow to generate datasets of arbitrary size, which still follow the distribution of the real (potentially smaller) dataset. Second, it allows to test the reliability of methods when the environment changes, by changing the parameters of the simulator. E.g., we can test what happens when the amount of fraud increases, or when fraudulent behaviour changes over time.
In this paper, we are particularly interested in (developing a simulator for) *on-line* fraud detection, specifically in the context of payment transactions. A central issue with fraud detection methods that are heuristics or trained off-line on fixed datasets is a lack of flexibility when deployed in practice. I.e., the authentication strategies are often *uniform* and *static*: the same rules apply to all transactions, and the method cannot adapt when customer behaviour (of genuine or fraudulent customers) changes over time. There exists a trade-off between customer satisfaction and fraud detection however. Genuine customers prefer convenient methods when possible, especially for low-risk, inexpensive or regular transactions. If too much authentication is asked from them frequently, they might get annoyed and divert to other payment options. We thus need models that can learn how genuine users react to authentication mechanisms in the long term, and adapt the authentication level given the current transaction and user, including her past behaviour. Regarding fraudulent customers on the other hand, static authentication rules give fraudsters the chance to adapt their behaviour to potentially circumvent detection methods (e.g., by exploiting that transactions below a certain threshold will not be checked).
Authentication methods should therefore be *dynamic,* i.e., adapt to changes in user or fraudster behaviour, and at the same time consider the long-term consequences of the authentication procedures. The goal of a fraud

detection method is thus to prevent fraud, but also to keep genuine users satisfied, so that they are likely to make future purchases via the same payment processing platform. Testing such dynamic on-line methods can be challenging in practice, because there is a high risk involved in real-life applications, and it takes time to observe long-term consequences (which might be unforeseeable), and we can usually only test one approach at a time.

We therefore present MultiMAuS, a simulator that offers a tool for on-line fraud detection research, by simulating credit card transactions with multi-modal authentication. The authentication process is dynamic in the sense that the number of authentication steps can vary for different transactions and users. The authentication algorithm can decide how many steps of authentication to request to verify a user's identity. E.g., for online credit card transactions the default authentication is just entering the security code of the card, which can easily be compromised. A second authentication step could be in form of a sms-verification, or calling a customer directly to verify their identity. We use an agent-based modelling framework for the simluator, in which customers, fraudsters and merchants interact by making transactions via a *payment processing platform*. The construction of the simulator consists of two steps: the core simulator is generated based on aggregated data from real online credit card transactions, and represents a system with only the default authentication step (presented in section 4.1). We then extend this to multi-modal authentication, and additionally model the patience and satisfaction of the users (section 4.2).

In the following section, we outline related work to simulators and fraud detection. In section 3, we introduce the properties of the real data we base our simulator on, and what exactly we aggregated from this to use as direct input to the simulator. Section 4 then describes in detail how the implementation was done. The performance of the simulators is analysed in section 5. We compare the real and simulated transaction logs to show that the simulator produces realistic data. Furthermore, we show the performance of a few simple authentication mechanisms for multi-modal authentication in terms of revenue.

2. RELATED WORK

Lopez-Rojas and Axelsson (2012a) highlight in their work the challenges of obtaining data in fraud detection research, and propose using synthetic datasets generated by simulation. They propose multi-agent based simulation, which allows to implement simple behaviour for the participating agents (like customers and fraudsters), while leading to a complex dynamic when these agents interact, exhibiting overall behaviour similar to real-world financial transactions.

Previous work on simulators for fraud detection relying on real data exists for, e.g., mobile money payments (Gaber et al. 2013), shoe retail stores (Lopez-Rojas et al. 2013), and money laundering (Rieke et al. 2013). We refer to Lopez-Rojas and Axelsson (2016) for a more extensive overview of the use of simulations in fraud detection research regarding financial datasets. Amongst others, Haq et al. (2016) present a novel technique to automatically generate synthetic data. This however only produces data for batch learning. For an on-line setting on the other hand, we explicitly need independent agents (customers and fraudsters) which make transactions sequentially, and where we can intervene with fraud detection methods.

To the best of our knowledge, our simulator is the first to support multi-modal authentication, and customers changing their purchase behaviour based on their experience. We think it is therefore uniquely suitable for developing and assessing dynamic, flexible on-line fraud detection algorithms.

Our simulator is intended for fraud detection research, and can be used to test both off-line and on-line methods. Surveys of off-line methods are, e.g., Kirkos et al. (2007) or Wang (2010). Research specifically into dynamic, adaptive methods is, amongst others, Alese et al. (2012) and Wheeler and Aitken (2000).

3. DATA

Our simulator is based on real credit card transaction data, provided by a company that processes online credit card payments (and whose identity we cannot disclose). The data consist of around 90,000 non-fraudulent (genuine) transactions and around 1,200 fraudulent transactions from 2016. Each transaction has the following attributes: credit card ID, merchant ID, amount, currency of purchase, date of transaction, card issuing country, fraud label. The labels were obtained by fraud reports made by the credit card companies.

3.1. Data Summary

The number of unique credit cards in the dataset is around 54,000 for genuine, and 800 for fraudulent transactions. Each card is used on average 1.65 times, and most credit cards are only used once (around 65%). Customer activity depends heavily on the hour of the day (most transactions occur in the evening), and the month of the year (most transactions occur in summer). On average, 10.18 non-fraudulent and 0.13 fraudulent transactions are made per hour, which is the time step we chose for our simulator. The transaction amounts range from about 0.5 to 7,800 Euro (after converting everything to the same currency). Purchases are made with credit cards from 126 countries (19 for fraudulent transactions) in 5 (3) different currencies. There are a total of 7 merchants (after removing merchants with less than 100 transactions), of which 6 are affected by fraud. A complete overview of the data properties is given in the results section 5 (table 1), where we compare the real data to simulated data.

3.2. Aggregated Data

To realistically model customer and fraudster behaviour, we aggregated information from the real dataset. Our simulator is built on these, and the aggregated data can be (unlike the raw transaction logs) shared publicly without compromising the privacy of the real customers. In this section we summarise what kind of data is used in our simulator.

Customers. The following information was extracted for genuine and fraudulent customers (separately).

- The total number of transactions in the entire dataset, i.e., in the year 2016.
- The (empirical, discrete) distribution of credit cards over countries (i.e., what is the probability of a customer/fraudster having a credit card from a certain country).
- The distribution over currencies, given the country.
- The distribution of transactions over the merchants, conditioned on the currency. This is to determine where a customer will purchase something.
- The expected number of transactions, given: the hour of the day, the day of the week, the day in a month, and the month in a year.
- The probability of a customer/fraudster making another transaction, given she just made one.
- The average number of days between two subsequent transactions (which is 12.99 days for genuine and 8.94 for fraudulent customers).

To estimate the probability of a customer making another transaction (after just having made one), we only looked at those that did a transaction in the months April and May of 2016, and then estimated how many of these did at least another transaction until the end of 2016. We did this irrespective of how often this card was used again to simplify the computations. The reason we only look at April and May is that the cards used during the first few months of 2016 were *never* used again. We believe this is due to some peculiarity of the data (there were not enough customers, or data is simply missing), since the repetition pattern starting around April is very different, and many genuine customers make several transactions with their cards. For the cards used later in 2016 we do not have enough future data to estimate whether more transactions were made. Due to these restrictions to April and May, we will have more average transactions per card than in the original dataset. This is intended and we think is realistic behaviour of customers. It is important for us to model the transaction patterns of unique customers over time, to simulate how they adapt their behaviour to different authentication models.

Additionally to the above, we estimated the fraction of compromised cards that were used both in fraudulent and genuine transactions in this dataset (around 33%). This will later be used to determine whether fraudsters steal credit card details from existing users, or users outside of the customer pool. Further, we estimated the probability of a genuine customer making another transaction after the card was compromised (which lies only at about 9%).

Merchants. For the simulator, we use the merchants from the true dataset with more than 100 transaction. All merchants with less than 100 transactions were not subject to fraud, and they do not play a noticeable role compared to the other merchants. We do not know whether this is because of a lack of data for these merchants, or whether there is an intrinsic property that makes these merchants less subjective to fraud. One of the remaining seven merchants was also not subject to fraud. For each of these merchants we model the distribution over the price range of their products/services (separately for genuine and fraudulent customers) by using histograms (with 20 bins).

This aggregated information is used as input for the simulator, and released together with our implementation. In the following section, we explain how our simulator utilises the information to simulate transaction behaviour of genuine/fraudulent customers and merchants.

4. MODEL AND IMPLEMENTATION

The simulator was implemented in Python 3 and uses the agent-based modelling framework mesa (Masad and Kazil, 2015). Our implementation is available at github.com/lmzintgraf/MultiMAuS.

In this section, we describe in detail how the simulator was implemented. We do so in two steps; we first present the base implementation with a uni-modal authentication model (section 4.1., UniMAuS), and then the extension of this to incorporate multi-modal authentication (section 4.2., MultiMAuS). We separate the description because the simulator with uni-modal authentication is based on the real transaction logs, and we can then directly analyse whether the simulated data has similar patterns (which we do in section 5.1.).

4.1. UniMAuS

In this section, we describe our base implementation UniMAuS (Uni-Modal Authentication Simulator) - i.e., when there is only one (implicit) authentication step, as in the real transaction logs. By default, all incoming

transactions are permitted and processed without further authentication. This part of the simulator then produces transaction logs that follow patterns similar to the original data. It can be used to produce data sets with realistic properties, of arbitrary size, over any time span. Additionally, different scenarios can be simulated, such as varying levels of fraud.
This base implementation can be used for on-line learning where transactions can either be permitted or denied (without further authentication), or for off-line learning with the produced transaction logs. In our context, its main intention is to serve as a realistic basis for incorporating multi-modal authentication (section 4.2.).

4.1.1. Entities

There are two active agents in the simulator, genuine customers and fraudulent customers (and later also the authenticator, see section 4.2., MultiMAuS). At each time step, these (autonomously) decide whether to make a transaction. Merchants are passive agents, who offer their products to customers, but do not actively execute any actions during the simulation. In the initialisation phase, merchants and customers are created, based on the real transaction data. We describe the details of these agents in the following.

Merchants. We set the number of merchants to the number of merchants observed in the real dataset, restricted to ones with more than 100 transactions. Each merchant further has a sampling function to generate prices. Whenever a customer/fraudster picks a merchant to purchase a product from, the amount for this product is sampled according to this distribution (which differs for genuine and fraudulent customers). Sampling takes place by first randomly selecting a bin for a price range according to the probabilities taken from the histograms of the real data. We then uniformly at random select an amount within this selected price range.

Customers (fraudulent/genuine). Each customer gets a unique credit card ID when instantiated. Further we assign a country to the customer, sampled according to the empirical (discrete) distribution over countries. Conditioned on this country, we select a currency (again, sampled from the empirical distribution). Finally, each customer has a scalar value we call the 'intrinsic transaction motivation'. This is for all customers set to 1 divided by the total number of customers in the simulation. This way we can make sure that the total number of transactions is approximately the same as in the true data, and have customers that do decisions *autonomously* (without looking at what other agents do, or having a central mechanism that decides which customers make transactions).

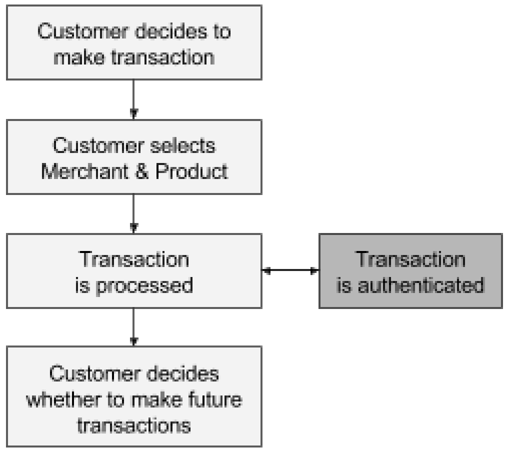


Figure 1: Action flow during one simulated hour. The step "Transaction is authenticated" is optional in the MultiMAuS model. In the UniMAuS model, each transaction is processed without further authentication.

Fraudulent customers. When credit card IDs for fraudsters are initialised, the card ID is sometimes (according to the fraction observed in the real data) taken from the pool of genuine customers. In that case, the country and currency are taken from that customer (the country must be the same since it the card issuing country, and the currency only changed in 3 out of 250 cases in the real dataset, so we chose to disable that option). Otherwise, country and currency are assigned like described above.

The intrinsic transaction motivation makes sure that the total number of transactions is (with the default parameters and little noise) similar to what we observed in the real data. By varying the starting number of customers in our simulator, we can influence how much time lies between two transactions of one customer (the more customers exist, the less often an individual customer will have to make a transaction so that we match the true total number of transactions). We found empirically that starting with 3333 genuine and 44 fraudulent customers works well in this respect. Note that the number of customers can change over time, but that the intrinsic transaction motivation (which in this case is 1/3333 and 1/44) stays the same. This means that when more customers use the payment processing platform, we also get more transactions.

4.1.2. Transactions

Transactions are made sequentially, where one time step represents one hour in Pacific Standard Time. Figure 1 shows a simplified overview of the steps during one such hour. In the following, we describe this in detail.

1. *Start transaction.* Each customer (genuine or fraudulent) decides whether to make a transaction or not. This is conditioned on the current *local* time of the customer (i.e., we

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

convert the current global time according to the customer's country) and the probabilities of making transactions from the true data (section 3.2). This is then weighted by the intrinsic transaction motivation. The resulting transaction probability is used to (stochastically) determine whether a user will make a transaction.

2. *Pick product.* Each customer/fraudster that decides to make a transaction then picks a merchant. This is done by sampling a merchant from the empirical distribution, conditioned on the currency. In a next step, the amount of the product is determined by sampling from the amount distribution of the merchant.
3. *Process transaction*. After all customers decided whether to make a transaction, these are processed in random order and log entries for each transaction are made. In UniMAuS, no fraud detection mechanism or extra authentication is executed or intervenes, i.e., all transactions will be processed. This also means that all attempted fraudulent transactions will succeed.
4. *Customer migration*. When all transactions have been processed, customers decide whether they will make a transaction in the future (then they stay in the customer pool) or not (in which case they are removed from the customer pool) - again, based on the empirical probability of making another transaction. For genuine users, staying also depends on whether their credit card has been used in a fraudulent transaction or not. There are also new customers added to the customer pool. To compute how many new users join, we use the expected number of users leaving, which is determined by the transaction probability given the *global* time, and the intrinsic motivation of customers/fraudsters. The size of the user pool therefore varies over time, but will stay around the initial numbers 3333 (genuine customers) and 44 (fraudulent customers) with the default parameters and little noise.

4.2. MultiMAuS

Our simulator is intended for the development of *dynamic fraud detection* systems. To this end, we extend our base implementation to support multi-modal authentication. In this section we explain the additional components that were used to realise this.

4.2.1. Entities

Authenticator. The authenticator handles the security measurements taken for each transaction. I.e., the authenticator can evaluate each transaction, and then decide to permit it, request additional authentication, or deny the transaction. For our current implementation, we give the authenticator the option to either permit the authentication, or ask for an additional authentication which is either provided by the user or denied (in which case the transaction gets cancelled). Whenever the second authentication is provided, the authenticator permits the transaction. Note that adaptations to this could be that the authenticator has the option to deny transactions, or that authentications have some sort of quality measure (e.g., how close is a signature to the original signature of the user?). We plan to implement and studies such scenarios in future work.

Genuine customers. Genuine customers are extended to support multi-modal authentication in the following way. During initialisation, a customer gets a *patience* and a *satisfaction* value (both values between 0 and 1). The patience is initialised randomly from a right-skewed beta distribution. The patience gives an indication of how likely the user is to accept more authentications, which will also depend on the transaction amount (see next section).
The satisfaction at the beginning of the simulation is set optimistically for all customers (for our example experiment, we used 0.9). For customers that are added to the pool of customers later in the simulation, it is set to the mean satisfaction of all customers in the pool to mimic the influence of existing customers on new customers via sharing opinions. The satisfaction of the user changes over time with the experiences the user makes with the service. The satisfaction then has an effect on how likely a user is to make additional transactions in the future. How exactly the patience and satisfaction influence the transactions is explained in the next section, 4.2.2.

Fraudulent customers. The properties of fraudster do not change compared to the uni-modal scenario, except that they can now also react to additional authentication (see next section). In the current implementation, fraudsters will always deny a second authentication, since we assume they cannot provide it.

4.2.2. Transactions

In this section, we describe what changes regarding the transactions described for the uni-modal authentication (section 4.1.2.).

1. *Start transaction.* What changes compared to the uni-modal authentication scenario is that the satisfaction of a genuine customer now also plays a role when deciding whether to make a transaction or not. In practise, the transition probability which is computed for the UniMAuS simulator is multiplied by the customer's satisfaction.

1. *Pick Product.* (Same as above.)
2. *Process Transaction.* For each transaction that a customer wants to make, the authenticator can now decide whether to request an additional transaction. Genuine customer will provide this with a probability that is computed by taking the average of the customer's patience and the current transaction amount divided by the maximum possible amount at the given merchant. Fraudulent customers will always cancel the transaction if asked for a second authentication.
3. *Agent migration.* The probability of making another transaction is now multiplied by the satisfaction for genuine customers. I.e., satisfied customers will be more likely to make additional transactions in the future.

The customer's satisfaction changes after every transaction, depending on their experience. If the transaction was successful and there was no additional authentication, the satisfaction goes up by 1%. If however the user provided a second authentication, the satisfaction goes down by 1%. If the user cancels a transaction instead of providing a second authentication, the customer's satisfaction goes down by 5%. We chose these numbers so that there is a clear challenge in making good authenticators that balance the customer satisfaction and pure fraud detection. We show the behaviour of simple authenticators in section 5.2. Unfortunately real data on customer behaviour regarding the reaction of second authentications (cancelling or providing the authentication) was not available to us during the development of this simulator, and we hope to get access to and incorporate real data regarding this in the future.

5. RESULTS

In this section we analyse the performance of our simulator; first to see how well the UniMAuS implementation fits the real data, and second how the MultiMAuS implementation reacts to different simple authentication mechanisms.

5.1. Evaluation: UniMAuS

We compare the real transaction logs (which are private and cannot be shared) to simulated transaction logs from one run with default parameters. These are published together with the implementation, see files 0_transaction_logs.csv and 0_parameters.pkl.
Table 1 shows an overview of the properties from the transaction logs, of the real and the simulated outputs. Running a simulation of one year takes around twenty minutes. We can see that the simulated transaction logs have properties that are close to the real dataset. The largest difference is in the maximum transactions for

Table 1: Overview of summarised data statistics from the real versus simulated transaction logs, taken from a single run of the simulation for the year 2016.

	Customers		Fraudsters	
	real	simul.	real	simul.
transactions	89,194	88,603	1,163	1,239
trans./hour	10.15	10.09	0.13	0.14
trans./month	7432.8	7383.6	96.9	103.3
num cards	54,133	44,817	799	826
num cards, single use	34,873	22,548	543	553
num cards, multi use	19,260	22,269	256	273
fraud cards in genuine	-	-	33%	34%
min amount (€)	0.44	0.44	1	1
max amount (€)	7,835.6	6,276.3	2,737.0	2,685.1
avg amount (€)	298.5	305.6	61.7	69.2
num merchants	7	7	6	6
countries	126	105	19	15
currencies	5	5	3	3
max trans/card	109	17	15	7
avg trans/card	1.64	1.98	1.46	1.50

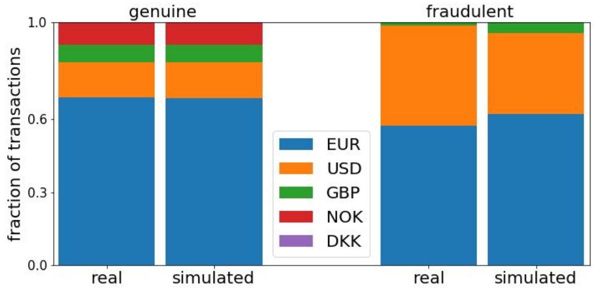


Figure 2: How transactions are distributed over the different currencies. The two columns on the left show this for genuine transactions, comparing the real vs the simulated data. On the right we show this for fraudulent transactions. Note that the currencies NOK and DKK are not present in the fraudulent dataset.

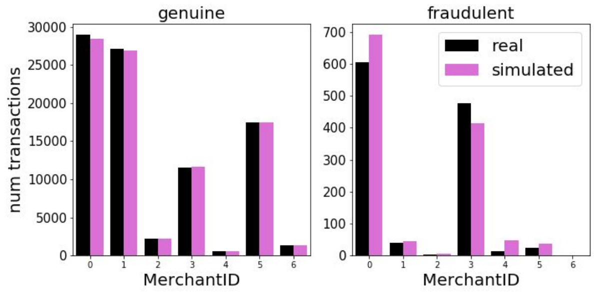


Figure 3: Total number of transactions per merchant in 2016, for genuine (left) / fraudulent (right) transactions. We compare the transaction logs of the real data set (black) and one simulated dataset (magenta).

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

one single card. Since after each transaction, a customer decides whether to make another transaction (based on true recurrence rates), making as much as 100 transactions per card is unlikely, which is in part due to the way we computed the recurrence rates (i.e., not looking at the number of recurrency but only binary indication, see section 3.2).

Figure 2 shows the distribution of all transactions over currencies. There is a clear difference between genuine and fraudulent transactions: the fraction of transactions in US dollars is much higher in fraudulent transactions. Further, fraud is only done in the three most popular currencies. We also see that the real and simulated data have matching behaviour, with the noise evening out more for (the large number of) genuine transactions.

Figure 3 shows the total number of transactions per merchant (for 2016). The transaction varies a lot over the merchants, with some merchants processing only a few hundred transactions in the entire year 2016, and others several tens of thousands. This is true for both the genuine and fraudulent transactions. The figure shows that this pattern remains also in the simulated data.

Figure 4 shows the the distribution over transaction amounts, summarised for the year 2016 (in Euro, after converting everything to this same currency). About half of all genuine transactions are worth under 100 Euro. Interestingly, the amounts of fraudulent transactions are generally lower, with only about 10% of all transactions over 1000 Euro. This could be due to fraud detection mechanisms (and detected fraud not showing up in our data), or because fraudsters anticipate higher security measures for higher payments.

Figure 5 and 6 show the transaction activity over time, for the month in a year (figure 5) and the hour of the day (figure 6). We can see that the true and simulated data show similar transaction behaviour over time on average. Note that the simulator follows the real data distribution on *average*, but that individual customers (green lines) can differ. This is because when initialising customers, we add some noise to the transaction probabilities to simulate individual behaviour. This noise is higher on small time scales (like hours or days in the week) than on large time scales (like months in a year). In practise this is achieved by sampling from a multivariate Gaussian distribution with the mean being the distributions observed in the real transaction logs and a small variance (of 0.1 in the default case).

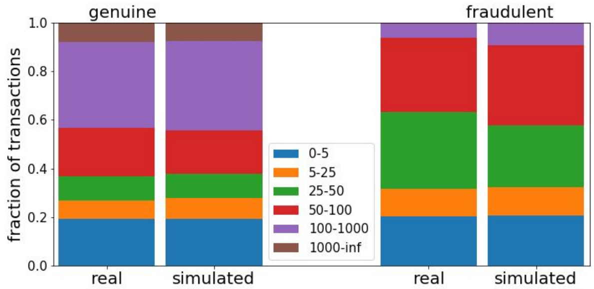


Figure 4: Distribution of transaction amount in Euro). The two bars on the left show the real and simulated data distribution for genuine transactions. The two bars on the right show this for fraudulent transactions.

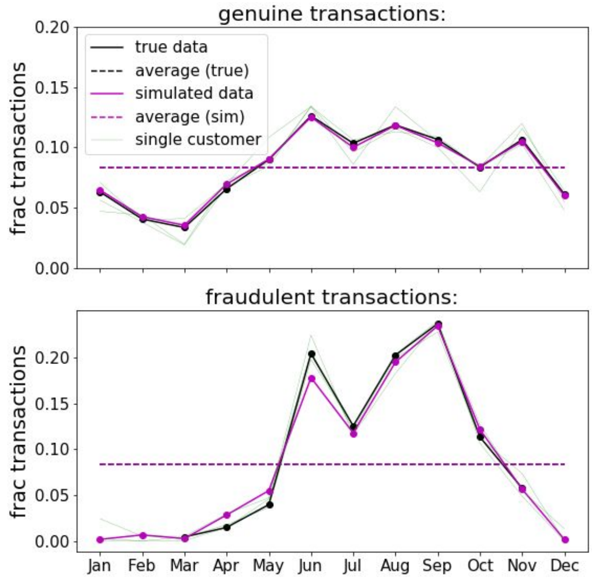


Figure 5: Fraction of transactions per month in a year, for the real data (black) and the simulated data from one run (magenta). We show a few examples of this distribution for a single user in the simulator.

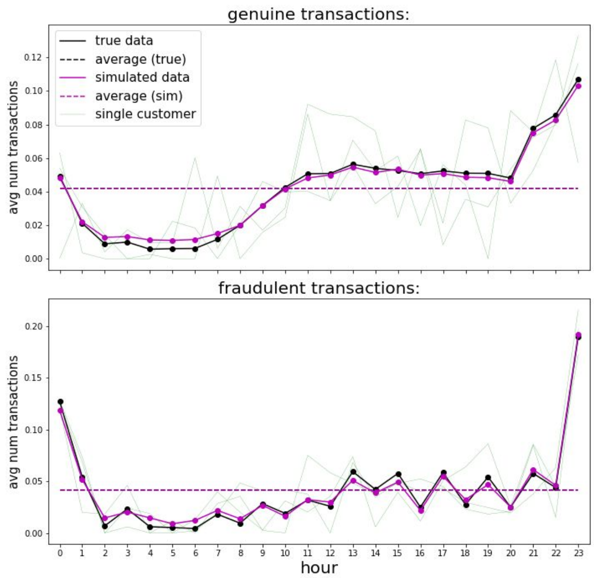


Figure 6: Fraction of transactions per hour in a day, for the real data (black) and the simulated data from one run (magenta).

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

5.2. Analysis: MultiMAuS

We tested several simple authentication mechanisms on the multi-step authentication simulator. We look at the problem from the view of the payment processing platform, i.e., in terms of business revenue, and calculate the reward of an authentication mechanism as follows. For each successful genuine transaction, the reward is 0.3% of the transaction amount, plus 1 cent (all calculated in Euro). If a fraudulent transaction is not detected, the reward is the negative amount of the transaction, since this has to be paid back to the genuine user whose credit card was compromised.

Figure 7 shows the results in terms of reward from one experimental run, where each run is initialised with the same random seed. We tested five different simple authentication mechanisms, which are the following.

- *Oracle.* The oracle agent has access to the true labels. It never asks a genuine customer for a second authentication, and always asks fraudulent customers for a second authentication. This means that fraudsters will always cancel the transaction (since they cannot provide further authentication), and that user satisfaction is as high as possible. In other words, this is an unrealistically good policy, that provides us with an upper bound on the attainable revenue.
- *Random.* The random authenticator asks for a second authentication with a probability of 0.5. If the second authentication is provided, the transaction is authorised.
- *Heuristic.* The heuristic authenticator only asks for a second authentication if the transaction amount lies above 50 Euro.
- *Never second.* This agent permits all incoming transactions without further authentication. This leads to high satisfaction in the users and therefore relatively more transactions. However as fraud increases (as it does in the summer), the monetary loss outweighs the gain.
- *Always second.* This agent always asks for a second authentication for all transactions. This means that all fraud will be prevented, but that customers also get unsatisfied over time.

As we can see in figure 7, the agent that never asks for a second authentication has the lowest cumulative reward. This is due to the increased number of fraud during summer (where it starts to decline) and because the customer satisfaction does not grow fast enough to make up for this. The authenticator which always asks for a second authentication gains more cumulative reward, but is still significantly below the oracle agent. This is due to the lower customer satisfaction, leading to customers making less transactions.

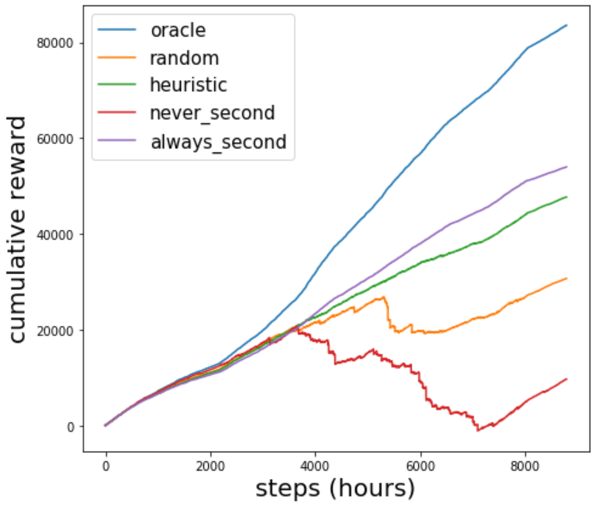


Figure 7: Cumulative reward for different simple authentication agents over one simulation run (one year, where each time step is one hour).

6. CONCLUSION AND FUTURE WORK

We introduced MultiMAuS, a simulator supporting multi-modal authentication systems, which can be used for fraud detection research. We used an agent-based framework, including merchants and customers (genuine and fraudulent) which interact with merchants via a payment processing platform. The simulator is based on real-world data for uni-modal authentication, and on top we added the option for multi-modal authentication. Our results show that the simulated transaction logs produce data which follows similar patterns as the original data. We conclude from this that our implementation realistically simulates real transaction behaviour of genuine and fraudulent customers, and is therefore suited for fraud detection research.

Looking at several simple authentication mechanisms, we see that there is need for more sophisticated (learning) algorithms that fill the gap between the oracle and the other authenticators. We believe that reinforcement learning (RL) algorithms are well suited for this, since they can take into account long-term effects of their actions (like the effects on customer satisfaction and resulting transaction behaviour). There are several interesting directions for future research using the presented simulator that we want to investigate.

One interesting line of future work is in the context of safe RL (Garcia and Fernández 2015). When deploying learning algorithms in practise, we want to make sure the policy acts within reasonable bounds and follows certain constraints regarding security. E.g., we need to have some guarantees on their performance in the sense that we do not want them to perform (much) worse than the policy currently used in practise (e.g., a heuristic policy) which we want to replace (Thomas 2015).

One way of approaching the fraud detection problem with multi-modal authentication is treating it as a multi-objective decision problem (see, e.g., White and Kim 1980, Roijers and Whiteson 2017). I.e., we explicitly try optimising not only the cumulative reward but at the same time the satisfaction of the user. User satisfaction can indirectly be measured by the number of authentication steps required for transactions, or for example by a (sparse) reward signal via direct user feedback (e.g., through user surveys).

We also plan to extend the simulator by implementing adversarial fraudsters that actively adapt their behaviour to circumvent the security measures to commit fraud, leading to a multi-agent setting (Vlassis 2007, Nowé et al 2012, Littman 1994). This is important to study how algorithms can deal with concept drift.

Another aspect to consider is that the customer's satisfaction cannot be observed directly, and thus the problem can be formulated as partially observable (Kaelbling et al. 1998), possibly in combination with multiple objectives (Roijers et al. 2015) and/or multiple agents (Oliehoek and Amato 2016).

We believe there are more interesting opportunities for future work, and hope to provide a useful tool for the development of fraud detection algorithms. We would like to note that since the customer behaviour in terms of satisfaction and patience was not build on real data, we would advise researchers to test algorithms on different initial satisfactions, or update rules for this. We will try to in the future incorporate expert knowledge or information from real data on customer behaviour in this respect.

ACKNOWLEDGMENTS

This research was supported by Innoviris, the Brussels Institute for Research and Innovation, Brussels, Belgium. Diederik M. Roijers is a postdoctoral fellow of the Research Foundation - Flanders (FWO). We would like to give special thanks to the company (identity undisclosed) that gave us access to their data.

REFERENCES

Alese BK, Adewale OS, Aderounmu GA, Ismaila WO, Omidiora EO. 2012. Investigating the Effects of Threshold in Credit Card Fraud Detection System. In International Journal of Engineering and Technology, 2.

Gaber C, Hemery B, Achemlal M, Pasquet M, Urien P. 2013. Synthetic logs generator for fraud detection in mobile transfer services. In Collaboration Technologies and Systems, International Conference, pp. 174-179. IEEE.

Garcıa J, Fernández F. 2015. A comprehensive survey on safe reinforcement learning. Journal of Machine Learning Research, 16 (1), 1437-80.

Haq IU, Gondal I, Vamplew P, Layton R. 2016. Generating Synthetic Datasets for Experimental Validation of Fraud Detection. In Conferences in Research and Practise in Information Technology, 170. Fourteenth Australian Data Mining Conference, Canberra, Australia.

Kaelbling LP, Littman ML, Cassandra AR. 1998. Planning and acting in partially observable stochastic domains. Artificial intelligence, 101(1), 99-134.

Kirkos E, Spathis C, Manolopoulos Y. 2007. Data mining techniques for the detection of fraudulent financial statements. Expert systems with applications, 32(4), 995-1003.

Littman ML. 1994. Markov games as a framework for multi-agent reinforcement learning. In Proceedings of the eleventh international conference on machine learning, 157, 157-163.

Lopez-Rojas EA, Axelsson S. 2012a. Money laundering detection using synthetic data. In The 27th annual workshop of the Swedish Artificial Intelligence Society (SAIS), 071, 33-40. Linköping University Electronic Press.

Lopez-Rojas EA, Gorton D, Axelsson S. 2013. RetSim: A ShoeStore Agent-Based Simulation for Fraud Detection. In 25th European Modeling and Simulation Symposium, 25-34.

Lopez-Rojas EA, Axelsson S. 2016. A review of computer simulation for fraud detection research in financial datasets. In Future Technologies Conference (FTC), 932-935). IEEE.

Masad D, Kazil J. 2015. Mesa: An Agent-Based Modeling Framework. In 14th PYTHON in Science Conference, 53-60.

Nowé A, Vrancx P, De Hauwere YM. 2012. Game theory and multi-agent reinforcement learning. In Reinforcement Learning, 441-470. Springer Berlin Heidelberg.

Oliehoek FA, Amato C. 2016. A concise introduction to decentralized POMDPs. Springer International Publishing.

Rieke R, Zhdanova M, Repp J, Giot R, Gaber C. 2013. Fraud detection in mobile payments utilizing process behavior analysis. In Availability, Reliability and Security (ARES), 2013 Eighth International Conference, pp. 662-669. IEEE.

Roijers DM, Whiteson S, Oliehoek FA. 2015. Point-Based Planning for Multi-Objective POMDPs. In IJCAI 2015, 1666-1672.

Roijers DM, Whiteson S. 2017. Multi-objective decision-making. Synthesis Lectures on Artificial Intelligence and Machine Learning.

Thomas PS. 2015. Safe reinforcement learning. Thesis (PhD). University of Massachusetts Amherst

Vlassis N. 2007. A concise introduction to multiagent systems and distributed artificial intelligence. Synthesis Lectures on Artificial Intelligence and Machine Learning, 1(1), 1-71.

Wang S. 2010. A comprehensive survey of data mining-based accounting-fraud detection research. In Intelligent Computation Technology and Automation (ICICTA), 1, 50-53. IEEE.

Wheeler R, Aitken S. 2000. Multiple algorithms for fraud detection. Knowledge-Based Systems, 13(2), 93-9.

White CC, Kim KW. 1980. Solution procedures for vector criterion Markov decision processes. In Large Scale Systems, 1(4), 129-40.

SUGAR PLANT SIMULATOR FOR ENERGY MANAGEMENT PURPOSES

Cristian Pablos [a], L. Felipe Acebes [b], Alejandro Merino [c]

[(a)(b)] Systems Engineering and Automatic Control Department, EII, University of Valladolid, C/ Real de Burgos s/n, 47011, Valladolid, Spain
[(c)]Electromechanical Engineering Department, University of Burgos, Av. Cantabria s/n, 09006 Burgos, Spain

[(a)]cristian.pablos@uva.es, [(b)]felipe.acebes@eii.uva.es, [(c)]alejandro@cta.uva.es

ABSTRACT
This paper deals with the development of a simulator for energy purposes. The aim of the simulator is being used as a substitute of a real sugar plant in a testing phase of a Real Time Optimization system for the management of the power energy in the plant, and the use of steam in the process. An Equation Based Object Oriented Language (EBOOL) has been used, so different models of each part of the production process have been modelled separately and then joined together to build the final model. To do so, first principle models are combined with others based on experience depending on the accuracy needed, in order to maintain always equilibrium with computing performance. Several tests, changing between different operational points were performed with the simulator to verify its robustness and performance, and to check if it is suitable to be used by the optimization system.

Keywords: Sugar factory simulator, CHP plant, Grey Box Modeling, Energy optimization.

1. INTRODUCTION

Nowadays, while the complete development of renewable energies is hoped, cogeneration, which can be defined as the simultaneous generation of heat and power, has raised as a very important candidate to energy production in industrial processes. The high efficiency inherent to this process makes possible to save money reducing the amount of fuel used to obtain energy in comparison with conventional plants. Furthermore, this feature also implies a reduction in emissions, something that is very interesting as global leaders are putting their attention on new laws to stop global warming. As a production point of view, it also provides more security in the supply of heat and power energy to the process, so it can be very interesting study how processes associated with this type of technology can be optimized (Tina, G.M. Passarello G. 2012).

One problem related to this type of processes is that changes in energy prices, and restrictions in the way energy can be obtained makes them compulsory to be operated in real time. However, due to the amount of variables involved, it is very hard to find the optimum operational point which gives the maximum profits. Different approaches have been found trying to solve a problem where the energy costs wanted to be optimized in non-related industrial processes applications, such as district heating systems (Ristic M. et al. 2008) or power plants based on this technology (Mitra S. et al. 2013), (Sanaye S. and Nasab A.M. 2012). Other studies put their attention on petrochemical or other type of industries where the production is fixed and hence the power and heat demand (Ashok S., Banarjee R. 2003). However, few studies consider the possibility of changing the production in order to reduce the energy costs and make more profit. To solve this problem a RTO (Real Time Optimization) System (Darby M.L. et al. 2011) has been thought to help managing the production, taking into account considerations aspects like the possibility of selling or buying power energy to the external grid. As case study, a sugar process industry has been selected because of its traditional relationship with cogeneration and the extensive experience of the working team with this type of process industry.

Tools based on optimization must be proved on simulation before trying them on real processes plants. This way, security and production problems can be avoided saving time and money. Taking this into account, the global problem can be divided in two main parts. On the one hand, the optimization problem has to be study deeply in order to understand the different variables and aspects that take place. The formulation of the problem is a key part in the development of the global work, therefore different approaches and optimization strategies must be studied.

On the other hand, a sugar plant dynamic simulator focused on energy aspects has to be modelled in order to test the optimization and replace the real sugar plant. This simulator will be also useful to study the process, and make experiments related to the effect that different key variables may have in energy consumption and management. Previously, the working team developed a complete sugar factory simulator for operators training (Merino A., et al 2009). Nevertheless, because of the complexity of the models used, the simulation can hardly run faster than real time, which makes very difficult carry out tests in an agile way.

Furthermore, as was aforementioned this simulator is focused on training operators, so the process is not completely automated and the simulation cannot run without human intervention. Besides, the simulator does not include the beet storage and the factory power demands. All this problems makes recommended the construction of a new simulator based on the previous one, complex enough to represent the behaviour of the global process, but simpler and faster. So the objective of this work is the development of that simulator and perform tests in order to analyse whether it can be applied as a replacement of the complete sugar factory in the RTO system.

The rest of this paper is organised as follows: In the second section, a description of typical beet sugar process is given to help to understand the different parts of this process and the most important variables associated. Next, the third section deal with the plant mathematical model and its implementation to perform the simulator. Later, some simulation tests are shown to demonstrate the simulator features, and finally some conclusions are stated.

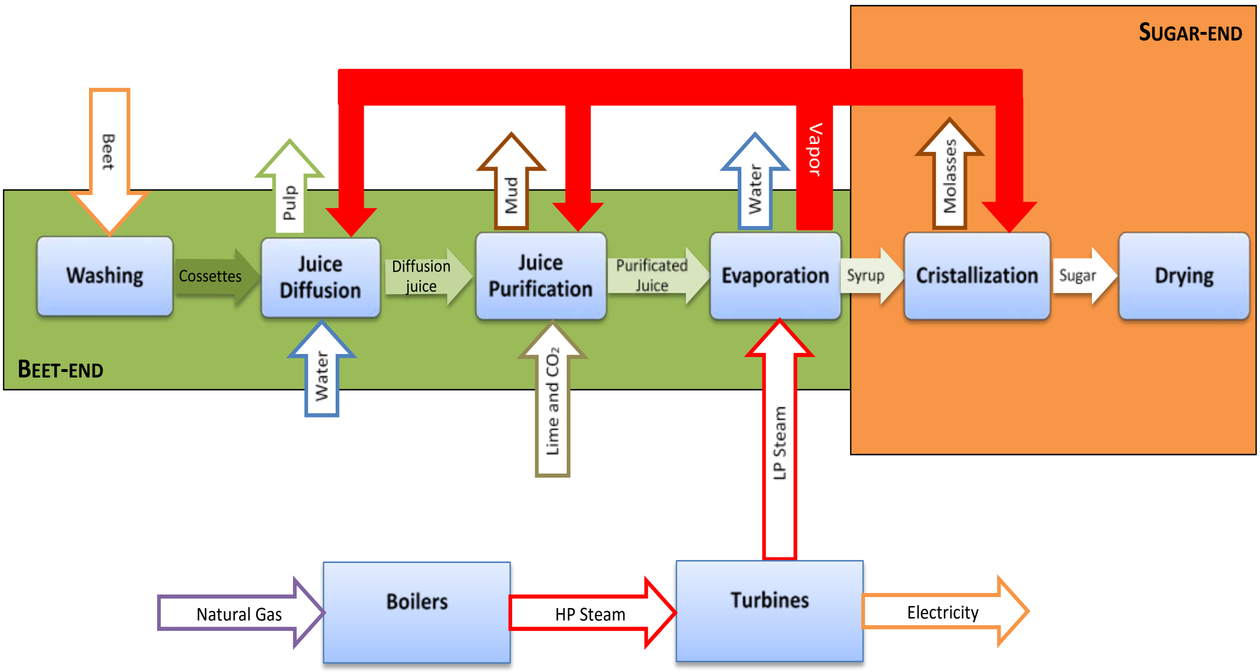


Figure 1. Simplified schematic of a typical sugar factory with a CHP plant

2. SYSTEM OVERVIEW

The manufacturing of raw sugar from beets is a highly technical process which requires large and costly manufacturing equipment. A typical beet-sugar factory is divided into two sections: beet-end or raw-side and sugar end or refinery-side (Asadi M., 2007). Besides, steam boilers and steam turbines are necessary to produce low pressure steam and electricity. Figure 1 shows a simplified block diagram of the production.

The factory takes delivery of the beets through lorries, which are unloaded and beets are piled and stored in beet-storage areas. The storage time is an important variable that affects sugar contain-reduction of the beets due to the beet respiration and microorganism. Other important constraint is the available surface of the storage areas. So, these issues affects to the beet feeding to the factory.

The production process begins when the beet is washed and cut to obtain thin slices called cossettes. In the diffusion section, cossettes are contacted with hot water flowing countercurrently to extract the sucrose from them and the diffusion juice is obtained. This process is carried out into the diffuser, whose electrical drives are one of the main power consumers of the factory. The pulp, or sugar-exhausted cossettes, is compressed, dried and processed as animal feed. The pulp presses are other important power consumers.

In the purification section, the diffusion juice is heated and lime added to control the juice pH. After two carbonation process, the juice impurities are removed by filtration and the obtained clear or purified juice is used to feed the evaporation section.

In the diffusion and purification sections the main steam consumption is due to the heat exchangers that heats the different juices.

The clear juice feed the evaporation section which purpose is to remove part of the water in the juice, increasing its Brix, or percentage of dry substance, to obtain syrup. This process is carried out in a multiple-effect evaporation, in which the juice is boiled in a sequence of vessels. The first vessel (at the highest pressure) requires steam from turbines and the vapor boiled off in each vessel is used to heat the next as well as to feed other steam consumers of the factory. So, this section is the largest steam consumer of the factory.

The obtained syrup feeds the sugar-end, where sucrose is crystallized to obtain granulated-refined sugar and molasses (a by-product). The process is divided in three stages, in each stage vacuum pans are used to crystallized the sucrose and, later, centrifuges separate the sugar crystals from honeys. The commercial sugar is obtained in the first stage, the next ones are recovery stages to minimize the sugar losses. The sugar-end is very complex to operate, because the majority of crystallizers and centrifuges are batch process units, and the product storage capacity is limited. Crystallizers are the greatest consumers of the evaporation steam, and, together with the centrifuges, consume a lot of power. Besides, as they are batch units these demands are discontinuous in time and magnitude.

To supply the demands of steam and power, the factory disposes of a cogeneration, or combined heat and power (CHP), plant. Using natural gas, or other fuel, boilers generate high pressure steam that is sent to steam turbines to generate electricity, and low-pressure steam for the evaporation section. The CHP plant is working to produce the steam and power that the factory needs. But, sometimes, there will be available power to sell, and other times it will be necessary to buy power to the electrical companies.

After this description, it is easy to imagine that a sugar factory is a good example to schedule the production, keeping in mind the prices of the fuels and power, the deliveries of sugar beets and the storage capacity of the factory. So, in order to test different kinds of scheduling strategies it is necessary to dispose of a good simulator to replace a real factory.

3. MODELING AND SIMULATION

3.1. Modeling and simulation environment

It is well known that the first stage of a simulation project is to establish the simulation aims. In this case, the final objective is to have a dynamic and realistic simulation of a well-defined industrial process to obtain a set of experimental data to identify reduced models of the system, develop a RTO algorithm, and test the implementation of the RTO in an industrial control system.

As scenarios of several days of length must be simulated, it is necessary a simulator simple enough to carry out a lot of experiments with short computing time. However, an excess on simplification can lead to a problem related to lack of accuracy, so it must be managed an equilibrium between accuracy and computation time.

Then, the mathematical model must be a first principles dynamic one, but in case the difficulty of the model exceeds the intended use, sufficiently validated empirical relations will be used. This implies that the formulation of the model must be based on ODEs, DAEs, algebraic equations, tables and events.

Given the magnitude of the system to be simulated, its mathematical model must be developed incrementally, using models that can be instantiated from a library and composed in a hierarchical way.

A requirement for the simulation tool is the capacity to develop model libraries from scratch. Although using commercial libraries can be helpful sometimes, since a gain in time and no errors are expected, sometimes they limit the characterization of the simulated process and in certain cases, like the sugar industry, these libraries does not exist. Another requirement is that the modeling tool should help in the task of symbolic manipulation of the global mathematical model to obtain in an easier way the simulation model based on the input variables.
These requirements imply to choose a modeling environment that implements an equation object oriented modeling language (EOOBML), and a graphical interface that allows the modeling based on the connection of components.

The modeling environment must be linked to a simulation environment that facilitates model execution, changes in parameter and input signals, and visualization of output signals. Also, it must allow the recording of the results, so that the data can be exploited with another tools.

The simulation must be able to be used standalone the modeling and simulation environment. In particular, it is desired to use a modeling and simulation environment that allows to obtain standalone simulators with OPC (OLE for Process Control) connectivity, since OPC is supported by industrial control systems and by other tools such as MATLAB.

So, the modeling and simulation environment selected is EcosimPro (EA Internacional 2017), since it disposes of the aforementioned specifications. In particular, its modeling language shares many characteristics with languages based on Modelica (Modelica Association et al 2017).

In this project, a set of libraries of sugar process unit models, developed by the authors in previous work, has been used (Merino A. et al 2009). However, new component models have been added (like power turbines), and power related equations have been included to some existing ones.

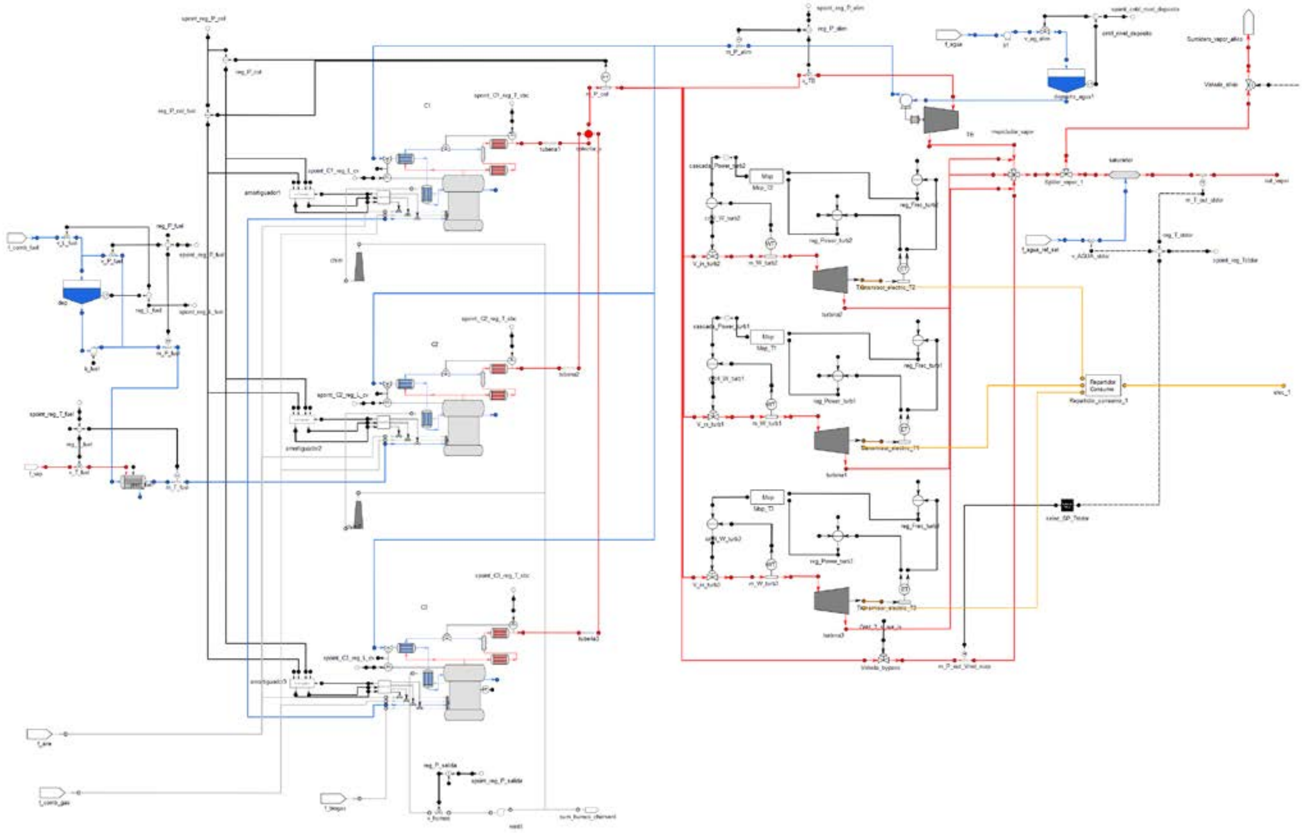


Figure 2: Example of EsosimPro graphical interface. CHP plant.

3.2 Model description

The mathematical model of the simulator is a grey one which does not represent a concrete sugar factory, but reflects a general factory. The complete model uses detailed first principle models where more accuracy is needed (cogeneration plant and evaporation stage), and reduce ones when only is necessary calculate heat and power demands (beet storage, diffusion, purification, crystallization stages and power consumption).

Power consumption

One of the simplest models developed has been the power consumption of the entire factory. A typical sugar factory is composed by many process equipment which needs power supply to work. According the aim of the simulator, modeling the individual power consumption of each one is not a suitable option because some components of the model libraries would have to be modified, each one of the model instances parametrized and, when the electrical connections were added, the flow process diagrams would be more complicated. Thus, to reduce the complexity of the model, energy consumption is calculated for each of the main stages of the plant (diffusion / drying pulp, purification, evaporation, beet and cogeneration plant). The overall energy consumption will be the sum of the consumption of each stage.

The average total power consumption depending on the production has been considered (Urbaniec K. 1989). As it can be seen in Figure 3, when the production is increased the specific power demand is reduced. The consumption of the main stages of the plant has been calculated applying typical percentage values over the global consumption (Frankenfeld T., Voβ C. 2004). So, it is possible to model the power consumption of each stage depending on the flow of products that manage. It is a natural way to incorporate the dynamic of the global power consumption. For instance, when the beet feed is increased, first the power consumption of the diffusion increases, later the power consumption of the purification increases, and so on.

Furthermore, regarding the RTO system, this feature will allow in future work to study a mixed-integer problem where the optimizer will have to decide which stages of the plant must be connected to the external grid.

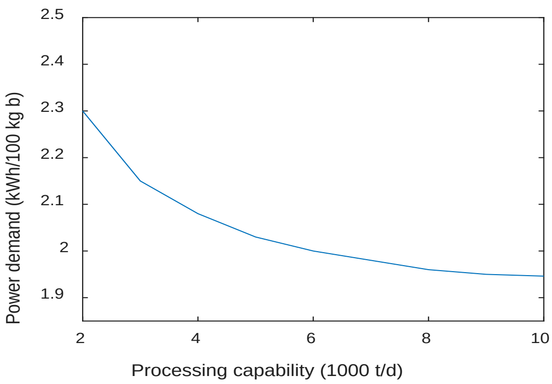


Figure 3: Specific power demand as a function of the processing capability.

Beet reception and storage

As mentioned above, beet is transported from the beet fields to the industry by trucks and is discharged into storage areas, with a limited capacity. The residence time beet spends there is a very important variable, since the percentage of sucrose in the beet decreases with this time.

To model these issues of the storage area (SA), a dynamic mass balance has been implemented. Where m_{SA} is the beet mass in the SA, W_{Bi} the input beet flow to the SA, W_{Bo} the output beet flow from the SA, $Brix_{Bi}$ the sucrose content of input beet flow to SA, $Brix_{SA}$ the sucrose content of beet in SA, $Brix_{Bo}$ the sucrose content of output beet flow from SA to diffusion, T_r is the beet residence time in SA, f_l is a function to model the sucrose losses that depends on the T_r and U_{SA} is the percentage of SA used.

$$\frac{dm_{SA}(t)}{dt} = W_{Bi}(t) - W_{Bo}(t) \quad (1)$$

$$\frac{dBrix_{SA}(t)}{dt} = \frac{W_{Bi}(t)\big(Brix_{Bi}(t) - Brix_{SA}(t)\big)}{m_{SA}(t)} \quad (2)$$

$$T_r(t) = \frac{m_{SA}(t)}{W_{Bo}(t)} \quad (3)$$

$$Brix_{Bo}(t) = Brix_{SA}(t) \cdot \big(1 - f_l(T_r(t))\big) \quad (4)$$

$$U_{SA}(t) = \frac{m_{SA}(t)}{m_{SA\_max}} \quad (5)$$

To model the arrival of trucks, W_{Bi} was modelled with a pulse function with a random frequency to represent problems in sugar beet crops (Acebes et al. 1999).

Diffusion

In this stage, the most important equipment is the diffuser, in which many complex processes are carried out: mass and heat transfer, reverse osmosis ... Besides, the diffuser is the largest process unit of a sugar factory, and, strictly, it should be modelled as a distributed parameter system. This type of models require Partial Differential Equations (PDEs) that must be discretized as a set of ODEs or DAEs, which involves an increase on computation time.

Fortunately, as part of another project, a complete dynamic simulator of a DDS extractor was developed by the authors (Merino A., Acebes L.F. 2003). Using this simulator, it was estimated that, for our purposes, the diffusion operation performed within the diffuser can be replaced by a delayed first order system, without much loss of information. This type of model allows calculate the flow and Brix of the green juice from W_{Bo} and $Brix_{Bo}$.

With respect to the steam demand of this stage, it can be assumed that it is mainly due to the action of some heat exchangers to reheat part of the juice recirculated from the extractor. These heat exchangers, that use steam from the evaporation, have been modeled using first principles to calculate the steam demand with accuracy (Merino A. 2008).

Purification

This stage which is also very complicated, has been modelled as the diffusion. As it was mentioned in section 2, essentially, its aim is to eliminate the impurities that the green juice contains through liming, carbonation and filtration processes. To handle the efficiency of the chemical reactions, the juice must be heated with some heat exchangers. Thus, the purification steam demand is calculated in a manner similar to the diffusion stage.

Evaporation

As the evaporation stage is the heat centre of the factory, it has been modelled using detailed first principle models of the process units (evaporators, heat exchangers, tanks, pumps, valves and tubes) and the complete control structure has been include. The mathematical model was built from a library of model components developed by the authors (Merino A. 2008) using the graphical capabilities of the modeling tool.

The modelled evaporation section counts with a tank to store the clear juice from the purification, four heat exchangers to increase the temperature of the juice and six evaporators to increase the Brix. Besides, there are valves, pumps and controllers to manage the levels of the tank and evaporators, some steam pressures, and the Brix of the produced syrup.

The steam economy of the factory falls on how the evaporation and the crystallization stages work. In particular the steam demand of the crystallization depends on the Brix of the syrup produced by the evaporation and the steam pressure required by the crystallizers. So, the modelled control structure allows considering the set points of controllers of the mentioned variables as input variables.

Crystallization
According to the previous description of the production process, the detailed modeling of the crystallization section requires a lot of batch process units. Besides, a right modeling of the process of formation of sugar crystals is quite complicated and it requires the use of models of population balances. Also, it is complicated the detailed modeling of the separation, by centrifugation, of sugar crystals from honeys. In previous work these problems were studied thoroughly (Mazaeda R. 2012) and the simulators had a high computational load.

So, according to the aims of the simulator, a simplified model of this section has been developed. This model calculates the average steam demand of the crystallization as a function of the Brix and flow of syrup produced by the evaporation and the steam pressure to crystallizers:

$$W_{SC}(t) = K \cdot W_{syr}(t) \cdot \frac{100 - Brix_{syr}}{100} \cdot \frac{H_{ss}(1.5)}{H_{ss}(Boiling\ Pressure)} \quad (6)$$

Where:

- W_{SC}: average steam mass flow demanded by crystallization
- W_{syr}: syrup mass flow
- $Brix_{syr}$: Syrup Brix
- H_{ss}: Enthalpy of saturated steam at a determined pressure

The value of the K parameter was obtained by experimentation using the detailed simulator of crystallization previous cited.

Additionally, a first order dynamic is added to the previous calculated average value. Finally, as the steam demand of the crystallization depends on the cycle time of the crystallizers, to make a more realistic approach, some negative and positive pulses have been added to the average value. The magnitude and frequency of pulses can be selected from the simulator.

Cogeneration
In the cogeneration station is where the steam and power demanded by the factory are produced. Due to its importance, the CHP plant has been modelled with the same method than the evaporation stage. This CHP plant mainly consists of boilers where steam is produced by heating water using the energy obtained from the combustion of some kind of fuel, and turbines where steam is expanded to obtain power energy. As showed in Figure 2 in our process this stage is formed by three boilers which can use natural gas or fuel-oil as fuel. There are also three different backpressure steam turbines where electric power is obtained. Apart from this, steam can pass through a bypass from the boilers directly to the evaporation section, this is the usual configuration in real plants. To protect the evaporation from huge steam pressure, a relief valve has been modelled. The amount of steam which passes through the bypass and the relief valve is the manipulated variable of the controller that assures a determined steam pressure in the evaporation first effect.

The model used to represent the behaviour of the boilers is very complex and based in first principles (Pelayo S. 1999). Mainly, besides the steam generator, it considers a preheating of the feed water before entering in the boiler, a superheating of the steam obtained through heat exchangers fed by the combustion fumes, and the typical control system of boilers.

Regarding the turbines, the objective of this model is to simulate the steam expansion and the conversion of its thermal energy into electric power. Furthermore it is necessary simulate the possibility of sending power to the external grid or generating completely the amount of power demanded by the factory.

If the CHP plant produces the power demanded by the factory, the turbines control system must assure that the rotational speed of each turbine axis is constant and equal to 50 Hz. However, if the CHP plant is connected to the external grid, the controlled variable is the generated power. Both cases, the steam flow which is expanded in the turbine is the controlled variable.
Hence, the turbine model must represent not only the power generation, but also the rotational speed of the turbine. The model is based on the papers of (Thomas P. 1999) and (Chaibakhsh A., Ghaffari A. 2008), and on a library of thermodynamic properties of the steam (Merino A. 2008).

3.3 Simulator size
The great order of magnitude of the simulator can be estimated taking into account that it is formed by 6485 equations, being 449 of those dynamic and 7 algebraic. Furthermore, it counts with 2131 parameters and 29 boundaries.

A test was carried out to measure the simulation speed. In this test 48 h (172000 s) of production were simulated, changing between different operational points. The real time used by the simulation was 274 s, so a time ratio between simulation and real time of 628 was obtained.

As inputs of the model, changes in the production and in the syrup Brix obtained from evaporation were considered. As outputs, besides of different operational variables related to the process, others associated with energy consumption were selected, as heat and power demand, cogeneration performance, fuel consumption…
It is noteworthy that all the control system of the process has been modelled in detail, so changes in the inputs lead to new stationary points without human intervention, which will be very useful in future work with the RTO system.

4. SIMULATION RESULTS

As shown in the previous section, the presented simulator is a very complex model composed by many little models joined all together. As was aforementioned, it does not represent any existing sugar factory in particular, so a quantitative validation of the global model is impossible. However, a qualitative validation of the simple models and the global one is possible observing the system response with expert's guidance. And this is how the simulator has been validated.

In this section different experiments, to show the behaviour of the model and test if it is ready to substitute a real sugar factory for the RTO system, are provided. Essentially, the optimization will have to find the optimal working point where the operational cost is the lowest respecting the constraints related to the process. Taking this into account, the most important variable that will be manipulated in order to reach that point is the production which in our process is the quantity of beet that leaves the silo where it is stored.

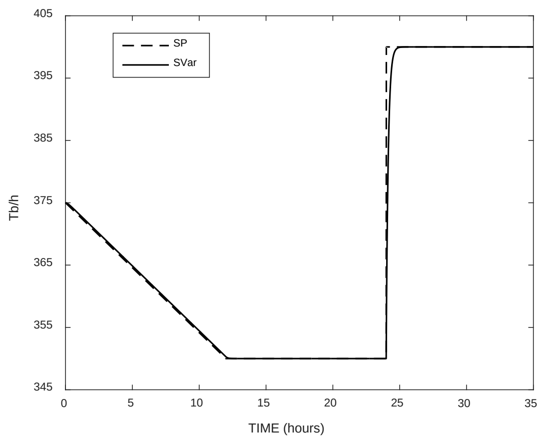


Figure 4: Changes in the production.

The simulator is designed to work in a working range from 300 Tb/h (7200 Tb/d), to 500 Tb/h (12000 Tb/d). As nominal working point a production of 375 Tb/h (9000 Tb/d), has been selected. This way, three different boilers are being used to obtain 82.1 Tvap/h with a pressure of 37 bar, using natural gas as fuel. The steam is expanded to 2.45 bar using three backpressure turbines which provide the power needs of the plant, one-third each. As process variables in the evaporation a Brix of 63.2º is demanded, and the pressure inside the fourth effect is desired to be 1.33 bar.

Two experiments in 35 h of simulation have been carried out to see the response of the model when the production is changed. First of all, as it can be seen in Figure 4, the experiment has started from the nominal point and a slowly change in the production has been introduce, so in 12 h it has gone down to 350 Tb/h. Later, to test the model robustness, a step of 50 Tb/h was carried out, reaching a production of 400 Tb/h. As a result, the most important variables related to energy purposes are shown in the next Figures.

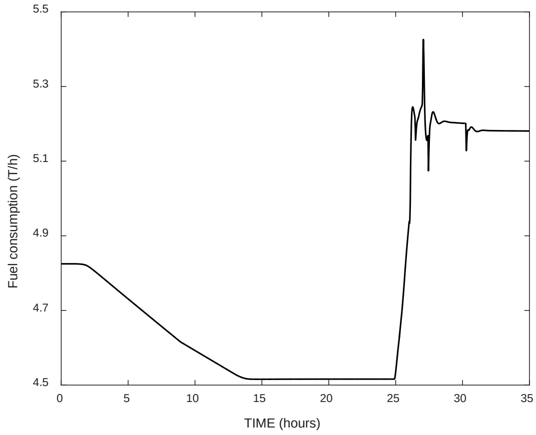


Figure 5: Fuel (Natural gas) consumption

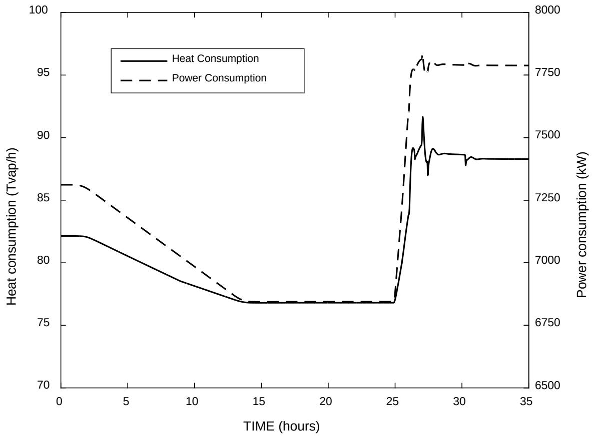


Figure 6: Heat (solid orange line) and power (dotted blue line) consumption.

In Figure 5 and Figure 6 the energy needs of the factory can be observed. As expected, while the production goes down the fuel consumption and the heat and power load decrease. A delay close to a couple of hours is found in both graphs since the change in the production is done due to the process dynamic. It takes some time since a change in the production was made, till the different parts of the plant start to consume less energy. The same behaviour can be observed when the step in the production is done at time 24 h. This time the energy demand does not increase until hour 26 approximately. It can be seen that the dynamic is not so smooth, this is due to the control system is not well tuned for big changes in the production, so it needs more time to reach a stationary point again.

With this Figures it is demonstrated that the simulator can estimate not only the steady state point, but also the transient. This will be very useful in the optimization problem, as sometimes reaching a different stationary point means passing through not permissible operation points. To show one example of how the process respond with these changes on the production, the signal output of the pressure split range controller, which was mentioned in section 3.2 when the cogeneration was described (see Figure 7), has been selected.

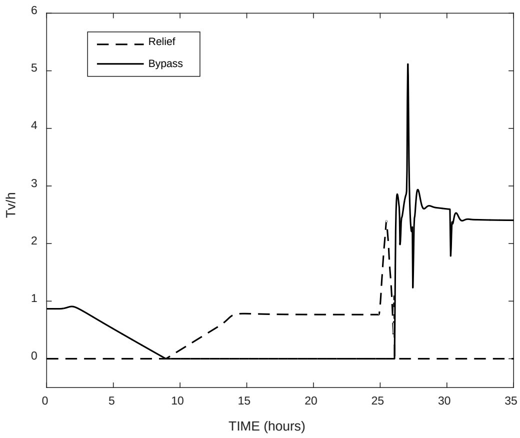


Figure 7: Relief and bypass valve response to changes in the production

Due to a decrease in energy needs when the production is reduced, it can be seen how the controller firstly tries to maintain the pressure closing the bypass valve to not disturb the turbines performance. As this is not enough, when the production is too low the controller has to open the relief valve and throw steam to the atmosphere. In this situation, the heat load of the plant is not sufficient to accomplish the power needs, and since the process is not connected to the external grid, turbines must supply the entire power demand. Therefore, the excess of steam must be released to the atmosphere to protect the evaporation from an increase in the working pressure. Obviously, this will not be a desire operational point. When the production is increased to 400 Tb/h the opposite behaviour is seen, now the relief valve must be closed and the bypass valve open. This is due to the fact that now the heat demand is greater than the one steam turbines need to generate the power required.

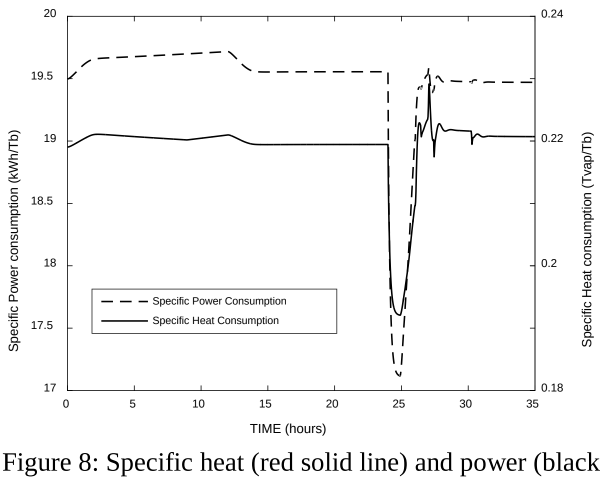


Figure 8: Specific heat (red solid line) and power (black solid line) consumption.

In Figure 8 the specific heat and power consumption of the entire factory can also be analysed and seen how linear it is to changes in the production. In the starting point, the specific power consumption is 19.5 kWh/Tb, which is close to the average value of 20 kWh/Tb found in (Frankenfeld T., Voβ C. 2004) for a typical sugar factory. Regarding the specific heat consumption, it can be observed that it is close to 0.22 Tvap/Tb, so it is between the typical limits for this type of industry, 0.2-0.22 Tvap/Tb, found in (Van der Poel P.W. et al 1998).
As the production decreases, both heat and power specific consumptions grow a little. However, when the stationary point is reached it can be observed in Figure 8 that heat and power specific consumptions are almost the same as in the beginning. Only steam specific consumption increases a little because some steam is being thrown to the atmosphere as it can be seen in Figure 7. When the production increases, both the specific power and steam consumption decrease a lot in the transitory state because of the delay inherent to the system, but when the stationary state is reached it can be observed in Figure 8 that the specific power is essentially the same again. This indicates that the relationship between heat and steam specific consumption and production is linear.

Finally, in Figure 9 it is showed how typical performance parameters for cogeneration plants can be evaluated using the simulator. The global efficiency (η_G) is defined as how much useful heat (Q_u) and power (W) are being obtained instantaneously from the combustion of natural gas (F).

$$\eta_G = \frac{Q_u + W}{F} \tag{7}$$

Typical values for this coefficient on cogeneration systems with backpressure turbines are between 80 – 90%. In our starting point, a global efficiency is closed to 91.5 %, which is almost constant until the 9th hour when the relief valve is opened and the coefficient starts to decrease. This can be explained since the useful heat has been defined as the steam used by the process, the evaporation in our case. Because more fuel has to be burned in order to obtain steam that is not being used in the evaporation, the global efficiency of the process decreases. When the production is increased, values around 91.5% are obtained again.

This behaviour can be explained in a better way observing the power (η_e) and thermal (η_t) efficiency separately, and taking into account that the sum of them yields the global performance defined on the previous equation.

$$\eta_e = \frac{W}{F} \tag{8}$$

$$\eta_t = \frac{Q_U}{F} \tag{9}$$

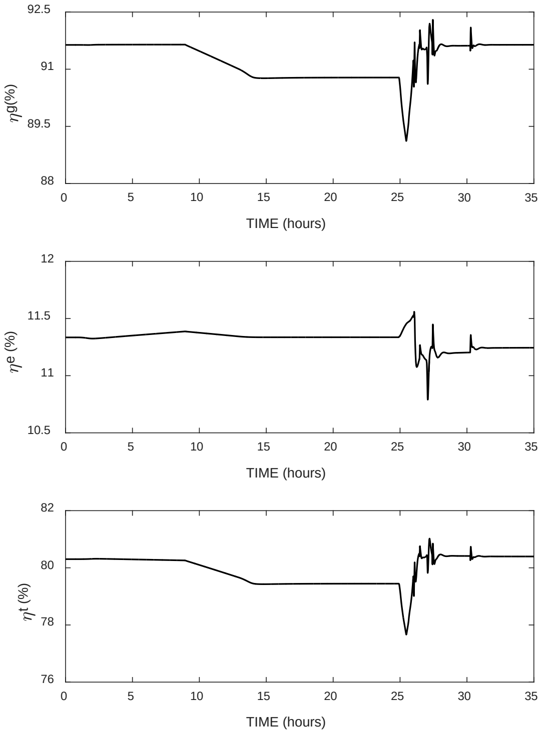


Figure 9: (Top) Global efficiency of the cogeneration system. (Middle) Power efficiency. (Bottom) Thermal efficiency.

In Figure 10 it can be seen how the power efficiency remains almost constant during operation excepting transient points. However, the heat efficiency decreases one point when the relief valve is opened.

Changing the Syrup Brix

Now, a change in a key variable of the process is proposed, as it is the set point of the controller of the Brix value of the syrup at the output of the evaporation section, to see how this change affects the energy consumption of the complete factory. The brix degree has been changed from 63.2° to 66° at time 1h, the evolution of the most relevant variables related to energy management are shown in Figure 10.

As can be seen, with an increase in the Brix degree it is obtained, after a transitory state, less heat and power global consumption. Firstly one could expect an increment in the heat consumption because evaporation needs to remove more water from the juice. However, now crystallizers have to deal with a more concentrate syrup from evaporation so globally the process needs less steam. The same explanation can be applied to the power consumption, now crystallizers work in a more relaxed way, so they need less power supply to work.

As less heat is demanded, the fuel consumption of the factory also decreases, and as the production was not changed but less heat and power is being consumed, the specific power and heat consumptions also decrease. Regarding the global efficiency of the cogeneration plant, it can be seen how it is almost the same as before. This can be explained by the fact that less steam and power is being consumed, but also less fuel. That means that the relationship between the heat and power production and fuel consumption is almost linear.

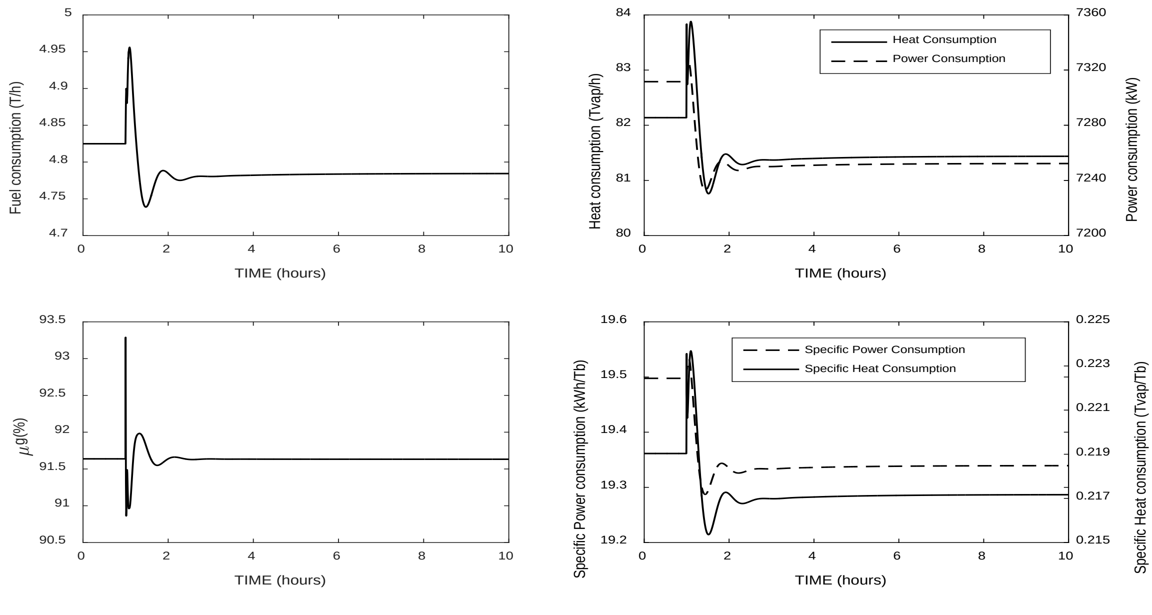


Figure 10: Simulation results for changes in the Brix degree

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

5. CONCLUSIONS

In this paper, a simulator of a typical sugar factory with enough accuracy to study problems related to energy, heat and power, has been provided. A grey box model, which combines first principles models with others based on data and the existing literature has been outlined. The model has been implemented in a general purpose simulation tool for continuous systems that is based in an EBOOL and allows the graphical process modelling. The simplified model provides well enough the global process dynamic and balances, as it was shown in the simulation results, and decreases the simulation time compared with previous simulators.

Both simulator features will allow, first, to obtain by identification the simple models of the process that needs the aforementioned RTO system for energy purposes, second, to perform the RTO tests in simulation. For the first aim, it is critical that the model accurately reflects the industrial process and, for the second one, it is very convenient to reduce the simulation time because, due to the process dynamic, the length of the simulation tests will be in the range of several days or weeks.

6. ACKMOWLEDGMENTS

The authors wish to express their gratitude to the Spanish Government for the financial support through the project "Integration of Optimization and Control in Process Plants" (DPI2015-70975-P).

REFERENCES

Acebes L.F, De Prada C., Gorostiaga L., 1999. Evaluating cane feeding control. International Sugar Journal (Cane Sugar Edition), Vol. 101, No. 1210, 495-500.

Asadi. M, 2007. Beet-Sugar Handbook. New Jersey: Wiley & Sons, Inc.

Ashok S. Banarjee R., 2003.Optimal Operation of Industrial Cogeneration for Load Management. IEEE Transactions on Power Systems Vol. 18, 931-937.

Chaibakhsh A., Ghaffari A., 2008.Steam turbine model. Simulation Modeling Practice and Theory Vol. 16, 1145-1162.

Darby Mark L., Nikolaou M., Jones J., Nicholson D., 2011. RTO: An overview and assessment of current practice. Journal of Process Control Vol. 21, 874-884.

EA Internacional. EcosimPro. Available from: http://www.ecosimpro.com/ [Accessed 03 July 2017]

Frankenfeld T., Voβ C., 2004. Electrical power consumption – an European benchmarking-exercise. Zuckerindustrie, Vol. 129, Num. 6, 407-414.

Mazaeda R., Merino A., de Prada C., Acebes L.F., 2012. Sugar house training simulator. International Sugar Journal Vol. 114, 42-48.

Merino A., 2008. Librería de Modelos del Cuarto de Remolacha de una Industria Azucarera para un Simulador de Entrenamiento de Operarios. Thesis (PhD). Universidad de Valladolid.

Merino A., Acebes L.F., 2003. Dynamic simulation of an RT extractor. Zuckerindustrie, Vol. 128, Num. 6, 443-452.

Merino A., Acebes L.F., Mazaeda R., de Prada C., 2009. Modelado y Simulación del Proceso de Producción del Azúcar. RIAI Vol. 6, Num. 3, 21–31.

Mitra S., Sun L., Grossmann I.E., 2013.Optimal scheduling of industrial combined heat and power plants under time-sensitive electricity prices. Energy Vol. 54, 194–211.

Modelica Association et al. Modelica. Available from: https://www.modelica.org/ [Accessed 03 July 2017]

Pelayo S., 1999. Modelado y Simulación Dinámica de una Caldera de Vapor Industrial. Final Career Project. Universidad de Valladolid.

Ristic M., Brujic D. Thoma K., 2008. Economic dispatch of distributed heat and power systems participating in electricity spot markets. Proceedings of the institution of mechanical engineers, Part A: Journal of Power and Energy Vol 222, Issue 7, 743-752.

Sanaye S., Nasab A.M., 2012. Modeling and optimizing a CHP system for natural gas pressure reduction plant. Energy Vol 40, 358-369.

Thomas P., 1999. Simulation of Industrial Processes for Control Engineers. BH.

Tina, G.M., Passarello G., 2012. Short-term scheduling of industrial cogeneration systems for annual revenue maximization. Energy Vol. 42, 46-56.

Urbaniec K., 1989. Modern energy economy in beet sugar factories. Elsevier.

Van der Poel P.W., Schiweck H., Schwartz T. 1998. Sugar Technology. Berlin: Bartens.

VERIFICATION OF INDUSTRIAL CONTROL ALGORITHMS IN VIRTUAL LABORATORY STANDS

Tatyana Liakh[(a)], Vladimir Zyubin[(b)]

[(a),(b)] Novosibirsk State University,
Institute of Automation and Electrometry SB RAS

[(a)]antsys_nsu@mail.ru, [(b)]zyubin@iae.nsk.su

ABSTRACT
The complexity of industrial control systems increases every year. The imposed requirements should provide the reliability of the developed algorithm. But today the common practice of industrial automation is characterized by the following: generally testing of control algorithms starts only when you run the software on a new facility.
In the article the control algorithms development method was put forward. The method uses virtual laboratory stands for control algorithm creation and its verification. Virtual laboratory stand consists of the five components: graphical representation of the automated object (GRAO), control algorithm module (CA), virtual plant module (VP), scripts control module (SCM) and verification module (VM). All modules are described as a hyper-process with the Reflex language. SCM imitates different situations on the object: environmental changes and breakdowns. VM watches over the CA during the simulation and automatically checks whether the algorithm meets the given specification under the certain script.

Keywords: control algorithms, industrial automation, process-oriented programming, virtual plants, language Reflex, verification

1. INTRODUCTION

Testing and verification problem of industrial control algorithms is one of the key problems in industrial automation area. Industrial control algorithms are logically complicated because of the great number of dependences between different parts of an algorithm. But the cost of errors in such algorithms is extremely high: errors in control algorithm leads to malfunctions, abnormal situations and accidents on the technological plant.

Today many approaches and mathematical models for the control algorithms development exists. Domain-specific languages (DSL) are perspective instrument for this issue. Reflex language is the DSL for the control algorithms development in industrial automation and robotics. The Reflex language is based on the hyper-process model. Hyper-process is an extension of the classical state-machine model (Zyubin 2007). The basic concept of the Reflex language – process. The program on the Reflex language is a set of parallel executable processes. They can trigger each other, stop, and monitor states. This allows to process signals from the technological plant in parallel. Reflex also allows operations with time intervals and offers means for the interaction with sensors and control devices. Ease of use of the Reflex language and its adequacy to industrial automation tasks was confirmed in automation projects for complex technical objects.

But testing and verification of the algorithm is still a serious issue. It is impossible to test the algorithm without plant. In most cases control algorithms testing starts only when you run the software on a new plant. As a result, the testing of the algorithm is postponed until the start-and-adjustment works begin. Such practice leads to high risks, emergency situations or even to accidents at the facility.

The contribution is structured as follows. First, we discuss specificity of control algorithms for complex technical objects. Second, we describe shortly the hyper-process mathematical model and the Reflex language. After that, we demonstrate the approach of the control algorithms verification via virtual plants conception in virtual laboratory stands.

2. THE SPECIFICITY OF INDUSTRIAL CONTROL ALGORITHMS

Due to the specificity of the automation tasks, especially in the case of complex technical objects, it is extremely hard to develop control algorithms for such issues. The reason is that industrial algorithms have a number of properties, unique to the field of industrial automation (Zyubin 2005, Kof 2003):

1. Presence of an external environment to interact with.
2. Cyclic and event-driven functioning.
3. Synchronism – control algorithm implies synchronization of its functioning with physical processes in the external environment.
4. Mass logical parallelism – it reflects existence of a large set of concurrent (or to be precise – independent and weakly connected) physical processes in the controlled objects. As the events appear in an arbitrary consequence, any

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

attempt to describe the system reaction within a monolithic block leads to a combinatorial task with exponential increase of complexity, so called the combinatorial explosion of complexity

5. Hierarchical structure. Any complex control algorithm has to have hierarchical structure that reflects artificial nature of the external environment, the designer plan that is implemented in form of the facilities (Zyubin 2004). Because of the logical parallelism, the hierarchical structure consists of chains independently executed in parallel. It means that the divergence and convergence of control flow are a significant part of control algorithm.

3. HYPER-AUTOMATON MODEL OF CONTROL ALGORITHMS

The variety of different formalisms for specifying control and reactive systems were introduced (Hoare 1985, Harel 1987). One of these models - hyper-automaton model, which is introduced as a useful formalism for control algorithms reasoning.
The hyper-automaton model is an extension if the the finite state automata (FSA). A hyper-automaton is an ordered set of processes, which are cyclically stirred to activity with a period of activation TH. A hyperautomaton is a triplet:

$$H = <T_H, P, p_1> \quad (1)$$

- T_H is a period of activation.
- P is a finite nonempty and ordered set of processes.
- ($P = \{p_1, p_2, \ldots, p_M\}$), where M is the number of processes.
- p_1 is the first marked process, $p1 \in P$, which is the only non-passive when hyper-automaton starts.

A process is a state machine where the states are functions. Mathematically, i-th process is
a quintuple:

$$p_i = <F_i, f_{1i}, fcur_i, T_i>, \text{ where} \quad (2)$$

- F_i is a set of mutually exclusive functions.
- f_{1i} is the first function, ($f_{1i} \in F_i$).
- $fcur_i$ is the current function, ($fcur_i \in F_i$).
- T_i is the current time.

The F_i and f_{1i} elements characterize static features of the process. The $fcur_i$ and T_i elements give us means for reasoning about a process dynamics.
A function of a process is a set of events and reactions to the events. Mathematically, j-th functions of i-th process is a twain:

$$f_{ji} = <X_{ji}, Y_{ji}>, \text{ where} \quad (3)$$

- X_{ji} is a set of events ($X_{ji} = \{x_{ji1}, x_{ji2}, \ldots, x_{jiL}\}$).
- Yji is a set of reactions ($Y_{ji} = \{y_{ji1}, y_{ji2}, \ldots, y_{jiL}\}$).

Hyper-automation model can be implemented by means of general-purpose programming languages, but in case of control algorythms one more often choose domain-specific languages (DSL), designed for industrial automation issues. One oh these languages is the Rexlex language. The Reflex language was used in a number of automation protects for complex technical plants (Liakh 2016) and shows its adequacy to industrial automation tasks.

4. INDUSTRIAL CONTROL ALGORITHMS TESTING WITH VIRTUAL PLANTS

But the description problem of control algorithms is no the only problem in industrial automation, Verification of the algorithm is also a serious challenge for the developers.
The concept of Virtual Plant (VP) was put forward to solve the control algorithms testing issues (Zyubin 2007). This concept was offered in the Institute of Automation and Electrometry. VP – is the program imitator of the automated technological process. VP code and control algorithm (CA) code executes separately (Fig. 1). Unified data ex-change between VP and CA ensures the connections to be saved when the CA is changed. This approach allows us to use the iterative development model and debug the algorithm code before start-and-adjustment works begin.

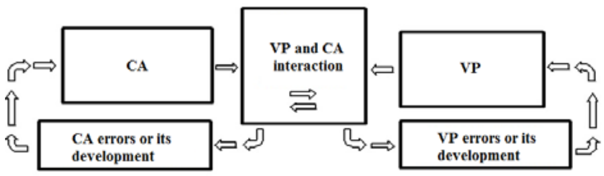


Figure 1: Control Algorithm Iterative Development Model via the Virtual Plant Conception

For correct simulation of the plant one need to describe signal processing and parallel operations on the object. That's why CA and VP were both created on Reflex language. VP operates in multiple modes: correct operation mode, or malfunction-imitation modes. The testing software was created with means of LabVIEW package. This software allows running VP and CA simultaneously, testing VP and CA and imitating data exchange and external events (Liakh and Zyubin 2014). The CA of the Large Solar Vacuum Telescope (LSVT, Baikal Astrophysical Observatory) was developed via this software (Liakh and Zyubin 2016).
But some disadvantages of such approach were noticed. Many operations had to be made manually by the operator. Operator controlled VP modes, imitated GUI commands and environmental changes. Also the operator had to check the correctness of algorithm

reactions manually. All this disadvantages led to the long testing process. Also the possibility of missed errors and unexplored behavior grows.

5. METHOD EXTENTION WITH THE SCRIPTS CONTROL MODULE AND VERIFICATION MODULE

To increase the reliability of the algorithm and to speed up the testing process, one offer the modification of the method. One notice that in the method mentioned above operator has multiple issues. First, he has to manage the sequence of different events for the CA and the VP. Also he has to check whether the CA meets a given specification under a certain script.

That's why one offer to supplement the existing scheme with two modules: scripts control module (SCM) and verification module (VM). The interaction between all modules is described on the Fig. 2.

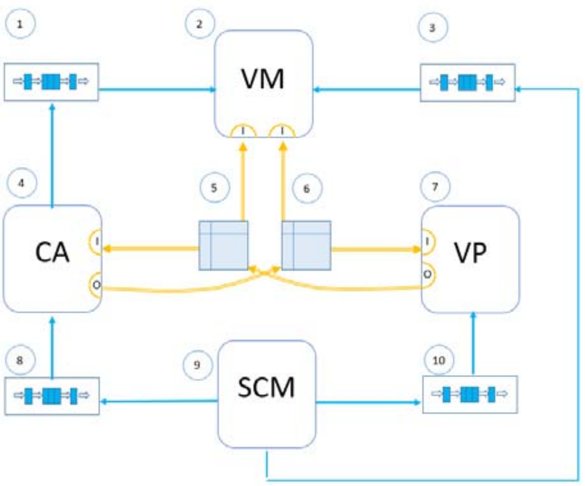
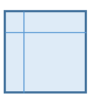
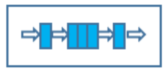
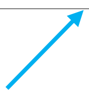


Figure 2: Control Algorithm Iterative Development Model via the Verification Module and Scripts Control Module

Table 1: Conventional Notations

 	Event-Driven Algorithmic Module
	Digital Ports (Input/Output)
	Message Queue
	Digital Ports Values Cache
	Digital Data
	Messages



Four interacted algorithmic modules (hyper-processes) are used to control algorithm verification. They run in the determined order, form message queues, digital ports values and factual parameters values.

1. Control algorithm (CA, 4) – CA module implements the control algorithm logic.
2. Virtual plant (VP, 7) – VP module implements the plant operation logic. VP imitates data flow to CA from ADC and data flow from digital input devices.
3. Verification module (VM, 2) – VM automatically checks whether the CA meets a given specification under a certain script. VM alternately checks a set of temporal requirements and reports about malfunctions. VB verifies only CA block.
4. Script Control Module (SCM, 9) – SCM manages the sequence of different events for CA, VM and VP, such as: operator actions, environment states and malfunctions of devices. SCM sends commands through the message queues. SCM imitates operator commands for CA block. For VP SCM sets operation modes: it forces VP to imitate different accidents and breakdowns – or operate in correct mode. Also SCM sends environmental data for the VP. SCM reports VB which of scripts is actual now. According these information, VB choses properties to verify.

Algorithmic blocks interact according to the scheme:

1. SCM launch. SCM forms three message queues: for CA (8), VP (10) and VB (3). For the CA SCM sends messages, thereby imitating operator commands. SCM informs VP about the current operation mode (regular operation mode, equipment failure or an accident). Also SC informs VP about environmental state (e.g. temperature or pressure) – but it is for VP to decide, how process these information – or ignore it. In the VB message queue SCM puts messages to identify current script: this information allows

VP to determine, how CA should operate in case of correct operation.

2. CA launch. In the module the control code of the technological control system executes. CA takes input digital data from the ports cache (5). Also CA handles messages from SCM queue (8). At the end of the cycle CA puts its digital output to the CA output ports cache (6). Output messages CA puts into the VM input message queue (1).
3. VM launch. VM analyses all data flow from the CA module: digital input and output values from port cashes (5, 6), and its output messages (1). With this data VM is able to verify whether the CA meets all imposed requirements. From the SCM message queue VM takes information about current script – in order to determine the correctness of the algorithm reactions. VM sends testing results to the operator GUI.
4. VP launch. VP takes information about current mode and environmental changes from the SCM message queue. VP takes input digital data from the CA output ports cache (6). This imitates the data exchange between the CA and a real plant. At the end of the cycle VP puts its digital output to the CA input ports cache (5).

Development of industrial control algorithms occurs iteratively:

1. A functionally separate part of the CA code is created in Reflex language, i.e. a set of processes responsible for processing a specific task.
2. A functionally separate part of the VP code in Reflex language is created. This code simulates the operation of those elements of the control plant that are controlled as described in the algorithmic module.
3. The SCM block is described. SCM sends three types of messages for other algorithmic blocks:
   - Messages for the CA block. These messages simulate the actions of the operator.
   - Messages for the VP block. These messages determine the scenarios set for all parts of the plant already described and for the whole virtual plant.
   - Messages for the VM - these messages notify the VM about the current verification mode. According to these messages, the VM determines which scenario is being executed at the moment and what reaction to expect from the CA.
4. The VM block is described. In the VM block, the response to the messages from the SCM, described in step №3, is added. Also, the requirements imposed on the created algorithmic control unit are described. Since the development happens iteratively, these requirements can also verify the operation of the algorithm parts created in the previous steps – in this case, the joint interaction of various parts of the algorithm is verified.

The Reflex language was chosen as a programming language for the all event-driven modules. The language Reflex based on the hyper-process model. Hyper-process model reflects openness, event-driven nature, cyclicity, synchronism, and mass logical parallelism of a control algorithm. The temporal dependences of the hyper-processes allows do describe the temporal requirements.

3. CONCLUSION

In the contribution the scheme of the iterative CA development was put forward. The scheme for the automated verification of the developed algorithm was created. Such approach to the CA testing and development allows to give the strict answer if the algorithm satisfies all necessary requirements. Also virtual stands are helpful for the engineering students training.

ACKNOWLEDGMENTS
The research has been supported by Russian Foundation for Basic Research (grant 17-07-01600).

REFERENCES

Zyubin V. E., Hyper-automaton: a Model of Control Algorithms // Proceedings of IEEE International Siberian Conference on Control and Communications, SIBCON-07. Russia, Tomsk April 20-21, 2007, PP.51-57

Zyubin, V. E., Multicore Processors and Programming. // Open Systems J., №7-8, 2005, PP.12-19

Kof, L., Schätz, B., Combining Aspects of Reactive Systems. // Proc. of Andrei Ershov Fifth Int. Conf. Perspectives of System Informatics. Novosibirsk, 2003, PP.239-243

Zyubin, V. E., Text and Graphics: What Language Does Programmer Need? // Open Systems J., №1, 2004, PP. 54–58

Hoare, C. A. R., Communicating Sequential Processes. // PrenticeHall Int., 1985

Harel, D. Statecharts: a Visual Formalism for Complex Systems. // Science of Computer Programming 8. Elsevier Science Publishers B.V., North-Holland,1987, PP. 231–274

Liakh T. V., Zyubin V. E., Application of the virtual plant for industrial automation issues // Proc. Ershov informatics conference-14, satellite “Workshop on Science Intensive Applied Software”, Saint Petersburg, Russia, 2014, 43-48.

Liakh T. V., Zyubin V. E. The Reflex Language Usage to Automate the Large Solar Vacuum Telescope // 17th International Conference of Young

Specialists on Micro/Nanotechnologies and Electron Devices (EDM). (Erlagol, Altai Republic, Russia, June 30 2016-July 4, 2016). pp. 137-139.

AUTHORS BIOGRAPHY

Zyubin V. E. Lead researcher of the Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences. In 1992 he graduated from the Novosibirsk State Technical University, majoring in "Automation and Remote Control". The degree of Doctor of Technical Sciences was awarded in 2014. Associate Professor in the specialty.

He has authored over 90 scientific publications.

Interests – Languages of technological programming, programming psychology, human factors, complex control algorithms, the theory of finite automata, hyper-automata model, event polymorphism, semiotics and pragmatics, the massive parallelism of logic, Reflex language, languages of IEC 61131-3 standard

Liakh T. V. is a postgraduate student of the Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences. In 2013 Tatiana graduated with honors from the Novosibirsk State University (Physical Faculty, Automation of physical and technological researches department).

From 2009 to 2011 Tatiana was engaged in the researches in the field of computer simulation and artificial intelligence in the SoftLab-NSK company.

Since 2011 Tatiana develops industrial automation systems in the Institute of Automation and Electrometry. She has participated in many projects of automation, such as the development of the automated control system of the Large Solar Vacuum Telescope – the biggest Eurasian telescope and main instrument of the Baikal Astrophysical Observatory.

Interests – Complex control algorithms, methods of industrial algorithms development, programming languages for industrial automation and industrial algorithms verification.

OPTIMAL STOCHASTIC CONTROL OF AN ALUMINIUM RECYCLING UNIT IN REVERSE LOGISTICS

Franck Éric KEMI YOTAT[(a)], Jean Pierre KENNE[(b)], Victor SONGMENE[(c)]
[(a),(b),(c)] Mechanical Engineering Department, Production Technologies Integration Laboratory
University of Quebec, École de Technologie Supérieure,
1100 Notre Dame West, Montreal, Quebec Canada, H3C1K3
[(a)]franck-eric.kemi-yotat.1@ens.etsmtl.ca, [(b)]Jean-Pierre.Kenne@etsmtl.ca, [(c)]Victor.Songmene@etsmtl.ca

ABSTRACT
This paper deals with the problem of control of an aluminum recycling unit in reverse logistics. The unit studied is in open loop. It's assimilated to a machine producing a single type of products and is subject to random failures and repairs, fueled by a random return rate of aluminum and supplying a constant customer demand rate. Due to the presence of random breakdowns of the machine and the constraints related to customer satisfaction, it is imperative to determine an optimal policy of production in order to insure the customer demand satisfaction. The objective of this study is to determine the production and disposal policies in order to minimize the overall production cost. A corresponding optimal stochastic model has been developed and leads to Hamilton-Jacobi-Bellman equations describing the optimality conditions. A numerical solving method has been used and led to an optimal policy which is of the Hedging Point Policy type (HPP).

Keywords: Reverse logistics; Stochastic processes; Dynamic programming; Supply chain management: Numerical methods.

1. INTRODUCTION

This paper addresses the problem of optimal control in reverse logistics system in open loop with random return of end of life aluminum products. Previously, consumers were concerned only with quantity, quality and price of products consumed. The conventional product cycle ranged from production sites to landfills (disposal). However, mentalities have changed today, and they are increasingly concerned about the preservation of the environment and the possibility of recycling products made available to them, as mentioned by Thomas et Wirtz (1994). The authors showed the advantages and the importance of the recycling of aluminum against their complete destruction because it allows recovering their potential wealth. Many authors have shown that it's better to recapture the value of end of life products than dispose them (Rogers and Tibben-Lembke, 1998). The need for recycling of aluminum doesn't only increase for economic, environmental and legislative reasons, but also for social reasons. The optimal management of the logistics network and their activities is very complex in reverse logistics due to the wide variety of decisions of different scopes, disturbance factors and attention to the desired time horizon.

One of the challenges facing manufacturing companies is achieving production targets that meet customer demand. Although it is often possible to predict the state of machines (operational or failed) over a short time horizon with a degree of confidence, it is difficult to predict their behavior over a long time horizon given that they are generally prone to breakdowns and repairs. The need for optimal production planning tools to deal with these hazards has prompted and several authors addressed the corresponding issues.

Stochastic dynamic programming method has been applied for many systems under different initial considerations. Indeed, the joint optimization of production and preventive maintenance for a non-homogeneous Markov process was treated by Gharbi and Kenne (2000). The extension of production policy optimization to larger systems subjected to random phenomena such as machine failures and repairs have also been studied by Gharbi and Kenne (2003). Out of the control of machine failure and repair, Kenne (2004) integrated tool wear and rejection rates in order to determine the optimal production policies and preventive replacement of tools to minimize the incurred costs (storage, repairs and replacements). Nodem et al. (2011) developed a policy of production, repair and preventive maintenance of a manufacturing system subject to random failures and repairs; They also obtained a suboptimal policy of critical threshold type.

In order to optimize production for a closed loop hybrid system, Kenné et al. (2012) studied a production unit consisting of two machines in parallel producing a single type of products. They determined the optimal production policies of each machine in order to supplement customer demand while minimizing the incurred costs (inventory and shortage).

The rest of this paper is organized as follows: in section 2, the problem statement is presented. In section 3, the formulation of the control problem and the numerical approach used to solve it are presented. In section 4, the results of a numerical case are presented. In section 5, the conclusion of the paper is presented.

2. PROBLEM STATEMENT

This section presents the problem statement of the stochastic optimal control problem under study. The recycling unit under our study consists of a single machine producing a single product so as to supply a constant demand rate of aluminum ingot customer.

End of life aluminum products in the form of bales arrive at the recycling unit at variable rates r_1 and r_2 as shown in figure 1 and then stored in the warehouse so as to constitute stock x_1 of material necessary for the production. The return rate is described by the stochastic process $\alpha_1(t)$. Sometimes, the storage costs of these collected products become very high or the amount of material collected reaches the maximum storage capacity of the recycling unit denoted by x1max. In these cases, returned products collected are disposed from the warehouse at a variable and unknown rate denoted u_{d_α}.

The manufacturing machine is supplied with raw material of the stock x_1 so as to build stock x_2 of finished products (aluminum ingots). It produces with an unknown variable production rate u_α in order to meet the constant demand d of the customers. However, the machine (M) is subject to random breakdowns and repairs. The availability of the machine is described by the stochastic process α_2 (t). The decision variables are respectively the production rate uα and also the elimination rate u_{d_α} while state variables are respectively the inventory level x_1 of returned products and the inventory level x_2 of the finished products.

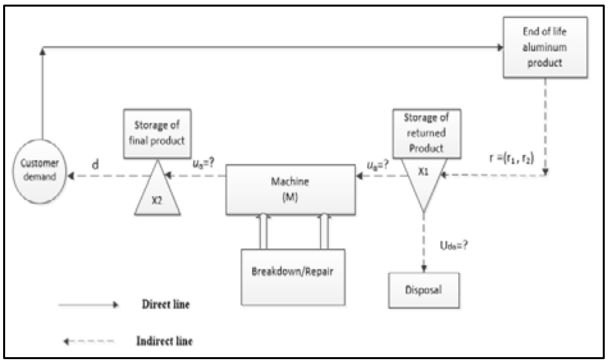


Figure 1: Structure of the recycling unit fueled by a variable return rate

The main assumptions that support our model are:
(1). The process is a homogeneous Markov chain process; (2). The machine is prone to random breakdowns and repairs described by known constant rates; (3). The customer demand rate is constant and known; (4). The maximum production rate of the system is known and constant; (5). Different return rates are known and also transition rate between them; (6). When the machine breaks down, a corrective maintenance activity is immediately implemented.

3. PROBLEM FORMULATIONS

Considering a discrete-state stochastic process $\{\alpha_1(t), t \geq 0\}$ which describes the return rate of end of life products at each time t taking values $\{1, 2\}$ such that: $\alpha_1(t) = 1$ when $r = r_1$ and $\alpha_1(t) = 2$ when $r = r_2$.

Considering a discrete-state stochastic process $\{\alpha_2(t), t \geq 0\}$ which describes the availability of the machine at each time t taking values $\{0, 1\}$ such that: $\alpha_2(t) = 1$ when the machine is available and $\alpha_2(t) = 0$ when the machine is broken down.

Finally, $\{\alpha(t) = \alpha_1(t) \times \alpha_2(t),\ t \geq 0\}$ is the discrete-stochastic state process that describes the whole state of the system at each time with value in B = {1, 2, 3, 4} depending on whether the machine is available or not and supplied at a certain return rate as resume in the table 1 below.

Table 1 Dynamic of the system

$\alpha_1(t)$	1	1	0	0
$\alpha_2(t)$	1	2	1	2
$\alpha(t)$	1	2	3	4

The state of the reverse logistic unit is modeled by a discrete-state continuous-time Markov chain with an ergotic 4×4 matrix of transition rates $Q = [\lambda_{ij}]$. The relationship between the transition rate λ_{ij} and transition probability from mode i to mode j is given by equation (1).

$$\Pr[\alpha(t+\delta t) = j / \alpha(t) = i] = \begin{cases} \lambda_{ij}\delta t + o(\delta t) \text{ if } i \neq j \\ 1 + \lambda_{ij}\delta t + o(\delta t) \text{ if } i = j \end{cases} \quad i, j \in B = \{1, 2, 3, 4\} \tag{1}$$

With $\lambda_{ij} \geq 0$ for all $i \neq j$, $\lambda_{ii} = -\sum_{i \neq j} \lambda_{ij}$ and $\lim_{t \to 0} \frac{o(\delta t)}{t} = 0$

Different possible transitions among the state of the system are given by fig. 2.

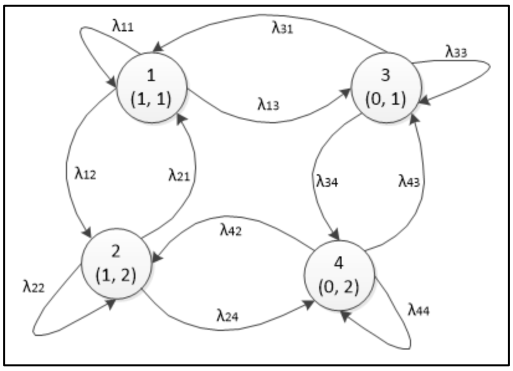


Figure 2: State transition diagram

Let $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ be the vector of limiting probabilities at each mode α (α =1,..,4) of the system. Those limiting probabilities π_i (i=1, 2, 3, 4) are the solution of the equation (2).

$$\begin{cases} \sum_{i=1}^{4} \pi_i \times \lambda_{ij} = 0 \quad \textit{for all } j = 1,..,4 \\ \sum_{i=1}^{4} \pi_i = 1 \end{cases} \tag{2}$$

The feasible control policies are given by the set $\Gamma(\alpha)$ define as follow:

$$\Gamma(\alpha) = \begin{Bmatrix} (u_\alpha,\ u_{d_\alpha}) \in R^2 / \\ 0 \leq u_\alpha \leq u_{\max} \ , 0 \leq u_{d_\alpha} \leq r_\alpha ,\ \alpha = 1,..,4 \end{Bmatrix}$$

For example, π_2 represent the probability that the machine is available and fueled at the returned rate r_2. Considering the fact that in our study, returned rate r_2 is lower than the maximal production rate u_{max}, the system is feasible only if the equation (3) is verified so as:

$$(\pi_1 \times u_{\max} + \pi_2 \times r_2) \geq d \quad (3)$$

The variations of different stocks are described by equations (4). These equations take in account the fact that, at each time t, in the given state of the system, the inventory level of the storage x1 is increased by the return rate r_α (α = 1,..,4) while the same stock is decreased by the rate of elimination u_{d_α} and also by the production rate of the machine u_α. Similarly, the stock level of finished products is increased by the production rate u_α and decreased by the demand rate d.

$$\begin{cases} \frac{dx_1(t)}{dt} = r_\alpha - u_{d_\alpha} - u_\alpha, & x_1(0) = x_{10} \\ \frac{dx_2(t)}{dt} = u_\alpha - d, & x_2(0) = x_{20} \end{cases} \quad (4)$$

where x_{10} and x_{20} are respectively the initial stock of returned and final products.

The running cost of our model depends on the storage costs of the returned products c1, the storage cost of finished products c2p, the shortage cost of finished products c2m, the disposal cost of excess returned product cd, the manufacturing cost cm, the corrective maintenance cost cα, the environmental cost c_{env} and the cost penalizing the lack of raw materials during shortages of finished products. Its expression is given by equation (5).

$$\begin{aligned} g(x_1, x_2, \alpha) &= c_1 x_1 + (c_{2p} x_2^+ + c_{2m} x_2^-) \\ &+ c_d u_{d_\alpha} + c_m u_\alpha + c^\alpha (ind\{\alpha(t)=3\} + ind\{\alpha(t)=4\}) \\ &+ c_{env}(x_{1\max} - x_1) + c_{ecart} \left|(|x_2| - x_1)\right| ind\{x_2 < 0\} \end{aligned} \quad (5)$$

where $x_2^+ = \max(0, x_2)$; $x_2^- = \max(-x_2, 0)$ and

$$ind\{\Theta(.)\} = \begin{cases} 1, & if \ \ \Theta(.) \ is \ true \\ 0, & otherwise \end{cases}$$

By taking in account the discount rate ρ (Garceau, 1996), the discounted total cost is given by the relation (6):

$$J(x_1, x_2, \alpha) = \mathrm{E}\left\{ \int_0^\infty e^{-\rho t} g(x_1, x_2, \alpha) \mathrm{d}t \,\middle|\, x(0) = x, \alpha(0) = \alpha \right\} \quad (6)$$

In this equation, E(A/B) symbolizes the conditional expectation operator and $x=(x_1, x_2)$ is the vector of stock levels.

Our objective is to obtain the optimal control policies that will minimize the discounted cost (6). In another word, the objective is to find the production rate u_α and the disposal rate u_{d_α} so as to minimize the expected discounted cost given by (6). The value function of such a problem is defined by:

$$v(x_1, x_2, \alpha) = \inf_{(u_\alpha,\ u_{d\alpha}) \in \Gamma(\alpha)} J(x_1, x_2, \alpha), \quad \forall \alpha \in \mathrm{B} \quad (7)$$

The value function (7) satisfies the Hamilton–Jacobi–Bellman (HJB) equations (8) which describes the optimality conditions and can be found in chapter 8 of Gershwin (1994). In the work of Rivera-Gómez et al. (2016), they showed that the value function is continuously differentiable and viscosity solution to the Hamilton-Jacobi Bellman (HJB) equations. Such HJB equations, which integrate the dynamics of the stock as well as the machine modes, are given by:

$$\rho v(x_1, x_2, \alpha) = \min_{(u_\alpha, u_{d_\alpha}) \in \Gamma(\alpha)} \begin{bmatrix} g(x_1, x_2, \alpha) + \\ \sum_{j \in B} \lambda_{\alpha j} v(x_1, x_2, j) + (r_\alpha - u_{d_\alpha} - u_\alpha) \frac{\partial v(x_1, x_2, \alpha)}{\partial x_1} \\ + (u_\alpha - d) \frac{\partial v(x_1, x_2, \alpha)}{\partial x_2} \end{bmatrix} \quad (8)$$

The solution of equation (8) has been obtained by Akella et Kumar (1986) but, for the simplest single machine under certain assumptions which are not valid for our model. Since an analytical solving method does not exist for the HJB equations, we adopted a numerical methods methodology to obtain the structure of the optimal control policies.

Numerical approach based on Kushner et Dupuis (1992) method is used to obtain a sub-optimal solution which is a good approximation of value function at a predetermined precision degree δ (in our study, we fixed δ=0.01).

That numerical method consists in simplifying the HJB equations by approximating the value function $v(x_1, x_2, \alpha)$ by the function $v^h(x_1, x_2, \alpha)$ and the approximation of the gradient of the value function $\partial v(x_1, x_2, \alpha) / \partial x_i$ (i = 1,2) respectively by the following expressions.

$$\frac{\partial v(x_1, x_2, \alpha)}{\partial x_1} = \begin{cases} \frac{v^{h_1}(x_1 + h_1, x_2, \alpha) - v^{h_1}(x_1, x_2, \alpha)}{h_1}, & if \ \ r_\alpha - u_{d_\alpha} - u_\alpha \geq 0 \\ \frac{v^{h_1}(x_1, x_2, \alpha) - v^{h_1}(x_1 - h_1, x_2, \alpha)}{h_1}, & otherwise \end{cases}$$

$$\frac{\partial v(x_1, x_2, \alpha)}{\partial x_2} = \begin{cases} \frac{v^{h_2}(x_1, x_2 + h_2, \alpha) - v^{h_2}(x_1, x_2, \alpha)}{h_2}, & if \ u_\alpha - d \geq 0 \\ \frac{v^{h2}(x_1, x_2, \alpha) - v^{h_2}(x_1, x_2 - h_2, \alpha)}{h_2}, & otherwise \end{cases}$$

where hi (i = 1, 2) is the discretization step of the state variable. Considering such approximations, equations (8) becomes equations (9):

$$\rho v^h(x_1, x_2, \alpha) = \min_{(u_\alpha, u_{d_\alpha}) \in \Gamma(\alpha)} \begin{bmatrix} g(x_1, x_2, \alpha) + \lambda_{\alpha\alpha} v^h(x_1, x_2, \alpha) + \sum_{j \neq \alpha} \lambda_{\alpha j} v^h(x_1, x_2, j) \\ + \frac{|r_\alpha - u_{d_\alpha} - u_\alpha|}{h_1} v^{h_1}(x_1 + h_1, x_2, \alpha) ind\{r_\alpha - u_{d_\alpha} - u_\alpha \geq 0\} \\ + \frac{|r_\alpha - u_{d_\alpha} - u_\alpha|}{h_1} v^{h_1}(x_1 - h_1, x_2, \alpha) ind\{r_\alpha - u_{d_\alpha} - u_\alpha < 0\} \\ + \frac{|u_\alpha - d|}{h_2} v^{h_2}(x_1, x_2 + h_2, \alpha) ind\{u_\alpha - d \geq 0\} \\ + \frac{|u_\alpha - d|}{h_2} v^{h_2}(x_1, x_2 - h_2, \alpha) ind\{u_\alpha - d < 0\} \\ - \frac{|r_\alpha - u_{d_\alpha} - u_\alpha|}{h_1} v^{h_1}(x_1, x_2, \alpha) - \frac{|u_\alpha - d|}{h_2} v^{h_2}(x_1, x_2, \alpha) \end{bmatrix} \quad (9)$$

After several manipulations as appeared in the works of Kenne et al. (2003), equation (9) becomes:

$$
v^h(x_1,x_2,\alpha)=\min_{(u_\alpha,u_{d_\alpha})\in\Gamma(\alpha)}\left[\begin{array}{l}(\rho+|\lambda_{\alpha\alpha}|+\frac{|r_\alpha-u_{d_\alpha}-u_\alpha|}{h_1}+\frac{|u_\alpha-d|}{h_2})^{-1}\times \\ \left\{\begin{array}{l} c_1x_1+c_{2p}x_2^{+}+c_{2m}x_2^{-}+c_du_d+c_mu_\alpha+c^\alpha(ind\{\xi(t)=3\}+ind\{\xi(t)=4\}) \\ +c_{env}(x_{1max}-x_1)+c_{ecart}\left|(|x_2|-x_1)\right|ind\{x_2<0\}+\sum_{j\neq\alpha}\lambda_{\alpha j}v^h(x_1,x_2,j) \\ +\frac{|r_\alpha-u_{d_\alpha}-u_\alpha|}{h_1}v^{h_1}(x_1+h_1,x_2,\alpha)ind\{r_\alpha-u_{d_\alpha}-u_\alpha\geq 0\} \\ +\frac{|r_\alpha-u_{d_\alpha}-u_\alpha|}{h_1}v^{h_1}(x_1-h_1,x_2,\alpha)ind\{r_\alpha-u_{d_\alpha}-u_\alpha<0\} \\ +\frac{|u_\alpha-d|}{h_2}v^{h_2}(x_1,x_2+h_2,\alpha)ind\{u_\alpha-d\geq 0\} \\ +\frac{|u_\alpha-d|}{h_2}v^{h_2}(x_1,x_2-h_2,\alpha)ind\{u_\alpha-d<0\}\end{array}\right\}\end{array}\right] \quad (10)
$$

4. NUMERICAL ANALYSIS AND CONTROL POLICIES

In this section, to solve numerically the HJB equations in order to determine the optimal control policy of the recycling unit, we are going to use algorithm of Yan et Zhang (1997) and also matlab software. Furthermore, discretization step used is $h_{x1} = h_{x2} = 0.5$. A finite grid D defined below is necessary to circumscribe the domain for the state variables.

$$
D=\{0\leq x_1\leq 20;\ -10\leq x_2\leq 20\} \quad (11)
$$

The set of parameters needed during the simulation are presented in the table 2 below.

Table 2: Parameters of the numerical example

Variables	c_m	c_d	c_1	c_{2p}	c_{2m}	c^α	c_{ecart}
Value	14	1.5	1.8	3.5	55	13	55
Units	\$/product/UT	\$/product/UT	\$/product/UT	\$/product/UT	\$/product/UT	\$/break/UT	\$/product/UT

C_{env}	ρ	r_1	r_2	d	u_{max}
1	0.1	0.55	0.45	0.4	0.5
\$/product/UT	/UT	produit/UT	produit/UT	produit/UT	produit/UT

After running the simulation program, we have obtained the structure of optimal production and disposal policies, which are both, hedging point policy type. In all the following figures, x_1 represents the stock of returned product while x_2 represents the stock of finished products.

During mode 1 (machine available, return rate of end of life products is equal to r_1), the optimal production policy shown in fig.4 stipulates to produce at the maximum rate u_{max} when the inventory level of final products is strictly below the critical threshold $\psi(x_1)$ which is not constant, but depends also on the inventory level of returned product.

It stipulates to set the production rate at demand rate d when the inventory level of final products is equal to the critical threshold $\psi(x_1)$

Finally, it is needed to set the production rate at zero rate (no production) if the inventory level of final products is bigger than the threshold $\psi(x_1)$. That policy is summarized by equation (12).

$$
u_1(x_1,x_2,1)=\begin{cases} u_{max} \;\; if \;\; x_2<\psi(x_1) \\ d \;\; if \;\; x_2=\psi(x_1) \\ 0 \;\; if \;\; x_2>\psi(x_1)\end{cases} \quad (12)
$$

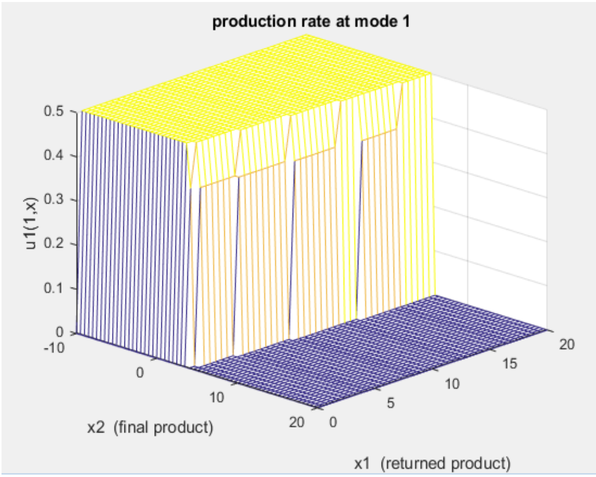


Figure 3 Production rate at mode 1

At the same mode 1, the optimal disposal policy u_{d_1} shown in fig.5 stipulates to dispose the returned products at the maximum rate r1 when the returned products level is strictly higher than the critical threshold $z_{12}=\sigma(x_2)$, which is a function that gives the threshold depending on inventory level of final products x_2. It also stipulates to set the disposal rate at the rate r1-umax when the inventory level of returned products is equal to the critical threshold $z_{12}=\sigma(x_2)$.

Finally, it is needed to set the disposal rate at zero rates if the inventory level of returned product is lower than the threshold $z_{12}=\sigma(x_2)$; in other words to store the entire returned products that conveyed to the recycling unit. This disposal policy is summarized by equation (13).

$$
u_{d_1}(x_1,x_2,1)=\begin{cases} r_1 \;\; if \;\; x_1>\sigma(x_2) \\ r_1-u_{max} \;\; if \;\; x_1=\sigma(x_2) \\ 0 \;\; if \;\; x_1<\sigma(x_2)\end{cases} \quad (13)
$$

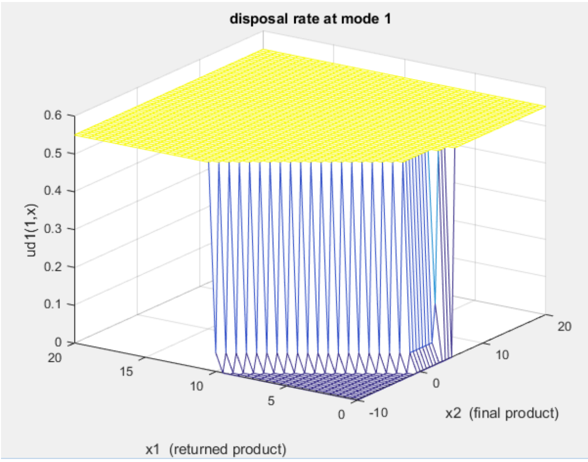


Figure 4 Disposal rate at mode 1

At the mode 2 of the unit (machine failed, return rate equal to r_2), the optimal production policy shows in fig.6 stipulate the following rule for the machine:

- Set the machine rate to r_2 when the stock level of final products x2 is under the threshold $z_{21} = \varpi(x_1)$, where function $\varpi(x_1)$ is a function depending on level of returned products x_1.
- Set the machine rate to the demand rate d when the level stock x_2 is equal to the threshold $z_{21} = \varpi(x_1)$.
- Stop the production of the machine when the stock level x_2 is greater than the threshold $z_{21} = \varpi(x_1)$. The production policy at state 2 is summarized by equation (14).

$$u(x_1, x_2, 2) = \begin{cases} r_2 & if \quad x_2 < \varpi(x_1) \\ d & if \quad x_2 = \varpi(x_1) \\ 0 & if \quad x_2 > \varpi(x_1) \end{cases} \qquad (14)$$

In the same way, disposal policy at mode 2 is shown in fig.7 and is summarized by equation (15). In this equation, $z_{22} = \delta(x_2)$ is the threshold for the disposal rate depending on the inventory of final products x_2.

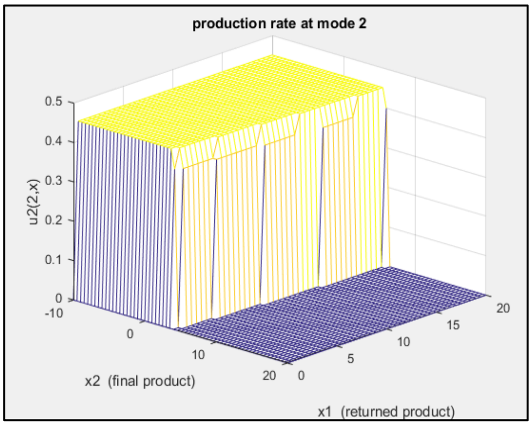


Figure 5 Production rate at mode 2

$$u_{d_2}(x_1, x_2, 2) = \begin{cases} r_2 \; if & x_1 \geq \delta(x_2) \\ 0 \quad if & x_1 < \delta(x_2) \end{cases} \quad (15)$$

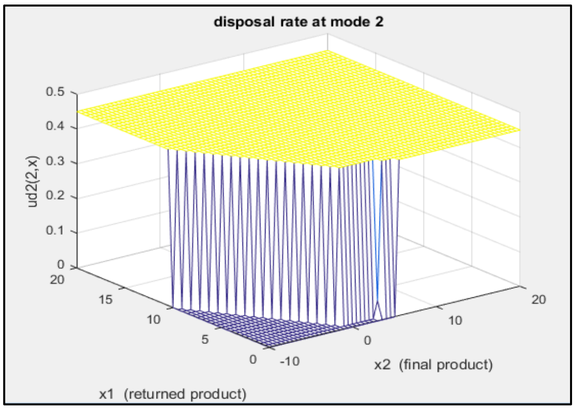


Figure 6 Disposal rate at mode 2

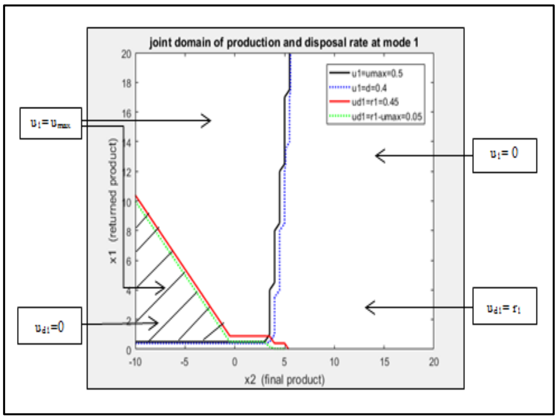


Figure 7: Feasible domain at mode 1

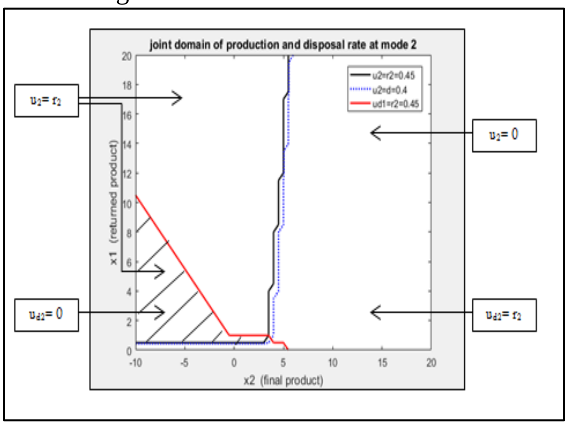


Figure 8: Feasible domain at mode 2

During the modes 3 and 4, the machine is broken, there is no possible production, and hence disposal policies are summarized respectively by equation 16 and 17 and can be interpreted as previously. In those equations, $z_3 = \eta(x_2)$ and $z_4 = \iota(x_2)$ represent respectively the function that give the threshold of disposal policy at mode 3 and 4 depending on the level of the final products storage.

$$u_{d_3}(x_1, x_2, 3) = \begin{cases} r_1 \; if & x_1 \geq \eta(x_2) \\ 0 \quad if & x_1 < \eta(x_2) \end{cases} \qquad (16)$$

$$u_{d_4}(x_1, x_2, 4) = \begin{cases} r_2 \; if & x_1 \geq \iota(x_2) \\ 0 \quad if & x_1 < \iota(x_2) \end{cases} \qquad (17)$$

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The hedging point policy structure obtained in this study is an extension of woks of Akella et Kumar (1986). Although they have obtained a fixed threshold, we obtained variable thresholds for the control policies obtained in this paper.

Having obtaining a disposal and a production policies simultaneously at the operational modes 1 and 2, we legitimately want to know where the raw material required for the production at those modes would come from since at the same modes, returned product is disposed ? To answer this question, we have studied the joint domain of the two policies (disposal and production rates) illustrated by fig. 8 and 9 (hatched area), which clearly shows the existence of a joint domain for production at a non-zero rate ($u_1 = u_{max}$, $u_2 = r_2$) and the elimination of returns at the zero rate ($u_{d_1} = 0,\ u_{d_2} = 0$). This intersection corresponds to the feasibility domain, meaning the field where the machine will operate while being sufficiently fueled with material.

5. CONCLUSION

Attending to the end of our study, we can conclude that, the objectives have been achieved. We have considered the problem of control of production and disposal policies for an aluminum recycling unit in open loop. The unit considered was considered as a single machine producing a single type of product and subject to stochastic random breakdowns and repairs. After modeling the system using a homogeneous Markov chain, we have determined the optimality conditions through the HJB equations. We were able to solve these equations numerically by using the Kushner and Dupuis (1992) method and determined the optimal controls. We have obtained the optimal policies which are of critical type (Hedging Point Policy) at each mode of the machine (running or broken). In our study, we showed that, the optimal critical threshold of finished products depends on the inventory level of returned products which is quite realistic.

REFERENCES

Akella, Ramakrishna, et P Roshan Kumar. 1986. « Optimal control of production rate in a failure prone manufacturing system ». *Automatic Control, IEEE Transactions on,* vol. 31, n° 2, p. 116-126.

Garceau, J. J. 1996. *Analyse de rentabilité en ingénierie* (1996). Trois-Rivières: Éditions SMG.

Gershwin, Stanley B. 1994. *Manufacturing systems engineering*. Prentice Hall.

Gharbi, A., et J. P. Kenne. 2000. « Production and preventive maintenance rates control for a manufacturing system: An experimental design approach ». *International Journal of Production Economics,* vol. 65, n° 3, p. 275-287.

Gharbi, Ali, et Jean Pierre Kenne. 2003. « Optimal production control problem in stochastic multiple-product multiple-machine manufacturing systems ». *IIE transactions,* vol. 35, n° 10, p. 941-952.

Kenne, J. P., E. K. Boukas et A. Gharbi. 2003. « Control of production and corrective maintenance rates in a multiple-machine, multiple-product manufacturing system ». *Mathematical and Computer Modelling,* vol. 38, n° 3–4, p. 351-365.

kenne, Jean-Pierre. 2004. « Commande optimale stochastique appliquée à la gestion de production d'une unité de fabrication ». *Journal européen des systèmes automatisés,* vol. 38, n° 3-4, p. 269-290.

Kenné, Jean-Pierre, Pierre Dejax et Ali Gharbi. 2012. « Production planning of a hybrid manufacturing–remanufacturing system under uncertainty within a closed-loop supply chain ». *International Journal of Production Economics,* vol. 135, n° 1, p. 81-93.

Kushner, Harold, et Paul G Dupuis. 1992. *Numerical methods for stochastic control problems in continuous time*, 24. Springer Science & Business Media.

Nodem, FI Dehayem, JP Kenné et A Gharbi. 2011. « Simultaneous control of production, repair/replacement and preventive maintenance of deteriorating manufacturing systems ». *International Journal of Production Economics*, vol. 134, n° 1, p. 271-282.

Rivera-Gómez, Héctor, Ali Gharbi, Jean-Pierre Kenné, Oscar Montaño-Arango et Eva Selene Hernandez-Gress. 2016. « Production control problem integrating overhaul and subcontracting strategies for a quality deteriorating manufacturing system ». *International Journal of Production Economics*, vol. 171, p. 134-150.

Rogers, D.S, et R.S. Tibben-Lembke. 1998. « Going Backwards : Reverse Logistics Trends and Practices ». *Reno, Nevada: Reverse Logistics Executive Council.*

Thomas, MP, et AH Wirtz. 1994. « The ecological demand and practice for recycling of aluminium ». *Resources, Conservation and Recycling*, vol. 10, n° 1, p. 193-204.

Yan, H, et Q Zhang. 1997. « A numerical method in optimal production and setup scheduling of stochastic manufacturing systems ». *Automatic Control, IEEE Transactions on,* vol. 42, n° 10, p. 1452-1455.

AGENT BASED MODELING ARCHITECTURE WITH BPMN AND DEVS NETWORK

Mariem SBAYOU, Youssef BOUANAN, Gregory ZACHAREWICZ, Julien FRANCOIS

Univ. Bordeaux, IMS Laboratory, 351, Cours de la Libération
Talence, 33400, France

{firstname.lastname}@u-bordeaux.fr

ABSTRACT

The adoption of Business Process (BP) can deal with the (re)development of information process, for instance it can help healthcare providers structuring the way information system and people have to interact. Business Process Management (BPM) is known as a methodology that aims to give a structured way of representing processes of systems. At the same time, the human resources are organized in identified or implicit structures that allows individual to exchange information either related to their work function or not. Nevertheless, the human organizations structure and communication channels are not, up to now, fully captured by the information systems. It may lead to lose part of useful information exchanged by participants. Accordingly, this paper focuses on multi-agent solutions representing social networks in the healthcare domain associated with BPM of patient pathways. The purpose is to combine BP with agent-based models in order to better improve performance and manage resources.

Keywords: Business Process, Business Process Management, multi-agent, performance, resources.

1. INTRODUCTION

Healthcare processes or pathways are typically described informally by text description or by semi-formal languages (Eshuis et al. 2010) . A process is generally defined as a sequence of events that uses inputs to produce outputs. Besides, a business process is an activity or set of activities that will accomplish a specific organizational goal. It is also considered as a sequence of performed steps that drives information to produce goods and/or provides services. The business process needs to be managed and controlled; it attests the need to use a Business Process Management (BPM) methodology. According to (Van Der Aalst, Ter Hofstede, and Weske 2003) the business process management (BPM) includes methods, techniques, and tools to support the design, enactment, management, and analysis of operational business processes. It can be considered as a formalization of classical Workflow Management (WFM) systems and approaches. BPM process solutions enable enterprise to measure and standardize processes and also provide reusable processes that can be networked. In this context, several languages are used for BPM; the most used one nowadays is Business Process Modeling Notation (BPMN). BPMN has been proposed by the Business Process Modelling Initiative (BPMI) and is currently maintained by the Object Management Group (OMG 2003) that provides this standard for IT and business actors. BPMN is an increasingly significant standard for process modeling and has received high attention and uptake in BPM practice (Recker 2008). It is an established standard for business process modelling in industry and economy and frequently supported by a computer program which enables a quite easy graphical description of complex processes (Scheuerlein et al. 2012).

As argued in (Antonacci et al. 2016), the use of BPM in healthcare sector is becoming a key enabler for the improvement of healthcare processes, since the healthcare environment is becoming more dynamic and volatile, and follows more complex processes, the combination of agent based model with business process may be efficient for resolving simulation's limitations in BP in terms of resource allocation, in the other hand, managing the availability of resources in healthcare sector is a very challenging problem with very little research attention. For this reason, the use of business process modeling combined with the adoption of simulation-based analysis provides a cost effective, accurate, and rapid way to evaluate alternatives before committing the required effort and resources (Tumay 1996; Nakatumba, Rozinat, and Russell 2009).

Furthermore, the analysis of performance of BPMN is crucial, it helps Business analysts to predict whether the goal can be achieved or not. For this reason, simulation has been identified as a key technique for BP performance analysis (Antonacci et al. 2016). It can be important for business processes as it helps and supports the decision making process, reduces cost that can occurs in the case wrong decisions haven't been anticipated and as final objective provides a good quality of services.

A viable approach to address the problem of managing the resources allocation is Modeling and Simulation. In this paper we propose a new approach which combines BPMN with agent based model in the case of healthcare systems, the proposed approach adopts DEVS (Discrete

Event System Specification) formalism to analyze the healthcare process behavior. The reminder of this paper is structured as follows first a state of art and background are presented, then we justify our motivation. Section 4 provides the application of the approach in the emergency cases. Finally, conclusions and plans for future works are presented in Section 5.

2. RELATED WORK AND BACKGROUND

BPMN is a graphical notation for drawing business processes (OMG 2003). It provides a standard notation that is easily understandable by all stakeholders; and also bridges the communication gap that frequently occurs between business process design and implementation. Nowadays BPMN becomes widely used by different organizations, as it is simple to learn, but yet powerful enough to depict the potential complexities of business processes. One of the main complexity of BP is the limitation of its simulation, which is due to the following reasons: lack of simulation know-how of BP analysts, costs and difficulties in retrieving and analyzing the data required for simulation model parameterization, large semantic gap between the business process model and the simulation model and finally, the use of models that may be (partially) incorrect or may not be at the right level of abstraction (Bocciarelli, D'Ambrogio, and Paglia 2014). In this context, the use of BPMN models in healthcare sector helps to facilitate the management of complex hospital BPs (Antonacci et al. 2016).

In addition, simulation proved its value in the manufacturing sector and has been used to evaluate process problems in healthcare as well (Harrell and Price 2000). It has been used for over a decade by health services for improving patient care. In this respect, there are a lot of examples in the literature which provide simulation tools for healthcare organization, in (Laskowski and Mukhi 2008), authors have developed a tool for managing emergency department, by planning capacity resources utilization and staff capacity. Moreover, authors in (Norouzzadeh et al. 2015) show how modelling and simulation of internal medicine practice process can help on decision making. The results of simulation, based principally on patient waiting time were able to give an idea about improving resources utilization. Another example is given by Gűnal and Pidd in (Gehlot, Matthew, and Sloane 2016) which describes a model of the process flow of patients, that represents the multitasking behavior of medical staff (doctors and nurses), the only issue is that the ignorance of other possible factors such as: doctors' interactions with patients, other medical staff and their working environment lead to incomplete consideration of the problem (Jain et al. 2011).

Moreover, popular approaches include decision analysis, Markov process, mathematical modeling, systems dynamics and discrete event simulation (Fone et al. 2003) (Kanagarajah et al. 2010). The limitation of these approaches is that they ignore the effect of naturalistic human decision-making and behaviors on the performance of healthcare processes. In order to overcome such limitation the use of Agent Based Modeling (ABM) offers complementary perspectives to model the process of health care domain (Wang 2009). Agent based modeling (ABM) is formed by a set of autonomous agents that interact with their environment and other agents through a set of internal rules to achieve their objectives (Onggo 2010) (Grundspenkis and Pozdnyakov 2006).

The purpose of this paper is to present architecture for modeling and simulation in healthcare domain, the proposed architecture aims to overcome BPMN limitations and drawbacks such as allocation of resources. To this goal, we propose new approach which uses DEVS simulation for emergency department and that considers both healthcare participants and actors.

In this context, there are some studies In the literature using BPMN model with DEVS model, these studies are based on meta-model approach (OMG 2003), which is one of the most used transformation techniques that includes the mapping of BPMN concepts to DEVS concepts. Based on the proposed approach of BPMN to DEVS (Cetinkaya, Verbraeck, and Seck 2012) which presented a Model Driven Development (MDD) framework for modeling and Simulation (MDD4MS), and where a set of transformation rules were defined: some BPMN concepts (Pools, Lane, SubProcess) were mapped to DEVS coupled component while task, event (start, end and intermediate) and Gateway were mapped to DEVS atomic component. This proposal doesn't cover the intervention of different resources like (human resources, devices and/or Software services) which may affect the execution of the task. To overcome such limitations, authors in (D'Ambrogio and Zacharewicz 2016) proposed new approach, by introducing reliability analysis that takes into consideration unexpected failures of the resources that execute the process tasks (unavailability of a resource allocated to task).

Discrete Event system Specification (DEVS) (Zeigler, Praehofer, and Kim 2000) is formalism for modelling Discrete Events Systems. The hierarchical and modular structure of DEVS allows defining multiple models that are coupled to work together in a single and model by connecting their input and output through messages (Wainer 2009). In the same way, the resulting model can also be coupled with others models defining multiple layers in the hierarchical structure. In DEVS, atomic models define the behaviour of the system, and coupled models describe the structure of the system. The DEVS formalism has several advantages. The reason behind using DEVS is that it is based on dynamical systems theory and provides well defined concepts for coupling components, hierarchical and modular model construction, and an object oriented substrate supporting repository reuse. Modular construction is one of the most important characteristics of DEVS because it allows the modeler to design and construct each model independently for optimal efficiency. As long as models adhere to certain protocols, they can interact which each other (Pérez et al. 2010). Accordingly, we use VLE (Virtual Laboratory Environment) for the implementation.VLE is an open source software and API under GPL which supports multi-modeling and simulation by implementing the DEVS abstract simulator(Quesnel, Duboz, and Ramat 2009). It integrates specific models developed in most popular programming languages into one single multi-model. It also proposes several simulators for particular formalisms; for example, cellular automata, ordinary differential equations (ODE), differential equations, various finite state automata (Moore, Mealy, Petri-Nets, etc.) and so on. VLE can be used to model, simulate, analyse and visualize dynamics of complex systems. Its main features are: multi-modeling abilities (coupling heterogeneous models), a general formal basis for modelling dynamic systems and an associated operational semantic, a modular and hierarchical representation of the structure of coupled models with associated coupling and coordination algorithms, distributed simulations, a component based development for the acceptance of new visualization tools, storage formats and experimental frame design tools (Bouanan et al. 2015). In the next sections, we give an overview about the architecture and its application.

3. CONTRIBUTION

3.1. Problem Statement

BPM depends on a very important notion which is a workflow. Workflow can be any business process, which consists of two or more tasks performed in social on concurrently by two or more people. It should assure the right people at the right time. It also provides general information about the business process: individuals and teams needed to complete task, information and resources needed to complete task, finally; Dependencies and deadlines for task completion (Grundspenkis and Pozdnyakov 2006). The use of workflow in healthcare domain is thriving, but since healthcare environment is evolving and complex to manage, it faces some drawbacks including: limited flexibility during process enactment (Bolcer and Taylor 1998), inability to cope with dynamic changes in resource levels and task availability, limited ability to predict changes, due to external events, in both the volume and composition of work, lack of performance, scalability and reliability as well (Pang 2000). For overcoming some of the mentioned limitations, an agent based oriented approach is proposed, this approach aims to involve agent network in coordination of BP. The main idea is to connect only agents who perform required tasks for achieving the goal; this connection is established through the flow information exchanged in the workflow.

3.2. Agent Based Healthcare Process architecture

The proposed architechture aims to combine BPMN with DEVS in ordre to manage resources. Fo this, we classify entities by considering their roles and interactions they have within a multilayer network (Sbayou et al. 2017). This network uses information from both an XML file generated from BPMN diagram and database information for input simulation. Thus, the general model is divided into two parts:

- Coupled model of agents that perform tasks.
- Decisional tasks model to orchestrate network.

This method allows us the integration of different relationships between actors in BPMN that are represented with a node in network representation (Bouanan et al. 2015). Actors are defined as agents, which are represented in the agent based model as individuals or group of individuals. Each agent is described by a set of attributes distinguished into two categories:

- Static attributes i.e., id, gender, and status.
- Dynamic attributes (variables) i.e., availability and state.

Static attributes are intrinsic or unchanged parameters, i.e., time has no effect on them. Dynamic attributes evolve with time or events (Bouanan et al. 2016) (Ruiz-Martin et al. 2016). For instance, doctors can be reached depending on their availability based on worktime and number of patient. The DEVS model is then used for simulation. Final results can be compared to those estimated by experts. We aim also to cover by our architecture various existing situations or structures in healthcare organizations thanks to BPMN and DEVS M&S in order to make it general and not only available for special cases.

In order to apply our approach, at first the user draws a BPMN process of the studied case, in order to clearly represent the role of each agent and to make the process unambiguous. Then the tool generate an XML file which contains information about decisional tasks, these tasks are responsible of the orchestration of the general model. In parallel, it extracts input information required for simulation from a database that contains information about population and healthcare stakeholders.

Once the XML file is generated and data are collected, the DEVS network is created, and it is represented as a multilayer network where each layer describes a level of connection between agents (Sbayou et al. 2017). The DEVS network is created within VLE; which uses the XML file that contains elements which assure connection between represented agents in the network. It also uses R-Studio tool for visualizing results. This method allows us the integration of different relationships between actors in BPMN that are represented with a node in network representation.

We propose an overview picture of our architecture in Figure 1. Actors are defined as BPMN resources (lanes in Figure 1 left part) and then M&S agents (represented by nodes in the figure 1 center part).

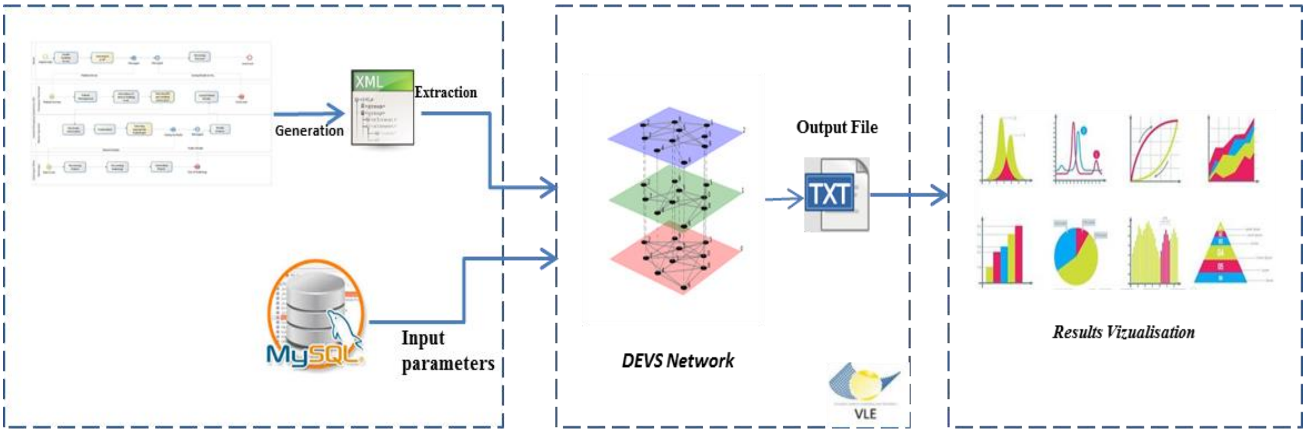


Figure 1: Agent Based Healthcare Process architecture

The BPMN model is describing the medical pathway of patients. It integrates the connection of patient with the different medical resources. Medical resources are including: General Practitioners (GP), Hospital (Fr: CHU) and medical specialists. In some cases, the patient can directly go to specialists without passing through GP. In the study case, these specialists are: gynecologists, ophthalmologists, dentists and psychiatrists (for patients under 26 years old). Then the agent-based model is considering agents as individuals or group of individuals. It represents the social interaction between agents.

3.3. Controller atomic model

In order to implement the architecture, we propose an atomic model as a controller which is based on the XML file generated from the BPMN diagram; it is considered as the dynamic part of our architecture this atomic model is in charge of orchestrating the DEVS network of different agents who performs the process. The proposed model is shown in figure (5), it has input "xml_file", and number of outputs ports which depends on the process, these ports are used to send information to the general model, the general model is a coupled model of atomic models, where each atomic model describes a role in the BPMN process.

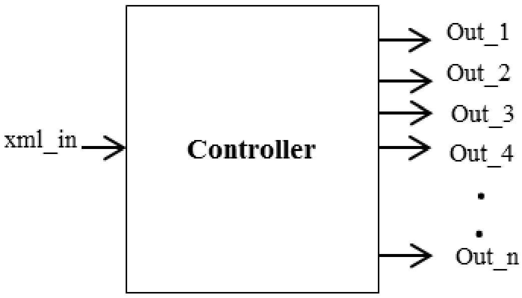


Figure 2: Basic atomic model of controller

3.4. Healthcare system in France

In order to apply our approach we briefly introduce the specific context of health related sector in France. It will permit to understand the process of the healthcare system we describe in our example in France.

The French healthcare system covers both public and private hospitals, doctors and other medical specialists who provide care services to French resident. It is accessible for all residents, independently of their age, income or status. French resident has to register a General Practitioner (GP) "Fr: *Medecin Traitant*", in order to ensure full eligibility to reimbursement of health costs. While following this process, the General Practitioner (GP or G), becomes the principal route which follows the patient care pathways "Fr: *Parcours de soin*". In emergency cases, there is no appointment needed, in addition, Emergency Departments (ED) are the most complex system in healthcare sector. They usually require a lot of resources, at any time.

4. CASE STUDY

4.1. Data Collection

In order to apply our architecture, we have localized the different French hospitals which includes emergency department (Figure 3) according to (CNOM 2017), and we localize them in google maps, also we generate a small population of patient, which is near the hospitals and also those who are living far. In our first study we will only take the case of the patient who are living near to the chosen hospital. We focused our study in the southwest region of France (*Nouvelle Aquitaine*); we consider the case of the University Hospital Center (*CHU of Bordeaux*). Considering the CHU of Bordeaux; we have collected some statistical information about the patients as well as their care in the emergency department. The hospital is visited daily by about 149 patients, the completeness of reception is 100%, the sex ratio representing the number of men and number of women at birth during a given age group is 1,17. 1% of patients visiting the hospital are under one year old, 14% are patients under 18 years of age, patients aged 75 years and over account for 16%, and 7% of patients are not from the region. The average duration of process for each patient is about 4 hours according to ("ORU" 2017). The figure 3 display several Emergency Departments (Green ED Bubbles) of the region of Bordeaux (including CHU of Bordeaux) and the set of Patients (Red P Bubbles).

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

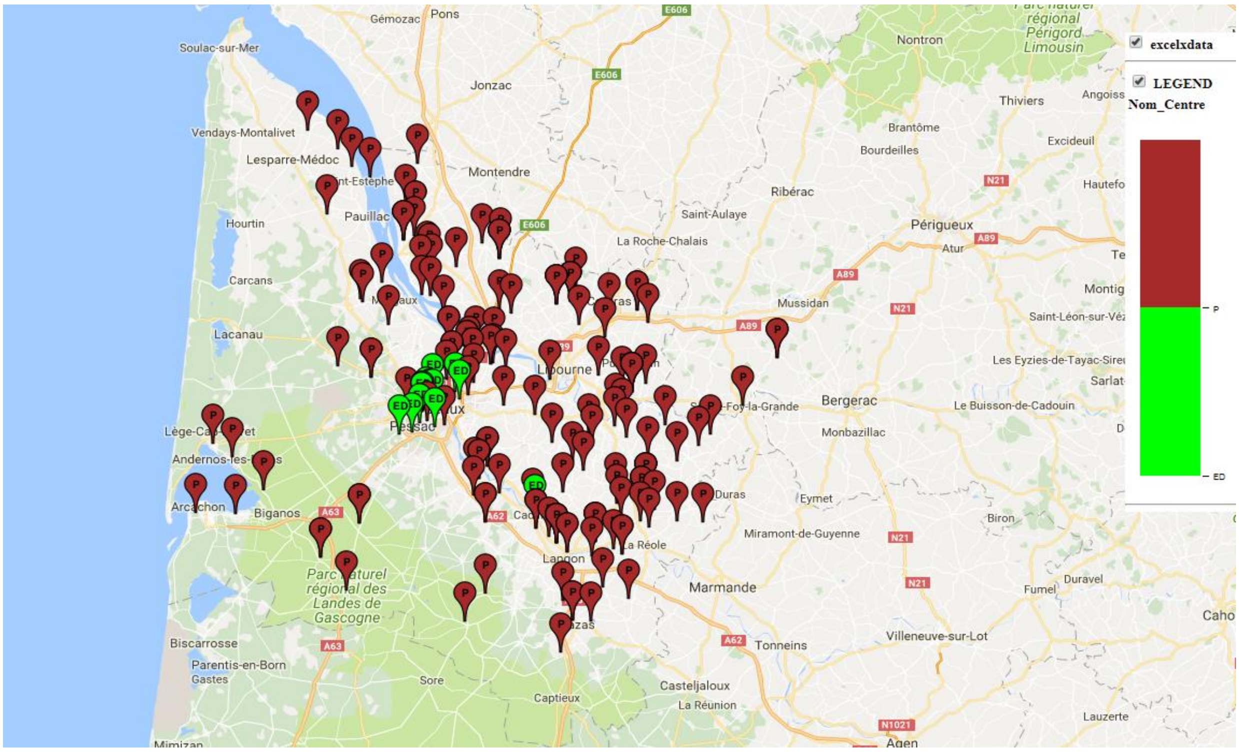


Figure 3: Geographic data in southwest of France

4.2. Emergency Department workflow
In this section we analyze the patient flow in emergency department of CHU of Bordeaux. Figure 4 shows in BPMN a sequence of steps followed by patient and health resources agents in the situation of emergency at ED, the diagram contains agent's roles, and each pool describes tasks to be performed by each agent. The process starts when a health problem occurs, once the patient has a health problem, he has to decide whether to visit his referred GP or to go the hospital. In our case the choice is focused on emergency cases, so the patient selects the hospital and generally the nearest one. Once the patient is received at the hospital, the administrative agent creates admission file for registration, redirects him to the waiting room, then the agent selects the next available practical nurse (Fr: *Infirmier Organisateur de l'Accueil, IOA*). It is described by a BPMN massage flow connection between patient and the CHU.
The priority of patient care is determined by the severity patient's health state. This degree of urgency is evaluated by the practical nurse (IOA) who evaluates the patient state, selects the available emergency physician and then provides him preliminary patient's information. Depending on the patient's health state, the emergency physician orients the patient, who may be referred to resuscitation (case of vital distress), examination boxes (priority patients) or consultations (ambulatory patient). After the installation in boxes, a waiting time of 2 hours minimum is envisaged (surveillance, collection of the results, possibility of recourse to specialist's opinion). In the other hand a hospitalization may be considered (conventional hospitalization, emergency hospitalization (short-term)).
The waiting time is often a source of incomprehension, anxiety and sometimes annoyance. This expectation may depends on number of patients, the arrival of other patients in severe conditions, the availability of resources (practical nurses, physicians and material resources such as beds...) (CHU Bordeaux 2017).

4.3. Operation atomic model for ED controller
In order to capture the detail of the collaboration between agents, the BPMN diagram (Figure 4) described 3 orange tasks (select H or GP; select practical nurse and provide information; select EP and provide information) which we named respectively (T1; T2; T3). These tasks are called decisional tasks, they are used for simulation and are responsible of verifying the availability of resources and connecting them. The operation of the controller in the case of ED has four basic states then set in the DEVS model: «IDLE, "State1", "State2" and "State3" described in Figure 5.
According to the BPMN, the DEVS model (Figure 5) is initialized in "IDLE" state where it reads the XML file and checks for decisional tasks. Then the model can change to the next state according to the convenient task given by the BPMN. The time advance of each tasks depends on the parameters set by the modeler at the BPMN building step at the initialization step. In case described Figure 4, the general model is a coupled model of four atomic models (Patient, Administrative Agent, Nurse and emergency physicians) which can be reused for other case studies, each model is linked with a specific output of the 'Controller' model.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

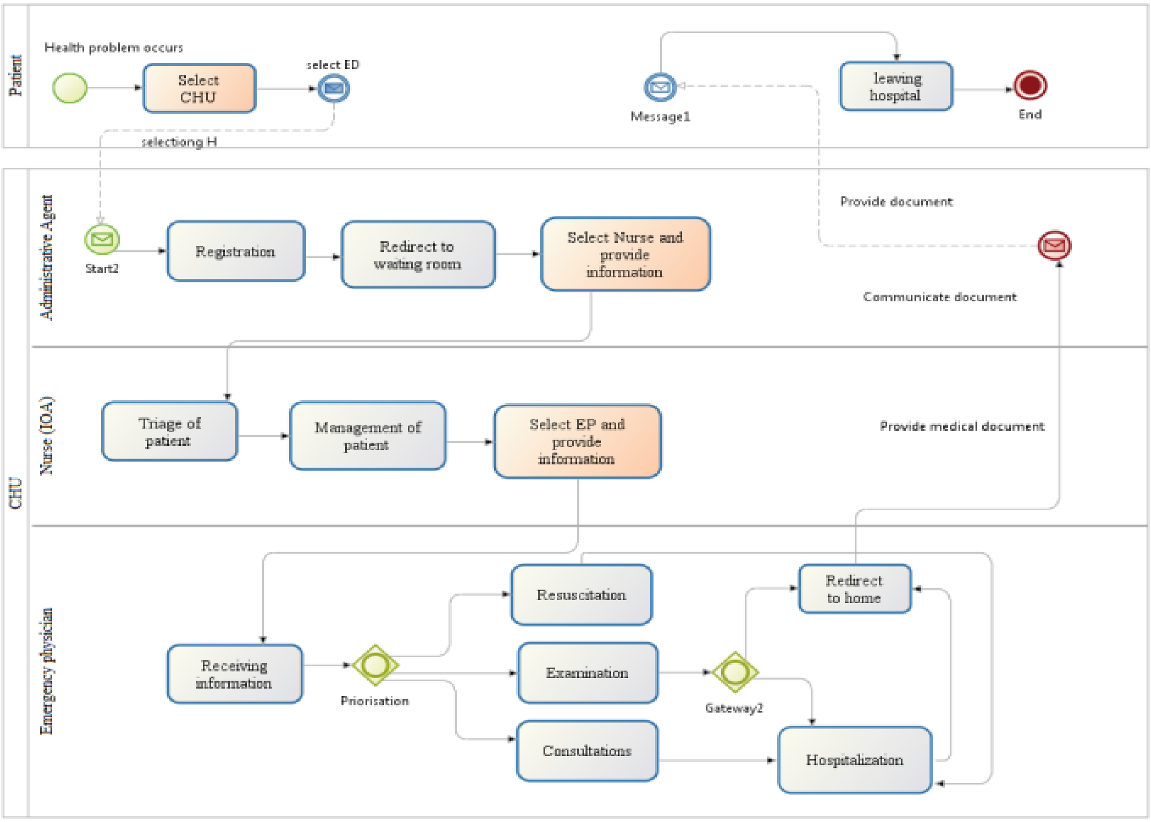


Figure 4: ED patient pathway

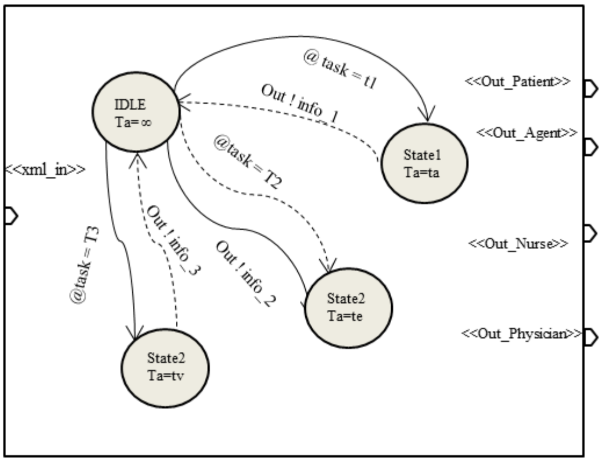


Figure 5: Operation of the controller model

4.4. Results and discussion

In this section we present results based on some data (ORU 2017) (CNOM 2017). Our aim is to simulate the allocation of resources according to some input parameters (number of arrived patient, priority, number of resources and their availability) and also the xml generated file.

For this we take the case of 10 arrived patients, at a specific time, each patient is connected to the administrative agent according to his priority, the arriving of a lot of patient at close time may affect the length of stay (LOS) which we are going to include in the next works. Once the patient is registered, information is sent to the available nurse, the nurse selects the available physician, and affect it the prior patient, and the connection between available resources and seek patient is established.

At the end; each patient has his own health network which includes (administrative agent, nurse, and emergency physician) who participate on performing the patient emergency flow.

In order to apply the approach and ensure its feasibility, we have chosen the case of 1 administrative agent, 3 nurses and 4 emergency physicians. In order to simulate the number of patients who reached the end of the process. The output file of the simulation contains the id of agents and their states, the figure 6 shows simulated DEVS network of one agent (A) connected to several resources (patients (Pi), nurses (Ni) and physicians (Di)). It details all connected resources in the case of arrival of 10 patients at a specific time.

We can observe that some patients are only registered because there is not enough health resources which may take them in charge. We can also observe that that administrative agent is connected to all coming patients that are still waiting for their turn in order to join the network.

This study simulates the allocation of available resources, the main idea is to test and verify the proposed architecture in this paper. In that objective, we use a DEVS network which is helpful in multidimensional structures where interaction between agents is complex.

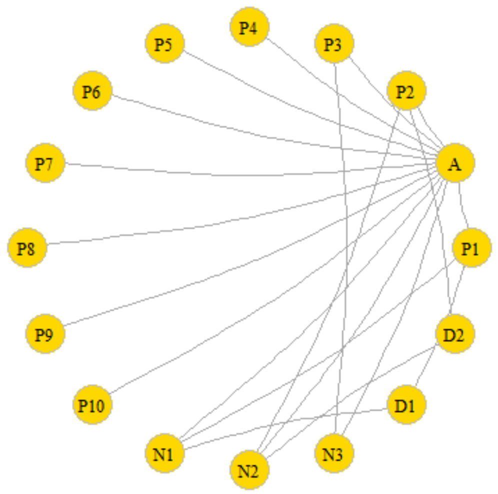


Figure 6: ED department resource network

Our objective in the next work is to play scenarios based on the number of resources and the performed tasks in the BPMN, and also include time parameters in order to improve waiting time in different steps of the process, We are going to focus now on resource allocation time and time to reach the proper resources (so including the geolocation information of agents). In particular we are studying the different time duration of different tasks also integrating performance of different resources depending on their solicitation.

5. CONCLUSION AND PERSPECTIVES

The paper has reported a new approach that combines BPMN with DEVS, applied at first in healthcare modelling sector. The idea presented in this paper was to demonstrate the interest of coupling a health care patient pathway workflows modelled with BPMN with different healthcare resources organized in social networks described with DEVS. Such an approach can be proposed to study a territory in terms of sufficient or insufficient resources available in a specific area. The case study, located in the region of Bordeaux, has been possible thanks to data coming from a national repository that publishes the list of healthcare resources with their geographic location. The presented case study aims to manage resources and shows their impact on the performance of the process. In the coming works we will include new parameters as new resources (equipment) in order to manage the overcrowding of ED. We also want to apply our architecture in other sectors in order to make it functional for different application domain.

6. REFERENCES

Antonacci, G., A. Calabrese, A. D'Ambrogio, A. Giglio, B. Intrigila, and N. Levialdi Ghiron. 2016. "A BPMN-Based Automated Approach for the Analysis of Healthcare Processes." In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2016 IEEE 25th International Conference on*, 124–129.

Bocciarelli, Paolo, Andrea D'Ambrogio, and Emiliano Paglia. 2014. "A Language for Enabling Model-Driven Analysis of Business Processes." In *Model-Driven Engineering and Software Development (MODELSWARD), 2014 2nd International Conference on*, 325–332.

Bolcer, Gregory Alan, and Richard N. Taylor. 1998. "Advanced Workflow Management Technologies." *Software Process: Improvement and Practice* 4 (3): 125–171.

Bouanan, Youssef, Mathilde Forestier, Judicael Ribault, Gregory Zacharewicz, Bruno Vallespir, and Nejib Moalla. 2015. "Simulating Information Diffusion in a Multidimensional Social Network Using the DEVS Formalism (WIP)." In *Proceedings of the Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium*, 63–68.

Bouanan, Youssef, Gregory Zacharewicz, Bruno Vallespir, Judicael Ribault, and Saikou Y. Diallo. 2016. "DEVS Based Network: Modeling and Simulation of Propagation Processes in a Multi-Layers Network." In *Proceedings of the Modeling and Simulation of Complexity in Intelligent, Adaptive and Autonomous Systems 2016 (MSCIAAS 2016) and Space Simulation for Planetary Space Exploration (SPACE 2016)*, 8:1–8:8. MSCIAAS '16. San Diego, CA, USA.

CHU Bordeaux. 2017. *Accueil Aux Urgences*. Accessed July 21. https://www.chu-bordeaux.fr/.

CNOM. 2017. "Cartographie Interactive de La Démographie Médicale." Accessed July 21. https://demographie.medecin.fr/.

D'Ambrogio, Andrea, and Gregory Zacharewicz. 2016. "Resource-Based Modeling and Simulation of Business Processes." In *Proceedings of the Summer Computer Simulation Conference*, 63:1–63:8. SCSC '16. San Diego, CA, USA: Society for Computer Simulation International..

Eshuis, P., E. ten Pas Sopheon, H. Rutten Sopheon, and J.-M. Verlinden Zorggemak. 2010. "State of the Art Clinical Pathway Definition: Gap Analysis."

Fone, David, Sandra Hollinghurst, Mark Temple, Alison Round, Nathan Lester, Alison Weightman, Katherine Roberts, Edward Coyle, Gwyn Bevan, and Stephen Palmer. 2003. "Systematic Review of the Use and Value of Computer Simulation Modelling in Population Health and Health Care Delivery." *Journal of Public Health* 25 (4): 325–335.

Gehlot, Vijay, Liji Matthew, and Elliot B. Sloane. 2016. "From Data to Processes- Use of Modeling and Simulation in Healthcare." In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 320–323.

Grundspenkis, Janis, and Dmitrij Pozdnyakov. 2006. "An Overview of the Agent Based Systems for the Business Process Management." In *Proc. Int'l Conf. Computer Systems and Technologies*.

Kanagarajah, Ashok Kay, Peter Lindsay, Anne Miller, and David Parker. 2010. "An Exploration into the

Uses of Agent-Based Modeling to Improve Quality of Healthcare." In *Unifying Themes in Complex Systems*, 471–478.
Laskowski, Marek, and Shamir Mukhi. 2008. "Agent-Based Simulation of Emergency Departments with Patient Diversion." In *International Conference on Electronic Healthcare*, 25–37.
Nakatumba, J., A. Rozinat, and N. Russell. 2009. "Business Process Simulation: How to Get It Right." In *International Handbook on Business Process Management*. Citeseer.
Norouzzadeh, Shaghayegh, Nancy Riebling, Lawrence Carter, Joseph Conigliaro, and Martin E. Doerfler. 2015. "Simulation Modeling to Optimize Healthcare Delivery in an Outpatient Clinic." In *2015 Winter Simulation Conference (WSC)*, 1355–1366.
OMG. 2003. "MDA Guide Version 1.0.1." Edited by Joaquin Miller and Jishnu Mukerji, June.
Onggo, Bhakti SS. 2010. "Running Agent-Based Models on a Discrete-Event Simulator." In *Proceedings of the 24th European Simulation and Modelling Conference*, 51–55.
ORU.2017. Accessed July 21. http://www.orulim.com/fichiers/PANORAMA_ORU_NA_WEB_2015.pdf.
Pang, Gian. 2000. "Implementation of an Agent-Based Business Process." *Technical Report, University of Zurich*.
Pérez, Eduardo, Lewis Ntaimo, Carla Bailey, and Peter McCormack. 2010. "Modeling and Simulation of Nuclear Medicine Patient Service Management in DEVS." *Simulation* 86 (8–9): 481–501.
Quesnel, Gauthier, Raphaël Duboz, and Éric Ramat. 2009. "The Virtual Laboratory Environment–An Operational Framework for Multi-Modelling, Simulation and Analysis of Complex Dynamical Systems." *Simulation Modelling Practice and Theory* 17 (4): 641–653.
Recker, Jan C. 2008. "BPMN Modeling–who, Where, How and Why." *BPTrends* 5 (3): 1–8.
Ruiz-Martin, Cristina, Gabriel Wainer, Youssef Bouanan, Gregory Zacharewicz, and Adolfo López Paredes. 2016. "A Hybrid Approach to Study Communication in Emergency Plans." In *Winter Simulation Conference (WSC), 2016*, 1376–1387.
Sbayou, Youssef Bouanan, Gregory Zacharewicz, Judicael Ribault, and Julien François. 2017. "DEVS MODELLING AND SIMULATION FOR HEALTHCARE PROCESS APPLICATION FOR HOSPITAL EMERGENCY DEPARTMENT." *SpringSim 2017*.
Scheuerlein, Hubert, Falk Rauchfuss, Yves Dittmar, Rüdiger Molle, Torsten Lehmann, Nicole Pienkos, and Utz Settmacher. 2012. "New Methods for Clinical Pathways—business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (T. BPM)." *Langenbeck's Archives of Surgery* 397 (5): 755–761.
Tumay, Kerim. 1996. "Business Process Simulation." In *Proceedings of the 28th Conference on Winter Simulation*, 93–98.
Van Der Aalst, Wil MP, Arthur HM Ter Hofstede, and Mathias Weske. 2003. "Business Process Management: A Survey." In *International Conference on Business Process Management*, 1–12.
Wainer, Gabriel A. 2009. *Discrete-Event Modeling and Simulation: A Practitioner's Approach*. CRC Press.
Wang, Lu. 2009. "An Agent-Based Simulation for Workflow in Emergency Department." In *Systems and Information Engineering Design Symposium, 2009. SIEDS'09.*, 19–23.
Zeigler, Bernard P., Herbert Praehofer, and Tag Gon Kim. 2000. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic press.

7. AUTHOR BIOGRAPHIES

MARIEM SBAYOU is a Ph.D. Student at IMS Laboratory of University of Bordeaux. Her research domain includes DEVS, Healthcare M&S. She received her Engineer degree as software engineer from the International Academy of Civil Aviation in Morocco. Her email address is mariem.sbayou@u-bordeaux.fr.

GREGORY ZACHAREWICZ is Associate Professor HDR at University of Bordeaux and IMS Lab with both competences in enterprise engineering and M&S. His email address is gregory.zacharewicz@u-bordeaux.fr.

YOUSSEF BOUANAN is a Postdoctoral researcher at University of Bordeaux. He received his Ph.D. degree in Production Engineering from University of Bordeaux, France. His research interests include modeling and simulation theory, social network and workflow. His email address is youssef.bouanan@u-bordeaux.fr.

JULIEN FRANCOIS is Associate Professor at University of Bordeaux and IMS Lab His research interests include supply chain, modeling and simulation in industrial sector. His email address is julien.francois@u-bordeaux.fr.

ROBOT SOCCER STRATEGY ADAPTATION

Václav Svatoň[a], Jan Martinovič[b], Kateřina Slaninová[c], Václav Snášel[d]

[a],[b],[c],[d]IT4Innovations, VŠB - Technical University of Ostrava,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[a],[d]Faculty of Electrical Engineering and Computer Science, VŠB - Technical University of Ostrava,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[a]vaclav.svaton@vsb.cz, [b]jan.martinovic@vsb.cz, [c]katerina.slaninova@vsb.cz, [d]vaclav.snasel@vsb.cz

ABSTRACT

The robot soccer game presents an uncertain and dynamic environment for cooperating agents. Robot soccer is interesting for its multi-agent research, real-time image processing, robot control, path planning and machine learning. In our work, we present an approach to describe the strategies of the robot soccer game and propose a method of strategy adaptation based on the information from the previously played games and therefore acquiring better results than with the original strategy. Because this robot soccer strategy describes a real space and stores physical coordinates of real objects, proposed methods can be used in strategic planning in different areas where we know geographic positions of the objects.

Keywords: strategy adaptation, strategy planning, robot soccer, time-series

1. INTRODUCTION

The game situation on the playground in robot soccer games is typically read in terms of the robot's postures and the ball's position. Using near real-time information of this dynamically changing game situation, the system of robot soccer game would need to continually decide the action of each team robot, and to direct each robot to perform a selected action.

A strategy, as understood by a game theory (Osborne 2004; Kim, Kim, Kim, and Seow 2010), is a complete set of options which are available to players in any given situation in order to achieve the desired objective. This principle can be applied to a number of areas from the real world and is generally called strategy planning (Ontanón, Mishra, Sugandh, and Ram 2007). Games can in general represent any situation in the nature. Game theory could be applied for adversarial reasoning in security resource allocation and scheduling problems. Our approach is to use the strategies to describe a space and objects in it, and to use the subsequent search for the optimal path or relocation of these objects in order to achieve the desired goals.

The very objective of the robot soccer game is as simple as in real soccer. Win the game over an opponent by a higher number of scored goals. To achieve this goal, the best possible cooperation of team players and the adaptability to the opponent's actual strategy is necessary.

This paper is organized as follows: First, basic overview of robot soccer strategy and rule selection will be introduced in Section 2. Then, the proposed approach for the strategy adaptation based on the information from played games will be described in detail. In the Experiments and Results section, results of proposed approach will be demonstrated and evaluated. At the end, both advantages and disadvantages of the approach will be summarized and the future work will be outlined.

2. ROBOT SOCCER STRATEGY AND RULES

Our robot soccer architecture uses two types of representation of the game field (Martinovič, Snášel, Ochodková, Zoltá, and Wu 2010). Very accurate abstract coordinate system representation used to control the robots with the high precision and the grid coordinate system used for the strategy definition and the underlying rules (see Figure 1).

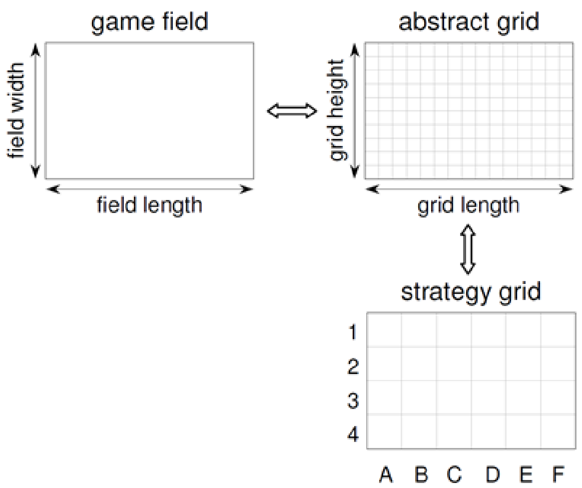


Figure 1: Game Field Representation

By using grid coordinates we reduce the accuracy of the mapping of physical coordinates of the robots into the logical ones, but on the other hand, we dramatically

Proceedings of the European Modeling and Simulation Symposium, 2017
ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

reduce the number of the rules needed to create a strategy. If necessary, it is possible to convert the grid coordinates back to the physical and to use them for the aforementioned robot control.
Strategy is a finite set of rules that describes the current situation on the game field. Each rule can be easily expressed as the quaternion containing the grid coordinates of our robots, the grid coordinates of opponent's robots, grid coordinates of the ball and the grid coordinates of where our robots should move in the next game step.

2.1. Rule Selection

The selection of the appropriate rule from strategy depends on the current game situation. This situation is represented by the current state of the game that contains information about the position of the robots and the ball on the game field and also their rotation and speed. On the basis of a priori defined metrics the most similar rule from the strategy is selected. This rule also contains the destination coordinates of the players for the following game step.

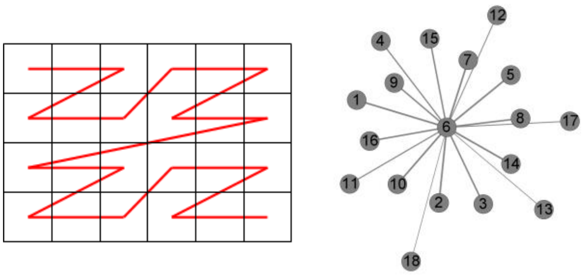


Figure 2: Z-order Mapping and Rule Graph

Because our robot soccer system architecture does not include the robots which are uniquely identifiable by ID or assigned roles, the method based utilizing the graph and Z-order curve (see Figure 2) was devised (Svatoň, Martinovič, Slaninová, and Snášel, 2014). Z-order is a function mapping the multi-dimensional space into the one-dimensional space while preserving the locality of data points. Due to its properties, it is used for converting two-dimensional matrix representing the playing field into one dimensional array of the coordinates of the individual robots. Before the start of the game the graph representing the similarities between the rules is precomputed. The set of vertices consists of the individual rules from the strategy and the edges contain the evaluation which corresponds to a distance between the two neighboring vertices (rules). As a distance is considered a normalized value of Euclidean distance computed from two sorted sequences using the above mentioned Z-order applied to the neighboring vertices which contain the robots grid coordinates.
Using this approach to a rule selection from the strategy, we were also able to utilize the more accurate description of strategies using substrategies. Substrategies allow the author to create a strategy, which will be performed during the game by such way, by which it was intended during its creation. Substrategies represent game situations such as left wing offence or right wing defense and ensures that players will perform the strategy actions more continuously and in a faster way then using the original approach.

3. STRATEGY ADAPTATION

Robot soccer game is dynamic and fast changing environment. In order to properly execute number of different tasks towards the given objective, the system should be able to evolve and have the flexibility to adapt to an opponent´s strategy. Different approach to a similar problem illustrated the use of the fuzzy decision making system (Huang and Liang, 2002) or the use of evolutionary algorithms inspired by biological evolution, such as reproduction, mutation, recombination, and selection to approximate the solution to a problem of strategy selection and adaptation. The work of Nakashima et al. (Nakashima, Takatani, Udo, Ishibuchi, and Nii, 2006) proposes an evolutionary method for acquiring team strategies of RoboCup soccer agents. They define a chromosome as a concatenated string of action rules for all agents. Larik et al. (Larik and Haider, 2016) used evolutionary algorithms for strategy optimization problem, Tominaga et al. (Tominaga, Takemura and Ishii, 2017) proposed an approach using SOM neural networks, Shengbing et al. utilized approach based on swarm intelligence methods (Shengbing, Gang and Xiaofeng, 2016) and for example decision making methods like the work of Akiama et al. (Akiyama, Tsuji and Aramaki, 2016).
There exist many different types of robot soccer games with many different architectures used in these games so it is understandable that the proposed methods for strategy description, rule selection and strategy adaptation are heavily dependent on the used robot soccer game architecture. Also the common problem is that different approaches have different understanding of the term strategy. Some of them see the problem as a low level description of how to move the robot, others as a high level abstraction that is used for the robot control with the higher granularity or from the global point of view.

Using our own architecture, the proposed method for strategy adaptation consist of the following steps:

1. Extract the information about the game progress from a log of a played game.
2. Detect the relevant game situations for strategy adaptation.
3. Analyze the rules preceding the relevant situation.
4. Create aggregated anti-rule/s for the relevant situation.
5. Include the newly created rules to the original strategy.

The basic idea is to reveal weak spots within our own strategy based on the progress of the game played against an opponent and to be able to adapt our strategy to it. In the end having the strategy with the possibility of gaining better results than with the original one. The proposed method can be used for a one-time adaptation from the logs of the played games or for a real-time adaptation during the game.

3.1. Game Information Extraction

The experiments described in Subsection 3.3 were created using our robot soccer architecture and 3D robot soccer simulator (see Figure 3). The game was played between the two teams each consisting of five robots. The standard game lasts for 2 minutes and simulator logs the current state on the game field every 20ms. The final log file consists of 6,000 records containing strategy rule selected for each team, grid coordinates of every robot and the ball, score and game time.

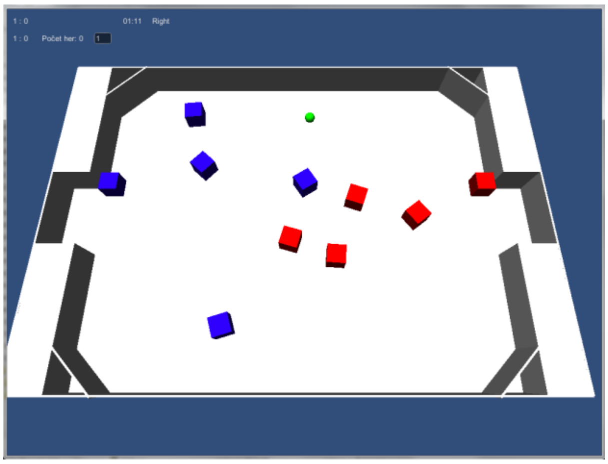


Figure 3: 3D Robot Soccer Simulator

Therefore, we are able to parse a log file for information relevant to a game situation (opponent’s scored goal, etc.) and to extract the strategy rules and grid coordinates of robots preceding this game situation.
The rule selection is invoked every 20ms during the game but during those 20ms the robots do not have enough time to move too far on the game field. Keeping that in mind, it is necessary to extract a wide enough time interval from the log preceding the detected relevant situation. One robot soccer game lasts for 2 minutes therefore it is sufficient to analyze the past 3 seconds which are represented by 150 log entries that will be used for the strategy adaptation. This should be long enough time to actually react to a current situation on the game field.

3.2. Strategy Adaptation

Game information extracted from the log is used to create anti-rules. There are 150 log entries for each detected game situation (strategy weakness) that need to be transformed to new rules that should prevent this kind of situation in the future.
Log entries are therefore divided by the offset of 50 entries (1 second of a game) and aggregated to a single rule resulting in 3 new rules for a given game situation.
The aggregation is performed by averaging the grid coordinates of the selected 50 log entries by all the robots and the ball therefore creating one aggregated rule describing 1 second of the given game situation.
In the final phase of the rule adaptation, the modification of the Move coordinates within the rule is applied. These coordinates are overwritten for the two robots to move to a position of the ball. By this way, we are able to create a new anti-rule covering the weak spot of our strategy (missing game situation within the strategy) that is trying to prevent this kind of situation in the future.

Overall process of the rule extraction and strategy adaptation is as follows:

1. Split the 150 log entries by 50 entries for every relevant game situation.
2. Compute average rule from the 50 log entries.
3. Change the grid Move coordinates for 2 of the robots to move towards the ball.
4. If the created anti-rule is not already in the strategy, insert it as a new rule.

This process was used not only for the detection of relevant game situations in the defense but also for the situations occurring during the offensive play. In terms of the defense play, opponent’s scored goal is considered as a relevant game situation. This action can be detected from the log file of the played game or during the current game.
The way to improve our team’s offensive play is to detect a game situation where our team was attacking but was unable to score the goal; therefore the ball was in front of the opponent’s gate but the play did not ended with the scored goal and the ball was afterwards kicked off to some other part of the game field. The idea for the offensive game adaptation is to find these unsuccessful offensive situations and to change the destination coordinates of selected robots to move closer to the opponent’s gate to put more pressure on the goalkeeper.
The overall process of the strategy adaptation can be used in post-game analysis as a way to train the current strategy on a number of already played games or as a fully automatic process that is performed during the ongoing game. The static adaptation method has an advantage in a potentially big initial training set whereas the automatic adaptation process is much more dynamic and therefore able to quickly react to the opponent’s current behavior during the game. Both of these approaches were tested and the results summarized in the following chapter.

3.3. Experiments and Results

Experiments were performed using our robot soccer architecture and within the 3D robot soccer simulator. The game was played between the left team strategy showed in Table 1 representing the strategy with the weak spots (strategy does not contain rules covering the left wing defense and center defense and is mostly focused on offense) and reference strategy of the right team (complete strategy containing basic offense and defense game situations).

Table 1: Left Team Strategy

Substrategy	#	Coordinate	Rule Desc.
Offensive Middle	1	Mine	4,2 4,3 2,2 2,3
		Oppnt	4,2 4,3 5,1 5,4
		Ball	4,2
		Move	5,2 5,3 2,2 3,3
	2	Mine	5,2 5,3 2,2 3,3
		Oppnt	5,2 5,3 5,1 5,4
		Ball	5,2
		Move	6,2 5,3 2,2 3,3
	3	Mine	6,2 5,3 2,2 3,3
		Oppnt	5,2 5,3 5,1 5,4
		Ball	6,2
		Move	6,2 5,3 2,2 3,3
Offensive Left	4	Mine	3,2 3,3 3,1 2,3
		Oppnt	4,1 4,2 5,1 5,3
		Ball	3,1
		Move	4,2 3,2 4,1 3,3
	5	Mine	4,2 3,2 4,1 3,3
		Oppnt	4,1 4,2 5,1 5,3
		Ball	4,1
		Move	5,2 4,2 5,1 3,3
	6	Mine	5,2 4,2 5,1 3,3
		Oppnt	4,1 5,2 5,1 5,3
		Ball	5,1
		Move	5,2 4,3 5,1 3,2
	7	Mine	5,2 4,3 5,1 3,2
		Oppnt	4,1 5,2 5,2 5,3
		Ball	5,2
		Move	6,2 5,3 5,1 3,2
	8	Mine	6,2 5,3 5,1 3,2
		Oppnt	4,2 5,2 6,2 5,3
		Ball	6,2
		Move	6,2 5,3 5,1 3,2
Defensive Right	9	Mine	3,1 3,3 2,1 2,2
		Oppnt	3,1 3,2 5,2 4,3
		Ball	3,1
		Move	2,1 2,3 1,1 2,2
	10	Mine	2,1 2,3 1,1 2,2
		Oppnt	2,1 2,2 4,2 3,3
		Ball	2,1
		Move	2,2 1,3 1,2 2,3
	11	Mine	2,2 1,3 1,2 2,3
		Oppnt	1,2 2,2 4,2 2,3
		Ball	1,2
		Move	2,2 1,3 1,2 2,3

Ten robot soccer games were played between the original left team strategy and the referenced one for the right team. Result scores from these games are visible in Table 2. Resulting score sum is 5:9 in favor of the right team strategy. Left team was able to win 3 times, lost in 5 games and achieved 2 draws. The lack of good defense in the strategy was obvious during the games as most of the opponent's score goals were scored via left side of the game field.

Table 2: Score Table – Original vs. Reference

Game #	Left Score	Right Score	Result
1	0	1	Loss
2	1	0	Win
3	0	2	Loss
4	0	2	Loss
5	1	0	Win
6	1	1	Draw
7	1	0	Win
8	0	1	Loss
9	1	1	Draw
10	0	1	Loss
SUM	**5**	**9**	

Adaptation from the log files

First experiments were performed using the static strategy adaptation method. Log files generated from the games of a Left team strategy and the Right team reference strategy were used for the static strategy adaptation mechanism, described in Subsection 3.2. Each scored goal for the opponent's team was detected as relevant game situation and was used to create 3 anti-rules resulting in sum of 27 new and different rules (see Table 3).

These rules were included to the original strategy for the left team and played once again against the reference strategy. The results are shown in Table 4. The adapted strategy scored the same amount of goals as the reference strategy with two wins, one loss and seven draws. The adapted strategy is now able to compete with the reference one thanks to the new rules that are able to react to the previously unsuccessful game situations.

Table 3: Adapted Defense Rules

#	Mine	Oppnt	Ball	Move
1	32 51 62 33	62 13 52 62	13	32 51 13 13
2	32 51 62 43	62 13 52 62	13	32 51 13 13
3	32 51 62 53	62 13 52 62	13	32 51 13 13
4	21 32 52 33	52 32 42 53	21	21 32 21 21
5	21 42 52 33	52 32 42 43	11	21 42 11 11
6	21 42 52 33	52 22 42 43	12	21 42 12 12
7	51 53 32 52	63 62 52 44	44	51 53 44 44
8	41 53 33 62	63 62 52 33	33	41 53 33 33
9	42 53 23 62	63 62 52 23	13	42 53 13 13
10	32 62 51 24	33 14 22 53	14	32 62 14 14
11	32 62 51 14	33 13 22 53	13	32 62 13 13
12	32 62 51 24	33 13 22 53	13	32 62 13 13
13	22 33 32 24	52 43 42 53	32	22 33 32 32
14	22 33 32 34	52 32 32 43	23	22 33 23 23
15	22 33 33 43	53 32 32 33	13	22 33 13 13

16	22 33 52 43	52 23 22 43	23	22 33 23 23
17	22 23 52 43	52 33 22 53	23	22 23 23 23
18	22 23 52 43	52 22 32 53	12	22 23 12 12
19	22 44 62 33	62 53 52 54	34	22 44 34 34
20	22 44 62 34	62 53 52 44	23	22 44 23 23
21	22 54 62 23	62 53 52 44	13	22 54 13 13
22	22 53 33 52	62 53 52 63	52	22 53 52 52
23	22 52 33 52	62 53 42 63	42	22 52 42 42
24	22 53 33 62	62 53 42 63	23	22 53 23 23
25	62 32 53 24	24 52 62 53	14	62 32 14 14
25	62 22 53 14	13 52 62 53	13	62 22 13 13
26	62 22 53 23	13 52 62 43	13	62 22 13 13
27	32 51 62 33	62 13 52 62	13	32 51 13 13

Table 4: Score Table – Adapted Defense vs. Reference

Game #	Left Score	Right Score	Result
1	0	0	Draw
2	0	0	Draw
3	0	2	Loss
4	0	0	Draw
5	0	0	Draw
6	0	0	Draw
7	1	0	Win
8	1	0	Win
9	1	1	Draw
10	1	1	Draw
SUM	**4**	**4**	

Using the same training set of 10 previously generated log files, the strategy was further updated using the mechanism described in Subsection 3.2 with the intent to improve the offensive play of the Left team strategy. During the current robot soccer game, there is a number of unsuccessful goal attempts on the side of each competing team therefore also bigger number of detected relevant situations can be expected resulting in bigger number of generated offensive rules.

Proposed adapting mechanism generated 138 new and unique rules and because 3 rules are created for each detected relevant situation (3 seconds of game play preceding this situation) the number of detected situations can be roughly estimated to 46. This number is probably higher as the same or very similar rules are considered as duplicates and therefore they are discarded from the strategy addition. The new rules were added to the Left team strategy and again played against the reference one from the right team. The results are shown in Table 5.

Table 5: Score Table – Adapted Defense + Offense vs. Reference

Game #	Left Score	Right Score	Result
1	1	0	Win
2	2	0	Win
3	1	1	Draw
4	3	1	Win
5	0	0	Draw
6	2	0	Win
7	1	0	Win
8	1	1	Draw
9	0	1	Loss
10	2	0	Win
SUM	**13**	**4**	

The results show that the Left team strategy was able to win the match over the opponent's strategy in 6 cases, achieved 3 draws and lost only in one played game. The difference in the resulting strategy is most visible on the number of scored goals. The original Left team strategy lost the game over an opponent by a higher number of scored goals. The same strategy adapted by 27 defensive rules was able to achive more balanced gameplay but the strategy consisting of both deffense and offense adaptation rules, comprising of 176 rules, was able to score 9 more goals than the opponent's strategy over the course of 10 matches.

Adaptation during the game

The same mechanism of strategy adaptation from the defensive and also offensive point of view was used directly in our robot soccer simulator (see Figure 3) during the actual game between the original Left team strategy (Table 1) and the reference strategy used for the right team.

The whole adaption process was implemented as a part of our robot soccer library. This library is an integral part of our 3D robot soccer simulator and contains all functionality necessary for robot control, rule selection from strategy, path planning, tactics execution and much more. The pseudocode for the main part of the process is shown in Algorithm 1.

```
//called every game step - 20ms
AdaptStrategy(LogHolder log, Strategy str)
{
    //get defense rules
    List<Rules> defRules = AdaptDefense(log, str);
    //get offense rules
    List<Rules> offRules = AdaptOfense(log, str);

    foreach defRules
    {
        //if the rule is unique, insert into strategy
        if(SimilarRuleCheck(defRules, str))
        {
            AddRulesToStrategy(defRules);
        }
    }
```

```
    foreach offRules
    {
        //if the rule is unique, insert into strategy
        if(SimilarRuleCheck(offRules, str))
        {
            AddRulesToStrategy(offRules);
        }
    }
}
AdaptDefense(LogHolder log, Strategy str)
{
    //check current game situation
    LogInfo lastEntry = GetLastLogEntry(log);

    //relevant situation detection, did opponent score a goal?
    if(GoalCheck(lastEntry))
    {
        //extract previous 3s of gameplay, 150 log entries
        ExtractPreviousEntries();

        //generate new rule, move selected players closer to
        //left team gate
        foreach 50 entries
        {
            List<Rules> defRules =
                GenerateDefensiveRule();
        }
    }
    return defRules;
}
AdaptOffense(LogHolder log, Strategy str)
{
    //check current game situation
    LogInfo lastEntry = GetLastLogEntry(log);

    //relevant situation detection, was the ball before the
    //opponent's gate but the goal was not scored?
    if(UnsuccessfulGoalCheck(lastEntry))
    {
        //extract previous 3s of gameplay, 150 log entries
        ExtractPreviousEntries();

        //generate new rule, move selected players closer to
        //right team gate
        foreach 50 entries
        {
            List<Rules> offRules =
                GenerateDefensiveRule();
        }
    }
    return offRules;
}
```

Algorithm 1: Real-time Adaptation Pseudocode

The offensive adaptation mechanism is able to generate a large number of rules therefore it was necessary to use the effective algorithm for rule comparison part of the implementation that is checking whether the generated rule (or a very similar one) is already part of the current strategy. This effective rule comparison algorithm was already developed as a part of a rule selection mechanism described in Subsection 2.1 and thus simply put to use also in a proposed strategy adaptation process. Each generated rule is simply compared to every rule from the current strategy based on their Z-order coordinates. This implementation uses space filling curves to order the robots on the game field thus saving time during rule selection and rule generation in the strategy evaluation and adaptation phase of the robot soccer game.

The real-time adaptation experiments were performed for 5 sets of played games. Each set consists of 10 matches between the original Left team strategy (see Table 1) and the right team reference strategy. During this matches the left team was able to perform automatic defensive and offensive strategy adaptation therefore able to detect the relevant game situations and to add newly generated rules to its own strategy set. The adapted strategy remained stored over the course of these 10 games thus able to improve in each game iteration.

The results from these game sets are shown in Tables 6 to 10. Overall results represented by a number of scored goals show that the adapted Left team strategy is a balanced opponent (3 wins and 2 loses) for the reference strategy and is able to score a sufficient number of goals against it. The tables also contain the number of automatically added defensive and offensive rules for each game iteration.

Table 6: Score Table – Real-Time Adaptation, Set 1

Game #	Left Score	Right Score	Defense Adapt Rules	Offense Adapt Rules
1	0	0	0	10
2	1	2	6	12
3	0	1	8	20
4	0	1	11	24
5	2	0	11	27
6	4	0	11	31
7	0	1	13	32
8	2	2	17	37
9	0	2	20	40
10	1	0	20	41
SUM	**10**	**9**	**20**	**41**

Table 7: Score Table – Real-Time Adaptation, Set 2

Game #	Left Score	Right Score	Defense Adapt Rules	Offense Adapt Rules
1	0	3	5	3
2	0	1	5	12
3	1	1	5	19
4	2	1	5	20
5	0	1	5	23
6	1	0	5	25
7	1	0	5	34
8	0	1	5	39
9	0	0	6	41
10	1	0	6	42
SUM	**6**	**8**	**6**	**42**

Table 8: Score Table – Real-Time Adaptation, Set 3

Game #	Left Score	Right Score	Defense Adapt Rules	Offense Adapt Rules
1	1	2	3	8
2	1	0	3	11
3	4	1	6	12
4	2	1	8	12
5	1	3	14	18
6	0	1	14	28
7	0	1	16	31
8	1	0	16	32
9	1	0	16	40
10	1	0	16	44
SUM	**12**	**9**	**16**	**44**

Table 9: Score Table – Real-Time Adaptation, Set 4

Game #	Left Score	Right Score	Defense Adapt Rules	Offense Adapt Rules
1	2	2	1	11
2	0	1	3	14
3	1	2	4	17
4	2	0	4	22
5	1	1	6	22
6	0	0	6	24
7	0	2	8	30
8	2	1	10	35
9	0	0	10	35
10	0	1	12	35
SUM	**8**	**10**	**12**	**35**

Table 10: Score Table – Real-Time Adaptation, Set 5

Game #	Left Score	Right Score	Defense Adapt Rules	Offense Adapt Rules
1	1	1	3	6
2	1	0	3	13
3	0	0	3	19
4	0	1	4	27
5	0	1	7	30
6	1	0	7	32
7	0	1	10	38
8	0	0	10	40
9	2	0	10	42
10	1	1	11	44
SUM	**6**	**5**	**11**	**44**

Following the proposed adaptation mechanism, the results show that the new defense adaptation rules are added only in the case of opponent's goal thus enabling the left team to learn the opponent's attack pattern and devise a rules to block it.
The largest number of rules is also added in the first game iterations but that is understandable because the smaller the rule set of a strategy the smaller chance that the newly generated rule will be the same or similar to some other rule from the strategy. As the game progresses the number of rules in the strategy also grows, more generated rules are discarded because of it during the game.

3.4. Validation and Visualization

The proposed approach could be validated using the sequence extraction and comparison which is also very useful for the visualization. This method proceeds from the original social network approach with the adaptation to robot soccer games. The sequences or game profiles can be extracted from the log that contains data related to rules selected from the game strategy during the game.

The definition of a game profile is as follows:

Let $U = \{u_1, u_2, \ldots u_n)$, be a set of games, where n is a number of games u_i. Then, sequences of strategy rules $\sigma_{ij} = <e_{ij1}, e_{ij2}, \ldots e_{ijm_j}>$, are sequences of strategy rules executed during a game u_i in the simulator, where $j = 1, 2, \ldots, p_i$ is a number of that sequences, and m_j is a length of j-th sequence. Thus, a set $S_i = \{\sigma_{i1}, \sigma_{i2}, \ldots \sigma_{ip_i}\}$ is a set of all sequences executed during a game u_i in the system, and p_i is a number of that sequences.
Sequences σ_{ij} extracted with relation to certain game u_i are mapped to set of sequences $o_l \in S$ without this relation to games: $\sigma_{ij} = <e_{ij1}, e_{ij2}, \ldots, e_{ijm_j}> \rightarrow \sigma_l = <e_1, e_2, \ldots, e_{m_l}>$, where $e_{ij1} = e_1$, $e_{ij2} = e_2$, ..., $e_{ijm_j} = e_{m_l}$.
Define matrix $B \in N^{|U| x |S|}$ where

$$B_{ij} = \left\{ \begin{array}{r} \text{frequency of sequence } \sigma_j \in S \text{ for game } u_i \text{ if} \\ \sigma_j \in S_i, \\ \text{else } 0 \end{array} \right\}$$

A base game profile of the games $u_i \in U$ is a vector $b_i \in N^{|S|}$ represented by row i from matrix B.

Each sequence is labeled with a sequence number and number determining the possession of the ball (0 - none, 1 - left team, 2 - opponent's team). Each sequence contains a list of rules selected for the left team in every game step until the team possession of the ball has changed. These sequences are then compared using the method for sequence comparison – LCS (longest common substring), LCSS (longest common subsequence) and T-WLCS (time-warped longest common subsequence). Results from the sequence comparison can be visualized by clustering of similar sequences.
Visualization of LCSS method for the original Left team strategy and the strategy adapted by defense rules from the log files (Table 3) can be seen in Figures 4 and 5. The clusters of similar sequences often represent the game situations defined within the original strategy. Therefore, the original strategy should contain the original game situations and the adapted strategy should also contain new clusters consisting of newly created anti-rules.
In most cases, the strategies are created manually with some specific game situations in mind. For example, the first X number of rules representing the left wing defense, next Y number of rules representing right wing

offense etc. Therefore, the above-described sequence extraction method and subsequent sequence comparison and visualization could be also used for the strategy validation and visualization of the progress of the robot soccer game in relation to strategy adaptation.

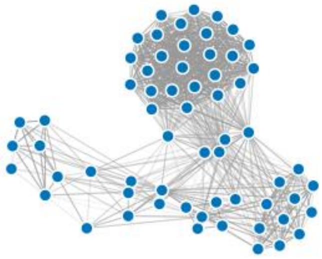


Figure 4: Clusters of Similar Sequences – Original Left Team Strategy LCSS

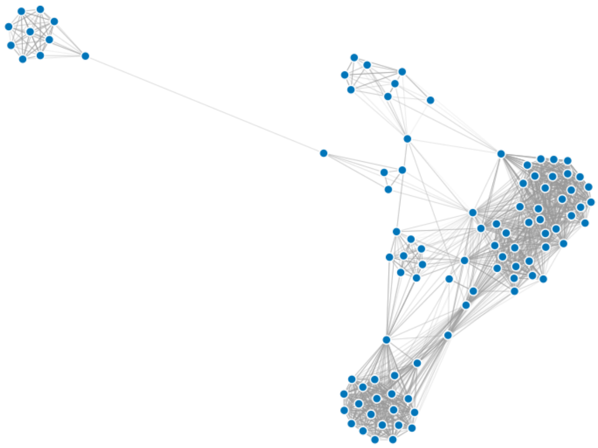


Figure 5: Clusters of Similar Sequences – Adapted Strategy LCSS

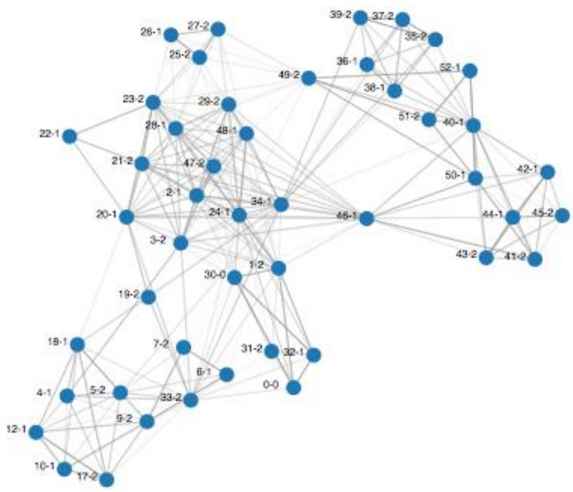


Figure 6: Clusters of Similar Sequences – Set 1, #1 LCSS

At closer examination of the extracted sequences from the played games of Set 1 (Table 6), the correlation between clusters and the rules can be seen. The clusters of similar sequences for selected iteration games in Set 1 are shown in Figure 6 to 10.

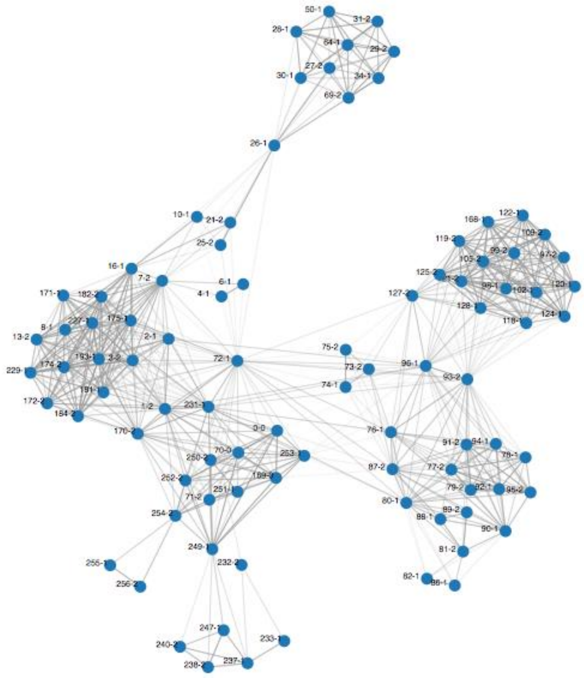


Figure 7: Clusters of Similar Sequences – Set 1, #2 LCSS

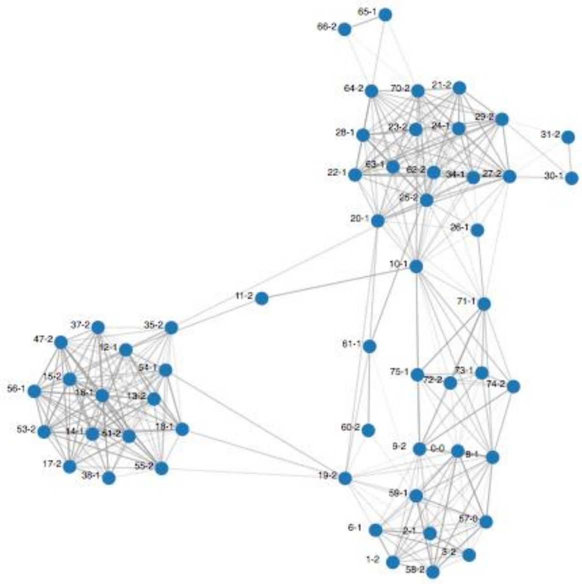


Figure 8: Clusters of Similar Sequences – Set 1, #4 LCSS

With the increasing number of game iterations, there is also increasing number of rules in the strategy. These rules are being selected during the game thus being part of a newly created sequences that represent specific game situations that occurred during the game.

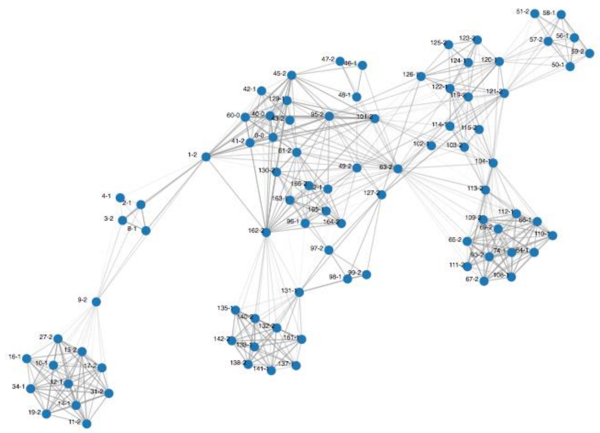


Figure 9: Clusters of Similar Sequences – Set 1, #6 LCSS

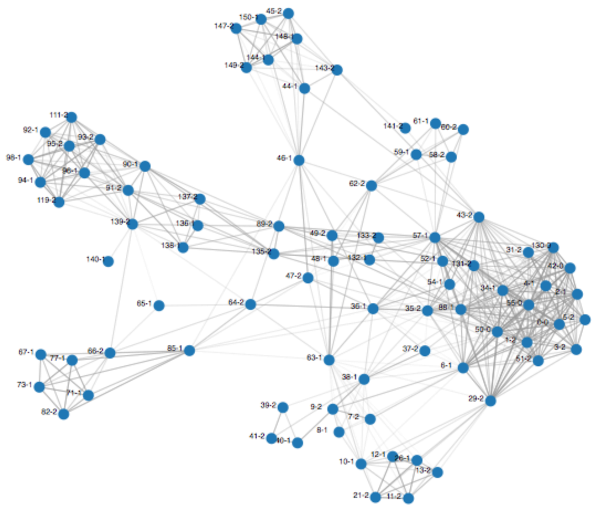


Figure 10: Clusters of Similar Sequences – Set 1, #8 LCSS

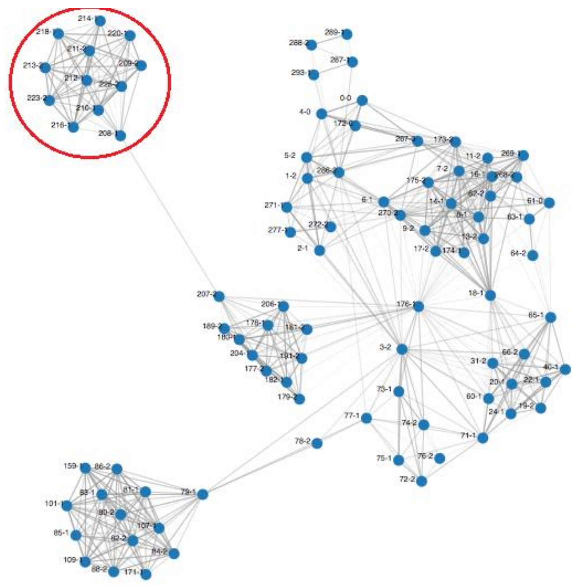


Figure 11: Clusters of Similar Sequences – Set 1, #10 LCSS

Each node represents a sequence comprised of selected strategy rules. Each sequence is labeled with a sequence number and with a number determining the possession of the ball. Therefore, each sequence number could be mapped to a sequence of strategy rules that were performed during some game situation. The selected sequence cluster from the Figure 11 is described in Table 11.

Table 11: Set 1, #10, Selected Sequence Cluster

Sequence #	Rules
208 - 1	34,34,34,34,34,34,34,34,34,34,34,34,34, 34,34,34,34,34,34,34,34,34,34,34,34,34, 34,34,34,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8
209 - 2	8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8
210 - 1	34,34,34,34,34,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8
211 - 2	8,8,8,8
212 – 1	8,8
213 - 2	8,8,8,8,8,8,8,8
214 - 1	8
216 - 1	8,8,8,8,8
218 - 1	8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8
220 - 1	8,8,8,8,8,8,8,8,34,34,34,34,34,8,8,8,8,8, 8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8, 8,8,8,8,8,8,8,8,8
223 - 2	8,8,8,8,8,8
226 - 2	8,8,8

Table 12: Selected Sequence Cluster - Rules

#	Mine	Oppnt	Ball	Move
8	62 53 51 32	42 52 62 53	62	62 53 51 32
34	51 32 62 64	62 63 52 53	54	51 32 62 63

This cluster consists of 12 sequences containing rules number 8 and 34 (see Table 12). The rule number 8 is the rule originating from the original Left team strategy and the rule number 34 is the offensive rule that was created during the strategy adaptation. Our robot soccer architecture does not require for each robot to be uniquely identifiable during the game. Therefore the robots' roles are mutually interchangeable. In relation to this, it is apparent the similarity of the rule 8 and 34. Further analysis of the extracted sequences will be the focus of our future work and the theme of the follow-up articles.

CONCLUSION AND FUTURE WORK

In this work, the strategies of the robot soccer game were discussed. The description of approach for the strategy adaptation was presented. The main part of the article discussed the static strategy adaptation using the log files of a previously played games and also the real-time strategy adaptation performed directly during the robot soccer game. The adaptation was performed for the defensive and also offensive part of the strategy.

The adaptation method performed after the game using the log files has an advantage in a potentially big initial training set whereas the real-time adaptation process is much more flexible and able to quickly react to the opponent's current behavior during the game. On the other hand, the real-time adaptation process might also need a sufficient number of iterations to achieve the desired results.

The future work will be mainly focused on improvements in the area of strategy adaptation and evaluation, and of course on the overall improvements of the developed robot soccer game architecture.

ACKNOWLEDGMENTS

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project „IT4Innovations excellence in science - LQ1602".

REFERENCES

Osborne M. J., 2004. An introduction to game theory. New York Oxford, Oxford University Press.

Kim J-H, Kim D-H, Kim Y-J, Seow K. T., 2010. Soccer Robotics, Springer Tracts in Advanced Robotics.

Ontanón S., Mishra K., Sugandh N. and Ram A., 2007. Case-Based Planning and Execution for Real-Time Strategy Games. Lecture Notes in Computer Science, Volume 4626, 164-178.

Martinovič J., Snášel V., Ochodková, Zoltá L., Wu J., Abraham A., 2010. Robot soccer - strategy description and game analysis. Modelling and Simulation, 24th European Conference ECMS.

Svatoň V., Martinovič J., Slaninová K. and Snášel V., 2014. Improving Rule Selection from Robot Soccer Strategy with Substrategies. In Computer Information Systems and Industrial Management - 13th IFIP TC8 International Conference (CISIM).

Huang HP, and Liang CC, 2002. Strategy-based decision making of a soccer robot system using a real-time self-organizing fuzzy decision tree. Fuzzy Sets Syst. 127, 1.

Nakashima T., Takatani M., Udo M., Ishibuchi H., and Nii M., 2006. Performance evaluation of an evolutionary method for robocup soccer strategies. In RoboCup 2005: Robot Soccer World Cup IX, Springer.

Larik A. S. and Haider S., 2016. On using evolutionary computation approach for strategy optimization in robot soccer. 2nd International Conference on Robotics and Artificial Intelligence (ICRAI)

Tominaga M., Takemura Y., and Ishii K., 2017. Strategy Analysis of RoboCup Soccer Teams Using Self-Organizing Map.

Shengbing Ch., Gang Lv, and Xiaofeng W., 2016. Offensive strategy in the 2D soccer simulation league using multi-group ant colony optimization. International Journal of Advanced Robotic Systems 13.

Akiyama H., Tsuji M., and Aramaki S., 2016. Learning Evaluation Function for Decision Making of Soccer Agents Using Learning to Rank. Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems, 2016 Joint 8th International Conference on. IEEE

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

BUILDING INFORMATION MODELLING AND SIMULATION INTEGRATION FOR MODULAR CONSTRUCTION MANUFACTURING PERFORMANCE IMPROVEMENT

Beda Barkokebas[(a)], Youyi Zhang[(b)], Chelsea Ritter[(c)], and Mohamed Al-Hussein[(d)]

[(a),(b),(c),(d)] Hole School of Construction Engineering, University of Alberta, Edmonton, Canada

[(a)]barkokeb@ualberta.ca, [(b)]youyi@ualberta.ca, [(c)]critter1@ualberta.ca, [(d)]malhussein@ualberta.ca

ABSTRACT

Building Information Modelling (BIM) technology and simulation modelling can be successfully integrated to support modular construction manufacturing (MCM). This integration can benefit the MCM project by improving estimation accuracy, production line scheduling, resource allocation, inventory control, and by providing informative shop drawings. This paper demonstrates a case study where BIM technology is implemented to maximize the effective usage of available data and structure information. The extracted information is then stored in a database and used as the input for a simulation model. The material quantity for a product can be generated automatically for different work stations on the production line and will benefit the inventory control. By assigning detailed task durations to the material in the BIM model, the new task duration given by the simulation model can allow for increased accountability, which provides direct improvement to labour resource allocation on the line.

Keywords: BIM, simulation, estimation, modular construction manufacturing

1. INTRODUCTION

Due to the high level of specialization required, construction projects involve multiple disciplines with a high level of interaction and exchange of information among the various stakeholders. For instance, accurate and detailed drawings, developed during the engineering design phase, not only provide support to the estimation department, but also enable an increased level of detail for production scheduling. In light of this, Building Information Modelling (BIM) technology offers significant benefits to the current construction industry due to its capacity to promote effective information exchange while also automating several steps in the design, estimation, and planning phases. During the engineering design phase of the project, accurate and detailed building information not only provides support to the estimation department, but also enables an increased level of detail for production scheduling. This function is especially valuable to modular construction manufacturing (MCM) due to its special requirements in design & drafting for manufacturing (Moghadam, Alwisy and Al-Hussein 2012). Despite its advantages, current MCM remains heavily dependent upon the conventional construction production operation, where there is little innovation regarding industrialization techniques. Accuracy of material and labour estimation is limited due to the amount of manual work required as well as the unpredictable nature of construction. AutoDesk Revit is a key medium for BIM implementation in the industry; however, promoting productivity and improving the construction process on the production line using information from the project model in Revit remains a challenge. Conversely, simulation is a tool that is often used for production line performance assessment, revealing operational barriers as well as offering guidance for improvement; however, regardless of the design of the simulation model, the quality of the input data for the model constrains the quality of data that the simulation can provide to decision makers in the factory. The level of detail and level of accuracy of the collected information using traditional construction methods is not sufficient for MCM.

This paper presents a case study where BIM technology and simulation are integrated and implemented to maximize the effective usage of available data from the model. The collected information, which was previously considered difficult to collect due to the labour-intensive operations in construction manufacturing, can be collected more accurately with the assistance of a BIM application. The simulation model, which uses this data, will provide a reliable production line assessment that will benefit MCM projects.

2. LITERATURE REVIEW

The use of BIM encompasses many benefits to construction processes, including increased productivity, reduced risk, increased sustainability through decreased waste, and increased collaboration, but it is typically used for large-scale projects as the potential for benefits increases as project value increases (Burt and Purver 2014). BIM can be employed in order to obtain and organize information about quantities, and has other benefits, including design validation and worksite cooperation (Ciribiniet, Mastrolembo Ventura and Paneroni 2016), as well as the additional ability to include production line scheduling and inventory control, which results in an increased benefit, possibly making BIM more feasible for smaller projects.

Liu, Al-Hussein and Lu (2015a) utilize outputs from BIM to create a detailed schedule for a panelized construction manufacturer by using BIM information with simulation to create an optimized construction schedule. This study illustrates some of the capabilities of integrated BIM and simulation, and demonstrates that there is an opportunity to bolster the benefits provided by the BIM model by integrating it with simulation and other tools.
Lu and Korman (2010) recognize the potential for BIM to be used in modular construction, and identify the impact of BIM implementation in a modular plant to include: visualization, modelling, code reviews, fabrication/shop drawings, communication, cost estimating, construction sequences, and conflict, interference, and collision detection. Of particular interest is construction sequence, for which it is stated that the BIM model will provide the construction schedule. While this is useful, there is potential for this use of BIM to be expanded further to represent the actual accomplishments of the company in terms of productivity and schedule.
Simulation is another commonly used tool in modular construction. Alvanchi et al. (2012) use discrete event simulation to model the offsite fabrication process for a bridge project, and then to test variations to the fabrication process to achieve a 10% reduction in fabrication time. In their study, the developed simulation model represents constraints, including crane time and construction space, to accurately reflect the conditions in the fabrication plant.
Velarde et al. (2009) use lean tools and simulation to improve production at a modular home construction plant, with the goal of assisting the company to become more competitive. They recognize the importance of simulation as a tool to test different alternatives prior to implementation to help predict the outcome of adapting a new practice. This use of simulation illustrates the opportunity for its use in predicting outcomes and production planning.
In all of the studied cases, a lack of interoperability among software applications is observed as a predominant limitation since a considerable amount of manual work is involved to transfer and prepare the information for analysis. The use of various applications is inherited in the construction industry as observed by Moghadam (2014), which states there is no single tool or application able to perform all the necessary tasks in a construction project. Thus, it can be concluded there is a need to establish a combination of various applications to improve the performance of the design and planning phases of construction projects which can be adapted on a per-project basis. Liu et al. (2015b) integrate BIM and simulation by extracting information from a BIM model and using it as entity properties in the simulation model. By doing this they are able to run a user-defined sequence of panels through a simulation for the pre-existing production line to receive an output representing the predicted process time; however, even in this case, there is significant manual input required to update the simulation model with up to date production times.
This paper proposes applications and workflows to reduce manual work thus allowing more time for analysis in order to improve the informed decision-making process.

3. METHODOLOGY

In order to successfully utilize both BIM and simulation for modular construction production line performance improvement, the following steps, which are described in Figure 1, are critical.

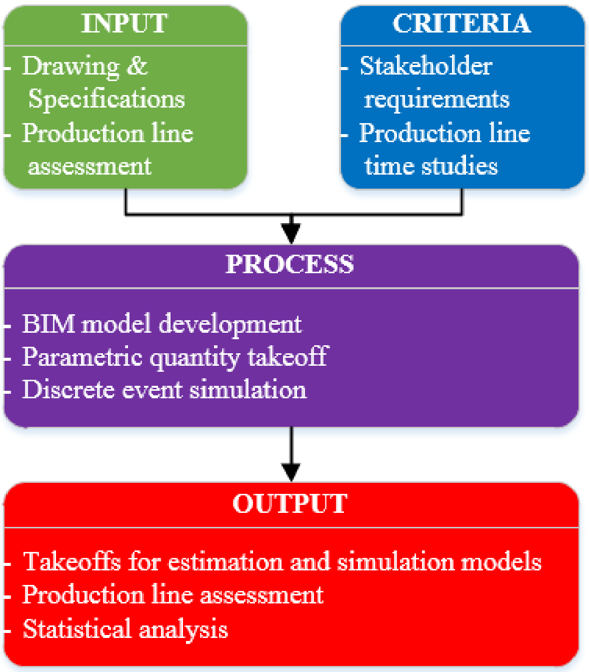


Figure 1: Proposed methodology

First, understanding the current production line operation process and the stakeholder requirements is the key to creating the connection between the drafting/design team, who uses Revit for BIM, and the operation team, who could benefit from simulation analysis. The raw data for either the Revit model or the simulation model needs to be collected during the site observation, but this can be time consuming, and the accuracy is often low. Therefore, with regard to the objectives of data collection, planning the collection method prior to the execution is required.
Second, it is important to know the level of detail required (through the production line time studies) to put into the BIM model and understand the objective for the production line assessment. For example, BIM models in conventional construction are not required to provide detailed information, such as the number of doors and windows per wall panel, due to the lack of simulation models to forecast the uncontrollable and unstandardized work site environment. For modular construction assessment and process improvement, this detailed information could promote a standardized

working process and make the time duration for each work station more predictable.
The third step is to build the BIM model with the targeted level of detail and generate the database for the simulation model. In addition to providing valuable information for the simulation model, the accuracy of quantity take-off for estimation increases by means of hidden material being revealed. For example, the backings for cabinet installation can be built into the BIM model, whereas previously, these details would not have been included in the drawings at the case study factory.
The next step is to connect the BIM model with the simulation model based on the current production line process and stakeholder requirements. By connecting the database generated from the BIM model to the simulation model, information will be provided as per the stakeholder requirements previously stated.
Finally, a statistical analysis is presented in order to provide more information about the production line and be utilized in decision making, such as inventory control, for each work station on the production line. The work duration at each work station can also assist in the allocation of labour resources on the line.

4. CASE STUDY

The case study factory is a modular home manufacturer whose current production line process is not only manually intensive, but also receives little support from the drafting team to reach a more efficient operation process on the line. The shop drawings appear to be the same type of drawing that is provided to workers on any conventional construction work site. The company is seeking a breakthrough to enhance their production line performance using machines to assist in labour-intensive activities, including a more supportive engineering team and better material and labour resource allocation on the line to increase their productivity.

4.1. Analysis of layout

Due to the scope of this paper, the framework will be applied to the wall framing stations since they comprise the central bottleneck and dictate the overall productivity of the production line. The use of machines is thus proposed to supplement labour-intensive manual framing at these stations. Since the initial investment for the purchase of such machinery is substantial, not to mention the adjustment in company culture, a simulation model is proposed to evaluate the feasibility of such an investment. The information extracted from the BIM model for machine-assisted wall framing will have a greater level of detail than that of traditional manual framing. For example, at the most detailed level, (1) the number of studs in each of the wall panels, and (2) the material used at each work station can be counted from the model. At a lesser level of detail, generally required for traditional manual framing, the total linear length of wall can be calculated instead of using the number of studs to represent the wall panel. At each level of detail, the simulation input database will be different. The algorithms in the simulation will be adjusted accordingly as well.
The primary output from the simulation model is expected to be the total duration for the project and total number of hours required to complete a typical project from the assessed company. This will assist decision makers in their assessment whether to invest in machinery for their production line while enhancing their current design and planning process. The simulated layout is presented in Figure 2 below and includes a total of three workers and one machine.

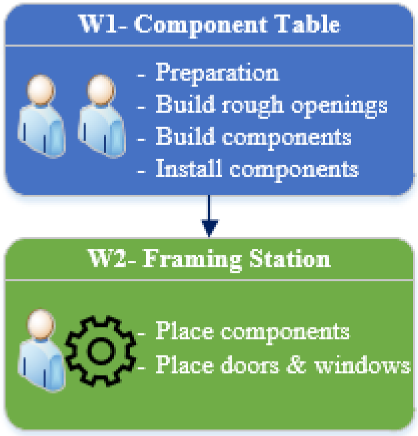


Figure 2: Simulated layout

The feasibility of the investment will be evaluated based on the simulated productivity increase promoted by the use of machinery in the future state. In order to predict the future state of the production line, a time study from another production line using similar machines and production system is used as a benchmark to forecast the system's productivity. Table 1 demonstrates data provided for this study according to each station. While some activities are dependent upon quantities extracted from wall components in the design, the fixed values are presented according to the number of panels. Possible delays are also presented according to the distributions acquired during the time study process.

Table 1: Information collection from performed time studies and input function for simulation model

Time study results		
Station	Task	Variable (minutes)
W1	Preparation	1.00
W1	Build rough openings (ext. walls)	4.75 × No. of windows + 5.60 × No. of doors + 8.00 × Fireplace
W1	Build rough openings (int. walls)	4.75 × No. of windows + 4.00 × No. of doors
W1	Build U-studs	1.00
W1	Build L-studs and triple studs	1.00
W1	Install bracing at openings	0.60 × No. of windows + 0.60 × No. of doors + 0.60 × Fireplace
W2	Place top-bottom plates	1.40
W2	Place singles studs	0.17
W2	Place double/L-studs	0.33
W2	Place multi-studs	0.50
W2	Place large windows (width >1,378mm)	1.33
W2	Place small/regular window (width <= 1,076 mm)	1.17
W2	Place large doors (width >1,134 mm)	1.33
W2	Place regular doors (width <= 1,134 mm)	0.75
W2	Place component	0.58
W2	Delay: Error correction	Triangular (0.08, 4.00, 0.75)
W2	Delay: Machine breakdown	Triangular (1.5, 10.00, 2.2)
W2	Delay: Material supply	Triangular (0.08, 3.75, 0.40)
W2	Delay: Worker away	Triangular (0.08, 1.16, 0.25)
W2	Delay: Other	Triangular (0.2, 1.5, 0.5)

4.2. Simulation model

The sequence of activities is presented in Figure 3 and developed using Simphony.NET to simulate activities observed during the time studies, thus reflecting its inter-dependency and values presented in Table 1. Since most activities involve a simple finish-to-start relationship between them, a discrete event simulation is chosen as the best approach for this problem due to its simplicity to be modelled in Simphony.NET. Delay activities at Station W2, the last station modelled, are modelled at the end since its order has no impact on the results of the model given its objective is to reflect the total manufacturing time of the project. As observed during the time study, each panel is added to the production line and progresses through the simulated stations and tasks. Some tasks are related to the type of panel (internal or external) and are differentiated by its time and sequence, while fixed values are deferred as per each panel regardless of its properties (e.g., preparation and the placement of top-bottom plates).

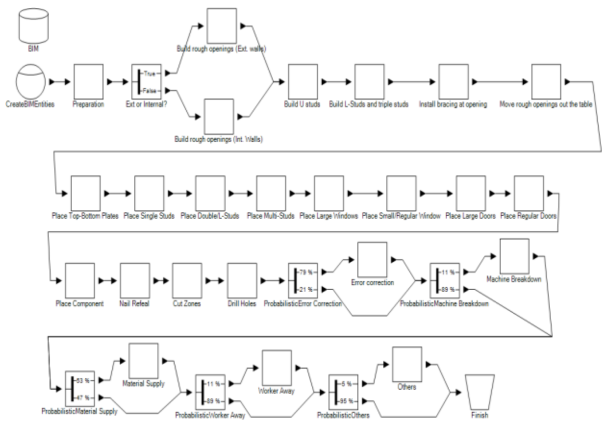


Figure 3: Simulation model

Each panel represents an entity and carries a set of properties imported from the BIM model automatically, thus establishing an interactive inter-dependence between the development of a product and assessment regarding its constructability. The imported data from the BIM model consists of design-related takeoffs addressed by consultants during the product development of the project and corresponds to the information contained in Table 1.

Such information is acquired simply by counting the elements in the model (e.g., number of windows and doors), while others are more time consuming as the number of elements increases such as the number of blockings per panel. Other panel properties must comply with certain logical relationships such as the geometry of each element (large windows > 1,378 mm) or the function of wood elements in the project (e.g., regular studs, corner and intersection components). Although the reasoning is simple, its implementation results in a significant amount of work as demonstrated in Figure 4 below.

Figure 4 describes the work required to transfer the information contained in one panel of the project. Framing components are modelled in Revit, with a total quantity of 46 on the sample panel, and it is sampled according to its use (e.g., single studs and blockings with a quantity of 10 and 28, respectively) by the research engine as properties of the panel entity to be used as inputs to the simulation model in one or more tasks. As projects have a significant number of panels with a large number of framing elements each, the activity to organize and sort the information contained in the BIM model becomes time consuming and error-prone, thus jeopardizing the practicality to simulate the production line. The BIM model and its connection with the simulation model are presented in the next section.

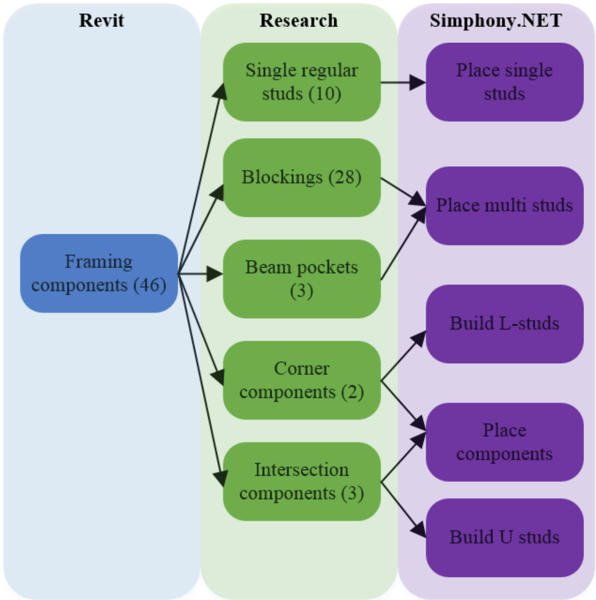


Figure 4: Information transfer required among BIM and simulation models

4.3. BIM model and connection with simulation

The case study project analyzed in this paper is presented in Figure 5 and consists of a modular single-family home with 25 wood-frame panels, one storey, and a total area of 71.64 m^2. The BIM model is developed in the Autodesk Revit environment and—due to the wood frame information required by the simulation model—the framing portion is developed by Framex, a Revit add-on developed by the University of Alberta.

Figure 5: Single-family case study module

Framex defines each framing panel based on existing walls addressed by architects in order to define the desired layout and models all framing components, such as regular studs, opening components, and blockings required for electrical outlets and kitchen cabinets, thus creating the information required to address the developed simulation model. Each panel is then considered an entity for the simulation model and carries distinct properties in order to calculate the total manufacturing time in the simulation model.

Figure 6 presents the workflow proposed in this paper. By using Excel, a software application common to all stakeholders during the development and manufacturing phases, the proposed research engine creates a two-way connection with the BIM model and prepares the information in an adequate format to be imported into Simphony.NET. In Excel, the user defines which parameters are to be extracted from elements (e.g., walls, framing elements, doors, etc.). This file is then imported into Revit generating the automated raw takeoffs, which are sorted as per the time study and formatted into Simphony.NET format through pre-set formulas in excel, comprehensible to all stakeholders for checking. The takeoffs can also be used for estimating, procurement, and inventory forecast according to stakeholder needs.

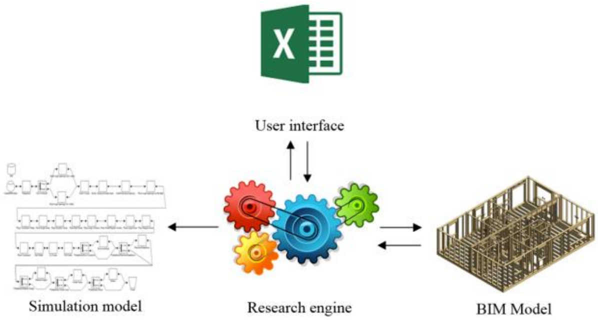


Figure 6: Proposed workflow

The research engine consists of an add-on programmed using Dynamo, an open source graphical programming tool. The development of this add-on is divided into three parts: (1) a stable connection between a pre-set user interface (in Excel) and the BIM model; (2) the import of the Excel file while gathering all required information; and (3) sorting of all gathered data as per the simulation model requirements.

4.4. Simulation results

Using Simphony.NET, the model is simulated 1,000 times, and its data is exported to EasyFit in order to determine an adequate fit for the data. In EasyFit, a histogram with the simulation results is produced while assigning the Log-Logistic distribution as the most adequate distribution as per Figure 7 below. As indicated in Figure 7, there is a variance in results of 8.30 minutes resulting in an approximate 11 % variation over the mean (74.84 minutes).

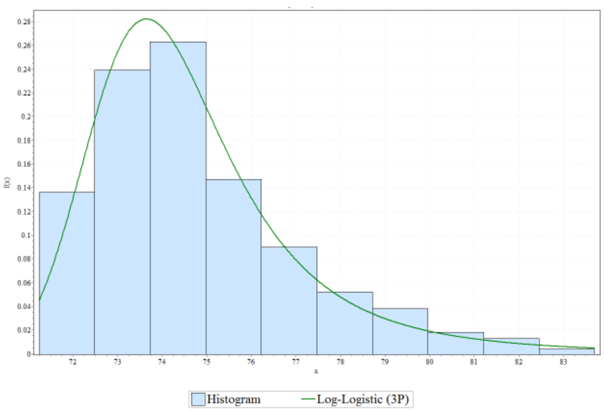


Figure 7: Histogram and fitted distribution on generated data

Moreover, the employment of each task is also addressed in the model. Figure 8 presents a summary of the most employed activities in the simulation model. In this figure, it can be observed that activities involving the manufacturing of components, such as openings, are employed the most on the production line (values of 32.4% and 45.30% for internal and external rough opening components, respectively). Also, the employment of preparation activities for each station are 28.90% and 40.50% for Stations W1 and W2, respectively.
For a better assessment of the production line as a whole it is recommended that the rest of the work stations be included in the model to generate a queue and better reflect a typical day at the manufacturing facility. These are the limitations of this work and will be addressed in future papers.

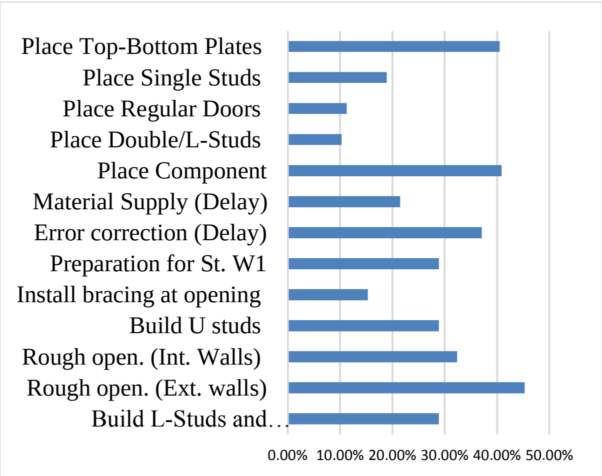


Figure 8: Summary of most employed activities during 1,000 simulation runs

4. CONCLUSION

The proposed methodology provides a solution to connect BIM technology with production line performance improvement using a simulation model. This research can also promote an industrialization transformation for modular manufacturing companies currently utilizing the traditional construction process. Innovated technology can improve the productivity by standardizing and automating the process, which requires support from BIM models.

ACKNOWLEDGEMENTS

We would like to acknowledge Dr. M. Sadiq Altaf for providing the time studies and other relevant information for the development of this work.

REFERENCES

Alvanchi A., Azimi R., Lee S., AbouRizk S., Zubick P., 2012. Offsite construction planning using discrete event simulation. Journal of Architectural Engineering, 18(2): 114-122.

Burt J., Purver K., 2014. Building information modelling for small-scale residential projects. Management, Procurement and Law, 167: 134-140.

Ciribini A., Mastrolembo Ventura S., Paneroni M., 2016. Implementation of an interoperable process to optimise design and construction phases of a residential building: A BIM Pilot Project. Automation in Construction, 71: 62-73.

Moghadam M., Alwisy A., Al-Hussein M., 2012. Integrated BIM/Lean Base Production Line Schedule Model for Modular Construction Manufacturing. Construction Research Congress, pp. 1271- 1280. May 21-23, West Lafayette (Indiana, USA)

Moghadam, M., 2014. Lean-Mod: An approach to modular construction manufacturing (MCM) production efficiency improvement. Thesis (PhD), University of Alberta.

Liu H., Al-Hussein M., Lu M., 2015a. BIM-based integrated approach for detailed construction scheduling under resource constraints. Automation in Construction, 53: 29-43.

Liu H., Atlaf M. S., Lei Z., Lu M., Al-Hussein, M., 2015b. Automated Production Planning in Panelized Construction Enabled by Integrating Discrete-Event Simulation and BIM. Proceedings of the 11th Construction Specialty Conference, pp 48:1-48:10. June 8-10, Vancouver (British Columbia, CA)

Lu N., Korman T., 2010. Implementation of Building Information Modeling (BIM) in modular construction: Benefits and Challenges. Proceedings of the Construction Research Congress, pp 1136-1145. May 8-10, Banff (Alberta, CA)

Velarde G.J., Saloni D.E., van Dyk H., Guinta M., 2009. Process flow improvement proposal using lean manufacturing philosophy and simulation techniques on a modular home manufacturer. Lean Construction Journal: 77-93.

ABOUT THE AUTHORS

Beda Barkokebas is a PhD student at the University of Alberta whose current focus is on the pre-panelized construction approach, and the development of innovative building information modelling (BIM) technologies.

Youyi Zhang is a Research Assistant at the University of Alberta, whose focus is on modular construction production line performance improvement, lean production, and simulation application.

Chelsea Ritter is a PhD student at the University of Alberta with a research focus on process improvement and production line layouts.

Mohamed Al-Hussein is a is a professor in the Hole School of Construction Engineering and Director of the Nasseri School of Building Science and Engineering at the University of Alberta, and holds the NSERC Industrial Research Chair in the Industrialization of Building Construction. A highly sought researcher and consultant in the areas of lean manufacturing, construction process improvement, CO_2 emission quantification and reduction, and building information modelling (BIM), Dr. Al-Hussein has successfully applied lean principles to improve work methods and productivity standards for various industries and projects. His research has developed best practices for panelized building systems, lean production, and modular construction, and has been published in approximately 200 peer-reviewed journals and conference proceedings.

ANALYSIS OF THE MANUFACTURING LEAD TIME IN A GRAPHIC ARTS COMPANY

J. Eduardo Morales López[(a)], Carlos Quintero Avilés [(b)], K. Karina Ramírez Mora[(c)], Leonardo J. Cureño Torres [(d)]

[(a),(b),(c),(d)] Technological University Fidel Velázquez

[(a)]moralesjeduardo@outlook.com, [(b)] carlos.quintero@utfv.edu.mx, [(c)]ramirezmorakarina96@gmail.com
[(d)]leojcrt@gmail.com

ABSTRACT
The study was made in a specialized printing folding pharmaceutical and cosmetic packaging company. This study was carried out to identify and propose solutions to the problems that afflict the organization, because, due to the inefficiencies in the delivery time of the work to the customers, they stopped making orders to the company with a direct impact on the revenues of the same. Previously the company had a total of 186 customers, the company has now lost 77% of them, now working only for 62 customers, reducing production demand by 50.2% compared to the same period, but 3 years ago. Therefore, the simulation of the process was carried out to identify the points susceptible to changes and be able to counteract delays and avoid the loss of more customers.
Keywords: Production times, methods, simulation, just in time.

INTRODUCTION
The company has been dedicated to offset printing for 46 years, has more than 307 m2 divided into 3 floors, which 154 m2 are for production and the 153 m2 remaining are used for administrative purposes. In addition, two more terrains are annexed which is one block away and a 5 blocks respectively which is rent and has an area of approximately 59.89m2 and 520m2.
Due to the inefficiencies of the deliveries, there has been a loss of customers, which has caused a decrease in the production orders that these generated.
In the following graph shows decrease in work orders respecting to the other semesters, where it can be seen that from the second half of 2016 there was a steep decline in the work orders sent to the production area for its realization. Likewise, it can be observed that the period between the end and the beginning of the year is where most production orders are lost, reflecting the loss of customers, with an average of 117 lost orders.

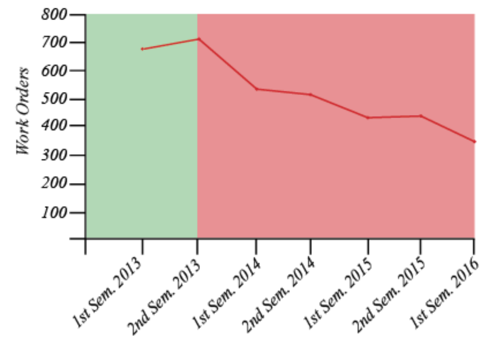


Image 1. Comparison of Work Orders by Semester. Own elaboration.

Due to this abrupt decrease in the orders that were generated towards the production department, it was decided to carry out an analysis of production times within the manufacturing process within the company.

DEVELOPMENT
A case to consider in the company and they are not given the necessary interest are the times of production and consequent the times of completion and delivery of the manufactured product, where the total orders of work per month that are issued a production of 80% presents the status of delay in the completion of the process, in the latter the average production is 2-5 days and sometimes the delay is usually more than 5 days.
It should be seen that sometimes the work orders at the delivery time of production and denote a certain range of delay, counting only 2 to 5 days for its conclusion, being that in some products the process is long and detailed.
An analysis made to the last 10 Work Orders registered from June 19 to July 30, showed the following results:

Product	*Printing process*	*Start date*	*Scheduled end date*	*Actual end date*	*Percent finished*	*Status*
40 Binnacles	Offset Printing	19th Jun.	26th Jun.	In process	90.80%	Delayed
30 Labels	Silk-screen printing	23th Jun.	10th Jun.	21th Jul.	100%	Delayed
250 Labels	Offset Printing	8th Jul.	7th Jul.	In process	0%	Delayed
35 Binnacles	Offset Printing	25th Jun.	2nd Jul.	In process	33.33%	Delayed
4,000 Labels	Offset Printing	15th Jul.	20th Jul.	In process	33.33%	Delayed
1,000 Order Sheets	Offset Printing	20th Jul.	21th Jul.	In process	33.33%	Delayed
50 Arrows	Silk-screen printing	15th Jun.	21th Jul.	23th Jul.	100%	Delayed
100 Labels	Silk-screen printing	15th Jun.	22th Jul.	23th Jul.	100%	Delayed
10 Rulers	Silk-screen printing	without date	23th Jul.	22th Jul.	100%	On Time
41 Binnacles	Offset Printing	23th Jun.	30th Jul.	In process	33.33%	On Time

On time production. Unfinished and delayed production. Delayed production.

Table 1. Status analysis of work orders. Own elaboration.

From Table 1 it can be seen that the last 10 orders sent to production, 30% were delivered to customers at wrong time, 50% of them have a delay in the process, between 30 and 90 percent of its culmination, the remaining 20% are orders completed in time and form.

This type of production delays have a direct repercussion on the company, showing a non-serious organization in the market, with customer losses due to non-compliance on the established delivery dates.

The workload that falls on the production department is very strong and therefore the pressure that this causes has a consequence in a decrease in the quality of the product that is manufactured, causing possible errors in the final product.

PRELIMINARY METHOD

To carry out the study of the process of transformation of the product will take 3 stages within a single product, this involves:

Image 2. Sequence of the elaboration process. Own elaboration.

Each of these stages was monitored, collecting data such as: times, distances, operations, inspections and adding notes to perceived events that are directly involved in the process, whether predicted or not.

Subsequently the data obtained will be plotted, interpreted and analyzed, evaluating the possible solutions to speed production times avoiding and predicting dead and/or unproductive times that affect the flow of the manufacturing process of the product.

DATA COLLECTION

As background:

Previously the product "tint box" was performed, which presented problems of variation in the register at the time of printing, of the total of the required printing, approximately 40% had this variation of registration in the yellow color. The problem is awarded to the supplier of CTP boards.

On August 19th, the printing was realized for the customer NX, printed work with 6 inks (CMYK + Pantone 1795C + Pantone 877C + Primer), the work contemplates two other processes for its completion: wiping and linear gluing. For the accomplishment of this work an analysis is made to the times of production and movements, using tools like chronometers, process diagrams and the observation of the processes.

First is showed the General Diagram of Operations Process which shows the complete process of the product. (See Attached A)

Subsequently it continued with the data collection of the processes that make up the general path to finalize the product. In order to locate the improvement points in the product manufacturing process, a time and motion study was carried out to determine unnecessary dead times and movements, which were recorded in process diagrams, see attached B (two processes were sequenced Which compose the development of the product in each of the diagrams; Prepress-Print and Die-Cut.)

Once collected the data of the process of impression of the packaging the analysis of them was realized, these data were summarized in a graph to make easier their understanding (Image 3), this graphic was divided basically in three sections: Pre-press, printing and downtime.

In this process the module, so to speak to each section of the process, that had the longest processing time was the pre-press which has 53.1% (2161 min); Here what most absorbed was not the work, if you do not wait for the approval and the lapse in time that the supplier of the CTP plates delayed in delivering them, since this type of plates are not made in the company.

Something that stands out above the print time is dead time, which with 34% is second in terms of time, which worries since more than 1/3 of the total time of the process is dead time or wasted time . The downtimes recorded in this process were due largely to carelessness and poor communication between departments

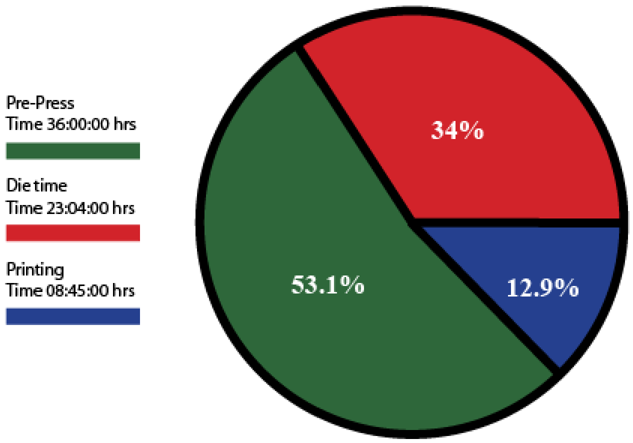


Image 3. Circular Diagram of Summary of Times of the Printing Process. Own elaboration.

As in the printing process, the analysis and graphing of the data collected during the punching and gluing process of the N.X. (Image 4), a very significant difference to be highlighted in this part of the process are the distances traveled by the product in only these two modules of the process, the total distance traveled is 708 m, just under one kilometer.

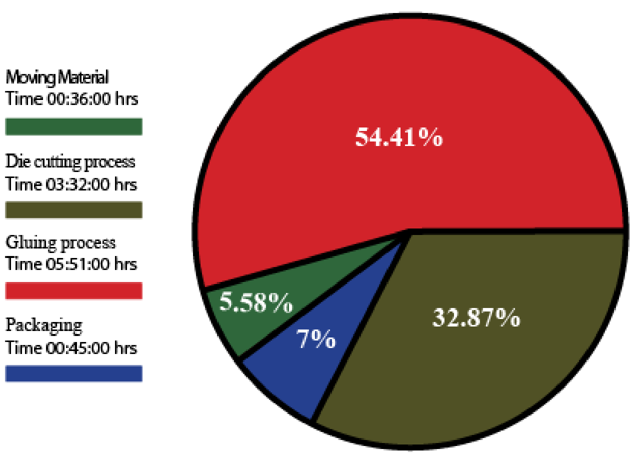


Image 4. Circular Diagram of Summary of the Die and Paste Process. Own elaboration.

This is because the three places where the product ultimately passes are not located in the home alone.

Integral Process Analysis

The data collected and analyzed above were concentrated to carry out a study of all the time and the movements that were used in the manufacture of the product, involving the 4 base modules for the completion of the work; Prepress, Print, Die Cut and Paste.

This general study of the manufacturing process involves all parts of the process and is summarized in a single main graphic to facilitate the analysis that takes place at the flow of the process.

First, the current flow diagram is presented, which is used in any work that is done within the company.

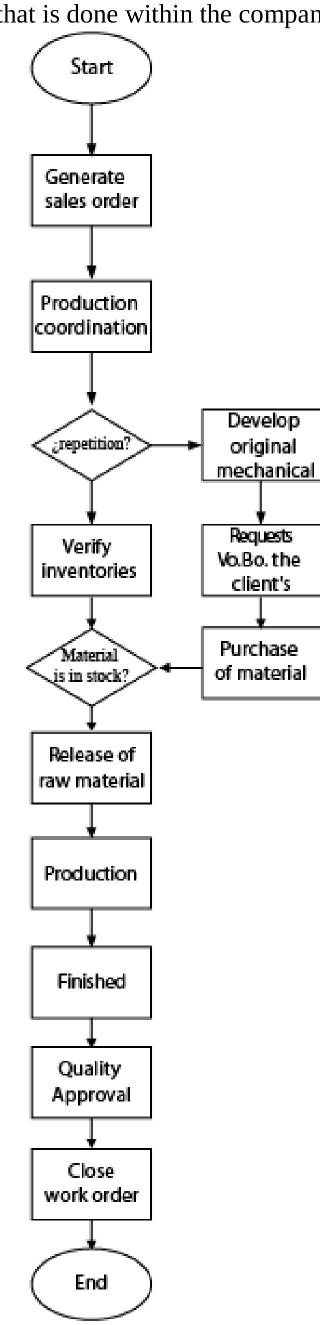


Image 5.General Process Diagram For Printing Of Folding Packaging. Own elaboration.

However, once the data were grouped for their general analysis, they were interpreted and a circular graph was obtained (Figure 6), where each of the modules by which

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

the manufactured product was transformed to be broken off, contains the time in each module were obtained and later analyzed separately.

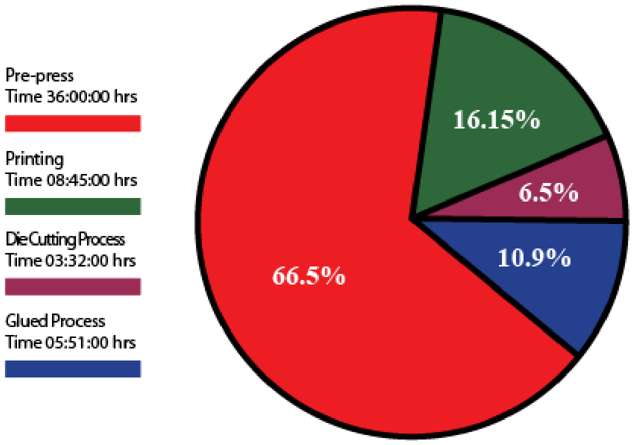


Image 6. General Time Graph in the 4 Process Base Modules. Own elaboration.

This general study showed that unproductive times span more than a quarter of the total time required to complete the process, a figure that is accurately reflected in Picture 7.

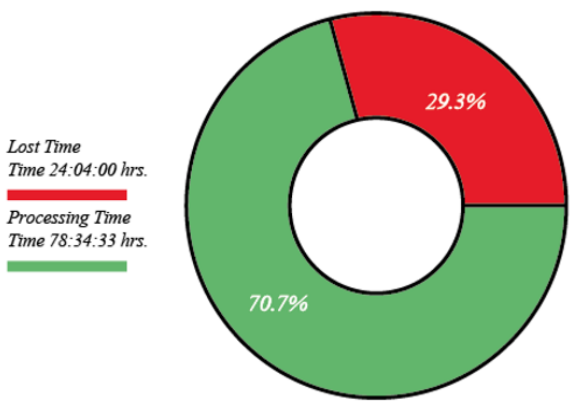


Image 7. Processing times and loss times. Own elaboration.

Modeling

The simulation of the system was used to experiment in the method commonly used to carry out production within the company. In the first instance, the current model was simulated to have a comparison reference for the future system, from which the data already mentioned and presented within the document were obtained. Here we can see the three addresses that the company has to carry out the production process of the folding packages. With this we observed the long movements that the material undergoes to take it from one processing site to another; As it happens to go from the place where the punching is carried out to where it will be glued and packed, for which you walk about 1.1 km occupying 26 minutes on average in moving from one home to another.

In the future modeling it was contemplated in the first instance to reduce the movements that were generated in the process, since after the transfer of the finished material in a process to its subsequent one that was in one of the other two areas with which the company counts The transfer times were very remarkable and did not generate any value within the system.

Image 8. Simulation model. Own elaboration.

Because in the printing process these movements represent 3.2% of the total time that was occupied for this process, along with pre-press. Similarly, in the gluing and die-cutting process, these movements amount to 5.6% of the total.

This is how it was supported by the simulation of the process to observe the behavior of the same, and then make an analysis of the obtained and with it generate alternatives for the best operation of the system.

Recommendations

In order to carry out this reduction of time it is suggested to transfer the machinery to the largest address with which it is counted, so that the process has a more fluid transit to the different areas through which it passes, thus at the end of a process Such as printing, the finished material can be moved to the next step, which would be sequentially to this and so on without having to transport large distances for its next process. Thus, in the proposed model, the mentioned option is modeled where an optimal plant distribution is implemented for the process flow.

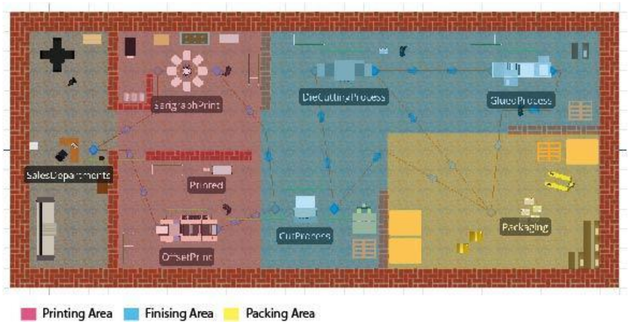


Image 9. Proposed plant distribution by sequential process. Own elaboration.

With this distribution it was sought that the process has a sequence as linear as possible, so that the material does not have to travel great distances which are unnecessary and do not contribute any value to the system, for this reason were formulated 3 zones or general stages of the process; Printing, finishing and packaging, these are the general steps that every product has to pass to its completion. That is why in the distribution these stages have sequence one after another, as seen in Image 9, where the process of printing (offset and screen) is as a pillar of the process, followed by the area of finishes; Comprised of cut or refined which is an average process through which all product passes, that is why it is in the middle of the route. From here you can transfer the product directly to the final packaging or if another process is required (punching or punching) these would be on the side having direct flow to the packaging area of the process for storage or shipment to be delivered to the client.

Thus, with this approach in the reduction of material transfers, the suggested plant distribution would reduce approximately one and a half hours of the general process, taking as comparison the data obtained in the first simulation that was carried out. This represents a reduction of 2.9% of the total time used for the overall process of the monitored product.

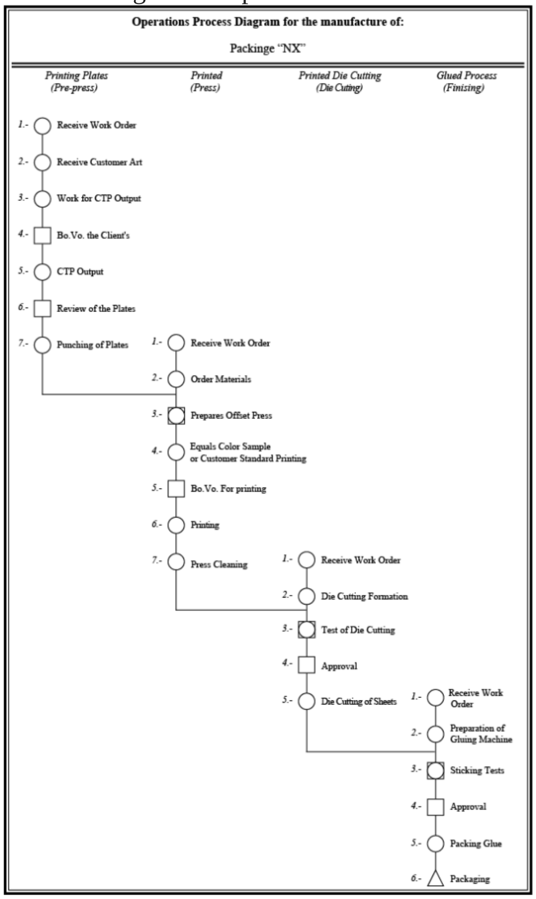
Thus, with the saving of 1.5 hrs for each product and based on the study that was performed to the 10 orders mentioned above (Table 1), there would be a saving in times of 15 hours, which is equal to almost two 8-hour shifts, so that orders that are in the process of being delayed could be terminated in a timely manner to meet customer demand.

Acknowledgments

To the Technological University Fidel Velázquez for the support granted to carry out this work. To COMECYT for their financial support for attending the event. And all the colleagues, workers and friends who were directly or indirectly part of the work.

Attached A

General Diagram of Operations Process.



Attached B

Process Diagrams For Making A Package.

1) Pre Press-Print Process Diagram.

PROCESS DIAGRAM OF: NX Print Packing

DIAGRAM : 3 | METHOD: Current | Page 1 of 1 | DATE: AUGUST 18th

PRODUCT: 10,000 Packings NX

OPERATOR: Javier | PRESS: XL | ~~FTE~~ | VTA

PRODUCT SPECIFICATIONS: 6-ink plus primer; CMYK + Pantone 1795C + Pantone 877C+ Primer

TOTAL TIME: 67 :49:53 hrs

TOTAL DISTANCE: 694 m

	DESCRIPTION	DIST. [m]	TIME [hrs]	●	➡	◗	■	▲
1.-	Pre-Press		24	●				
2.-	CTP ouput		12	●				
3.-	Punching plates		0.01.34	●				
4.-	Transfer material to plant	336	0.23.13		➡			
5.-	Deliver material (warehouse)		0.15.48	●				
6.-	Accommodate printing material		0.06.17	●				
7.-	Prepares Offset Press		1.33.21	●				
8.-	Equals color sample		0.35.11	●				
9.-	Black plate change		4.41.09			◗		
10.-	Insert plaque		0.01.13	●				
11.-	Equals color sample		1.03.21	●				
12.-	Change of mantilla		0.13.41			◗		
13.-	Printing (CMYK and first pantone)		1.25.12	●				
14.-	Machine cleaning		0.26.31	●				
15.-	Prepares Offset Press		0.33.12	●				
16.-	Change the Pantone 877c plate		18.10.0			◗		
17.-	Printing (second pantote and primer)		1.22.54	●				
18.-	Machine cleaning		0.18.21	●				
19.-	Move printouts to Die Cutting	358	0.38.55		➡			
			TOTAL	14	2	3	0	0

1) Die Cutting-Glued Process Diagram.

PROCESS DIAGRAM OF: NX Die Cuntin Process and Glued Process Packing

DIAGRAM : 2 | METHOD: Current | Page 1 of 1 | DATE: AUGUST 24th

PRODUCT: 10,000 Packings NX

OPERATOR: Margarito | M/C: Gluing machine | FTE | VTA

OPERATOR: Barush | M/C: Die cutter

TOTAL TIME: 10:44:40 hrs

TOTAL DISTANCE: 708 m

	DESCRIPTION	DIST. [m]	TIME [hrs]	●	➡	◗	■	▲
1.-	Move print to court	336	0:11:54		➡			
2.-	Cutting of printed matter		0:22:18	●				
3.-	Transfer of cut-to-die print	18	0:05:57		➡			
4.-	Prepare die cutting machine		0:36:44	●				
5.-	Die Cutting Testing		0:01:37	●				
6.-	Authorization of die cutting						■	
7.-	Die cutting of sheets		2:31:29	●				
8.-	Moving material to glue	354	0:18:41		➡			
9.-	Deburring of die-cut material		1:04:31	●				
10.-	Prepare the gluing machine		1:24:53	●				
11.-	Authorization of gluing						■	
12.-	Gluing process		3:21:22	●				
13.-	Packaging		0:45:14					▲
			TOTAL	7	3	0	2	1

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

REFERENCES

Báez O. et al. *Methodology decision support for production plan in flexible manufacturing* systems, Volume 2, No. 1, 2008.

Dr. Eduardo Marroquín Prado, *A scheme to monitor and control the statistical and technical capacity of a process* Volumen 5, México 2011**.**

Estudio del Trabajo, Recuperado de: http://www.ingenieria.unam.mx/industriales/descargas/documentos/catedra/libro_ET.pdf

Flores I., E. Cortes. *Simulation notes* Universidad Nacional Autónoma de México, Facultad de Ingeniería, México, 2006.

Hafezalkotob, A., Ketabian, H., Rahimi, H., 2014. *Balancing the Production Line by the Simulation and Statistics Techniques: A Case Study*. Research *Journal of Applied Sciences,* Engineering and Technology, 7(4): 754-763.

Kelton, W. David, Smith, Jeffrey S., Sturrock, David T. *Simio and Simulation: Modeling,* Analysis, Applications 2nd. Ed.2012

Kenett, S. Ron, Shelenyahu Z., 2000. *Estadística Industrial Moderna.* 2da edición. Thomson editores.

M.C. Ortiz F et al. Comparison of Standard *Costing System and the Theory of Constraints Control Flow of Materials Using a simulation model,* Volume, No. 1, México 2008.

Meyers F. (2001) *Estudios de Tiempos y Movimientos*. Recuperado de: https://books.google.com.mx/books?id=cr3WTuK8mn0C&pg=PA1&dq=tiempos+y+movimientos&hl=es-419&sa=X&ved=0CBsQ6AEwAGoVChMIitLr1OTUxwIViwySCh28TAEw#v=onepage&q=tiempos%20y%20movimientos&f=false

Quintero C., Flores I., *Simulation of the design area in a printing company*. EMSS 2013.

Senem, B., 2011. *Assembly Line Balancing in Garment Production By Simulation*. http://www.intechopen.com/books/assembly-line-theory-and-practice/model-sequencing-andassembly-line-balancing-in-garment-production-by-simulation-worker-transfer-syst

Vaughn R. *Introducción a la Ingeniería Industrial*. Recuperado de: https://books.google.com.mx/books?id=udFwMwT4xDMC&pg=PA385&dq=tiempos+y+movimientos&hl=es-419&sa=X&ved=0CCUQ6AEwAmoVChMIitLr1OTUxwIViwySCh28TAEw#v=onepage&q=tiempos%20y%20movimientos&f=false

AUTHORS' BIOGRAPHY

J. Eduardo Morales López. He studied Engineering in Production Systems in the Area of Graphic Arts in UTFV, participated in SIMIO Simulation Student Competition and in the National Award of Arts, organized by CANAGRAF, in the category of Craft Screen Printing.

Carlos Quintero Aviles. studied Industrial Engineering in the ITTLA, Technological Institute of Tlalnepantla. He taught at UTFV University and has participated in national and international conferences He Studied a Masters in Operational Research in the UNAM. His line of research is simulation and optimization of productive processes.

Karla Karina Ramírez Mora. Studied Top Technical Degree in Graphic Arts Area at Technological University Fidel Velazquez; actually, she is studying Production Systems Engineering, and she has participated in Simio Simulation Competition.

Leonardo Jesús Cureño Torres: Studied Computer technician at Scientific and Technological Studies Center, Studied Computing and Graphic Design at Professional Computing Center of Mexico, Studied Top Technical Degree in Graphic Arts Area at Technological University Fidel Velazquez; actually, he is studying Production Systems Engineering. He participated at Sinapse packaging contest 2016 and Simio Simulation Competition 2017.

SIMULATION OF THE MEXICAN AIRPORT NETWORK FOR ADDRESSING A GROUND DELAY PROGRAM

Wellens A. [(a)] **Miguel Mujica Mota**[(b)]

(a) Facultad de Ingeniería, UNAM-MÉXICO
(b) Aviation Academy, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

[(a)] wann@unam.mx, [(b)] m.mujica.mota@hva.nl

ABSTRACT

Air traffic in Mexico has grown at a high pace, despite the economic downturns the country has suffered recently. In turn, Mexico City airport is located close to the centre of the city and is Mexico's busiest airport and is considered congested. One of the consequences of airport congestion are flight delays which in turn decrease costumer's satisfaction. Air traffic control has been using a ground delay program as a tool for alleviating the congestion problems, particularly in the most congested slots of the airport. This paper describes the application of a simulation model to analyse the effectiveness of the ground delay program. The use of the simulation model will enable the decision makers to analyse the effectiveness of the ground delay policy as well as to evaluate different policies for coping with the increasing demand in the Mexican network of airports.

Keywords: simulation model, airport capacity, flight delay, airports network

1. INTRODUCTION

Air transportation has grown very fast in the last century, especially in high and middle-income countries. Even in conservative scenarios, this growth is expected to continue in the future (EUROCONTROL, 2013; Campanelli et al., 2016). As a result, congestion problems and flight delays are becoming more acute in many airports. A flight is considered delayed when it arrives 15 or more minutes after the scheduled time (Federal Aviation Administration, 2009). According to the Federal Aviation Administration (FAA) in the United States, flight delays increased by more than 58 percent from 1995 until 2002 and cancellations by 68 percent (Nilim and El Ghaoui, 2004).

In airports with important capacity constraints, such as Frankfurt (FRA), London-Heathrow (LHR) and London-Gatwick (LGW), there is virtually no idle capacity available for growth and/or unscheduled flights such as general aviation, military or governmental flights (Bubalo, 2011). This is also the case of Mexico City International Airport, which was declared saturated between 7:00 am and 10:59 pm, observing on more than 52 occasions in 2013, at certain times, that operations in the Mexican air space exceeded the maximum number that can be attended per hour (SEGOB, 2014).

A study conducted in 2010 by the FAA estimates that flight delays cost the airline industry $ 8 billion annually, mainly for concepts such as increased crews, fuel and maintenance costs (Ball et al., 2010). The delays cost passengers even more, almost $ 17 trillion, according to the same author. Due to the high costs of delay, airlines and airport service providers are constantly looking to optimize flight times and resource utilization.

Traffic flow management initiatives can be used to control air traffic demand and mitigate demand-capacity imbalances. These can include ground stops, ground delay programs, rerouting, rescheduling, airborne holding, miles-in-trail restrictions (Chatterji and Sridhar, 2004; Ball et al., 2007; Swaroop et al., 2012; SESAR, 2012; Brunner, 2014). Applicable policies can be classified according to their time horizon (Terrab, 1990; Leal de Matos and Ormerod, 2000):

- *Long term* policies (several years) include the construction of new airports or the expansion of existing ones, as well as an improvement in air traffic control technologies which lead to time reductions.
- *Medium term* policies (up to 1 year) include modifications to and/or temporary redistributions of the flight planning, and changing departures to off-peak times to avoid periods of excessive demand.
- *Short term* or *tactical* policies (24 hours) as ground delay programs (GDP) are applied to diminish acute delay related costs and safety problems.

Implementation of ground or pre-departure delay programs (Luo and Yu, 1997; Dell'Olmo and Lulli, 2003; Agustin et al. 2010) is one of the most popular management initiatives throughout the globe: this corresponds to tactically match demand with capacity in the arrival airport by imposing a delay on the ground for a reduced number of flights at the airport of departure. Originally, it was implemented to avoid problems due to inclement weather. For example, in the US, when weather conditions deteriorate, the FAA can determine that part of the expected arrivals at an airport will exceed the airport's capacity and thus implement a GDP, specifying new arrival slot assignments for the affected set of flights (Luo and Yu, 1997). Besides the use of a

GDP to cope with bad weather, it can also be used to balance a congested airport's demand and capacity. This practice is theoretically cheaper, less polluting and less complicated, than allowing the aircraft to take off and put it on holding when it approaches its final destination (Guest, 2007). However, it is a disruptive tactic for air operators, whose schedules are set up with tightly connected operational resources and can therefore lead to excessive delays for the affected flights.

According to SENEAM, the Mexican air traffic control authority, the airport of Mexico City can only receive a maximum amount of 40 arrivals per hour (BNAmericas, 2014), so a GDP is currently used to reduce capacity problems during peak-hours. However, local airlines claim that this is causing them more inefficiencies, coupled with high costs and a declining reputation.

2. THE IMPORTANCE OF MEXICO CITY AIRPORT

The total number of operations in the Mexican air transport system reached more than 1,750,000 in 2016 (SCT, 1017a). Correspondingly, over 92 million passengers were transported in that year, which is an increase of 9.4% compared with the previous year (SCT, 2017b) while passenger flights constitute almost 90% of Mexico's air transport.

The domestic sector transported 53 million passengers (58% of the total) while international carriers moved 39 million passengers. Figure 1 shows the demand of the 9 existing commercial passenger airlines in Mexico in 2016. It can be noticed that the biggest national airlines in terms of transported passengers are Volaris, Interjet, Aeromexico, Aeromexico-Connect and VivaAerobus, which moved respectively 14.3, 11.1, 11.1, 8.5 and 6.2 million passengers. Together, they move over 95% of the flights served by Mexican carriers.

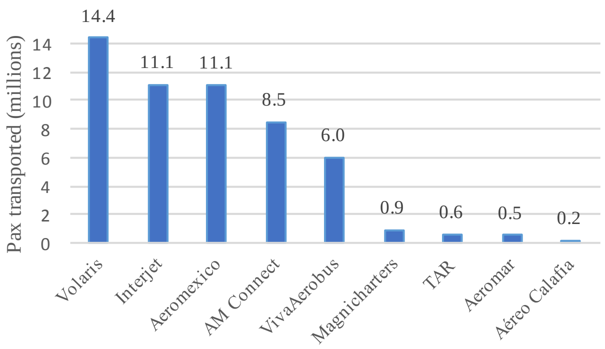


Figure 1: Passengers Transported by National Airlines in Domestic and International Routes in 2016

Mexico's flag carrier, Aeromexico, has had a steady growth since 2009, as can be observed in figure 2. However, Mexican low-cost carriers (LCC) are growing quite fast. In 2005-2006, Interjet, Volaris and VivaAerobus started operations, of which Volaris has presented the biggest growth until 2016. In 2016, the low-cost sector had already accounted for almost 80% of the market share. Other smaller airlines as Magnicharters and Aeromar have been operating for at least 15 years in the sector, although their market share is low and constant. Transportes Aereos Regionales (TAR) and Aéreo Calafia are two small LCC that just started operations three years ago.

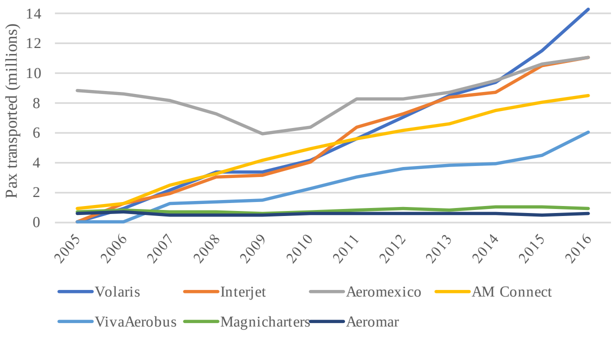
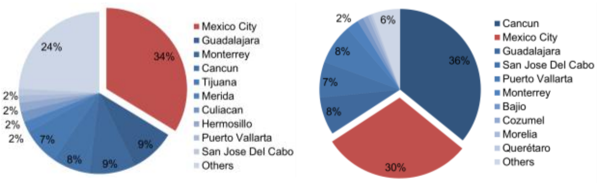


Figure 2: Main Development of Mexican Airlines Since 2005

Mexico has 76 airports, 58 of them are international airports and 18 national; in addition, there are 1,914 aerodromes registered in the country (SCT, 2017b). This places Mexico as one of the countries with the major airport network (CIA, 2017). Figure 3 presents the 10 top airports by passenger traffic within Mexico from January until May 2017. It can be noticed that Mexico City International Airport (IATA Code: MEX) moves 34% of the total domestic traffic of the country, followed by four other airports: Guadalajara (9%), Monterrey (9%), Cancun (8%) and Tijuana (7%), respectively. In the international context, Cancun International airport is a good competitor for Mexico City airport, moving 36% and 30% of the total, respectively. Considering both domestic and international passengers, MEX has a market share of approximately 32% of the total of transported passengers (SCT, 2017b), which makes it the busiest airport in the country. It also conforms, since 2003, the pillar of the Metropolitan Airport system, together with Queretaro, Puebla, Toluca and Cuernavaca.



(a) (b)

Figure 3: Air Passenger Traffic by Main Airports in Mexico, Jan-May 2017. (a) Domestic, (b) International (AICM, 2017)

The total number of operations (including less than 3% cargo flights and 8.5% of general aviation flights, mainly domestic) reached almost 450,000 in 2016, 73% of which corresponded to national flights and 27% to international ones (AICM, 2017).

Mexico City Airport is considered key for the development of the metropolitan region of Mexico city and the rest of the country. Recently, the government has announced the development of a new airport in Mexico

City which will have a final capacity of 120 mill pax/year. However, the first phase for this airport will not be operative until 2020. In the meantime, Mexico City as a destination is still growing and the country has also gained importance as a tourist and business destination. Since its important position in terms of number of operations as well as its functionality of the Hub operations of certain carriers, MEX reveals as an important node whose operation affects the complete national network of airports. Therefore, the understanding of efficient ways of managing the airport will affect not only the airport itself and the stakeholders that participate in it but also the complete national network.

3. LITERATURE REVIEW

Leal de Matos and Ormerod (2000) expose the application of operational research initiatives in the European Air Traffic Flow Management, detailing several of these strategies.

At the tactical level, the goal of GDPs (also called ground holding programs) is to avoid airborne delays by transferring them to the ground. The beginning of these policies goes back to 1973, when the oil crisis generated an increase in fuel costs that made air delays much more expensive. Consequently, the FAA adopted a policy to prevent the departure of an aircraft when its arrival at the destination airport could not be guaranteed and thus prevented the endless increase in the number of aircraft flying around the destination airport. Initially, the air traffic controllers made the decision based on their experience. However, advances in science have led to the development of operational research methodologies that allow finding an optimal or suboptimal solution (Agustin et al., 2010).

Most studies in the field focus on the optimal allocation of a GDP, as part of the Air Traffic Flow Management (ATFM) problem (Odoni, 1987; Andreatta et al., 1998; Leal de Matos and Ormerod, 2000; Inniss and Ball, 2004; Lulli and Odoni, 2007). In this sense, we can distinguish between the Single Airport Ground Holding (SAGH) problem, studied since the late 1980s (Andreatta and Romanin-Jacur, 1987; Terrab and Odoni, 1993; Richetta and Odoni, 1993; Dell'Olmo and Lulli, 1993), and the Multi Airport Ground Holding (MAGH) problem, studied since the early 1990s (Vranas et al., 1994; Richetta, 1995; Andreatta and Brunetta, 1998; Bard et al., 2001; Zhang et al., 2007; Glover et al., 2013).

Most studies model US applications, with congestion limited to airports. In-air congestion problems were not originally included in the analysis, because in the United States, where the problem was first studied, congestion only occurs at airports and not in the airspace. Early studies are generally deterministic (Terrab and Odoni, 1993), while recent studies, such as the ones from Mukherjee and Hansen (2007), Andreatta et al. (2011) or Agustin et al. (2012) consider the stochastic nature of the problem. Agustin et al. (2010) present an interesting and detailed review on optimization by mathematical programming models for air traffic flow management.

Since traffic flow management decisions are typically made 30 minutes to several hours in advance of anticipated congestion, the predictions are subject to significant uncertainty (DeArmon et al., 2008) and the solution to the described optimization problems are needed quickly. Documented solution mechanisms include branch and bound methods (Bard and Mohan, 2008), other exact methods (Andreatta et al., 1998), GRASP (Argüello et al., 1997), TSP (Vasquez-Marquez, 1991) and tailored heuristics (Luo and Yo, 1997), among others.

In addition, simulation has been used to represent and predict the air traffic system's capacity, demand and related congestion problems (Frolow and Sinnot, 1989; Winer, 1993) and to explore different strategies and system improvements (Frolow and Sinnot, 1989; DeArmon and Lacher, 1996). More recently, Fleurquin et al. (2013) used a simulation model to test a ground delay mechanism to a set of airports affected by weather perturbations. Delgado et al. (2013) used the FACET tool developed by NASA-Ames (Bilimoria et al., 2000) and the Airbus PEP program to assess cruise speed reduction for GDP.

This paper describes a simulation model to assess the current GDP in Mexico City. Stochasticity of the flight duration, on-time performance and turnaround times are included in the model to analyse how the effectiveness of the GDP is influenced by its parameters.

4. METHODOLOGICAL APPROACH

In this section, the proposed methodology is described together with the modelling approach for the different elements of the model.

Simulation using Discrete Event Systems (DES) is a special type of dynamic systems approach for modelling systems. The state of the system is a collection of variables that represent different values of the system under study. Hence the state of the system under study is defined by a combination of the values of the variables used. In the DES approach the "state" of these systems changes only at discrete instants of time and the term "event" is used to represent the occurrence of discontinuous changes at possibly unknown intervals (Flores de la Mota et al. 2017). Different discrete event systems models are currently used for specification, verification, synthesis as well as for analysis and evaluation of different qualitative and quantitative properties of existing physical systems such as manufacturing ones, port and airport systems.

In DES, the operation of a system is represented as a chronological sequence of events. Each event occurs at an instant in time and marks a change of state in the system; for this reason, this methodology suits the best for modelling a network of airports where the entities represent the aircraft that go from one place to the other following a specific sequence of steps where uncertainty affects mainly the speeds and processing times but not the structure of operations.

4.1. The Mexican network model

The simulation model used in this work corresponds to DES and was developed using the SIMIO software system (SIMIO 2017). SIMIO uses a process-object oriented approach which suits perfectly for the type of operations performed by the aviation industry, where everything happens at scheduled times and the control of uncertainty is one of the main goals of the operation.

The model involves aircraft moving between airports in a network of nodes connected by paths of a length proportional to the flight's travelling time. In the model only one destination is considered, which in this case is MEX; all direct flights to MEX and corresponding departure airports are included in the model. The first version of the model considers 98 departure airports, 26 carriers and 22 equipment codes; the latter are subdivided in medium, large and heavy aircraft, according to their maximum take-off weight (MTOW).

MEX airport has 56 direct boarding gates in two terminals, as well as 40 mobile contact positions in 6 remote platforms, making a total of 96 contact positions for air operations (SENEAM, 2015). Although flights are assigned to a specific terminal and/or contact position depending on the carrier and aircraft type, the model considers a total of 96 positions without distinguishing between carriers, aircraft type or terminal used.

The events in the simulation model are triggered by the information specified in the provided flight schedule, including origin airport, flight operator, aircraft type, departure time, arrival time and flight duration. Flights are generated in the model at the time of departure; the flight time to MEX is determined from the scheduled arrival time. Other data used by the model includes aircraft specific (for instance maximum take-of weight and wake category), airline specific (for instance on time performance, average arrival delay, type of operator) and airport specific (for example country of origin) information. Aircraft and airport specific data is used to be more accurate in the model logic, while airline data is used to be able to take into consideration the stochastic character of flight duration and delay.

4.2. Modelling the demand

Most of the data processing was done using the R software environment. The model was set up with flight information retrieved from OAG (2017), corresponding to the first week of 2013. The data includes a total of almost 200,000 registers, corresponding to the information of flights arriving to MEX airport in one or to flight legs. With information on the initial and last date where each flight is scheduled, and filtering the days of the week when a specific flight operates, daily flights were extracted from Jan 1 to Jan 8, 2013. Table 1 presents an example of the data used.

Table 1: Example of the Flight Data Used in the Model

Origin	Carrier	Equip	Flt No	Deptarture Time	Arrival Time
BJX	5D	ERJ	123	31/12/12 22:50	01/01/13 00:01
CUN	VB	733	3147	31/12/12 21:40	01/01/13 00:05
GDL	VB	733	2708	31/12/12 22:50	01/01/13 00:05
TIJ	Y4	320	816	31/12/12 20:44	01/01/13 00:05
PTY	CM	738	194	31/12/12 20:16	01/01/13 00:06
CJS	VB	733	3177	31/12/12 21:50	01/01/13 00:20
DFW	AA	M80	409	31/12/12 21:45	01/01/13 00:20
REX	VB	733	3219	31/12/12 22:55	01/01/13 00:25
CUN	VB	733	3149	31/12/12 22:25	01/01/13 00:50

4.3. Estimation of actual flight schedules

According to statistical information published by MEX (AICM, 2017), the number of flights in this airport have increased since 2013 with approximately 4% each year. While in January 2013 on average 490 flights were arriving at MEX, this number had increased to 575 in January 2017, registering a total increase of 17%. To take into account this increase and at the same time make the simulation model flexible enough to evaluate the GDP at different times, random flights were generated with the same origin, carriers, equipment and frequency distribution as registered flights. These additional flights were assigned to a specific hour-period according to the used time slots published by AICM (2017) for the first four months of 2017, and respecting the difference between different weekdays.

From Monday until Friday, an average of 1063 daily operations was registered in the analysed four-month period. Half of these, on average 531 flights, are assumed to correspond to arrivals, the rest to departures. On Saturday and Sunday, the number of daily arrivals diminishes with respectively 12% and 8% (See figure 4). Considering the weekend, the average was 1032 slots per day, thus approximately 516 arrivals and the same amount of departures.

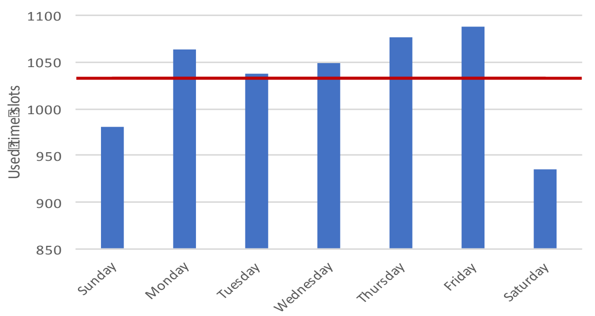


Figure 4: Average Used Time Slots in MEX per Weekday, Jan-Apr 2017

On the other hand, a variation of used slots according to the time of the day can be observed. Figure 5 presents the arrival slots for the less occupied and the busiest weeks in the analysed period. Analysis of the used AICM data indicated that the least busy week was just after the Eastern holidays, from April 23 to 29, with a total of 3437 arrivals for the whole week. The busiest week corresponded to March 19^{th} to 25^{th} in 2017; the total of 3932 arrivals in this week can be explained due to the spring break in the US (an increase of 14.4%).

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Under the slot scheme presented by AICM (2013), that indicates a maximum number of 58 slots assigned to airlines and 3 to official aviation, but giving priority to passenger transport, it can be shown (see figure 5) that the airport is working at high capacity most of the day. Analysis of the graph suggests that when operations increase, slot use is increased early in the morning or late at night, when still some capacity is available. See for example the difference in blue and green lines for the periods between 04:00 and 06:00, or after 22:00.

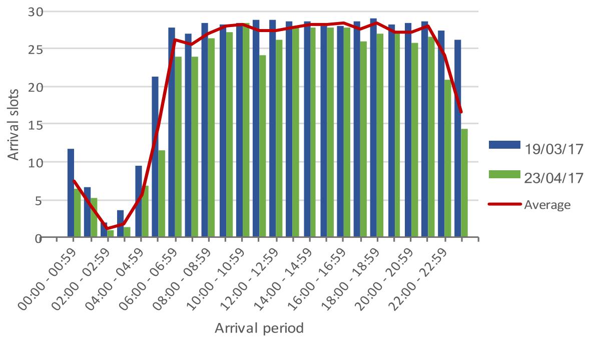


Figure 5: Used Time Slots per Hour in MEX, Jan-Apr 2017.

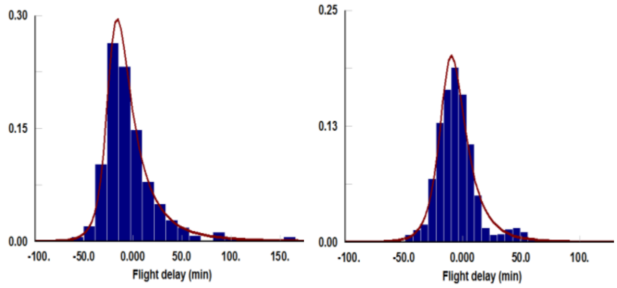
4.4. Analysis and modelling of flight times

The available departure and arrival times correspond to information scheduled before the flight takes place. Delay distributions were analysed from public flight information (Airportia, 2017) for the airlines flying to MEX in order to estimate in a more realistic fashion the arrival times. A total of 6221 flights operated between May 23 and June 10, 2017 were analysed.

As the highest share of analysed flights corresponds to Mexican airlines (26% Aeromexico, 23% low cost carriers), the delay distributions of these airlines were determined separately. All other airlines were grouped according to the continent where they were operating. Corresponding delay distributions were fitted using Stat::Fit®. In all cases, a Johnson S_U distribution could be fitted; this is a four-parameter family of distributions proposed by Johnson (1949) as a transformation of the normal distribution. Table 2 presents the values of the parameters for the fitted distributions. It is worth to note that negative values can occur, which correspond to flights arriving early; figure 6 presents two examples.

Table 2: Parameters for Fitted Delay Distributions, All Flights (Minutes)

Carrier	Johnson S_U Parameters			
	ξ	λ	γ	δ
Aeromexico	-22.0	18.7	-0.38	1.18
Aeromexico Connect	-23.0	18.9	-0.50	1.54
Aeromar	0.48	14.0	0.27	1.30
Interjet	-13.1	15.3	-0.49	1.23
Viva Aerobus	-34.0	12.0	-2.95	1.79
Volaris	-39.4	17.0	-1.38	1.47
Latin American carriers	-23.9	12.8	-0.85	1.14
North American carriers	-22.8	14.7	-0.87	1.07
European carriers	-16.1	15.7	-0.49	1.04



(a) (b)

Figure 6: Fitted Delay Distributions for (a) North American Carriers and (b) Interjet (min).

However, we perceived that, as the Johnson S_U distribution is unbounded, the use of the distributions presented in table 2 causes the model to sometimes estimate unrealistically large early arrivals. Also, flight delay for delayed flights was underestimated for all operators. To avoid these drawbacks, the data to be fitted was subdivided in two categories: on-time flights, i.e. flights being delayed less than 15 minutes and where both positive and negative delays can be observed due to randomness, and flights delayed more than 15 minutes due to a specific although not necessarily known reason. For on-time flights, most carriers or carrier groups could be fitted to a Weibull or gamma distribution. Only Viva Aerobus was fitted to a Johnson S_B distribution (see Table 3). The fitted distributions resulted to be rather symmetric, with mean values around -12 (aircraft arriving 12 minutes early) and standard deviations around 14 minutes. The high standard deviation explains the fact that, on long flights, there are planes arriving up to 60 minutes early. Our findings are consistent with the flight distributions used for example by Dorndorf (2016) and Pérez-Rodríguez (2017).

Table 3: Parameters for Fitted Delay Distributions, On-Time Flights (Minutes)

Carrier	Gamma Parameters		
	Location	Shape	Scale
Aeromexico	-249	202	2.15
Aeromexico Connect	-64	11.1	4.33
Volaris	-62	6.85	6.14
	Weibull Parameters		
	Location	Shape	Scale
Aeromar	-43	3.82	43.1
Interjet	-62	4.83	58.6
Latin American carriers	-55	3.42	45.9
North American carriers	-63	3.8	56.6
European carriers	-52	3.06	46.5

	Johnson S_B Parameters			
	ξ	λ	γ	δ
Viva Aerobus	-39.5	74.8	0.348	1.38

For delayed flights, a second distribution was fitted using the same logic as explained above. In all cases, a Weibull distribution was fitted for late flights (see table 4).

Table 4: Parameters for Fitted Delay Distributions, Late Flights (Minutes)

Carrier	Weibull Parameters		
	Location	Shape	Scale
Aeromexico	15	0.82	23.2
Aeromexico Connect	15	0.67	20.9
Aeromar	15	1.36	19.3
Interjet	15	1.35	28.9
Viva Aerobus	15	1.39	27.3
Volaris	15	1.05	37.0
Latin American carriers	15	1.43	38.3
North American carriers	15	0.92	30.3
European carriers	15	0.95	34.7

Having estimated distributions for both on-time and delayed distributions, in-flight delay was randomly assigned in the model to each incoming flight, according to the corresponding on time performance data (SCTb, 2017; BTS, 2017; Flightstats 2017). Published percentages of late flights (delays of more than 15 minutes) range from 3% for AVIANCA PERU to 28.03% for AVIANCA. Corresponding average arrival delay ranged from 30 minutes for Interjet to 71.3 minutes for Delta Airlines. The logic used in the simulation model to assign flight delay is presented in figure 7.

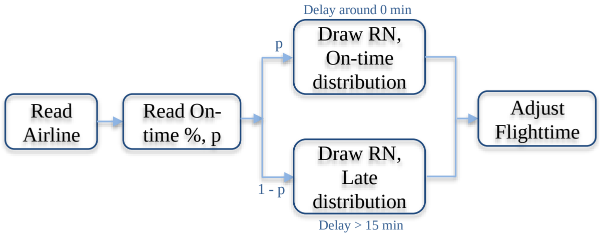


Figure 7: Assignment of Flight Delay in the Simulation Model.

As an example, Aeromexico Connect presented 76.5% of on-time flights according to the reviewed information. This means that the simulation model will assign a random positive or negative delay < 15 min, drawn from the corresponding gamma distribution in table 3 and figure 8, to 76.5% of the incoming flight operated by Aeromexico Connect. The other 23.5% of flights will have a randomly assigned delay > 15 min, drawn from the corresponding Weibull distribution in table 4 and figure 9.

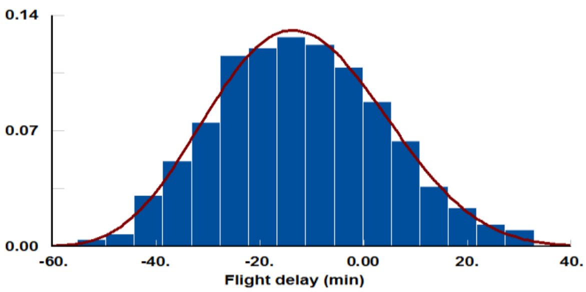


Figure 8: Fitted On Time Distribution for Aeromexico Connect

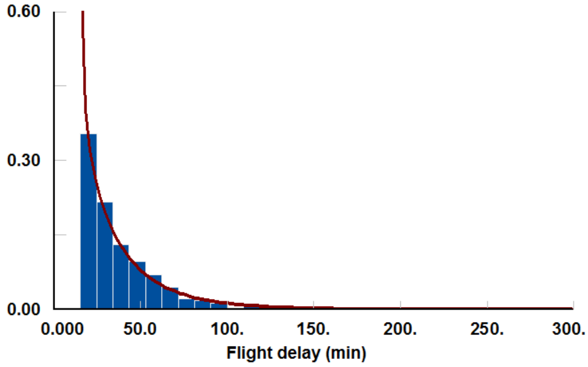


Figure 9: Fitted Late Distribution for Aeromexico Connect

4.5. Simulation of turnaround times

In order to represent the time that a plane is using the assigned gate, turnaround times were estimated from public flight data available from January 26th to February 15th, 2017. Turnaround times depend on several variables, among which the type and size of the aircraft, the degree of saturation and the type (hub or non-hub) of the arrival airport and airline strategies (full-cost or low cost carriers) (Kolukisa, 2011) and its determination is of vital importance to simulate the arrival and departure process correctly.

To obtain an estimated turnaround distribution, different aircraft types were selected for Mexican carriers, typically of the type flying to MEX airport. Through analysis of the aircraft's history, turnaround times were obtained for the Mexican flag carrier and for the 3 major low cost carriers. Airbus 320 and 321 (IATA codes 320 and 321), as well as Boeing 737-700, 737-800, 777-200 and 787-800 (IATA codes 737, 738, 777 and 788 respectively) were included in the analysis. Of the previous, only 777 and 788 are heavy aircraft; the rest are classified as aircraft with wake category M (*medium*).

The fitted distributions are presented in table 5. It can be observed that for medium size aircraft, the turnaround time was generally between 30 and 165 minutes, with distribution mode around 85 minutes. All analysed medium aircraft presented a similar pattern and the corresponding data was merged to obtain a generic loglogistic turnaround time distribution for other medium aircraft not considered in the analysis. For heavy aircraft, mean turnaround times were around 250 minutes (4.2 hours); variability seems so increase with aircraft size, presenting ranges from 2.8 to 5.3 hours for 787 aircraft and 2 to 7 hours for 777 aircraft. As less than 4% of the flights in the analysed flight schedule correspond to heavy aircraft, available information was insufficient to obtain more detailed results. The distribution obtained for Aeromexico's 787 aircraft was used for aircraft with wake category *heavy* when no distribution could be obtained.

Table 5: Proposed Distribution for Turnaround Times (Seconds)

Carrier	Aircraft	Turnaround Time Distribution (s)
Aeromexico	737	1980+7920*Beta(3.18, 4.18)
Aeromexico	738	3420+LogLogistic(3.97, 3030)
Aeromexico	777	8040+Weibull(1.78, 8640)
Aeromexico	788	8220+Weibull(4.52, 6760)
Interjet	320	2040+Lognormal(7.68, 0.508)
Interjet	321	3360+8040*Beta(3.59, 6.72)
Volaris	320	4140+ LogLogistic(1.95, 1480)
Volaris	321	3540+7070*Beta(1.79, 4.88)
Generic	Medium	1980+LogLogistic(3.66, 3390)

4.6. Verification phase

To analyse if the model is working as intended, and to ensure that it produces satisfactorily accurate and consistent results, it was verified thoroughly. Verification activities include comparison of used gates with published slot usage information, verification of the average arrival delay simulated with the proposed distributions and review of the results of simulation trials to check if calculated quantities correspond with the expectations. After the verification was performed and the model was ensured to behave qualitatively according to real reported results we continued to the next phase.
To verify the delay distributions used for late flights, mean and standard deviations were determined for the corresponding Weibull distributions to determine the simulated average arrival delay for all airlines flying to MEX airport. Its value ranged from 32.7 to 52.3 minutes. Although slightly underestimated, the simulated average arrival delays are in the expected range.

5. SCENARIOS AND FINDINGS

The simulation model was used to evaluate the current GDP at MEX airport. When the GDP is not active, the simulation model processes incoming flights on a first come first served base and on-time flights are processed immediately.
In its current form, when delay information on departing flights and/or meteorological conditions suggests that the maximum airport capacity will be reached in the following hours, MEX airport authorities decide to implement the GDP and aircraft coming from nearby airports are delayed by a fixed 15 minute-period to try to decrease the level of saturation. At present, the GDP does not apply to flights coming from abroad. Apparently, the current GDP is based on experience and there are no clear rules on how to implement it. As the airport saturation level in MEX persists to date, changing the GDP parameters might improve its effectiveness.
Two basic scenarios were analysed: the base case considers the current situation, were aircraft are included in the GDP depending on their origin, while the other scenario considers the selection of included aircraft based on their flight time. For both scenarios, 15 different thresholds were tested: the objective pursued was to identify the sensibility of the system to the modification of the threshold value where the GDP program is triggered. Each experiment is executed for a period of seven days; 10 replications were made in all cases. Specific parameters for the simulation runs are:

- *Airport arrival capacity or acceptance rate* where the GDP starts to operate: In our model, this limiting capacity is fixed during the simulation run, as it depends on the saturation of the airport; it varied from 25 to 40 arrivals per hour. However, when a GDP program is due to weather conditions, this value could change during its implementation (see for example Ball and Lulli, 2004).
- *Type of GDP*: The base case considers only flights departing from Mexican airports and operated by Mexican carriers to be included in the program (scenario 1). The alternative scenario includes in the GDP flights with less than 2h of programmed flight time (scenario 2). Longer flights are not included as they are assumed to arrive when the saturation crisis might be solved already.
- *Imposed delayed time*: Currently, the model considers imposed delays in the departure airport in blocks of 15 minutes; however, this can be changed in future runs.

Figure 10 presents how the GPD is implemented in the simulation model. Note that a specific aircraft can enter the GDP program more than once. If desired, a maximum number of delays can be specified in the model, however this has not been evaluated yet.

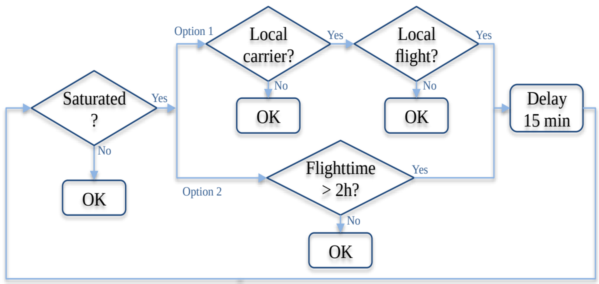


Figure 10: Simulation Logic for Implementation of the Ground Delay Program

Figure 11 displays the simulated total percentage of delayed flights, both including GDP and flight delays, as a function of limiting arrival capacity. It can be observed that for triggering airport arrival rates above 34, the total percentage of delayed flights is maintained almost constant for both scenarios; its value is around 26.7%, which is the same value as the percentage of flight delays. In other words, in this case the implementation of a GDP does not have a significant influence on the percentage of total delayed flights. However, for small limiting arrival rates, such as 25 arrivals per hour, this percentage increases to 40% for scenario 2 and up to 44.5% for scenario 1. These high percentages are due to the long periods were aircraft are obliged to remain on the ground.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

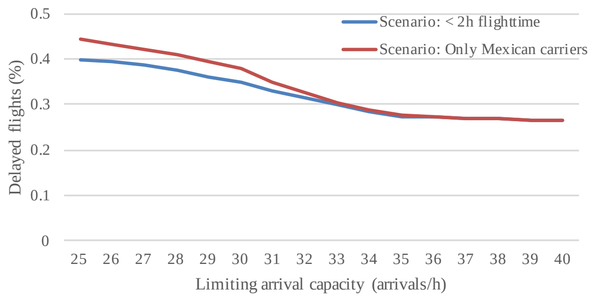


Figure 11: Average Total Delay (Sum of GDP and Flight Delay) as a Function of Limiting Arrival Capacity

Figures 12 and 13 show, respectively, the simulated average GDP delay imposed per aircraft (in minutes) and number of aircraft affected by the ground delay program in terms of limiting arrival capacity, for both the scenario where only Mexican carriers are included in the GDP and the one where flights are included depending on whether their flighttime is less than 2 hours. They indicate, respectively, that the average delay imposed on an aircraft due to the implementation of the GDP varies from around 70 minutes when the GDP is activated at 25 arrivals per hour, to around 20 minutes at 40 arrivals per hour. The number of affected aircraft increases dramatically when the GDP is triggered at lower arrival acceptance rates. The proposed value of 34 arrivals per hour as the limiting airport capacity would affect 200 to 300 aircraft in one week, depending on which scenario is chosen.

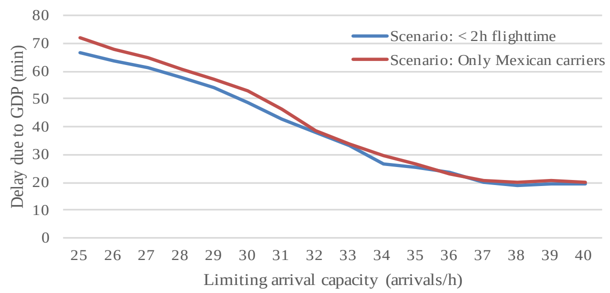


Figure 12: Observed Average GDP Delay per Aircraft (min) as a Function of Limiting Arrival Capacity

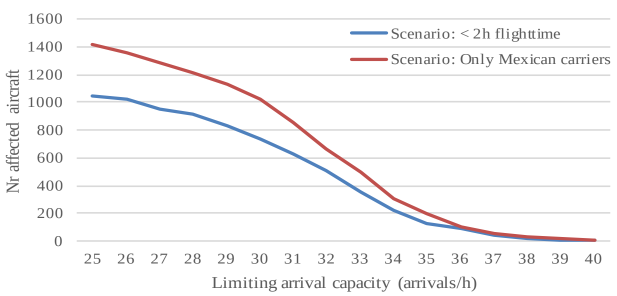


Figure 13: Observed Number of Aircraft Affected by the Ground Delay Program

Based on the previous information, preliminary simulation results suggest the following:

- A GDP at MEX airport seems to work best for acute congestion problems. Since not all aircraft are included in the program, under conditions of severe and chronic congestion, aircraft continue to arrive despite the GDP, which can increase total delay unacceptably for affected flights. In this case, cancellation of flights in combination with the GDP could be an option.
- As aircraft arrive on average about 10 minutes early, delaying an aircraft in a GDP by a 15 minute-period might not be effective, as this delay can be recovered during flight time.
- If the GDP is implemented with smaller limiting airport capacities (e.g. from 30 arrivals per hour), peaks can be observed during the simulation runs in the number of flights arriving per hour. This can indicate a shift in the peak: as more GDP delays are imposed, the saturation seems to decrease in the aimed peak; however, if after some period, for example 1 hour, this apparently lower saturation allows all delayed aircraft to take off, a new saturation peak can be observed in the next two hours, when all these aircraft arrive.
- Starting a GDP at limiting arrival capacities of 37 or more arrivals per hour does not seem to be effective, since almost no aircraft would be included in the program and saturation levels will not be improved.
- If the GDP is activated when 34 arrivals per hour or more are observed, the average percentage of delayed flights (including both GDP delays and flight delays) seems to be maintained almost constant (see figure 11). However, although the number of delayed flights does not increase, the affected flights are delayed by longer time periods. Although not explored at the time, in a future research a fuzzy logic rule can be implemented to select the triggering arrival capacity in the operation of the GDP.
- The scenario in which flights are imposed a GDP delay based on their flight time, seems to be more effective than when only Mexican carriers on flights from Mexican airports are considered. However, the differences between both scenarios decrease with the increase in limiting airport capacity.

6. LIMITATIONS AND FUTURE WORK

The conclusions that were drawn in this study correspond to a first version of our simulation model. As we found some limitations in the available data, our conclusions should be considered as preliminary. Examples of limitations are:

- On line statistics change considerably from month to month. On the other hand, values published by different sources do not necessarily correspond to each other.
- Availability of information in Mexico is more limited than is the US or European aviation

networks, so several parameters had to be estimated. Using a larger amount of data, model performance is expected to increase.

- As the departure time and therefore the use of airport facilities depends to a large extent on the turnaround time of the aircraft when it arrives at MEX, a deeper analysis of turnaround times in terms of carrier, aircraft type, arrival airport and flight route is desirable.
- For the moment, the model does not consider the situation in which carriers do not respect the capacity of the 58 assigned slots for landings and take-offs.

Future work includes a deeper analysis of model input data and an improved representation of its stochastic aspects, especially in relation to different causes of delay, updated flight information and probability distributions for delay and turn-around times per aircraft, airline and destination. The model's accuracy can also be improved by taking into account more detailed information on actual airport operations in MEX, such as specifying the arrival terminal and/or specific contact positions depending on the carrier and the type of aircraft.

7. CONCLUSIONS

Mexico City International Airport is a critical facility for the development of different aspects in the country, ranging from tourism to business. In this paper, we present a discrete-event-based simulation model that we used to analyse the effectiveness of the ground delay program currently imposed by Mexican airport authorities as a measure to address capacity imbalances. Stochasticity of the flight duration, on-time performance and turnaround times are included in the model to analyse how the effectiveness of the ground delay program is influenced by its parameters.

Simulation runs over several scenarios suggest that by activating the ground delay program with 34 arrivals per hour combined with a decision rule on which aircraft to include in the program could decrease its impact on carriers to some extent. In a future study, the stochastic nature of delays and turnaround times will be addressed more deeply, and different types of delay affecting MEX airport congestion will be included in the study.

ACKNOWLEDGMENTS

The authors would like to thank the Aviation Academy of the Amsterdam University of Applied Sciences and DGAPA at the Universidad Nacional Autónoma de Mexico for the support in this research, as well as the Dutch Benelux Simulation Society (www.DutchBSS.org) and EUROSIM for disseminating the results of this work. In addition, we thank Nico De Bock for his technical support.

REFERENCES

Agustín A., Alonso-Ayuso A., Escudero L.F., Pizarro C. (2010). Mathematical optimization models for Air Traffic Flow Management: A review. In: A. Bui, I. Tseveendorkj (Eds.), *Combinatorial Optimization in Practice*, Studia Informatica Universalis, Hermann Informatique 8, 141-184.

Agustín A., Alonso-Ayuso A., Escudero L.F., Pizarro C. (2012) On air traffic flow management with rerouting. Part II: stochastic case, *Eur. J. Oper. Res.* 219, 167–177.

AICM (2013). Press release. Available at: https://www.aicm.com.mx/inicia-nuevo-esquema-de-slots-en-aicm/25-10-2013

AICM (2017). Statistical and slot information, https://www.aicm.com.mx/categoria/estadisticas and https://www.aicm.com.mx/negocios/slots

Airportia (2017). https://www.airportia.com

Andreatta G., Brunetta L., Guastalla G. (1998). The flow management problem: recent computational algorithms, *IFAC Proceedings Volumes* 30 (8), 195-200.

Andreatta G., Brunetta, L. (1998). Multi-airport ground holding problem: A computational evaluation of exact algorithms, *Operations Research* 46 (1), 57-64.

Andreatta G., Dell'Olmo P., Lulli G. (2011). An aggregate stochastic programming model for air traffic flow management. *European Journal of Operational Research* 215, 697-704.

Andreatta G., Romanin-Jacur J. (1987). Aircraft Flow Management under Congestion, *Transportation Science*, 21 (4), 249–253.

Argüello M. F., Bard J. F. (1997). A GRASP for Aircraft Routing in Response to Groundings and Delays. *Journal of Combinatorial Optimization* 5, 211–228.

Ball M., Barnhart C., Dresner M., Hansen M., Neels K., Odoni A., Peterson E., Sherry L., Trani A., Zou B. (2010). Total delay impact study. Final Report. Nextor.

Ball M., Barnhart C., Nemhauser G., Odoni A.R. (2007). Air transportation irregular operations and control. Volume 14 of Handbooks in Operations Research and Management Science. Elsevier.

Ball M., Lulli G. (2004). Ground delay programs: optimizing over the included flight set based on distance. *Air Traffic Control Q*. 12 (1), 1–25.

Bard J.F., Mohan D.N. (2008). Reallocating arrival slots during a ground delay program. *Transp. Res. Part B* 42, 113–134.

Bard J.F., Yu G., Arguëllo M.F. (2001). Optimizing aircraft routings in response to groundings and delays. *IIE Transactions on Operations* Engineering 33 (10), 931–947.

Bilimoria K.D., Sridhar B., Chatterji G.B., Sheth K.S., Grabbe S. (2000). FACET: future ATM concepts evaluation tool. *Air Traffic Control Quarterly* 9 (1), 1–20.

BNAmericas (2014) http://www.bnamericas.com/en/news/privatization/mexico-city-airport-overloaded-last-year-dgac

Brunner J. O. (2014) Rescheduling of flights during ground delay programs with consideration of passenger and crew connections. *Transportation Research Part E* 72, 236–252.

BTS (2017). Bureau of Transportation statistics. On-time performance and flight cancellation statistics for 2016. Available at https://www.transtats.bts.gov/ONTIME/Airline.aspx

Bubalo B. (2011). Airport Punctuality, Congestion and Delay: The Scope for Benchmarking, German Airport Performance Research Project, Working Paper Series GAP.

Campanelli B., Fleurquin P., Arranz A., Etxebarria I., Ciruelos C., Eguíluz V.M., Ramasco J.J. (2016). Comparing the modeling of delay propagation in the US and European

air traffic networks, Journal of Air Transport Management 56 Part A, 12-18.
CIA (2017). US Central Intelligence Agency World Factbook. Available at https://www.cia.gov/library/publications/the-world-factbook/
Chatterji G.B., Sridhar B. (2004). National airspace system delay estimation using weather weighted traffic counts, in *Proc. AIAA Guid., Navig. Contr. Conf.*, San Francisco, CA.
DeArmon J., Lacher A.R. (1996) Aggregate flow directives as a ground delay strategy: Concept analysis using discrete-event simulation, *Air Traffic Control Quarterly* 4 (4), 307-326.
DeArmon J.S., Stalnaker S.E., Katkin R.D., McKinney M. (2008). Benefits analysis of an air traffic flow management capability. Conference paper.
Delgado L., Prats X., Sridhar B. (2013). Cruise speed reduction for ground delay programs: A case study for San Francisco International Airport arrivals. *Transportation Research Part C* 36, 83–96.
Dell'Olmo P., Lulli G. (2003). A dynamic programming approach for the airport capacity allocation problem. *IMA Journal of Management Mathematics* 14, 235–249.
Dorndorf U., Jaehn F., Pesch E. (2016). Flight gate assignment and recovery strategies with stochastic arrival and departure times. OR Spectrum, Springer.
EUROCONTROL (2013). Challenges of Growth 2013. Task 7: European Air Traffic in 2050, European Organisation for the Safety of Air Navigation. 40 p.
FAA (2009). Federal Aviation Administration, ORDER JO 7210.55F: Air Traffic Organization Policy. US Department of Transportation.
Fleurquin P., Ramasco J.J., Eguiluz V.M. (2013). Data-driven Modeling of Systemic Delay Propagation under Severe Meteorological Conditions. *Proceedings of the 10th USA/Europe Air Traffic Management R&D Seminar*. Chicago, USA.
Flightstats (2017). Available at http://www.flightstats.com/company/monthly-performance-reports/airlines/
Flores de la Mota I., Guasch A., Mujica Mota M., Piera M.A. (2017). Robust Modelling and Simulation: Integration of SIMIO with Petri nets. Springer, 1st Ed.
Frolow I., Sinnott J.H. (1989). National airspace system demand and capacity modeling, *Proceedings of the IEEE* 77, 1618-1624.
Glover C.N., Ball M.O. (2013). Stochastic optimization models for ground delay program planning with equity–efficiency tradeoffs. *Transp. Res. Part C* 33, 196–202.
Guest T. (2007): Air traffic delay in Europe. Trends in Air Traffic Vol. 2, Brussels, Belgium, EUROCONTROL.
Inniss T., Ball M.O. (2004). Estimating one-parameter airport arrival capacity distributions for air traffic flow management. *Air Traffic Control Quart.* 12(3), 223–252.
Johnson N.L. (1949). Systems of frequency curves generated by methods of translation, *Biometrika* 36 (1/2), 149-176.
Kafle N., Zou B. (2016). Modeling flight delay propagation: A new analytical-econometric approach. Transportation Research Part B 9, 520-542.
Kolukısa A. (2011). Evaluating Aircraft Turnaround Process in the Framework of Airport Design and Airline Behaviour. *Ph.D. Dissertation*, Universidade do Porto, Portugal.
Leal de Matos P., Ormerod R. (2000). The application of operational research to European air traffic flow management – understanding the context, European Journal of Operational Research 123 (1), 125-144.
Lulli G., Odoni A. (2007). The European air traffic flow management problem. *Transportation Science* 41, 431–443.
Luo S., Yu G., (1997). On the airline schedule perturbation problem caused by the ground delay program. *Transp. Sci.* 31 (4), 298–311.
MukherjeeA., Hansen M. (2007). A dynamic stochastic model for the single airport ground holding problem. *Transp. Sci.* 41 (4), 444–456.
Nilim A., El Ghaoui L. (2004), Algorithms for Air Traffic Flow Management under Stochastic Environments, In: Proc. of the American Control Conference, Boston.
OAG (2017). Available at https://www.oag.com
Odoni A. R. (1987). The Flow Management Problem in Air Traffic Control, in *Flow Control of Congested Networks* (Odoni A. R., Bianco L. and Szego G., editors), Springer-Verlag, Berlin, 269–288.
Pérez-Rodríguez J.V., Pérez–Sánchez J.M., Gómez–Déniz E. (2017). Modelling the asymmetric probabilistic delay of aircraft arrival, *Journal of Air Transport Management* 62, 90-98.
Richetta O. (1995), Optimal algorithms and a remarkably efficient heuristic for the ground-holding problem in air traffic control, *Operations Research* 43, 758-770.
Richetta O., Odoni A.R. (1993). Solving optimally the static ground-holding policy in air traffic control. *Transport. Sci.* 27 (3), 228–238.
SCT (2017a). Traffic statistics by region/Statistics by airport and Directorio de aeropuertos y Catálogo de Aerodromos y helipuertos. Retrieved from http://www.sct.gob.mx/transporte-y-medicina-preventiva/aeronautica-civil/
SCT (2017b). Aviación Mexicana en cifras (1991-2016), Secretaria de Comunicaciones y Transportes, Dirección General de Aeronáutica Civil.
SEGOB (2014). DECLARATORIA de saturación en el campo aéreo del Aeropuerto Internacional de la Ciudad de México. Diario Oficial de la Federación. 2014-09-29.
SENEAM (2015). Plano de estacionamiento y atraque de aeronaves del aeropuerto internacional Benito Juárez. AIP de México SCT-DGAC-SENEAM. 25-jul-2015 AMDT AIRAC 08/15.
SESAR (2012). D23.2 ComplexWorld Position Paper. SESAR WP-E ComplexWorld Research Network. Available at http://www.complexworld.eu
SIMIO (2017). www.simio.com
Swaroop P., Zou B., Ball M., Hansen M. (2012). Do more U.S. airports need slot controls? a welfare based approach to determine slot levels. Transp. Res. Part B: Methodological 46 (9), 1239-1259.
Terrab M. (1990). Ground holding strategies for air traffic control. PhD dissertation, Civil Engineering, MIT.
Terrab M., Odoni A. R. (1993) Strategic flow management for air traffic control. *Oper. Res.* 41, 138–152.
Vasquez-Marquez A. (1991), American airlines arrival slot allocation system (ASAS), *Interfaces* 21(1), 2–61.
Vranas P.B., Bertsimas D.J., Odoni A.R. (1994). The multi-airport ground-holding problem in air traffic control. *Oper. Res.* 42 (2), 249–261.
Winer D.E. (1993). Simulation and optimization in flow planning and management, in: *Large- Scale Computation in Air Traffic Control*, L. Bianco, A.R. Odoni (Eds.), Springer, Berlin.
Zhang X., Zhou Y., Liu B., Wang. Z. (2007). The air traffic flow management with dynamic capacity and co-evolutionary genetic algorithm. *IEEE Transactions on Intelligent Transportation System Conference*.

STIMULATION MODEL – A MULTILAYER SOCIAL NETWORK GENERATOR

Alexander Tarvid

University of Latvia

atarvid@inbox.lv

ABSTRACT
The paper presents the first two stages of building the Stimulation model. The model is aimed at generating a multilayer social network with four layers of size identified in a recent review of empirical studies. The model introduces narrow environments, where all members know everyone, thus, forming cliques in the weighted directed graph representing the social network. The weight of an arc connecting agent v to agent w depends on w's position in the unit square of v and w's layer in v's social network. The unit square is defined by the relative frequency of interactions of v with w and closeness of both agents in quality. At the second stage, stable distribution of layer size and relative time spent interacting with layers is achieved.

Keywords: social network, network generator, multi-layer network

1. INTRODUCTION

A review of empirical studies on the composition of social networks (Zhou et al 2005) identified a stable hierarchical multilayer structure. Four layers were found: support clique (3–5 vertices), sympathy level (9–15 vertices), band level (30–50 vertices) and community level (around 150 vertices). The frequency of contact decreases with hierarchy level, with the support clique being a source of personal advice or help in severe circumstances. To the best of my knowledge, no paper has attempted to re-create this empirical fact yet, and this paper is the first attempt to do it.

Formally, a *multilayer network* is a family of graphs, one graph per layer, with intra-layer edges connecting vertices of the same graph and inter-layer edges connecting vertices in different graphs (hence, layers) (Boccaletti et al 2014). A type of multilayer network is a *multiplex network*, where all graphs from the family contain the same set of vertices (Boccaletti et al 2014). Multilayer networks can, thus, represent both different groups of vertices (the membership of a *vertex* in a particular layer shows its membership in a particular group) and different types of connections between the same two vertices (a membership of an *edge* in a particular layer shows that the connection belongs to a particular type).

The model presented here generates a special type of a multiplex network, where, assuming that vertex v is considered the same vertex irrespective of the layer it appears in, any two vertices may be connected by not more than one arc with same direction (although it can also be viewed as a multilayer network with different vertices on each layer forming cliques). In other words, arc (v, w) can appear in not more than one layer of the network. Referring again to Zhou et al (2016), however, it is easy to see that a layer where a particular vertex lies is not its general position, but it is its subjective position from the point of view of another given vertex. For instance, person c may be very important for person a but not for person b. To complicate it further, person a may actually not be that important for person c. In other words, none of the two representations of a multilayer network mentioned above are applicable in this case.

The literature on multilayer/-level networks has been actively emerging in the last years. For instance, Cantor et al (2015) and Senior et al (2016) developed agent-based models of evolving animal multilayer networks. A multi-layer model of risk in financial markets is studied by Poledna et al (2015). Multi-layer social networks were used in modelling organized crime (Li et al 2015), diffusive processes (e.g., of innovations) (Li, Yan and Jiang 2015; Ramezanian et al 2015) and collaborative learning behaviour in organisations (Różewski et al 2015).

The rest of the paper is structured as follows. The next section introduces some conceptual foundations. Then Sects. 3 and 4 present the first and second stages of the model, respectively. The first stage introduces the main mechanisms but is limited to allow only growth of higher layers in terms of size. At the second stage, a balancing mechanism is introduced to limit the size of each layer around the empirically observed values. The last section concludes.

2. CONCEPTUAL FOUNDATIONS

Assume a population of agents, some of which are connected. Connections are directed, so the network is represented by a digraph. An arc's weight is one of the factors that determines the layer in which the arc exists in an agent's digraph. The higher the weight, the stronger the connection, the higher ('more intimate') the layer.

For the connections to be able to move across layers over time, the mechanism of *stimulation* is introduced, which is also the reason for which the model is called the *Stimulation model*. The stimulation of an arc occurs when its incident vertices (more concretely, the agents represented by these vertices in the social graph) interact. The more frequent the interactions become, the more the arc's weight increases.

This mechanism resembles that of reinforcement learning in that the agent does not decide purely randomly with whom to interact more, but bases this decision on the result of the history of interactions with different agents. More frequent interactions occur with those agents, the interactions with which bring most benefit to the agent.

The ultimate challenge of the model is to simultaneously create a layered structure of a network for every agent and keep the overall structure of the network scale-free (i.e., the distribution of the degrees of vertices should be governed by power law). The exponent of the power law should ideally be one of the parameters of the model, and it should be in the range typical for social networks. In this paper, however, the network is fixed and, by construction, does not have a scale-free structure. That structure should be attained at later stages, when more dynamics are introduced into the model.

The model is built around the following theoretical concepts. Firstly, *homophily* (McPherson, Smith-Lovin and Cook 2001) argues that personal networks are quite homogeneous on a large number of parameters characterising individuals. In other words, individuals tend to connect to similar individuals more frequently than to dissimilar individuals. Secondly, *structural constraint* (Fischer 1982) means that the pool of potential members of a social network is to a large extent determined by the social contexts where the individual participates. These social contexts will appear as narrow environments, where everyone knows everyone.

Besides Zhou et al (2016), the paper was also influenced by the empirical findings of Grossetti (2005) on the initial meeting contexts of network members and their current characterisations. It was shown that the majority of connections was initially met through family or existing friends, but also in other environments, such as at work or in organisations, during education and in the neighbourhood, while only 6% was met by chance. Most of these contexts will be abstracted in the model as narrow environments. It was also shown that the current characterisations typically differed from the initial context, allowing to group contacts into family, friends, work/organisations, neighbours and acquaintances. This (together with the results of Zhou et al (2016)) will be abstracted as moving up and down through the layers of the network.

3. STAGE 1: A GROWTH-ONLY MODEL

A fixed number of N agents, which will be called persons, reside in a system with M narrow environments. Each person is characterized by quality and a list of narrow environments it is a member of.

At this stage, the number of narrow environments is fixed and the membership of a person in narrow environments is immutable. Upon entering the model, a person is assigned to $\mathrm{Pois}(\lambda_e) + 1$ narrow environments.

The quality of person i is a fixed number $q_i \in [0, 1]$ generated from normal distribution $\mathcal{N}(\mu, \sigma)$. It will determine the overall similarity between two given persons.

The social network is represented by a weighted digraph $G = (V, A, w)$, where V is the set of vertices, A is the set of arcs and $w: A \to \mathbb{R}$ is a function mapping arcs to their weights.

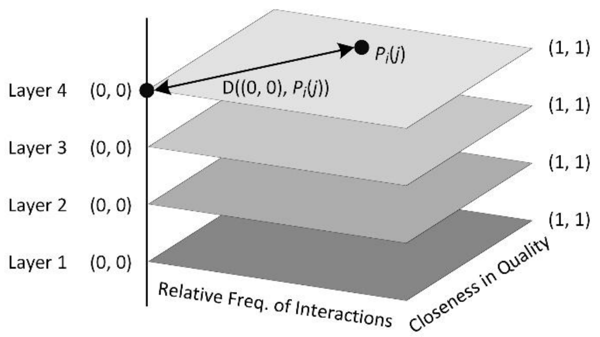


Figure 1: A Stack of Unit Squares Illustrating the Three Components of Arc's (i, j) Weight: Relative Frequency of Interactions, Closeness in Quality and Layer

Note: $P_i(j)$ is the position of j in i's unit square and $D(\cdot,\cdot)$ is the distance between two points.

A *narrow environment* represents a narrow community where everyone knows everyone, which is a typical situation in a small business, a department of a larger business or a group of students in a university. This is also why the environment is called *narrow* – it cannot represent the whole university or a big multinational corporation, where the assumption that everyone knows everyone does not hold. Thus, if a person becomes a member of a narrow environment, it connects with *all other* members of that narrow environment. At this stage, persons cannot leave a narrow environment.

Let persons i and j be connected. Denote the arc connecting i to j as (i, j) and its weight by w_{ij}. The arc's weight represents the importance of the vertex's neighbour to the vertex – in this case, the importance of j to i. It depends on three factors. The first is the closeness of both persons in their quality, $1 - |q_i - q_j|$, which is fixed over time, as quality is immutable.

The second factor is the relative frequency of interactions of person i with person j in the last period of time, I_{ij}/I_i, where $\sum_{j:(i,j)\in A} I_{ij} = I_i$. Thus, the closer the quality and the more time spent interacting, the heavier should be the connection. Because the closeness in quality and relative frequency of interactions are both in $[0, 1]$, the position of j for i on these two factors can be illustrated by a position in the unit square, see Fig. 1: the closer the point to the top right corner $(1, 1)$, the heavier the arc should be.

The third factor is the layer where j resides in i's network. From empirical analysis, it is known that there are four layers. Hence, this is also the number of layers in the model. As shown in Fig. 1, each layer is associated with its own unit square, so that e.g. a point (x, y) in layer 1 will have a substantially lower weight than the point with the same coordinates in layer 4. The weight of the arc (i, j) is given by the following expression:

$$w_{ij} = 10^{L_i(j)-1} + D\big((0,0), P_i(j)\big),$$

where $L_i(j)$ is layer where j is located in i's network, numbered from 1 to 4, $D(\cdot,\cdot)$ is the distance between two points and $P_i(j)$ is the point representing j in i's unit square, see Fig. 1 for illustration.

```
Add N persons
Add each person to a random number of narrow
environments  ~(Pois(λ_e) + 1)
for t = 1 to T do
  Prepare for interactions
  I ← random number from Pois(λ)
  for i = 1 to I do
    p ← random person with Pr(choose p) ∝ w_pq
    p interacts with random connection q: Pr(choose q) ∝ w_pq
  end for
  Update arc weights
  for all Persons p do
    for all Layers l do
      Promote q: L_p(q) = l ∧ w_pq = max_{v:∃(p,v)∈A} w_pv to next layer with
prob. π_l
    end for
  end for
end for
```

Listing 1: General Algorithm of the Model

The algorithm of the model is shown in Listing 1. It has several particularities. Firstly, every period has a fixed number of interactions I drawn from Poisson distribution with mean λ. During that period, exactly I interactions happen, but although during the model calibration stage, λ will be set with a certain number of interactions per person in mind, the set-up does not guarantee that every person will perform interactions in that period. In fact, persons with higher out-degree (i.e. more connections) will have higher probability to perform any particular interaction. Once the person is chosen, it interacts with someone from its connections, again chosen with probability proportional to the arc's weight. With that, more connected persons interact more often than less connected persons and a person interacts more often with more important connections than with less important ones.

Secondly, arc weights are updated after every period based on the interactions data. Recall that quality does not change, so the only factor that can change the arc's weight at this point is a change in the relative frequency of interaction. Note that it means that as a result of a period, the arc's weight can decrease.

Thirdly, the increase will not be dramatic for two reasons. The first is that connections cannot be moved to lower layers, whatever is the change in the arc's weight. The second is that the higher the weight of the arc, the higher the probability that an interaction happens and, hence, that the relative proportion of interactions is higher and, thus, the arc's weight is around the previous period's value or higher than it.

Fourthly, in the end of every period, every person takes its best-performing connection (the one with the maximum weight) in every layer and promotes it probabilistically. The probability of promotion decreases with layer, so that it is much more difficult to be promoted from layer 3 to layer 4 than from layer 1 to layer 2.

Table 1: Parameter Values

Parameter Name	Notation	Value
Length of simulation, periods	T	100
Number of persons	N	1000
Number of narrow environments	M	35
Mean narrow environments of a person	λ_e	1.5
Quality distribution parameters	(μ, σ)	(0.5, 0.3)
Weight of a new arc		1.0
Mean global number of interactions	λ	10^5
$\Pr(\text{promotion}\mid\text{layer} = 1)$	π_1	0.60
$\Pr(\text{promotion}\mid\text{layer} = 2)$	π_2	0.20
$\Pr(\text{promotion}\mid\text{layer} = 3)$	π_3	0.05

Table 2: Out-Degree Statistics by Number of Narrow Environments (M)

M	Median Out-Degree	
	Mean	Range
20	249	$[233, 264]$
25	203	$[190, 218]$
30	172	$[161, 184]$
35	149	$[139, 161]$
40	132	$[123, 141]$
45	118	$[111, 129]$
50	107	$[98, 116]$

Note: Statistics over 100 runs reported.

The model was implemented in Repast Simphony. The values of model parameters are shown in Table 1. The parameters were set so that there are on average 100 interactions per person in every period. Every person is a member of 2.5 $(= \lambda_e + 1)$ narrow environments on average, so that most of them participate in two or three narrow environments, which is close to reality. The number of narrow environments, given that the number of persons is 1000, is set so that the median out-degree is around 150 (Dunbar 1993), as determined empirically by running the graph-construction part of the model (see Table 2). The probabilities of promotion to the next layer, given the current layer, $\pi_1, \ldots, \pi_3$, are set to values where by $t = 100$, the layered structure in terms of layer sizes on average reproduces the values reported in Zhou et al (2016).

After conducting 30 runs, the average sizes of the four layers at t = 100 are close to the target sizes reported in Zhou et al (2016), see Fig. 2.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

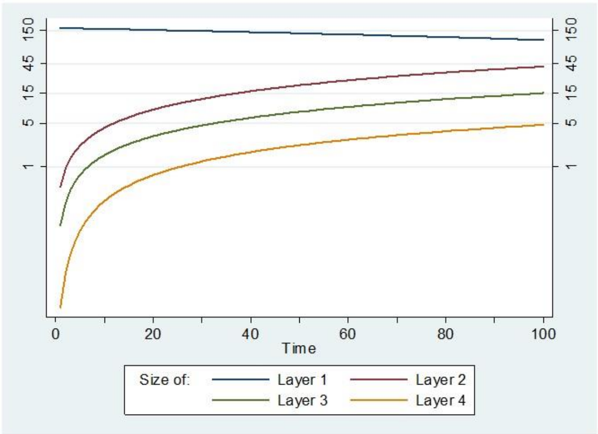


Figure 2: Size Dynamics of Layers over Time, Stage 1

4. STAGE 2: A BALANCING MECHANISM

In the previous stage, the model operated in the unrestricted layer size growth mode, which is an obvious disadvantage. After all, while the parameters were set to values where the expected average sizes of layers are attained at $t = 100$, the lack of a balancing force means that at large values of t (and with no changes in the membership in narrow environments), all connections will be in the highest layer of every person. In the second stage, a mechanism restricting the size of every layer starting from the second layer is introduced (the first layer cannot be restricted, as it depends on the number and sizes of narrow environments with which a person is affiliated).

There are several ways to introduce layer size balancing. In this model, the mechanism of attractors is applied. Experience-weighted attraction is used by Gemkow and Neugart (2011) to learn the best number of connections, but it is a machine-learning algorithm allowing to choose from several existing strategies, while in this model, attractors are built in the person. In summary, every person has a random attractor for every layer (starting from layer 2). The attractor is set to a value that keeps the respective layer size in the needed boundaries. After a new promotion to a layer, if that action resulted in a substantial upward deviation from the attractor of that layer, the person tries to demote the least worthy persons from the layer to the previous layer.

Attractors are a mathematical formalism applied in the analysis of dynamical systems. The attractors used in this model belong to a variety of attractors called *fixed point attractors*, because in every layer of a person's social network, there is exactly one point where the size of the network's layer will finally rest. Any deviation from that point leads to a pressure to change the network to return to it. See Vallacher and Nowak (2007) for details and examples of applying attractors in social psychology.

In detail, assume that person j is promoted to layer l in person's i social network. Denote the quality of correspondence between i and j as q_{ij} and i's attractor of layer l as $\alpha_i(l)$. Person i tries to keep the sum of quality correspondence with each connection at layer l, $Q_i(l)$, above but maximally close to $\alpha_i(l)$. Thus, when j is promoted to l, if $Q_i(l) > \alpha_i(l)$, person i moves all persons k from layer l with the lowest q_{ik} to layer $(l - 1)$ until removing any other person from that layer would move $Q_i(l)$ below $\alpha_i(l)$. Listing 2 formalises these actions.

$Q_i(l) \leftarrow \sum_{k:L_i(k)=l} q_{ik}$
$\Delta_\alpha \leftarrow Q_i(l) - \alpha_i(l)$
flag ← **false**
while $\Delta_\alpha > 0 \wedge$ **not** flag **do**
  $q \leftarrow \min_{k:L_i(k)=l} q_{ik}$
  $p \leftarrow k: L_i(k) = l \wedge q_{ik} = q$
  if $\Delta_\alpha - q > 0$ **then**
    $\Delta_\alpha \leftarrow \Delta_\alpha - q$
    Demote p to layer $L_i(p) - 1$
  else
    flag ← **true**
  end if
end while

Listing 2: Application of Attractors After Promoting j in i's Network

Assume for example that a certain layer of i's network contains persons with quality correspondence of 0.4, 0.5 and 1.0, which sum to 1.9, and that a person with quality correspondence of 0.8 is promoted to this layer. If the attractor is $\alpha_i(l) > 2.7$, no demotion occurs. If it is $2.3 < \alpha_i(l) \leq 2.7$, again no demotion occurs, because the worst correspondence is 0.4 but removing it from this layer would result in $Q_i(l)$ falling below $\alpha_i(l) > 2.3$. Finally, if $\alpha_i(l) \leq 2.3$, e.g., it is $\alpha_i(l) = 2.0$, the person with quality correspondence of 0.4 is moved to a lower layer, resulting in $Q_i(l) = 2.7 - 0.4 = 2.3$. Because the worst correspondence is now 0.5, and removing it from this layer would result in $Q_i(l) = 1.8 < 2.0 = \alpha_i(l)$, the demotion procedure stops.

For every person, attractors are chosen from uniform distributions: $[3, 5]$ for $\alpha(4)$, $[9, 15]$ for $\alpha(3)$ and $[30, 45]$ for $\alpha(2)$. These distribution boundaries were set according to empirical data in Zhou et al (2016).

The use of attractors means that a person has a built-in need for a certain overall quality of interactions at each level of its network. Overall quality in this case is approximated by the sum of closeness in quality of the person with its connections on a given level.

Too low overall quality of interactions is considered unsatisfactory, but the person is not ready to attain the attractor of a given layer immediately (by promoting a sufficient number of connections from lower layers). That means that persons are in a sense risk-averse, reluctant to quickly let others have close relationships with them.

Too high overall quality of interactions is also considered unsatisfactory, because it puts too much pressure on the person. It is, thus, assumed that interactions not only give satisfaction, but also require certain effort. Above the attractor, the person feels discomfort and becomes willing to decrease the overall quality of interactions.

At the second stage, the model was run 30 times with the same parameter values shown in Table 1 except for T, which was increased to 200 to check how fast layer size stabilisation occurs.

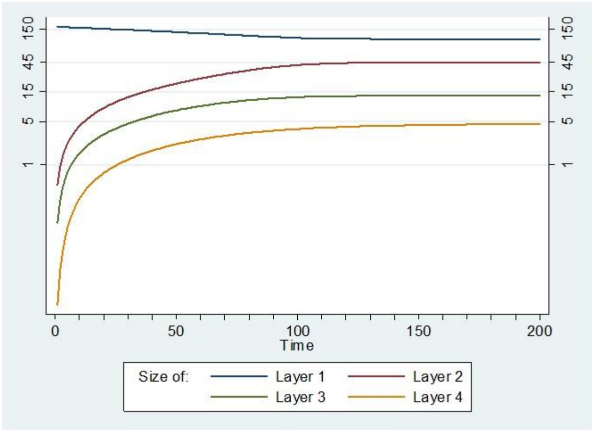


Figure 3: Size Dynamics of Layers over Time, Stage 2

As shown in Fig. 3, the use of attractors allows to stabilise the sizes of layers by period 150. By construction, the stable size levels mimic the empirically observed results.

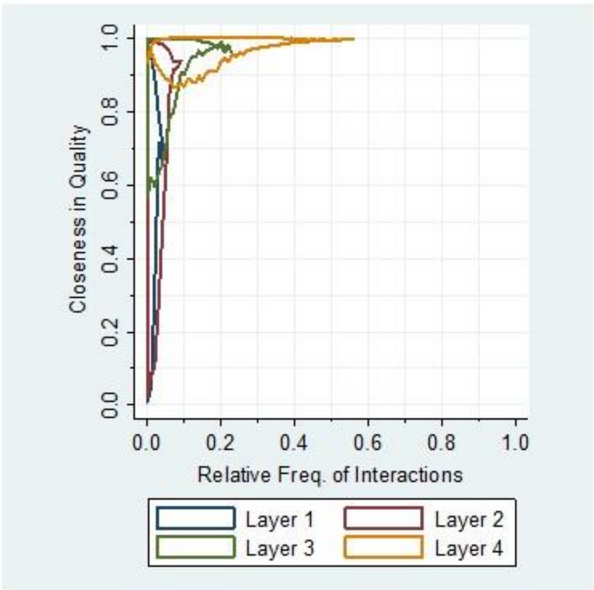


Figure 4: Distribution of Connections in Unit Squares by Layer at $t = 200$

Fig. 4 shows the regions of the unit squares where connections from each of the four layers are located. It is generally visible that, again by construction, more time is spent interacting with those connections which are closer in quality to the person. However, at higher layers, increasingly more time is spent with those with closest quality. In fact, it is clearly shown that the closeness in quality is not very important to be selected for promotion to the second layer. At the same time, no agent has on average lower quality correspondence than around 0.6 in the third layer and around 0.9 in the fourth layer. Neither closeness in quality nor membership in higher layers guarantees high relative frequency of interactions, as shown by all four layers intersecting around the top left corner of the unit square.

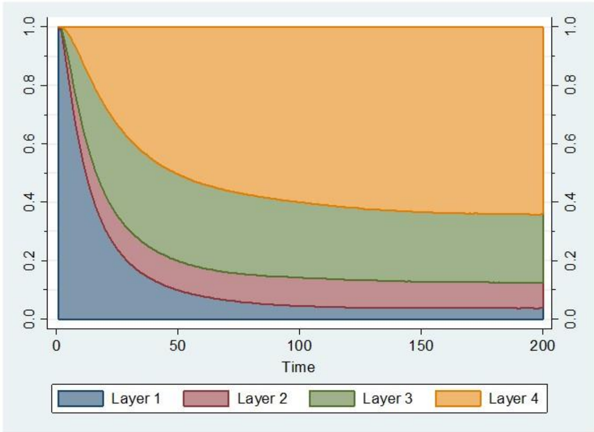


Figure 5: Average Time Share Spent on Interactions by Layer

The relative amount of time spent interacting with the four layers also stabilises by period 150, see Fig. 5. By the end of the simulation, more than 60% of time is spent on interactions with the highest layer, around 20% with the third layer, around 10% with the second layer, while only less than 5% of time is spent with the first layer.

5. DISCUSSION AND CONCLUSIONS

The paper presented only the first two stages of building the Stimulation model, but it already allows to reproduce sizes of four connection layers: support clique (3–5 vertices), sympathy level (9–15 vertices), band level (30–50 vertices) and community level (around 150 vertices) (Zhou et al 2016).

However, the contribution of this model was not only in reproducing that structure, but also in moving the focus from the 'skeleton' of the network – whether or not two given vertices are connected – to its 'meat' – the dynamics of the weights of arcs and the promotion of vertices to higher levels in the opinion of their connected vertices. The 'skeleton' part is focused on narrow environments, where everyone knows everyone, instead of purely random connections, as modelled usually. The 'meat' part is focused on the correspondence in quality and relative frequency of interactions as determinants of the arc's weight and becoming a candidate for a promotion to a higher layer.

Further work on this model should start with allowing persons to probabilistically enter and exit narrow environments instead of assigning them to a (randomly) defined number of narrow environments. It is expected that this change will allow to generate a vertex degree distribution that follows power law. As a result, the model will allow to generate a multilayer scale-free social network.

REFERENCES

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., Zanin, M., 2014. The structure and dynamics of multilayer networks. Physics Reports 544(1), 1–122

Cantor, M., Shoemaker, L.G., Cabral, R.B., Flores, C.O., Varga, M., Whitehead, H., 2015. Multilevel animal societies can emerge from cultural transmission. Nature Communications 6(8091)

Dunbar, R.I., 1993. Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences 16(4), 681–694

Fischer, C.S., 1982. To dwell among friends: Personal networks in town and city. The University of Chicago Press, Chicago

Gemkow, S., Neugart, M., 2011. Referral hiring, endogenous social networks, and inequality: an agent-based analysis. Journal of Evolutionary Economics 21(4), 703–719

Grossetti, M., 2005. Where do social relations come from?: A study of personal networks in the Toulouse area of France. Social Networks 27(4), 289–300

Li, B., Sun, D., Li, Z., Ou, C., Zhu, R., 2015. Understanding the dynamic of covert organization with multilayer networks model. In: European Intelligence and Security Informatics Conference (EISIC) 2015. pp. 37–44

Li, Z., Yan, F., Jiang, Y., 2015. Cross-layers cascade in multiplex networks. Autonomous Agents and Multi-Agent Systems 29(6), 1186–1215

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annual Review of Sociology 27, 415–444

Poledna, S., Molina-Borboa, J.L., Mart´ınez-Jaramillo, S., van der Leij, M., Thurner, S., 2015. The multi-layer network nature of systemic risk and its implications for the costs of financial crises. Journal of Financial Stability 20, 70–81

Ramezanian, R., Magnani, M., Salehi, M., Montesi, D., 2015. Diffusion of innovations over multiplex social networks. In: International Symposium on Artificial Intelligence and Signal Processing (AISP) 2015. pp. 300–304

Różewski, P., Jankowski, J., Bródka, P., Michalski, R., 2015. Knowledge workers' collaborative learning behavior modeling in an organizational social network. Computers in Human Behavior 51(B), 1248–1260

Senior, A.M., Lihoreau, M., Buhl, J., Raubenheimer, D., Simpson, S.J., 2016. Social network analyses and nutritional behavior: An integrated modeling approach. Frontiers in Psychology 7(18)

Vallacher, R.R., Nowak, A., 2007. Dynamical social psychology: Finding order in the flow of human experience. In: Social Psychology: Handbook of Basic Principles, pp. 734–758. The Guilford Press, New York

Zhou, W.X., Sornette, D., Hill, R.A., Dunbar, R.I.M., 2005. Discrete hierarchical organization of social group sizes. Proceedings of the Royal Society of London B: Biological Sciences 272(1561), 439–444

COMPACT INTEGRATED-OPTICAL AMPLIFIERS FOR THE TELECOMMUNICATION C - BAND ON THE BASIS OF POLYMER WAVEGUIDES WITH EMBEDDED $NALUF_4/YB^{3+}/ER^{3+}/CE^{3+}$ NANOCRYSTALS

I.M. Asharchuk[(a)], A.S. Akhmanov[(b)], I.O. Goriachuk[(c)], M.M. Nazarov[(d)], A.V. Nechaev[(e)], K.V. Khaydukov[(f)], V.I. Sokolov[(g)]

(a), (b), (c), (f), (g) Institute on Photonic Technologies, Federal Research Center «Crystallography and Photonics», Russian Academy of Sciences, Moscow, Russia
(d) National Research Center «Kurchatov Institute», Moscow, Russia
(e) Moscow State University of Fine Chemical Technologies, Moscow, Russia
(b), (g) Federal Research Center «Scientific Research Institute of System Developments», Russian Academy of Sciences, Moscow, Russia
(a) Ilya-asharchuk@ya.ru, (b) Asakhmanov@mail.ru, (c) Io.gorjachuk@physics.msu.ru, (d) Nazarovmax@mail.ru, (e) Chemorg@mail.ru, (f)Khaydukov@mail.ru, (g)Visokol@rambler.ru

ABSTRACT

Nano-sized β-$NaLuF_4/Yb^{3+}/Er^{3+}/Ce^{3+}$ core-shell crystals possessing intense photoluminescence in the 1530 – 1565 nm telecommunication C - band under 980 nm laser excitation were prepared by thermal decomposition approach. The synthesized nanoparticles (NPs) were dispersed in SU-8 photo-resist and single-mode polymer channel waveguides with embedded NPs were fabricated on the thermally oxidized silicon wafer using UV photolithography and wet etching techniques.

The relative optical gain of 2.5 dB was demonstrated at λ = 1549.2 nm wavelength in the 12 mm long waveguide, which means 2.1 dB/cm per unit gain. These results reveal that composite polymer material on the basis of SU-8 photo-resist and β-$NaLuF_4/Yb^{3+}/Er^{3+}/Ce^{3+}$ NPs is promising for fabricating compact integrated-optical amplifiers for the C – band.

Keywords: rare-earth doped nanocrystals, photo-luminescence, UV photolithography, polymer waveguides, optical amplifiers, telecommunication C – band.

1. INTRODUCTION

Erbium doped polymer waveguide amplifiers are attracting considerable attention due to their intense $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transitions of Er^{3+} ions, which match well to the low-loss telecommunication window of optical fibers in the 1530 – 1560 C - band (Wang 2015; Zhai, Li, Liu, Wang, Zhang, Qin, Qin 2013; Zhai, Liu, Liu, Wang, Zhang, Qin, Qin 2013; Zhao 2014). For real applications in the field of integrated optics the Er^{3+} doped NPs should possess small crystalline size, good dispersibility in the polymer matrix as well as highly efficient photoluminescence (PL) in the C - band.

In this work Er^{3+}, Yb^{3+}, Ce^{3+} co-doped β-$NaLuF_4$ NPs with the core/shell structure were synthesized and used for fabricating single-mode waveguide optical amplifiers with SU-8 polymer material.

2. SYNTHESIS OF β-$NALUF_4/YB^{3+}/ER^{3+}/CE^{3+}$ NANOCRYSTALS

Nano-sized hexagonal β-$NaLuF_4/Yb^{3+}/Er^{3+}/Ce^{3+}$ crystals with the core-shell structure (active β-$NaLuF_4$ core doped with Yb, Er, Ce and inert undoped β-$NaLuF_4$ shell) were prepared by thermal decomposition approach, which was developed in our research group earlier (Grebenik 2013). In Figures 1, 2 one can see TEM and high-resolution TEM photographs of synthesized NPs with the diameter in the range from 20 to 40 nm. Nanoparticles are covered with oleic acid thin film, which results in good dispersibility of NPs in the polymer matrixes.

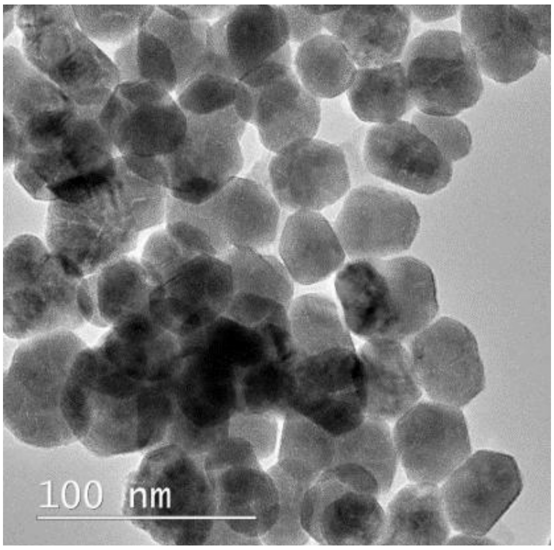


Figure 1. TEM microphotograph of the synthesized β-$NaLuF_4:Yb^{3+}:Er^{3+}:Ce^{3+}$ core/shell nanocrystals.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

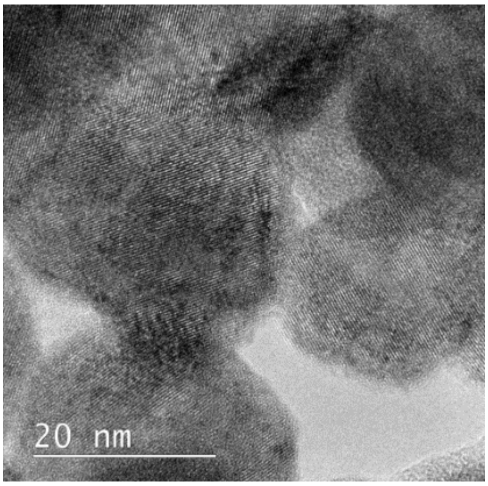


Figure 2. High-resolution TEM microphotographs of the hexagonal β-$NaLuF_4$:Yb^{3+}:Er^{3+}:Ce^{3+} core/shell nanocrystals.

The synthesized β-$NaLuF_4$/Yb^{3+}/Er^{3+}/Ce^{3+} nanocrystals possess intense PL in the 1530 – 1565 nm telecommunication C – band. In Figure 3 the simplified scheme of Yb^{3+}, Er^{3+} and Ce^{3+} energy levels is presented whereas Figure 4 shows PL spectrum of β-$NaLuF_4$/Yb^{3+}/Er^{3+}/Ce^{3+} NPs under 980 nm laser excitation. One can see from Figure 4 that FWHM of the PL spectrum equals 75 nm, which potentially permits to provide optical gain at any wavelength in the C – band.

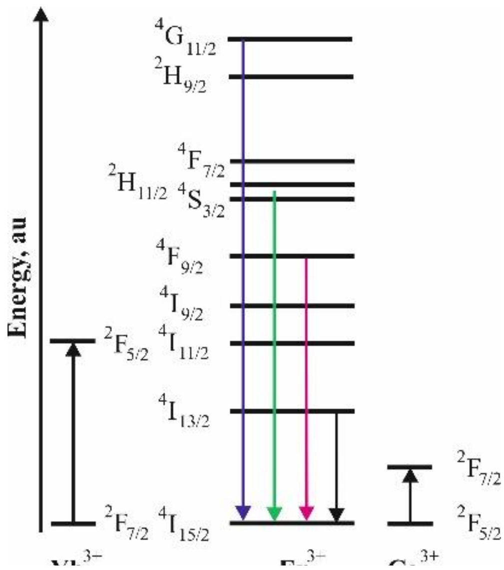


Figure 3. Simplified scheme of Yb^{3+}, Er^{3+} and Ce^{3+} energy levels. The energy transfer between ytterbium and erbium ions is stipulated by transitions $^2F_{5/2}$ (Yb^{3+}) + $^4I_{15/2}$ (Er^{3+}) → $^2F_{7/2}$ (Yb^{3+}) + $^4I_{11/2}$ (Er^{3+}), whereas that between erbium and cerium ions is due to $^4I_{11/2}$ (Er^{3+}) + $^2F_{5/2}$ (Ce^{3+}) → $^4I_{13/2}$ (Er^{3+}) + $^2F_{7/2}$ (Ce^{3+}) transitions.

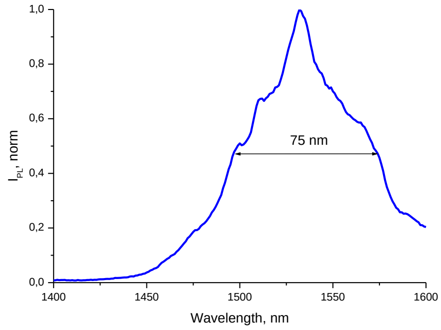


Figure 4. PL spectrum of β-$NaLuF_4$/Yb^{3+}/Er^{3+}/Ce^{3+} nanoparticles under 980 nm laser excitation. The intense photoluminescence band around 1532 nm is stipulated by $^4I_{13/2}$ → $^4I_{15/2}$ transitions in Er^{3+} ions.

3. NUMERICAL SIMULATION OF SINGLE-MODE POLYMER WAVEGUIDES

Single-mode propagation of light in the polymer waveguide is controlled by the geometrical parameters of the light-guiding core (width *w* and height *h*) as well as by its numerical aperture

$$NA = \sqrt{n_{core}^2 - n_{clad}^2} \qquad (1)$$

where n_{core} is refractive index of the core and n_{clad} – that of the cladding. One can see from Equation (1) that *NA* depends upon the refractive indices of polymer materials used for the core and cladding respectively. The typical numerical aperture, which provides single – mode operation of the waveguide is in the range *NA* = 0.1 – 0.16.

The design of the polymer channel waveguide we have chosen is shown in Figure 5.

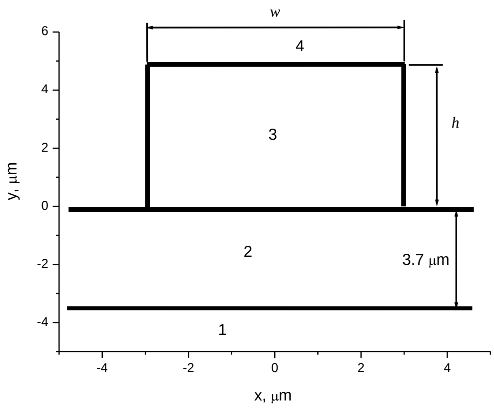


Figure 5. Cross-section of the polymer channel waveguide. 1 - silicon wafer, 2 - thermally grown 3.7 μm SiO_2 buffer layer, 3 - SU-8 rectangular core with embedded β-$NaLuF_4$/Yb^{3+}/Er^{3+}/Ce^{3+} NPs, 4 – amorphous polystyrene cap layer. *w* and *h* are the width and height of the light-guiding core respectively.

Using n_{core} = 1.575 (SU-8 photoresist), n_{SiO2} = 1.444 and n_{clad} = 1.564 (amorphous polystyrene) at 1550 nm we employed the beam propagation method in BeamProp software to calculate the mode field distribution in the rectangular channel waveguide, shown in Figure 5. The numerical simulation has revealed the range of geometrical parameters of the core, which provides single-mode operation of the waveguide in the telecommunication C - band. By taking into consideration the results of the simulation we have chosen the width and the height as w = 8 μm and h = 4 μm to have both single-mode light transmission and reasonably small losses when launching light from the optical fiber into the waveguide and vice versa. The computed transverse fundamental mode profile in the SU-8 polymer waveguide at 1550 nm is presented in Figure 6.

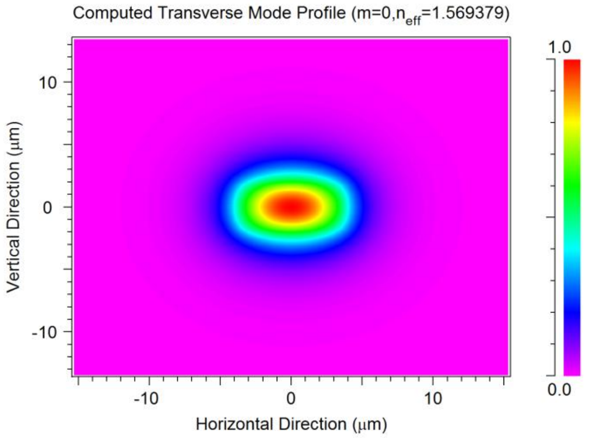
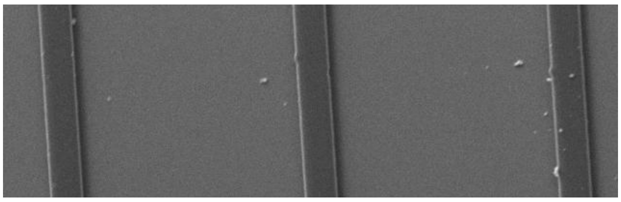


Figure 6. Computed transverse fundamental mode profile in the rectangular channel waveguide (see Figure 5) at 1550 nm. The width and the height of the light-guiding core are w = 8 μm and h = 4 μm respectively. The refractive indices of the core and cladding materials at 1550 nm are given in the text.

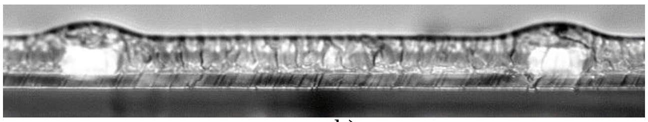
4. FABRICATION OF SINGLE-MODE POLYMER WAVEGUIDES WITH EMBEDDED β-NALUF$_4$/YB^{3+}/ER^{3+}/CE^{3+} NANOCRYSTALS

First, the synthesized β-NaLuF$_4$/Yb^{3+}/Er^{3+}/Ce^{3+} NPs were dispersed in SU-8 photo-resist at concentration 0.2%. Then, by using this nano-composite material the films with the thickness 3 – 5 μm were spin coated on Si wafers with the thermally grown 3.7 μm silicon dioxide layer. The thickness of the films was measured by m - lines technique using Metricon2010/M Prism coupler and original computer program, realized in FORTRAN (Asharchuk 2016). This program permits to calculate refractive index, extinction coefficient and thickness of multilayer thin film structures from the measures reflection coefficients for TE and TM polarized Gaussian light beams.

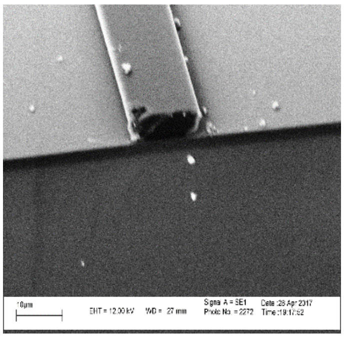
Single-mode polymer waveguides with embedded NPs were fabricated by using standard UV photolithography (365 nm mercury lamp) and wet etching techniques. Finally the waveguides were covered with amorphous polystyrene cap layer. In Figure 7 one can see the photographs of the fabricated array of polymer waveguides, made with the optical (a, b) and scanning electron (c) microscopes. The width of the waveguide cores is 8 μm, the height equals 4 μm. The length of the waveguides in the array equals 12 mm.



a)



b)



c)

Figure 7. Photographs of single-mode polymer waveguides with embedded β-NaLuF$_4$:Yb^{3+}:Er^{3+}:Ce^{3+} nanocrystals before (a, c) and after covering with polystyrene cap layer (b).

5. CHARACTERIZATION OF POLYMER WAVEGUIDE AMPLIFIES

The scheme of experimental setup for measuring optical amplification in the polymer waveguides in the telecom C - band is presented in Figure 8. We used Santec TSL-550 tunable semiconductor laser as signal source and MS9710B optical spectrum analyser as photoreceiver.

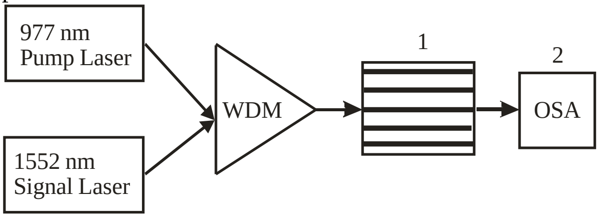


Figure 8. Scheme of experimental setup for measuring optical gain in the polymer waveguide amplifier. 1) Array of single-mode polymer waveguides with embedded β-NaLuF$_4$/Yb^{3+}/Er^{3+}/Ce^{3+} nanoparticles, 2) OSA - Optical Spectrum Analyser. WDM is 1550/980 nm fiber wavelength division multiplexer.

Figure 9 represents the measured optical gain in the polymer waveguide amplifier as a function of 980 nm pump power P_{pump}. The gain was measured by normalizing the output signal power with pump «on» and «off». One can see that the gain increases with P_{pump} and equals 1.6 dB (see the inset to Figure 9) for P_{pump} = 80 mW. Since the length of the waveguide is 12 mm, the per-unit gain of the device equals 1.3 dB/cm.

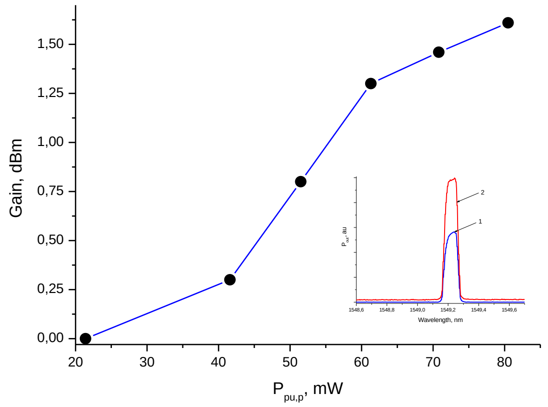


Figure 9. Optical gain in SU-8 polymer waveguide amplifier with embedded β-$NaLuF_4/Yb^{3+}/Er^{3+}/Ce^{3+}$ NPs at λ = 1549.2 nm as a function of 980 nm pump power P_{pump}. The signal power is 0.1 mW. Inset: output signal with pump «off» (1) and «on» (2).

By increasing the concentration of NPs in the polymer matrix up to 0.5% we have got the optical gain 2.5 dB (2.1 dB/cm) at λ = 1549.2 nm.

6. CONCLUSIONS

By using β-$NaLuF_4/Yb^{3+}/Er^{3+}/Ce^{3+}$ NPs and SU-8 photo-resist, single-mode waveguide amplifiers for the telecommunication C - band were fabricated. The relative optical gain of 1.6 - 2.5 dB was demonstrated at 1549.2 nm in the 12 mm long waveguide. These results reveal that composite polymer material on the basis of SU-8 and β-$NaLuF_4/Yb^{3+}/Er^{3+}/Ce^{3+}$ NPs is promising for fabricating compact integrated-optical amplifiers for the C – band.

ACKNOWLEDGMENTS

The work was partially supported by RFBR grants № 14-29-08265, № 16-02-00347, № 16-32-00934 and № 16-32-00935.

REFERENCES

Grebenik E.A., Nadort A., Generalova A.N., Nechaev A.V., Sreenivasan V.K.A., Khaydukov E.V., Semchishen V.A., Popov A.P., Sokolov V.I., Akhmanov A.S., Zubov V.P., Klinov D.V., Panchenko V.Ya., Deyev S.M., Zvyagin A.V., 2013. Feasibility study of the optical imaging of a breast cancer lesion labeled with upconversion nanoparticle biocomplexes. Journal of Biomedical Optics, 18, p. 076004.

Wang T., Zhao D., Zhang M., Yin J., Song W., Jia Z., Wang X., Qin G., Qin W., Wang F., Zhang D., 2015. Optical waveguide amplifiers based on $NaYF_4$: Er^{3+}, Yb^{3+} NPs-PMMA covalent-linking nanocomposites. Optical Materials Express, 5, p. 469.

Zhai X., Li J., Liu S., Liu X., Zhao D., Wang F., Zhang D., Qin G., Qin W., 2013. Enhancement of 1.53 µm emission band in $NaYF_4:Er^{3+}$, Yb^{3+}, Ce^{3+} nanocrystals for polymer-based optical waveguide amplifiers. Optical Materials Express, 3, pp. 270–277.

Zhai X., Liu S., Liu X., Wang F., Zhang D., Qin G., Qin W., 2013. Sub-10 nm $BaYF_5:Yb^{3+}$, Er^{3+} core-shell nanoparticles with intense 1.53 µm fluorescence for polymer-based waveguide amplifiers. Journal of materials chemistry C, 1, pp. 1525 – 1530.

Zhao P., Zhang M., Wang T., Liu X., Zhai X., Qin G., Qin W., Wang F., Zhang D., 2014. Optical amplification at 1525 nm in $BaYF_5$: 20% Yb^{3+}, 2% Er^{3+} nanocrystals doped SU-8 polymer waveguide. Journal of nanomaterials, 2014, Article ID 153028.

Asharchuk I.M., Marusin N.V., Molchanova S.I., Savelyev A.G., Sokolov V.I., Khaydukov E.V., Panchenko V.Ya. A spectroscopic prism coupler for measuring the refractive indices and thicknesses of dielectric films. Bulletin of the Russian Academy of Sciences: Physics. 2016. Vol. 80 (4), p. 413-416.

AUTHORS BIOGRAPHY

Ilya M. Asharchuk graduated from Volgograd state university in 2014. He is research scientist at Institute on Photon Technologies, Federal Research Center «Crystallography and Photonics», Russian Academy of Sciences, Moscow, Russia. The fields of scientific interests: upconversion nanoparticles, polymer composite materials, optical waveguides, waveguide lasers. The author of 8 paper.

Alexander S. Akhmanov was born in Moscow in 1955. He graduated from Physical Department of Moscow State University. He received the Ph.D. degree in Physics and Mathematics and is senior research scientist at Institute on Photon Technologies, Federal Research Center «Crystallography and Photonics», Russian Academy of Sciences, Moscow, Russia. The fields of scientific interests: high-speed telecommunications, polymer integrated optics. The author of 60 scientific paper.

Ivan O. Goryachuk was born in Moscow in 1993. He graduated from Physical Department of Moscow State University in 2017. He is research scientist at Institute on Photon Technologies, Federal Research Center «Crystallography and Photonics», Russian Academy of Sciences, Moscow, Russia. The fields of scientific interests: polymer waveguides, polymer integrated optics, numerical simulation of light transmission in thin film structures. The author of 3 paper.

Maxim M. Nazarov received the Ph.D. degree in Physics and Mathematics from Physical Department of

Moscow State University in 2002 on the topic of surface plasmon enhanced optical harmonics on metal grating. He was scientific researcher at MSU till 2012 on the topics of femtosecond lasers and THz spectroscopy. In 2012 he is a scientific researcher at Institute on Laser and Information Technologies of the Russian Academy of Sciences on the topic of polymer waveguides for THz and visible ranges, nanoparticles and integrated optics. In 2016 he is the head of laboratory of ultraintense laser pulses in National Research Center «Kurchatov Institute», Moscow, Russia. He is the author of more then 60 publications in peer-reviewed scientific journals.

Andrey V. Nechaev graduated from Moscow State University of Fine Chemical Technologies. He received the Ph.D. degree in Chemistry. Now he is Docent at this university. Specialization: synthesis of rare-earth doped nanoparticles. The author of 37 scientific paper.

Kirill V. Khaydukov graduated from Volgograd state university in 2014. He is research scientist at Institute on Photon Technologies, Federal Research Center «Crystallography and Photonics», Russian Academy of Sciences, Moscow, Russia. The fields of scientific interests: rare-earth doped nanoparticles, polymer waveguides, waveguide amplifiers, integrated optics. The author of 10 paper.

Victor I. Sokolov was born in Moscow in 1956. He received Ph.D. degree in Physics and Mathematics in 1991 at Moscow State University. In 1991 - 2016 he was Scientist, Senior Research Scientist, Head of the lab at Institute on Laser and Information Technologies of the Russian Academy of Sciences. Since 2016 he is the Head of the Institute on Photon Technologies, Federal Scientific Research Centre «Crystallography and Photonics». He is the author of more than 100 publications in the field of interaction of laser radiation with matter, photonics, integrated optics, polymeric materials and design of scientific instruments.

COMPARISON OF A MICROSCOPIC DISCRETE-EVENT AND A MESOSCOPIC DISCRETE-RATE SIMULATION MODEL FOR PLANNING A PRODUCTION LINE

Florian Gleye [a], Tobias Reggelin[b], Sebastian Lang[c]

[a]Salzgitter Mannesmann Großrohr GmbH
[b],[c]Otto von Guericke University Magdeburg

[a]f.gleye@gmx.de, [b]tobias.reggelin@ovgu.de, [c]sebastian.lang@ovgu.de

ABSTRACT
This paper compares a microscopic discrete-event simulation model and a mesoscopic discrete-rate simulation model to support the planning of a production line in the automotive industry in terms of simulation results and modeling and simulation effort. Mesoscopic discrete-rate models represent logistics flow processes on an aggregated level through piecewise constant flow rates instead of modeling individual flow objects like in microscopic discrete-event models. This leads to a fast model creation and computation.

Keywords: Mesoscopic simulation models, discrete-rate simulation, discrete-event simulation, production line planning

1. INTRODUCTION

Practitioners often prefer to use microscopic discrete-event simulation models because most production and logistics processes are of discrete nature (Scholz-Reiter et al. 2007) and because discrete-event models allow for a very high level of detail. Discrete-event models are state of the art in production planning and logistics planning in the automotive industry (Huber and Wenzel 2011). The term discrete-event modeling stands "for the modeling approach based on the concept of entities, resources and block charts describing entity flow and resource sharing" (Borshchev and Filippov 2004). Since discrete-event models are able to represent workstations, technical resources, carriers and units of goods as individual objects, they can depict production and logistics systems with a high level of detail and are also referred to as microscopic models (Borshchev and Filippov 2004, Pierreval et al. 2007). Models in this class can be very complicated and slow and their creation and implementation can be time and labor consuming (Pierreval et al. 2007; Law and Kelton 2007; Kosturiak and Gregor 1995; Huber and Dangelmaier 2009; Scholz-Reiter et al. 2008).
Mesoscopic simulation models based on the discrete-rate simulation approach have the potential for reducing modeling and computation efforts and allow for sufficient accuracy for many problems in production and logistics planning.

In cooperation with the MAN Truck & Bus AG at the Salzgitter plant, a microscopic discrete-event simulation model with Tecnomatix Plant Simulation was developed to compare different planning variants for the production line of truck parts. A mesoscopic simulation model with ExtendSim based on the discrete-rate simulation paradigm was implemented for the comparison.

2. MESOSCOPIC SIMULATION MODELS

The mesoscopic simulation approach proposed by the authors of this paper is situated between continuous and discrete-event approaches in terms of level of modeling detail and required modeling and simulation effort (Reggelin 2011, Reggelin and Tolujew 2011). It supports quick and effective execution of analysis and planning tasks related to manufacturing and logistics networks. The principles of mesoscopic simulation models to describe processes in logistics and production networks have been derived from and have been applied to the actual development of several mesoscopic models (Henning et al. 2016, Hennies et al. 2014; Hennies et al. 2012; Tolujew et al. 2010; Schenk et al. 2009; Savrasov and Tolujew 2008; Tolujew and Alcala 2004).
Mesoscopic models represent flow processes in production and logistics systems through piecewise constant flow rates. This assumption is valid since logistics flows do not change continuously over time. The control of resources is not carried out continuously but only at certain points of time like changes of shifts, falling below or exceeding inventory thresholds. The resulting linearity of the cumulative flows facilitates event scheduling and the use of mathematical formulas for recalculating the system's state variables at every simulation time step.
The simulation time step is variable and the step size depends on the occurrence of scheduled events. This leads to a high computational performance. The principles of event-based computation of linear continuous processes are employed in the discrete-rate simulation paradigm implemented in the simulation software ExtendSim (Krahl 2009, Damiron and Nastasi 2008) and the hybrid simulation approach described by Kouikoglou and Phillis (2001).

3. PRODUCTION LINE

The preferred planning variant of the production line is shown in figure 1. The key figures for evaluating the planning variant are:

- Throughput per hour [p/h]
- Cycle time [s]
- Workload [%]
- Technical degree of utilization N_T [%]

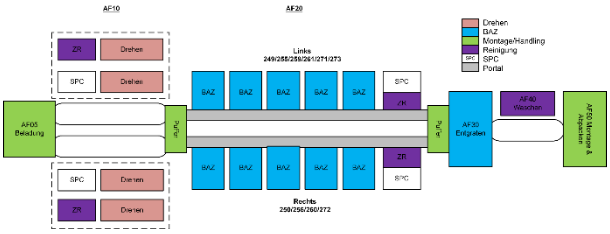


Figure 1: Preferred Planning Variant of the Production Line

The production process consists of the following six work sequences (WS):

- WS 05 Loading blank workpiece
- WS 10 Turning + cleaning
- WS 20 Milling + cleaning
- WS 30 Deburring
- WS 40 Washing
- WS 50 Assembly & discard

The production line consists of two strings (see figure 1), which produce different variants and alignments of a steering knuckle. These strings are brought together by one work sequence and are linked by fully automated facilities. Buffers are arranged between every work sequence. The production line operates in three shifts (22.5 hours per day).

4. MICROSCOPIC SIMULATION MODEL

The microscopic simulation model (see figure 2) was created with Tecnomatix Plant Simulation (V12.4) in conjunction with the VDA building blocks (V12.0-06.030). It depicts the process steps for the production of steering knuckles for commercial vehicles. It represents the flow of forged near net shape blank work pieces in the system until the discarding of finished parts.
The material flow is directed by the work plan feature of the VDA objects. The model offers the opportunity to manage different sets of settings like lot size, number of work piece carriers, processing times, breakdowns, tool change times and set up times.

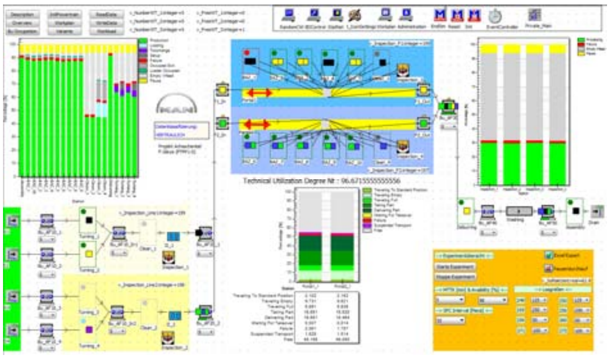


Figure 2: Microscopic Simulation Model in Plant Simulation

5. MESOSCOPIC SIMULATION MODEL

The mesoscopic model (see figure 3) is derived from the microscopic simulation model. For the implementation ExtendSim (V9.1) was used. The model uses objects from the rate (blue), item (green) and value (yellow) library. The item and value objects are only used to control the flow through the rate objects which represent the material flow. The value blocks are also needed for the evaluation of the simulation results.
The mesoscopic model consists of two strings of material flow which get reunited at one work sequence through a merge block. The model uses aggregated times for the different variants and not a work plan.
A discrete-rate model uses flow rates. The unit can be set freely in the model. One result is the throughput per hour. It is advisable to specify the flow unit in parts per hour (p/h). In this way, a target number can be determined directly. The concept of the discrete-rate model is conceivable in various levels of detail (Schenk, et al., 2008).

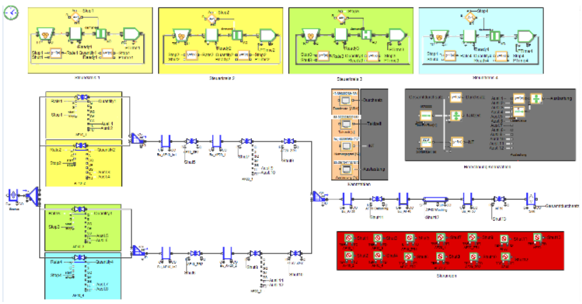


Figure 1: Mesoscopic Simulation Model in ExtendSim

In the model, the flow from the source to the sink is calculated by a linear program (LP) at each event (Krahl, 2009). The source (WS 05) can be represented as a tank. In the mesoscopic view, individual machines can be represented by a valve. A valve has a maximum and an effective rate. The effective rate is calculated from the inflow and outflow of a valve. The tool change, set-up procedures and breakdowns reduce the effective productive time of the individual stations. The results of the simulation correlate with the existence of these influences in the model. Work piece carriers as well as measuring stations are not considered in the mesoscopic simulation. The assumption is made that sufficient work

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

piece carriers are available in this configuration. The measuring stations can be neglected. They merely increase the mean throughput time because the process takes place only every fifty work pieces. The influence on the results is classified as low and is neglected.
Each machine is shown as a valve. The maximum rate of the valves corresponds to the average machining time from the work plan in the unit p/h for the respective lot size. The WS 20 is aggregated as a valve. The rate is calculated from the average machining time. The remaining work sequences are implemented as valves. Between the stations are buffers with the capacities from 5 to 10 pieces (like provided in the base model). A control circle of item elements is designed to model the tool changes and setup operations. Item elements are discrete-event blocks from the item library of ExtendSim, which can be used to control the flow in the mesoscopic model (figure 4). The control algorithm composes the delay time (PTime) from the manufactured quantity of the valve (Quantity). In a second control, the rate of the valve is controlled as a function of breakdowns (Shut_n), set-up and tool changing processes.

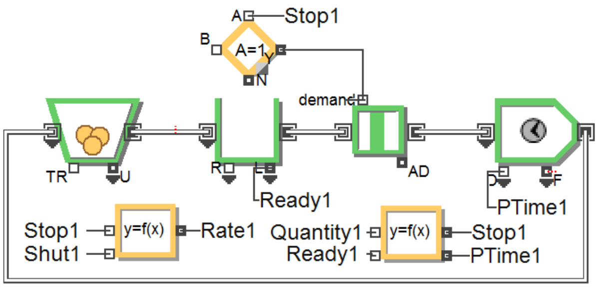


Figure 4: Control Circle

Distribution parameters of the Erlang distribution (MTTR) and the exponential distribution (MTBF) are prepared for the breakdowns. These distributions correspond to the selected distributions in the microscopic discrete-event model.
The breakdowns can be calculated in a shutdown block. The signals required to determine the key figures are processed from the information flow. The sink provides the current rate as well as the total throughput. Equities and Math blocks can be used to compile the key figures.
The implementation of the conception requires a tank as a source. The source (WS 05) is initialized with infinite initial content. The flow from the source is divided into four individual flows with a diverge block using the diverge mode "neutral" and fed into the valves which represent WS 10. The initial rate is set to the calculated rate in p/h. Between the work sequences, buffers are represented by a tank. The flows from the WS 10 are directed into a flow through the merge block. The flow runs in two strands into the valve for WS 20. The flows from the valves (WS 20) are merged into one flow through a merge block. The flow is routed through the remaining work sequences (WS 30 to WS 40) and destroyed in the sink (WS 50).
For the breakdowns, the output signals of the shutdown blocks are connected to the "maximum rate" connector of the valves. The key figures cycle time, throughput per hour and the technical degree of utilization are calculated from the total throughput, the time per run and the goal cycle time. The rate can be distinguished between = 0, > 0, limited or not limited. Overlapping, the value-adding time is assumed to be the time in which the rate is above 0. The signals from the "time-limiting" and "not time-limiting" connectors are processed through the equation block.

6. RESULTS

Due to the stochastic proportions, more than one replication is required. Using the confidence interval method, at least two replications are necessary to maintain a statistical confidence of at least 95 %. The average value with two replications does not differ greatly in comparison to three or more replications. If we perform more than two replications we only attain more effort. The simulation has a running time of 450 hours (20 days with 22.5 hours each day) like the microscopic model. Table 1 shows the full experiment set-up. The set-up is equivalent to the microscopic model.

Table 1: Experiment Set-up

Start time [h]	0
End time [h]	450
Run Time [h]	450
Replications	2

Table 2 and table 3 compare the results of both models.

Table 2: Simulation Results

	Discrete-event		Discrete-rate	
	Average	Standard deviation	Average	Standard deviation
Through-put [P/h]	42,53	0,14	41,10	0,13
Cycle time [s]	84,65	0,28	87,59	0,29
N_T [%]	97,81	0,32	94,64	0,31
Work-load [%]	76,39	0,27	85,52	0,28

Table 3: Deviation

	Through-put [p/h]	Cycle time [s]	Technical degree of utilization N_T [%]	Work-load [%]
Deviation	-1,43	+2,94	-3,17	+9,13
Deviation [%]	-3,35	+3,47	-3,24	+ 11,95

A survey of Schauf (2016) asked the production planners which margin of error they are willing to accept, see figure 5. The willingness to accept errors decreases with an increasing duration of a simulation project. Production planners are ready to accept errors of 5 % for simulation project which take less than a week. In simulation projects with a duration of more than six months they are willing to accept an error of about 1.5 %.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

These results could mean that production planners would accept to work with models which have not such a high level of detail but are capable of providing simulation results faster.

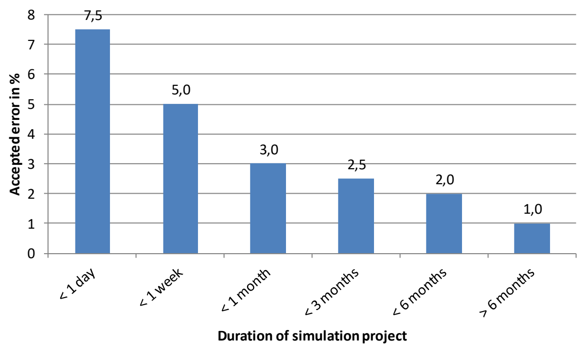


Figure 5: Accepted Errors by Production Planners in a Simulation Project (Schauf 2016)

There are different aspects to compare. A general statement regarding the total expenditure does not allow any statement about individual aspects. The following aspects were identified regarding modeling effort:

- Conception
- Data collection and processing
- Implementation
- Experiments

From the user's point of view, the assessment of the effort is subjective. Depending on the aspect, a rough classification of the effort can be made. Table 4 shows the assessment of the aspects for the DRS approach. This rating is in comparison to the DES approach.

Table 4: Effort for the DRS Approach

Phase	DRS
Conception	+
Data collection and processing	+
Implementation	+
Experiments	++

Scale: ++ very beneficial; + advantageous; 0 neutral; - disadvantageous; -- very disadvantageous

In the design phase the effort is reduced. Components of the system can be idealized, as well as times and events over the runtime can be considered as mean values. This has an impact on the data collection effort. Due to less required data and a mean of influencing data, advantages can be achieved at this point. In the implementation, the mesoscopic model has fewer required blocks, but the function and interaction of the blocks are limited. The experiments can be performed with less effort due to the greatly shortened run time of a replication (1 min runtime of the discrete rate model and 5 min runtime of the discrete-event model).

After the effort for the simulation approaches, the accuracy of the results must be assessed. Table 3 shows the deviations of the models. An average deviation of 5.5 % occurred across all results. With acceptance of a safety of 95 %, the suitability of the mesoscopic approach can be confirmed for the throughput, the cycle time and the technical degree of utilization. The deviation of the workload is clearly above the acceptance threshold with 11.95 %.

7. CONCLUSION

The results of the simulation experiments show that mesoscopic simulation models based on the discrete-rate simulation paradigm are capable to support planning tasks in production and logistics systems. For several key figures, their results differ only slightly from the results of a discrete-event simulation model. The results deviation stays within a margin that is accepted by production planners.

Mesoscopic simulation models can save enormous amounts of modeling and computational time compared to discrete-event models and thus comply with the requirements of production planners to receive simulation results within a short period of time.

8. REFERENCES

Borshchev, A., and A. Filippov. 2004. "From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools". In Proceedings of the 22nd International Conference of the System Dynamics Society, edited by M. Kennedy, G. W. Winch, R. S. Langer, J. I. Rowe, and J. M. Yanni. Oxford.

Damiron, C., and A. Nastasi. 2008. "Discrete Rate Simulation Using Linear Programming". In Proceedings of the 2008 Winter Simulation Conference, edited by S. J. Mason, R. R. Hill, L. Moench, O. Rose, T. Jefferson, and J. W. Fowler, 740–749. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hennies, T., T. Reggelin, and J. Tolujew. 2012. "Mesoscopic Supply Chain Simulation". In Proceedings of the international Conference on Harbour, Maritime and Multimodal Logistics Modelling and Simulation, edited by A. G. Bruzzone, M. Gronalt, Y. Merkuryev, M. A. Piera, and W. Talley, 85–90. Vienna: International Multidisciplinary Modeling and Simulation Multiconference, I3M.

Hennies, T., T. Reggelin, J. Tolujew, and P.-A. Piccut. 2014. "Mesoscopic supply chain simulation". In Journal of computational science 5 (3), 463–470. Amsterdam: Elsevier.

Hennig, A., T. Reggelin, and D. Wolff. 2016. Mesoscopic simulation models for the automotive industry In: The 28th European Modeling & Simulation Symposium, EMSS 2016: September 26 - 28, 2016, Cyprus - Genoa: DIPTEM, S. 309-

313 ; [Kongress: 28th European Modeling & Simulation Symposium, EMSS 2016, Cyprus, 26-28 September, 2016]

Huber, D., and W. Dangelmaier. 2009. “Controlled Simplification of Material Flow Simulation Models”. In Proceedings of the 2009 Winter Simulation Conference, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 839–850. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Huber, L., and S. Wenzel. 2011. “Trends und Handlungsbedarfe der Ablaufsimulation in der Automobilindustrie“. In: Industrie Management 27 (5), 27–30. Berlin: GITO-Verlag.

Kosturiak, J., and M. Gregor. 1995. Simulation von Produktionssystemen. 1st ed. Wien: Springer.

Krahl, D. 2009. “ExtendSim Advanced Technology: Discrete Rate Simulation”. In Proceedings of the 2009 Winter Simulation Conference, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 333–338. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kouikoglou, V. S., and Y. A. Phillis. 2001. Hybrid simulation models of production networks. New York: Kluwer Academic Plenum Publishers.

Law, A. M., and W. D. Kelton. 2007. Simulation Modeling and Analysis. 4rd ed. Boston: McGraw-Hill, Inc.

Pierreval, H., R. Bruniaux, and C. Caux. 2007. “A Continuous Simulation Approach for Supply Chains in the Automotive Industry”. In Simulation Modelling Practice and Theory 15 (2), 185–198. Amsterdam: Elsevier.

Reggelin, T. 2011. “Mesoskopische Modellierung und Simulation Logistischer Flusssysteme“. Ph.D. thesis, Institute of Logistics and Material Handling Systems, Otto-von-Guericke-Universität, Magdeburg, Saxony-Anhalt.

Reggelin, T., and J. Tolujew. 2011. “A Mesoscopic Approach to Modeling and Simulation of Logistics Processes”. In Proceedings of the 2011 Winter Simulation Conference, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 1508–1518. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Savrasov, M., and J. Tolujew. 2008. “Mesoscopic Approach to Modeling a Traffic System”. In: International Conference Modelling of Business, Industrial and Transport Systems, edited by E. Kopytov, H. Pranevicius, E. Zavadskas, and I. Yatskiv, 147–151. Riga: Transport and Telecommunication Institute.

Schauf, C. 2016. “Anforderungen von Produktionsplanern an die Simulation – Discrete-Rate Simulation als Ergänzung zur ereignisdiskreten Simulation in der Produktionsplanung?“. In Schenk, M. (Hrsg.) 21. Magdeburger Logistiktage »Logistik neu denken und gestalten«, edited by M. Schenk, 141–148. Magdeburg: Fraunhofer IFF.

Schenk M., Tolujew J., Reggelin T., 2008. Mesoskopische Simulation von Flusssystemen – algorithmisch steuern und analytisch berechnen. In: P. Nyhuis, Hrsg. Beiträge zu einer Theorie in der Logistik. Berlin Heidelberg: Springer Verlag, pp. 463-485.

Schenk, M., J. Tolujew, and T. Reggelin. 2009. “Mesoscopic modeling and simulation of logistics networks”. In IFAC Proceedings Volumes 42 (4), 582-587. Amsterdam: Elsevier.

Scholz-Reiter, B., F. Wirth, M. Freitag, S. Dashkovskiy, J. Jagalski, C. de Beer, and B. Rüffer. 2007. “Mathematical Models of Autonomous Logistic Processes”. In Understanding Autonomous Cooperation and Control in Logistics – The Impact of Autonomy on Mananagement, Information, Communication and Material Flow, edited by M. Hülsmann, and K. Windt, 121–138. Berlin: Springer.

Scholz-Reiter, B., C. de Beer, M. Freitag, T. Hamann, H. Rekersbrink, and J. T. Tervo. 2008. “Dynamik logistischer Systeme”. In Beiträge zu einer Theorie der Logistik, edited by P. Nyhuis, 109–138. Berlin: Springer.

Tolujew, J., and F. Alcala. 2004. “A Mesoscopic Approach to Modeling and Simulation of Pedestrian Flows”. In Proceedings of the 18th European Simulation Multiconference, edited by G. Horton, 123–128. Ghent: SCS International.

Tolujew, J., T. Reggelin, and A. Kaiser. 2010. “Discrete Rate Simulation als grundlegendes Paradigma bei der Entwicklung von mesoskopischen Flussmodellen“ In Integrationsaspekte der Simulation: Technik, Organisation und Personal. Karlsruhe : KIT Scientific Publ., 437-444. (ASIM-Mitteilung; AM 131).

A COMPARATIVE ANALYSIS OF DIFFERENT METHODS FOR IDENTIFICATION OF THE EVOLUTION OF NUMBER OF POSSIBLE CONFLICT-FREE AIRSPACE CONFIGURATIONS INCLUDING MULTIPLE AIRCRAFT AND SINGLE CONFLICT

Julia de Homdedeu [(a)], Maria del Mar Tous [(b)], Miquel Ángel Piera Eroles [(c)], Thimjo Koca [(d)], Marko Radanovic [(e)]

[(a),(b),(c),(d),(e)] Department of Telecommunications and Systems Engineering
Autonomous University of Barcelona
Barcelona, Spain

[(a)] juliade.homdedeu@e-campus.uab.cat, [(b)] mariadelmar.tous@e-campus.uab.cat,
[(c)] miquelangel.piera@uab.cat, [(d)]thimjo.koca@uab.cat, [(e)]marko.radanovic@uab.cat

ABSTRACT

Continuous increase in the traffic density over the certain en-route sectors provokes many situations in which a loss of separation minima (SM) between two aircraft occurs. Although, this loss is predicted well in advance, giving a proper look-ahead time (LAT) for a detection function, the resolution of such an event may lead to a new conflict situation due to dynamics of surrounding traffic aircraft. A multi-agent system framework can deal with these cases.

This work presents three different complexity indicators that can be used to shape the social behavior of the agents. Simulation results show that the proposed indicators can suggest drastically different nature of the same ecosystem, therefore further investigation of the correlation of the proposed indicators to the actual complexity is necessary.

Keywords: ecosystem, feasible solutions, opportunity costs, conflict maneuvers

1. INTRODUCTION

Continuous increase in the traffic density over the certain en-route sectors provokes many situations in which a loss of separation minima (SM) between two aircraft occurs (Bouarfa, Blom, Curran, and Everdij 2013; Kochenderfer, Holland, Chryssanthacopoulos 2012; Kuchar, Yang 2000). Although, this loss is predicted well in advance, giving a proper look-ahead time (LAT) for a detection function, the resolution of such an event may lead to a new conflict situation due to dynamics of surrounding traffic aircraft (Livadas, Lygeros, Lynch 2000; Murugan, Oblah 2010; Tang, Piera, Ling, Fan 2015). Namely, some resulting maneuvers of the conflicting aircraft can induce new loss of SM with nearby aircraft in which new LAT for detection can be significantly reduced. Consequently, a collision risk in this case is often at a higher level, which usually requires more demanding avoidance maneuver for the pilot-in-command, generating also inefficient trajectory segments.

At present, an upgraded Traffic Alert and Collision Avoidance System (TCAS II v7.1), has been designed for operations in the traffic densities of 0.3 aircraft per squared nautical mile (AGENT Project Team 2016; Tang, Piera, Ling, Fan 2015). It demonstrates excellent performances for the pairwise encounters, as well as the great improvements for multi-thread encounters, taking different flight configurations (cruising and evolving aircraft) into considerations. However, a TCAS logic shows some operational drawbacks due to limited number of resolution advisories, currently resulting in the vertical flight profile change only (Enea, Porretta 2012; Ramasamy, Sabatini, Gardi, Kistan 2014). Moreover, the well reported induced collisions in many traffic scenarios show a high probability of occurrence. Thus, there is a challenge to investigate and implement a new operational framework which will improve and extend TCAS functionalities at both tactical and operational level.

This paper relies on a new research in the ATM automation framework: the concept of ecosystems (AGENT project team 2016; Radanovic, Piera, Koca, Saez 2017). Ecosystem presents a set of aircraft, with self-automated capabilities, that form a cost-efficient separation management system for finding the best compromise in resolution trajectories. The goal is to transform the ecosystem aircraft into intelligent agents that can communicate with each other to safely make the best use of existing airspace capacity. The concept has been developed to handle a robust conflict management process considering aircraft performances, and consistent solutions in front of the scalability problems.

The study particularly compares three indicators of the evolution over time of the number of generated feasible resolution trajectories. All three of them address in common a maintenance of the SM criteria, the causal relationships, namely spatiotemporal interdependencies, and the metrics used for identification of the ecosystem aircraft through state space exploration.

The first indicator produces the accumulative number of conflict-free system configurations over time. The 4D time-space is discretized and the possibilities are counted.

In order not to count solutions which although feasible, they are under no criteria desirable, the second one is proposed. In this case, using a greedy search process the method identifies and counts the closest conflict-free configurations for each path of the search tree.

The third indicator is following a similar approach to the second one, but instead of counting the provided solutions, it measures the provided clearance for the given solutions.
In addition to this introductory section, the paper comprises five other sections. Section 2 includes some background, Section 3 describes three methods to generate the estimations, while Section 4 discusses simulation results and compares them. Section 5 gives concluding remarks and directions for the follow-up research.

2. BACKGROUND

2.1. Ecosystem concept

As said in the introduction this paper relies on concept of the ecosystem. An ecosystem is defined as a set that includes two aircraft in conflict between them and their surrounding traffic.
An aircraft A is part of the surrounding traffic if there exist a feasible maneuver that performed during the existence time (chosen to be 5 minutes) of the ecosystem by some member B of the ecosystem, or by aircraft A itself, will induce a new conflict in the ecosystem between aircraft A and B.

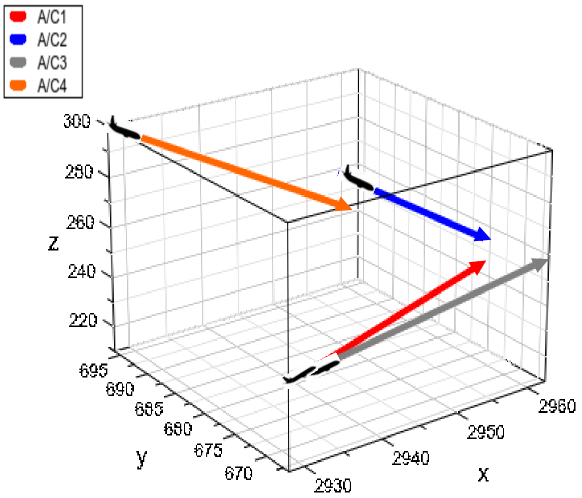


Figure 1 - Example of ecosystem

Essentially the framework tries to solve the conflicts by considering also the other aircraft that might be affected by a given deviation set of maneuvers in a pairwise conflict. In doing so, the ecosystem aircraft are treated as intelligent agents that can communicate with each other to safely make the best use of existing airspace capacity.

2.2. The necessity for complexity analysis

The multi-agent framework is capable to provide an acceptable solution, a conflict-free configuration of the system. However, the decision making is done in real time and therefore the system is time critical, therefore some information regarding the complexity of the scenario can be handy (Lyons 2012; Prandini, Piroddi, Puechmorel, Brázdilová 2011). A complex, not easy to solve scenario calls for a more collaborative behavior of the aircraft-agents, while on the other hand a not so complex ecosystem can allow agents to seek longer for a solution that fits better their criteria.
In these conditions, an indicator of the system's complexity is desired. Some attempts to quantify this complexity where done (Delahaye, Puechmorel 2000). This work proposes 3 more complexity indicators that can provide more explicit information.

2.3. Model Assumptions and Restrictions

In the considered scenarios, some assumptions are made. Firstly, the aircraft trajectories during the existence of the ecosystem are linear segments.
Secondly the maneuver space is discretized in space and time, which means aircraft can perform maneuvers with a certain deviation angle from the original trajectory and these maneuvers can be performed only at discrete time instances. The maximum angle deviation is assumed to be 30º and the increment 5º. The time increment is taken 1 second.

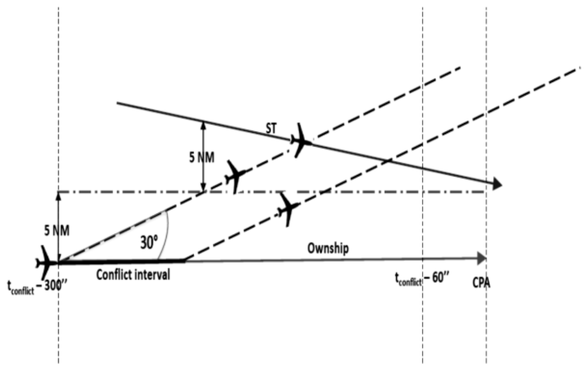


Figure 2 – Possible maneuvers performed by an aircraft and how they can affect interdependencies with another one

figure

Lastly the possible taken maneuvers should be synchronous, i.e. all the performed maneuvers that will be taken to resolve the conflict should be taken at the same time from all aircraft members.

3. INDICATORS DESCRIPTION

3.1. Full search of the discretized state space

The first discussed approach deals with the evolution in time of the number of possible configurations of the system that are conflict free. The pairwise interdependencies are identified in space and time and

using this information the configurations with at least a conflict are identified. Subtracting this value from the total amount of
possible configurations at each time instance will give us the desired quantity.

Indicator 1 Model

1. For each pair of aircraft identify the intervals during which interdependencies for different performed maneuvers are present
2. Using this information and the discretization parameters calculate the conflict-present system configurations
3. Calculate the total amount of possible system configurations
4. Calculate the number of conflict-free configurations by subtracting quantity calculated at step 2 from the quantity of step 3

3.2. Greedy search of the discretized state space

The previous indicator although it has some good properties, it has some drawbacks as well. One of them is that it needs to explore the whole discretized state space.
Moreover, some of the counted configurations, although feasible are at no way optimal, desirable, or necessary to consider. More specifically, let's consider an ecosystem with 3 members. AC1 And AC2 are in conflict and AC3 is surrounding traffic.
Let's further assume that AC1 going left 15º provides or AC1 going left 15º and AC3 climbing both provide conflict-free configurations of the system. Although both are visible, clearly option 1 is better than option 2.
To try to avoid this redundant information, indicator 2 is proposed.
Indicator 2, using a greedy search of the state space, gives an indication the number of conflict-free configurations of the system. More specifically, starting with the 2 conflict-aircraft, a maneuver is applied to one of them. If the system is conflict free, the configuration is counted as a solution and restarting from the original system configuration, other possibilities are sought. If on the other hand the maneuver produces a new conflict, a new maneuver is introduced to the traffic aircraft involved in the new conflict.

Indicator 2 Model

for each time instance:
1. Select 1 of the conflict aircraft
2. Perturb its trajectory using a maneuver from the set of maneuvers
3. if there is no conflict in the system:
4.   count 1 more solution
5.   if there are more unexplored solutions:
6.   repeat for a different maneuver
7.   else if there is unexplored aircraft left:
8.   repeat for the new aircraft
9.   else:
10.   end
11. else if there is unexplored aircraft left:
12.   apply perturbation to it and repeat
13. else:
14.   end

3.3. Accumulator of the conflict-free perimeters

Indicator 2 provides some information regarding the depth of the search by counting solutions. Another approach is introduced by indicator 2 and proposed in this section.
The key idea is that instead of counting the solutions themselves, the length of the arc formed taking as center the point where the aircraft performs the maneuver and as radius the distance that the aircraft will travel from that point until the end time of the ecosystem for each solution is measured. Given this information the evolution over time of the cumulative arc length is represented.

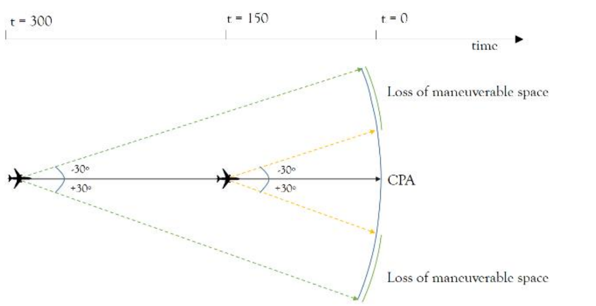


Figure 3 – Illustration of how a given maneuver performed at different time instances provides different amount of clearance

Indicator 3 Model

for each time instance:
1. Select 1 of the conflict aircraft
2. Perturb its trajectory using a maneuver from the set of maneuvers
3. if there is no conflict in the system:
4.   calculate the length of the formed arc
5.   Add the length to the total length
6.   if there are more unexplored solutions:
7.   repeat for a different maneuver
8.   else if there is unexplored aircraft left:
9.   repeat for the new aircraft
10.   else:
11.   end
12. else if there is unexplored aircraft left:
13.   apply perturbation to it and repeat
14. else:
15.   end

The main drawback of indicator 3 is that it does not provide any information regarding the number of solution at each time instance. However, a combination of indicator 2 and 3 can give us a clearer picture of the dynamics of the system.

4. SIMULATIONS RESULTS

The following section discusses some simulation results obtained from analysis of two ecosystems. The ecosystem trajectories are graphically presented in a 3D Euclidean space, with latitude and longitude measured in [km] and altitude in [ft]. The results show the cumulative number of the feasible trajectories over the ecosystem time.

A. Ecosystem 1

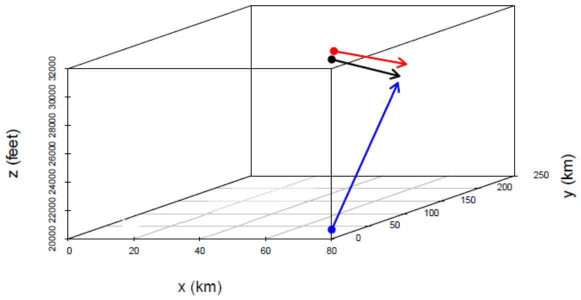


Figure 4 – Ecosystem 1

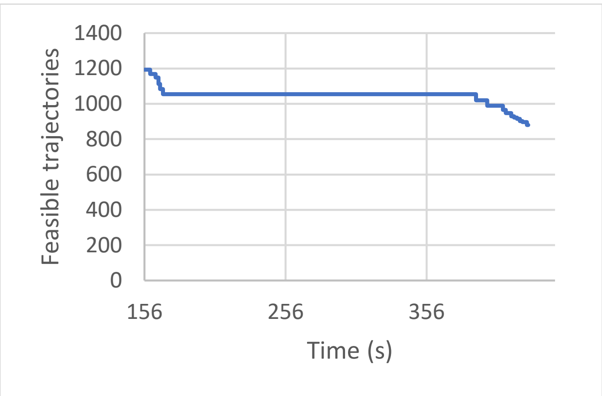


Figure 5 – Indicator 1 results for ecosystem 1

As can be seen from Figure 4, ecosystem 1 is composed of three aircraft in an evolving encounter: two of them in descending configuration and the third with a climbing approach. Figure 5 illustrates an amount of the feasible trajectories. It can be noted that this quantity follows a constant trend during the large percentage of the ecosystem time. The maximum number of 1200 trajectories is triggered at the time instant 156 seconds and it drops approximately to 1115 trajectories until 172 sec. The loss of maneuverable space for all three aircraft within these 16 seconds is significant. Another loss occurs at the time instant 400 seconds and decreases until the end of ecosystem time (430 sec) to 875 trajectories.

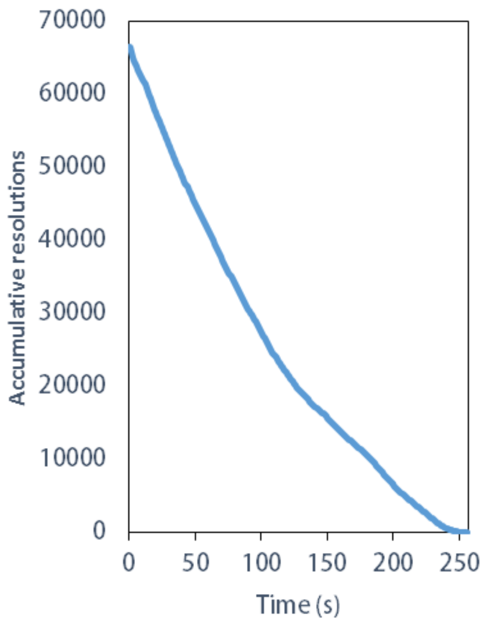


Figure 6 – Indicator 2 results for ecosystem 1

Figure 6 shows the number of solutions in time measured by Indicator 2. First thing that can be noticed is difference in the number of initial solutions. This can be justified by the different counting strategy of the 2 methods. The first one will count a maneuver on the left as unique solution, while the second will differentiate a solution that comes as a result of a 15º left maneuver from another one result of a 30º maneuver and count 2 distinctive solutions.
The second thing that can be seen is that the two curvatures are drastically different. Indicator 1 suggests for a simple ecosystem, where solutions are not fading quickly and therefore implies that agents can be more aggressive during the negotiations. Indicator 2 however suggests a sharper decrease rate which will call for a more collaborative negotiation between the agents.

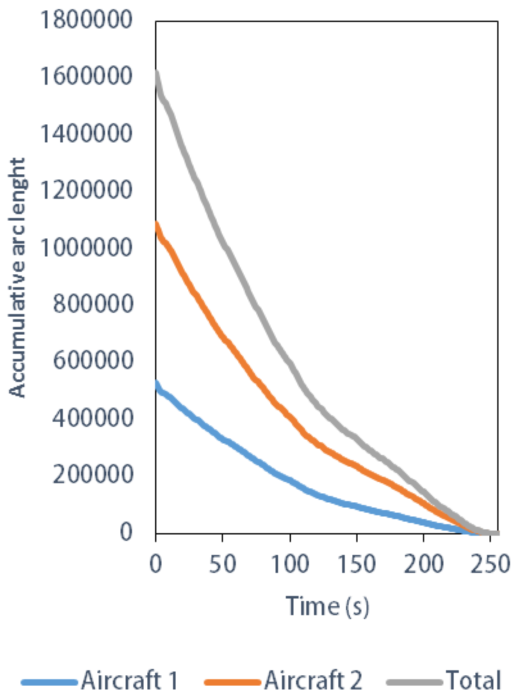


Figure 7 - Indicator 3 results for ecosystem 1. Arc lengths of conflict aircraft 1, 2 and their sum

What indicator 3 suggests furthermore is that aircraft 2 has a wider clearance. This suggests that it will probably be easier to convince this agent to reach an agreement than the other agent.

B. Ecosystem 2

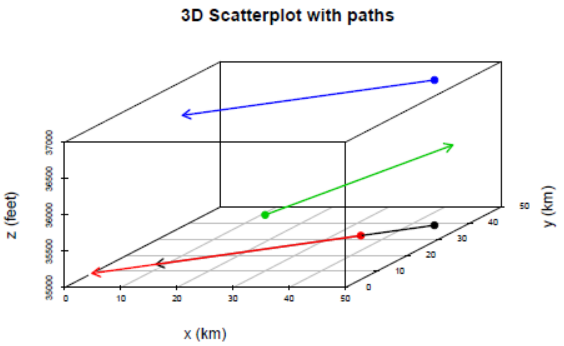


Figure 8 – Ecosystem 2

As can be seen from Figure 8, ecosystem 2 is composed of four aircraft. A conflict occurs between two aircraft in cruising configuration (red and black trajectories), with an overtaking scenario (loss of self-separation between two aircraft flying in the same direction). Figure 7 illustrates a rate of change in amount of the feasible trajectories over the time evolution, obtained at the time of triggering the ecosystem of approximately 6000. The rate, in this case, can be modeled as a step-wise function, with the smaller or greater quantity drops in discrete time intervals. It approached to zero value after 150 seconds of the ecosystem time.

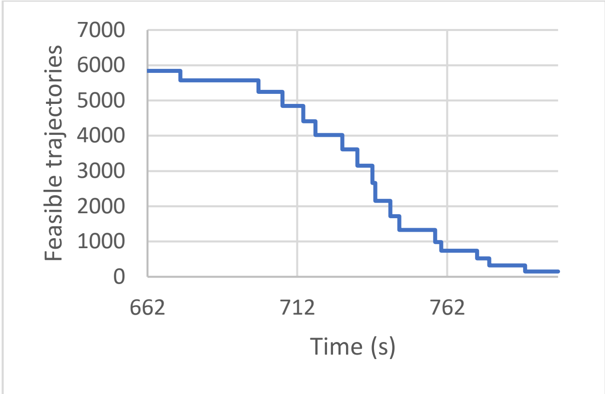


Figure 9 - Indicator 1 results for ecosystem 2

Indicator 1 here has a small decreasing rate initially and then changes to a more drastic one. This will suggest an initial selfish approach for the agent and a later collaborative one.

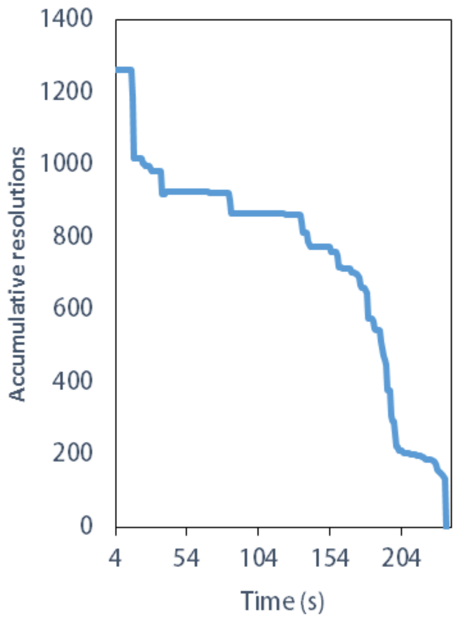


Figure 10 - Indicator 2 results for ecosystem 2

Indicator 2 in this ecosystem counts less initial solutions than indicator 1. Moreover a similar shape is detected between the 2 curves.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

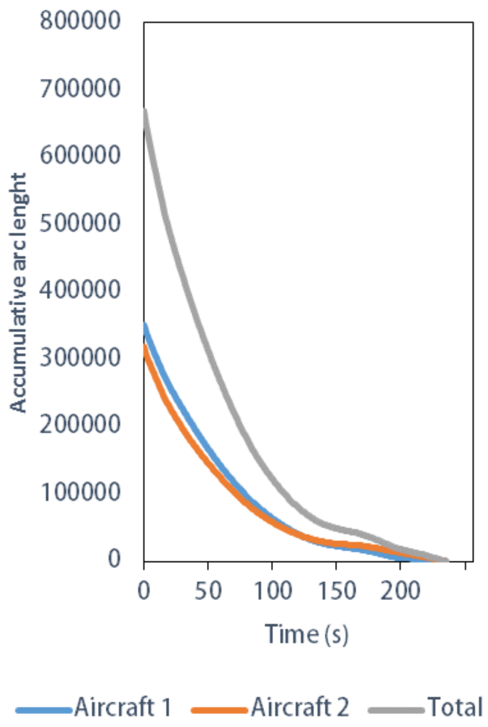


Figure 11 - Indicator 3 results for ecosystem 1. Arc lengths of conflict aircraft 1, 2 and their sum

Indicator 3 here suggests similar amount of clearance for the two conflict aircraft which should be materialized in similar behavior of them.

5. CONCLUSIONS

This study particularly analyze the ecosystem complexity through three indicators and compares the provided results. All three indicators support preservation of the SM criteria, generation of the spatiotemporal interdependencies, and the ecosystem identification metrics by agreed maneuverability.

The first indicator produces the accumulative number of conflict-free system configurations over time. The second implements a greedy search process that counts for each path of the search tree only the conflict-free system configurations with minimal introduced perturbations. Considerably, the third indicator has been introduced to measure the accumulated length of the conflict-free arcs for all aircraft that have to perform a resolution maneuver.

Each indicator individually provides information. It can be seen from the results that in some cases drastically different behaviors can be suggested. This calls the necessity for further investigation of them to understand individual correlation to the actual complexity of a given ecosystem and further use in a multi-agent environment.

REFERENCES

AGENT Project Team, 2016. Report on AGENT functional and non-functional requirements. Ref. 2015 (699313), Cranfield, United Kingdom.

Bouarfa, S., Blom, H. A., Curran, R. and Everdij, M. H., 2013. Agent-based modeling and simulation of emergent behavior in air transportation. Complex Adaptive Systems Modeling, 1(1), 15.

Delahaye, D., Puechmorel, S. 2000. Air traffic complexity: towards intrinsic metrics

Enea, G. and Porretta, M., 2012. A comparison of 4D-trajectory operations envisioned for Nextgen and SESAR, some preliminary findings. In: 28th Congress of the International Council of the Aeronautical Sciences, pp. 23-28, Sept, 2012.

Kochenderfer M. J., Holland, J. E. and Chryssanthacopoulos J. P., 2012. Next generation airborne collision avoidance system. Lincoln Laboratory Journal, 19 (1), 17-33.

Kuchar, J. K. and Yang, L. C., 2000. A review of conflict detection and resolution modeling methods. IEEE Transactions on Intelligent Transportation Systems, 1(4), 179-189.

Livadas, C., Lygeros, J. and Lynch, N. A., 2000. High-level modeling and analysis of the traffic alert and collision avoidance system (TCAS). Proceedings of the IEEE, 88(7), pp. 926-948.

Lyons, R., 2012. Complexity analysis of the next gen air traffic management system: trajectory based operations. International Journal of Industrial Engineering: Theory, Applications and Practice, 16(1), 4514-4522.

Murugan, S. and Oblah A. A., 2010. TCAS Functioning and Enhancements. International Journal of Computer Applications, 1 (8), 46-50.

Prandini, M., Piroddi, L., Puechmorel, S. and Brázdilová, S. L., 2011. Toward air traffic complexity assessment in new generation air traffic management systems. IEEE Transactions on Intelligent Transportation Systems, 12 (3), 809-818.

Radanovic, M., Piera, M.A., Koca, Th., Saez, F.J., 2017. Airborne Ecosystem Complexity Analysis for Tactical Conflict Management. IEEE Transactions on Intelligent Transportation Systems,

Ramasamy, S., Sabatini, R., Gardi, A. and Kistan, T., 2014. Next generation flight management system for real-time trajectory based operations. Applied Mechanics and Materials, 629, 344-349.

Tang, J., Piera, M. A., Ling, Y. and Fan, L., 2015. Extended traffic alert information to improve TCAS performance by means of causal models. Mathematical Problems in Engineering, 2015.

AUTHORS BIOGRAPHY

Júlia de Homdedeu is a graduate student from Aeronautical Management degree at Autonomous University of Barcelona. She has developed her final thesis of the degree in Multi-Agent Systems and is passionate for ATM research projects.

Maria del Mar Tous is a graduate student from Aeronautical Management degree at Autonomous University of Barcelona. She has developed her final thesis of the degree in Multi-Agent Systems and is passionate for ATM research projects.

Dr. Miquel Angel Piera Eroles is a full-time professor in the System Engineering Department at the Autonomous University of Barcelona. He graduated with excellence from UAB in Computer Engineering (1988), MSc from University of Manchester Institute of Science and Technology in Control Engineering (1991), and PhD from UAB in 1993. Dr. Piera is the UAB delegate for Technical Innovation Cluster, former deputy director of the UAB Engineering School and director of LogiSim (research group on Modelling and Simulation of Complex Systems). At present, Dr. Piera is active in the innovation sector, contributing as the scientific advisor to the company ASLOGIC.

Thimjo Koca is a PhD candidate within the Department of Telecommunications and System Engineering at the Autonomous University of Barcelona. He pursued his bachelor degree at Polytechnic University of Tirana in Computer Engineering (2014) and his master degree in a joint program between University of L'Aquila, Hamburg University and Universitat Autonoma de Barcelona in Mathematical Modeling, specialized in Stochastics and Optimization (2016). His research line is Multi-Agent System Simulations.

Marko Radanovic is a PhD candidate within Department of Telecommunications and Systems Engineering at the Autonomous University of Barcelona. He graduated from the University of Belgrade in air transport and traffic engineering (2007), and obtained master degree in satellite navigation from Politecnico di Torino (2012). His research lines are automation in air traffic management, modelling and simulation of aeronautical systems and air traffic safety.

MODIFIED RATE EQUATION MODELS FOR UNDERSTANDING PERFORMANCE LIMITS OF VERTICAL AND IN-PLANE COMPOUND CAVITY VCSELS

Naser F.G. Albugami[(a)], Eugene A. Avrutin[(b)],

Dept of Electronic Engineering, University of York, Heslington, York YO10 5DD, UK

[(a)] na668@york.ac.uk, [(b)]eugene.arutin@york.ac.uk

ABSTRACT

Modified rate equation models including both amplitude and phase properties are developed for advanced vertical cavity surface emitting lasers for multigigahertz modulation, and used to compare the performance of different designs and identify the corresponding limitations.

Keywords: vertical cavity laser, modified rate equation models, ultrafast modulation

1. INTRODUCTION

There are, fundamentally, two main approaches to increasing the modulation frequency range of vertical cavity semiconductor emitting lasers (VCSELs) at bit rates exceeding 40 GBit/s which can be used for high speed communication networks. The first one is through advanced developments in current (direct) modulation, which was demonstrated at bit rates up to 40 GBit/s (see e.g. Blokhin *et al.* 2009, Bobrov *et al.* 2015).

The second approach involves using some form of an advanced cavity structure. At least three solutions falling into the framework of this approach have been discussed. The first of them includes modulation of the photon lifetime in the passive cavity, rather than the curent (see e.g. Schchukin *et al.* 2014 and references therein) This approach promises substantially better dynamic properties than current modulation since it is free from the limitations of the direct modulation such as the the electron-photon resonance. The main laser design that allows this principle to be utilized is the copled-cavity VCSEL with one of the subcavities acting as an active one, and the other, as an electrooptic modulator (Van Eisden *et al.* 2008, Schchukin *et al.* 2008, Germann *et al.* 2012, Schchukin *et al.* 2014) or electroabsorption modulator (Chen et al. 2009) Modulation bit rates at least as good as those possible with direct modulation have been demonstrated in the references quoted, and rate equation analysis (Schchukin *et al.* 2008) predicts successful operation at bit rates up to and exceeding 80 GBit/s. However, rather than offering simple modulation of the photon lifetime, this design involves the laser oscillating in complex cavity modes, with the cavity decay time (photon lifetime), instantaneous frequency, and intensity all varying in time in a self-consistent way, which is not captured by rate equations and requires more sophisticated mathematical approaches.

A different type of an advanced cavity design involves using photon-photon resonance in some form, which comprises modulation of VCSEL arrays (Fryslie *et al.* 2015) and in-plane compound cavities, with amplified feedback from a slow wave waveguide resonator (see e.g. Dalir and Koyama 2011 and 2014, Park et al. 2016, and references therein).

A Lang-Kobayashi type delay-differential equation theory of the latter design has been presented in several papers (Dalir and Koyama 2011, Park et al. 2016), in the latter case with a good fit to the experiment achieved. This is despite of the fact that this model is, strictly speaking, intended for the case of weak optical feedback and as such used for the lasers in question (strong optical feedback) outside of its degree of applicability. The model was generalised to be more rigorous for the case of strong feedback, by introducing multiple delays, by Ahmed et al 2015, but the model is presented without derivation and appears to have a somewhat strange feature of the reflected light affecting he net gain rather than acting as an injected signal. Neither of the models includes the amplifying nature of the external feedback, it is not clear whether amplification in the feedback can help achieve further improvement in the laser performance. Finally, the question of what ultimately limits the laser performance does not appear to have been answered.

Thus, in order to understand the performance of both designs more accurately, and to compare those two approaches to improving the modulation properties, we have implemented a set of modified rate equation models each of which includes careful analysis of both amplitude and frequency (phase) of laser emission, as well as the spectrally selective nature of the laser cavity.

2. ELECTROOPTICALLY MODULATED LASER

2.1. The structure

Figure 1 shows the schematic of Cavity Laser which we used in our analysis. The laser is formed by two sub-cavities. The top, modulator, sub-cavity is a passive Fabry-Perot resonator whose reflectivity $R_m(t)$ is modulated via electro-optically varying the optical properties of a modulator layer contained within the sub-cavity by applying time-varying reverse bias voltage. In a typical design, the intermediate DBR has 25-35 periods of two layers of alternating composition, the top DBR consists of 15-25 periods; the resonator thus can be substantially asymmetric. Throughout the analysis, we consider only refractive index modulation (no absorption modulation). Indeed, it has been shown in the literature (see e.g. Schchukin *et al.* 2008) that in a practical design of this type, the electrorefraction effect plays a dominant role in modulating $R_m(t)$ as compared to electroabsorption.

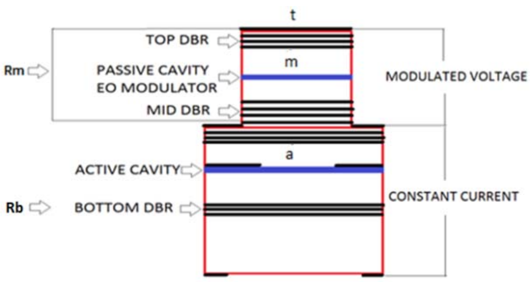


Figure 1: Schematic of a Compound Cavity Laser and model setup

2.2. The model

In designing the model (Albugami and Avrutin 2017), we follow the approach used previously for Distributed Feedback (DFB) lasers (see e.g. Wenzel *et al.*, 1996) and for multimode compound cavity lasers (see e.g. Avrutin *et al.*, 1999). Namely, the laser cavity is treated as a complex resonator, and a complex eigenfrequency of this resonator is found, which is then used to describe the laser dynamics. The electrooptically modulated VCSEL is very naturally suited for such an approach, because the complex resonator in this case can be defined by considering the active sub-cavity as a quasi-Fabry-Perot resonator terminated, on one side, by the bottom reflector with a (complex, generally speaking) amplitude reflectance r_b and on the other side, by the EO modulator subcavity treated as a passive, frequency-dependent reflector with a complex reflectance $r_m(\omega)$, where ω is the complex eigenfrequency sought. The value of the modulator reflectance is calculated from the equation

$$R_m = r_m r_m^* ; \qquad (1)$$

$$r_m = r_{am} + \frac{t_{am}\, t_{ma}\, r_{mt}\, e^{-j2K_m L_m}}{1 - r_{ma}\, r_{mt}\, e^{-j2K_m L_m}}$$

with the average wave vector in the form of:

$$K_m(\omega) = \frac{(n_m + \Gamma^t_m \Delta n_{EO})\omega t}{c} \qquad (2)$$

Here, Δn_{EO} describes the time-dependent correction to the refractive index of the modulator layer caused by electrooptic modulation. $\Gamma^t_m \approx \frac{d_m}{L_m}$ (assuming the modulator layer is thick enough that the standing wave factor is near one) is the confinement factor of the modulator layer, d_m being the modulator layer thickness and L_m, the total physical thickness of the modulator cavity, including any spacer layers between the modulator layer and the reflecting stacks but not including the penetration into mirrors. The complex eigenfrequency is then found by solving the usual threshold/resonant condition of a Fabry-Perot type cavity (the short cavity length ensures that the equation has one solution):

$$r_m(\omega)\, r_b(\omega) \times \exp\left[2\left(-j\frac{(n_a + \Delta n_a)\omega}{c} + \Gamma^t_a \frac{g}{2}\right)L_a\right] = 1 \qquad (1)$$

It is convenient to write the solution to this equation in terms of a frequency correction $\Delta\omega = \omega - \omega_{ref}$, where the (real) reference frequency ω_{ref} is arbitrary but can be conveniently taken, for example, as the position of the reflectance spectrum notch (**Errore. L'origine riferimento non è stata trovata.**) in either on or off-state. Then, the complex instantaneous frequency correction $\Delta\omega$ is given by a transcendental equation:

$$\Delta\omega = v_g \left[-\frac{j}{2}\left(\Gamma^t_a g - \frac{1}{L_a}\ln\frac{1}{r_m(\omega_{ref} + \Delta\omega) r_b}\right) - \frac{\Delta n_a\, \omega_{ref}}{c} + \left(\frac{q\pi}{L_a} - \frac{n_a\, \omega_{ref}}{c}\right)\right] \qquad (2)$$

Here, v_g is the group velocity, L_a is the geometrical thickness of the active subcavity, Γ_a is the confinement factor for the active area, g is the time-dependent gain, n_a is the refractive index of the active layer subcavity, averaged over the length in the same way as n_m is averaged over the modulator subcavity. The refractive index varies in time primarily due to self-phase modulation in the active layer; its time-dependent part can be quantified as:

$$\Delta n_a \approx \frac{c}{2\omega_{ref}} \Gamma^t_a \alpha_H (g - g_{th}) \qquad (5)$$

where α_H is the Henry linewidth enhancement factor in the active layer and g_{th} is the gain at threshold which we used as a reference value.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The choice of reference frequency near the modal frequency to ensure that $|\Delta\omega| \ll \omega_{ref}$ means that we can introduce a parameter q which is the number of half-wavelengths of light in material fitting (roughly) in the distance L_a. It depends on the VCSEL design, mainly the thickness L_a, and it is an integer number chosen in such a way that

$$q \approx \frac{2n_a L_a}{\lambda} - \frac{1}{2\pi}\arg(r_m r_b) \quad (6)$$

where $\lambda = 2\pi c/\omega_{ref}$ is the operating wavelength, and r_m can be estimated in the on- or off-state.

The frequency ω (or frequency correction Δω) has a real part, which determines the time-dependent spectral position of the lasing mode and thus the chirp of laser emission, and an imaginary part, which reflects the balance of gain and loss (the latter including the outcoupling loss, which is frequency dependent through $r_m(\omega)$). The imaginary part determines the dynamics of photon density (in the active cavity) N_p, giving a modified rate equation in the form:

$$\frac{dN_p}{dt} = -2 * \mathrm{Im}(\Delta\omega(t))N_p(t) + \frac{\beta_{sp}N}{\tau_{sp}(N)} \quad (7)$$

Where τ_{sp} is the carrier spontaneous recombination time and β_{sp} is the spontaneous emission factor. The dynamics of the carrier density N is determined by a standard rate equation:

$$\frac{dN}{dt} = \frac{\eta_i I}{eV} - \left(\frac{1}{\tau_{sp}(N)} + \frac{1}{\tau_{nr}(N)}\right)N - v_g\frac{g(N)}{1+\varepsilon N_p} \quad (4)$$

in which N and N_p are the the electron and photon densities, respectively, η_i is the internal quantum efficiency, I is the injected current, e is electron charge, V is the volume of the active region; $\tau_{sp} = [B_1 N^2/(1+b_1 N)]^{-1}$ and $\tau_{nr} = (A_1 + C_1 N^2)^{-1}$ are the spontaneous and nonradiative recombination times of carriers, respectively, $g(N)$ is the optical gain in the active layer, ε is gain compression factor (see Table 1).

Finally, the model includes a differential equation taking into account the electromagnetic resonance in the modulator cavity. The equation is derived by considering the modulator section in frequency domain and then substituting the imaginary part of the frequency correction by a time derivative. The result is conveniently expressed as:

$$\frac{d\tilde{E}_t}{dt} = \frac{\sqrt{\pi T_i T_t}}{4}\frac{v_g}{L_{eff,m}}E_a + \left(-\frac{1}{2\tau_{cm}} - j\left(\Delta\omega'(t) - \Delta\omega_{n0} + \Gamma_m\frac{\Delta n_{EO}(t)}{n_g}\omega_{ref}\right)\right)\tilde{E}_t \quad (9)$$

$$\tau_{cm} = \frac{L_{eff,m}}{v_g(1-\sqrt{R_i R_t})} \approx \frac{2L_{eff,m}}{v_g(T_i+T_t)} \quad (10)$$

Here, $\tilde{E}_t = E_t \exp(j\varphi_t)$ describes the complex amplitude of the output field emitted from the modulator subcavity (the top mirror), $E_a = \sqrt{N_p}$ is the field amplitude inside the active subcavity. Note that in the equation written in the form above, E_a is a *real* value, which means that the complex amplitude of the output light is actually $\tilde{E}_t^f = \tilde{E}_t \exp(j\int\Delta\omega' dt)$, where $\Delta\omega' = \mathrm{Re}(\Delta\omega)$ is the instantaneous frequency correction in the active cavity. Furthermore, $\Delta\omega_{n0} = \omega_{n0} - \omega_{ref}$ is the position of the notch in the modulator subcavity transmission in the absence of modulation ($\Delta_{nEO} = 0$), τ_{cm} is he effective photon lifetime in the modulator cavity, $R_i = |r_{mid}|^2$ and $R_t = |r_{out}|^2$ are the intensity reflectances of the intermediate and top reflectors (the two reflectors forming the modulator cavity) respectively, $T_i = 1 - R_i$, $T_t = 1 - R_t$ are the transparencies of the intermediate and top DBR stacks, and $L_{eff,m}$ is the effective thickness of the modulator section; it is typically a fraction of a micrometer greater than the physical one, as it considers the penetration of the field into the mirrors. The power emitted from the laser can be calculated as:

$$P \approx v_g \hbar\omega A_x E_t^2 \quad (3)(11)$$

Where A_x is the cross-section of the aperture

An alternative formalism for describing the laser dynamics with phase included would consist of writing out an equation similar to equation (**Errore. L'origine riferimento non è stata trovata.**) for the *active* subcavity, with an injection term representing light reflected from the modulator cavity, rather than solving a transcendental equation (**Errore. L'origine riferimento non è stata trovata.**) for the instantaneous frequency. A model of that type twould the laser as a system of two subcavities, one active, one passive but modulated, treated on the same footing. The results should be very similar to those of the current formalism so long as the dynamics of light *inside* the active cavity remain slower than the modulator cavity round-trip (which is the case for most realistic designs).

2.3. Small-signal analysis

The small-signal analysis was performed analyticallyAlbugami and Avrutin 2017. Typical results are shown in Figure 2.

The figure illustrates that the ultimate limit to the modulation speed is set by the photon lifetime in the modulator section, which in turn is determined by the reflectances of the top and intermediate reflectors. With the optimised laser design (a large R_i ensuring the external-modulator-type operation of the modulator section, and a more modest R_t ensuring the value of τ_{cm} of the order of picoseconds), this limit is of the order of hundreds of gigahertz, and thus not a concern for any realistic modulation scheme.

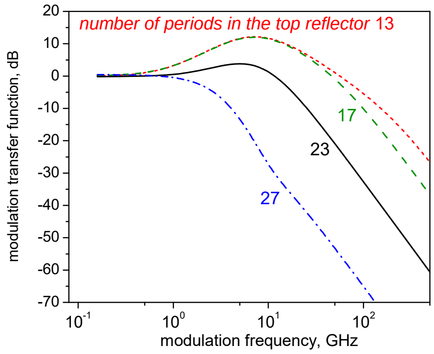


Figure 2. Calculated small-signal response of EO laser modulation. 35 periods in the intermediate reflector

The situation can be different however with a design of the modulator section less optimised for high speed operation. With a large (but perfectly technologically achievable) number of periods in the top reflector (hence R_t), when the notch in the reflectance of the modulator section (the peak in the transmittance) becomes narrow meaning a large photon lifetime in the modulator cavity, the 3dB frequency drops as low as ~10 GHz. As will be discussed later, these effects of the reflector design manifest themselves also in the results of *large* signal modulation simulationsLarge signal analysis

2.4. Large- signal analysis

The large signal analysis was implemented as a Matlab code; the transcendental equation (4) was solved at each step using the direct iteration method and with the value at the previous point in time used as the initial condition; typically only 3-4 iterations were required.

The large signal modulation laser behaviour was characterized using an eye diagrams (Figures 3-4) and quantified additionally by means of a Quality Factor

$$Q = \frac{\overline{P_{(1)}} - \overline{P_{(0)}}}{\sigma_{(1)} + \sigma_{(0)}} \qquad (11)$$

where $\overline{P_{(1)}}$, $\overline{P_{(0)}}$ are the mean values of the power corresponding to the logical one and zero states, respectively; $\sigma_{(1)}$, $\sigma_{(0)}$ are the corresponding standard deviations.

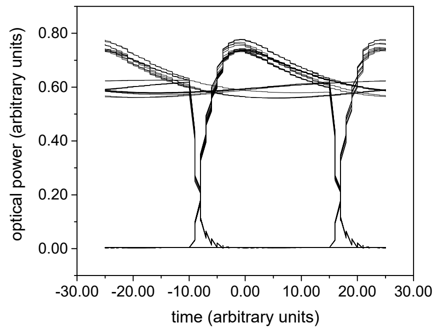


Figure 3. Eye diagrams for *i*= 10 mA, 40 GBit/s, 35 intermediate DBR periods, 17 top DBR periods

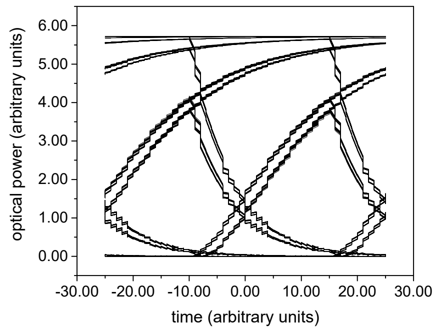


Figure 4. As figure 3, 23 top DBR periods

Figure 5 shows Q as function of the number of layer pairs (periods) in the *top* reflector DBR stack. As seen in the figure, there is an optimum number, in this case around 17.

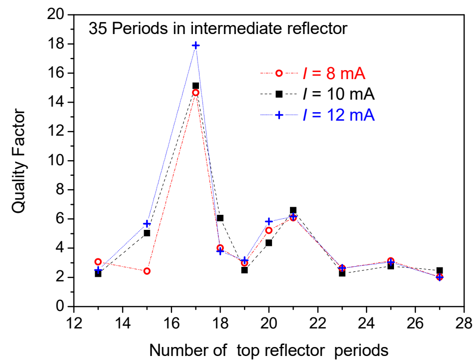


Figure 5: Modulation quality factor as function of the number of periods in the top reflector

When we increase the number of periods beyond that number, the photon lifetime in the modulator cavity is increased, leading to longer transients and closing the eye diagram, hence bad quality factor. When the number of top reflector period becomes too low, on the other hand, the power variation (P_1-P_0) becomes lower, hence lower quality factor. We observe that the optimum number of periods is not a strong function of

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

the current, so once optimised, a laser should be able to provide good modulation quality at all currents.

Figure 6. shows the modulation quality as a function of the number of periods in the *intermediate* reflector. Again, there is an optimum number of periods, in this case around 32.

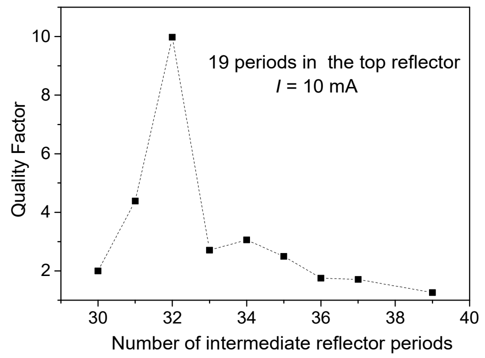


Figure 6: Modulation quality factor as function of the number of periods in the intermediate reflector Bit rate 40 GBit/s

An advantage of the electrooptic modulation is the relatively low chirp predicted. Despite a certain degree of phase modulation implied by Eqs, 4 and 7 (the phase and amplitude of the modulator reflectance are by necessity modulated simultaneously), the spectra of laser emission simulated (Figure 7) are almost transform limited.

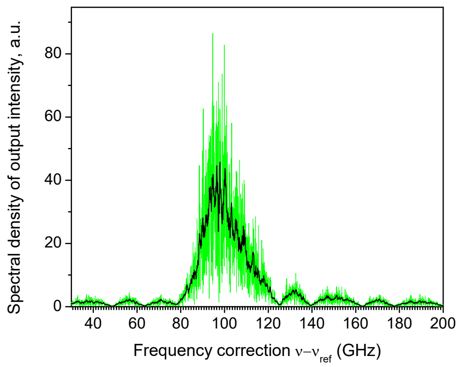


Figure 7: Amplitude Spectrum of output Field at 40 GBit/s

3. LASER WITH IN-PLANE FEEDBACK

In the first part of the work, we used a heuristic Lang-Kobayashi model similar to that used by Dalir and Koyama 2011 and adopted by some other authors (e.g. Park et al. 2016), with the rate equation for the electric field inside the cavity given by the rate equation with the delayed term:

$$\frac{dE(t)}{dt} = G^{net}(t)E(t) + KE(t-\tau_E) \qquad (12)$$

Where the spontaneous term has been omitted (it was introduced in simulations using a Langevin noise source but omitted for small-signal simulations),

$$G^{net}(t) \approx \frac{1}{2}\left[v_g\Gamma g(1+j\alpha_H) - \left(\frac{1}{\tau_{ph}^{V}} + \frac{1}{\tau_{ph}^{H}} + \frac{1}{\tau_{ph}^{in}}\right)\right] \qquad (13)$$

is the complex amplitude net gain, with the photon lifetime including contributions due to the vertical cavity, in-plane cavity, and internal losses; the rest of parameters have the same meaning as for the vertical cavity.

Finally, *K* is the heuristic optical feedback coefficient (in the case of a one-dimensional cavity with small feedback, it can be expressed in terms of (ii) the coupling between the laser and the external cavity and (ii) the reflectance of the distant reflector). The coefficient is complex, with its phase having a large effect on the laser dynamics.

When analysing the small-signal response, we use the impulse response method, applying a short (~2-5 ps) Gaussian current pulse to the laser and taking a Fourier transform of the resulting transient. This allows to make an express estimate of the 3dB cutoff frequency.

Typical results are shown in Figure 8. As in previous results, in-phase feedback increases the frequency of the relaxation oscillations and hence the 3dB modulation cutoff; however the limit to utilising this frequency increase by increasing the feedback is set by the onset of self-pulsations, with the relaxation oscillations becoming undamped.

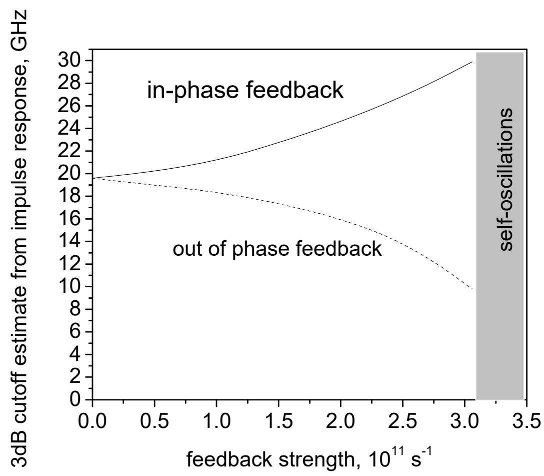


Figure 8. Estimate for 3dB modulation bandwidth using the impulse response method in the Lang-Kobayashi type model.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Including amplification in the external cavity by making the coefficient K time-dependent according to the model by Antonelli et al. 2015 does not appear to modify the result qualitatively.

Next, we attempted to extend the analysis from the heuristic approach of Dalir and Koyama 2011, based on Lang Kobayashi equations with a single delay, to a more electromagnetically justified model. In doing so, the complication is that we cannot restrict ourselves even approximately to one-dimensional analysis as in the vertical cavity problem, as the in-plane cavity problem is substantially multi-dimensional. In the case of a cavity with rectangular geometry as used by Park et al. 2016, we can restrict the analysis to two dimensions: the vertical dimension z and the direction of the in-plane cavity x, as shown in Figure 9 (in the third dimension y, the lateral one perpendicular to the external cavity orientation, a well-localised single lateral mode is assumed).

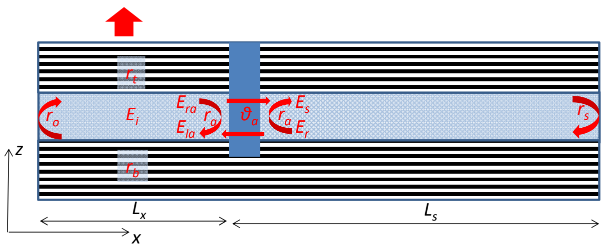


Figure 9: Schematic of the electromagnetic 2D model for a compound in-plane cavity.

In the in-plane direction, the laser is just a section of the slow-wave waveguide, of a length L_x, with gain due to the amplification in the active layer and loss due to both internal loss and emission through the top and bottom mirrors. The section is terminated on the left hand side by the outer reflectance r_o (typically $|r_o|$~1), and on the right, by an oxide stripe separating the laser from the amplifier (slow waveguide external cavity). The oxide stripe is assigned an amplitude reflectance r_a and an amplitude transmittance $\sqrt{\theta_a}\exp(j\varphi_a)$, θ_a being the intensity transmittance and $R_a = |r_a|^2$ the intensity reflectance. If there is no scattering of light out of the slow-waveguide mode in the oxide, $R_a + \theta_a$ =1; in reality it is likely that some scattering is present so $R_a + \theta_a < 1$. In the analysis, we assumed that the reflectance r_a is seen by the light travelling in both directions. It has to be pointed out that this applies better to the linear geometry cavity of Park et al. 2016, than to the more three-dimensional geometry of Dalir and Koyama 2014, in which the waveguide connecting the external and laser cavities is narrow, and the oxide reflectance is not necessary for determining the mode of the laser cavity.

In the model considered, the dynamic equation for the light amplitude inside the cavity takes the form

$$\frac{dE_l}{dt} = \frac{1}{T_x e^{g_{net}T_x}}\left\{\sqrt{f_x\theta_a}\, e^{j\varphi_a} E_r + \left(e^{g_{net}T_x} - 1\right) E_l\right\}$$
(14)

Here, T_x is the round-trip of the slow wave along the laser size L_x, and f is the geometric factor (~2 in a good quality cavity). The reflected light amplitude E_r in the passive external cavity limit is given by

$$E_r(t) = r_s E_s(t - \tau_s) \qquad (15)$$

Where $r_s = |r_s| \exp(j\varphi_s)$ is the field reflectance of the remote reflector (Figure 8), and

$$E_s(t) = \sqrt{\frac{\theta_a}{f_x}}\, e^{j\varphi_a} \frac{1}{r_a} e^{g_{net}T_x} E_l(t) + r_a E_r(t) \qquad (16)$$

is the field to the right of the oxide stripe travelling into the external cavity. The "recursive" boundary conditions (15-16) eliminate the need for summing the terms from multiple delays which needs to be done in the model by Ahmed et al. 2015, and also allow for the gain in the external cavity to be included along the lines of the paper by Antonelli et al. 2015. In the limit of weak external reflectance, the second term in (16) disappears, and the model reduces to the Lang-Kobayashi type one.

The results of the model, while quantitatively different from those of the Lang Kobayashi type model, predict a similar increase in the modulation frequency cutoff of the order of the frequency itself, in other words up to the values of 30-40 GHz but, at least in our simulations so far, not beyond.

4. CONCLUSIONS.

We have presented the two modified rate equations approaches suitable for analysis of advanced ultrafast lasers: the compound-cavity eigenmode model for the vertically integrated electrooptically modulated laser and a delay-differential one for the laser with in-plane compound cavity.

In terms of sheer modulation potential, the electrooptically modulated laser appears to be the more promising, with modulation cutoffs determined ultimately by the photon lifetime in the external cavity and thus, with proper design, extending into hundreds of gigahertz.

Practical considerations such as the ease and cost of fabrication and integration, ambient temperature tolerance, etc. will also contribute to determining the laser design to be used in each particular application.

REFERENCES

Ahmed, M., Bakry, A., Alghamdi, M. S., Dalir, H. & Koyama, F, 2015. Enhancing the modulation andwidth of VCSELs to the millimeterwaveband using strong transverse slow-light feedback. *Opt. Express* **23,** 15365

Antonelli C., Mecozzi A., Hu Z., Santagiustina M., 2015. Analytic study of the modulation response of reflective semiconductor optical amplifiers, J.Lightwave Technol., 22(20), 4367-4376

Albugami N.F. and Avrutin E.A. 2017, Dynamic modelling of electrooptically modulated vertical compound cavity surface emitting semiconductor lasers, Opt.Quantum Electron., accepted.

Avrutin, E.A., Marsh, J.H., Arnold, J.M., Krauss, T.F., Pottinger, H., De La Rue, R.M. 1999. Analysis of harmonic (sub)THz passive mode-locking in monolithic compound cavity Fabry-Perot and ring laser diodes. IEE Proc. Optoelectron. 146(1), 55–61

Blokhin, S.A., Lott, J.A., Mutig, A., Fiol, G., Ledentsov, N.N., aximov, M.V., Bimberg, D., 2009. Oxide confined 850 nm VCSELs operating at bit rates up to 40 Gbit/s. Electron. Lett. 45(10), 501–502

Bobrov, M.A., Blokhin, S.A., Maleev, N.A., Kuzmenkov, A.G., Blokhin, A.A., Yu, M.Z., Ustinov, V.M., 2015. Ultimate modulation bandwidth of 850 nm oxide-confined vertical-cavity surface-emitting lasers. J. Phys: Conf. Ser. 643(1), 012044

Chen C., Leisher P.O., Grasso D.M., Long C., Choquette K.D. 2009, High-speed electroabsorption modulation of composite-resonator vertical cavity lasers, IET Optoelectronics, 3(2), 93-99

Dalir, H. and Koyama F., 2011. Bandwidth enhancement of single-mode VCSEL with lateral optical feedback of slow light. IEICE Electronics Express, 2011. **8**(13), 1075-1081

Dalir, H. and Koyama F., 2014, Highly stable operations of transverse coupled cavity VCSELs with enhanced modulation bandwidth, Electron.Lett., 50 (11), 823-824

Fryslie S.T.M., Tan M.P., Siriani D.F., Johnson M.T., and Choquette K.D. 2014 . 37-GHz modulation via resonance tuning in single-mode coherent vertical-cavity laser arrays. IEEE Photon. Technol. Lett. 27(3), 415–418

Germann TD, H.W., Nadtochiy AM, Schulze JH, Mutig A, Strittmatter A, Bimberg D., 2012. Electro-optical resonance modulation of vertical-cavity surface-emitting lasers. Opt Express, **20**(5), 5099-107

Park, G.C., Xue, W., Piels, M., Zibar, D, Mork, J, Semenova, E., Chung, I-S, 2016. Ultrahigh-speed Si-integrated on-chip laser with tailored dynamic characteristics, Scientific Reports, 6, 38801.

Shchukin, V.A., Ledentsov, N.N., Lott, J.A., Quast, H., Hopfer, F., Karachinsky, L.Y., Bimberg, D. 2008. Ultrahigh-speed electro-optically modulated VCSELs: modeling and experimental results. Proc. SPIE 6889(1), 68890H .

Shchukin, V.A., Ledentsov, N.N., Qureshi, Z., Ingham, J.D., Penty, R.V., White, I.H., Novikov, I.I. 2014. Digital data transmission using electro-optically modulated vertical-cavity surface-emitting laser with saturable absorber, Appl. Phys. Lett. 104(5), 051125

Van Eisden J., Yakimov M, Tokranov V., Varanasi M., Mohammed E. M., Young I. A., Oktyabrsky S. R. 2008, Optically Decoupled Loss Modulation in a Duo-Cavity VCSEL, Photon.Technol. Lett., 20(1),42-44

Wenzel, H., Bandelow, U., Wunsche, H.J., Rehberg, J. 1996, Mechanisms of fast self pulsations in two-section DFB lasers. IEEE J. Quantum Electron. 32(1), 69–78

DEMAND PREDICTION MODEL OF AN ORGANIZATION

Oroselfia Sánchez[(a)], Idalia Flores[(b)]

[(a),(b)]Facultad de Ingeniería. División de Posgrado. Universidad Nacional Autónoma de México.

[(a)]oroselfia.sanchez@yahoo.com.mx, [(b)]idalia@unam.mx

ABSTRACT

Nowadays, the organization resources management should be more accurate to avoid delay risks and penalties. One start point for planning are historical data, which form the base for forecasting the future demand. This demand forecast is used to plan resources acquisition and resources management. Frequently, the management systems of the organization do not allow detailed historical record of activities, which causes insufficient data to forecast demand and other important aspects.

In this paper, data of a petroleum private enterprise are treated. The purpose of this work is to know all the feasible demand scenarios that may occur through a mathematical model based in a Poisson distribution from a scorecard. The objective of the model is to forecast demand of two types of services offered by the enterprise. Finally, a simulation model is used to validate this mathematical model.

Keywords: Poisson Distribution, Demand, Forecasting Scorecard, Risks, Simulation.

1. INTRODUCTION

In organizations, budgets for acquisition of material and human resources depend on demand of future services and products. A good prediction of such demand allows a more detailed planning of the required resources. When organizations do not have the necessary capacity to produce products or provide services failures appear, causing penalty costs. This is important, because costs of prevention are lower than those originated from corrections, since the possible impact may be difficult to assess.

Some common methods employed to forecast demands can be qualitative, such as expert opinions, surveys to costumers, Delphi methods (Middendorf,1973), and quantitative of classical and Bayesian kind (Niu, Zhao and Liu, 2009), such as random variables and time series models (Qiu, Suganthan and Amaratunga, 2016), causal and stochastic models (Ma, Wu, Khanwala, Li and Dang, 2015), Markov chains (Wan, Zhang and Dai, 2014), networks (Zou, Huang, Chen and Qu, 2011), simulations (Chen, Lu and Shao, 2010), among others. Moreover, some methods should be adapted in order to obtain data that helps to solve the particular issues of the organization, including constrictions addition and system properties.

In this work, we propose a mathematical model adjusted to the necessities of an organization with the purpose of obtaining expected demand events. The model uses as input data those historically recorded in a scorecard. It contains the number of two different kinds of services based on historical data and expert group predictions in a monthly frequency.

2. THE ORGANIZATION AND ITS DATA

In this work, a demand forecast model is established for a petroleum private organization, whose activity is oil well cementing. This company offers two types of services, oil well cementing (i) and pumping jobs (j). The first one consists in building a cement wall to stabilize and isolate the oil well, while the second one is the pumping of fluids from the surface to the bottom of the oil well. The resources of the organization are shared by both kind of jobs.

Jobs i and j performed each month are recorded in a scorecard format. It does not specify situations that occur by day, and then details of demand in worst scenarios are unknown. Scorecard data recorded for a year is shown in Table 1, scorecard data.

Managers would like to know, the maximum and minimum limits of jobs that probably occur by day and by job type using the scorecard data. With this information, the quantity of resources could be planned. Quantitative results obtained could help to analyze and improve predictions that otherwise are only based on beliefs.

3. MATHEMATICAL MODEL

3.1. Obtaining the model.

The objective of the mathematical model is to estimate the number of days that i cementing and j pumping jobs occur during a given month. According to expert group opinions and historical data recorded in scorecards, feasible scenarios are given by $i = \{0,1,...,6\}$ and $j = \{0,1,...,6\}$. Possible events are represented by a pair (i,j) that specify the number of cementing and pumping jobs in a given day. Calculation of occurrence probability of every event by month could be performed to obtain the number of days of each situation by month. We use a Poisson distribution to appraise probability of each kind of job in the period of one month. However, Poisson distribution do not consider constrictions of the system. In consequence, we add such restrictions to the model, to adapt it to data results required. In particular, Poisson distribution gives results of real type; however we need to count the number of days using natural numbers. Poisson distribution is then employed to establish inferior (d_{ij}^{-}) and superior (d_{ij}^{+}) bounds for the number of days that occur every event. These bounds allow the adjustment of data to integer numbers with the purpose to determine the Total Number of Days per month that each event can occur. The superior limit d_{ij}^{+} is given by $d_{ij}^{+} = round\left(\bar{d}_{ij} + \varepsilon\right)$ and the inferior is defined by $d_{ij}^{-} = round\left(\bar{d}_{ij} - \varepsilon\right)$, where ε is a value chosen, varying arbitrary between 0.2 and 0.3 and $\bar{d}_{ij}$ is the average number of days that event is presented per month, as obtained from the Poisson distribution.

Table 1: Total Number of each type of Jobs by month (Data from Scorecard Organization and Simulation results)

Month	**scorecard data**		**simulation results (average)**	
	(i)	**(j)**	**(i)**	**(j)**
January	33	51	32.9065	50.3874
February	29	46	28.9184	45.4272
March	42	56	41.9151	55.7194
April	48	53	47.517	53.301
May	47	47	46.6829	46.717
June	59	64	58.872	64.5
July	51	77	51.0167	77.4845
August	46	32	45.8025	31.5642
September	56	10	55.542	9.87
October	53	46	52.9759	45.7405
November	47	35	46.41	34.881
December	64	28	63.2989	27.962

In other words, the number of days that an event is expected to occur during a month is obtained through the determination of an integer-distribution that is closer to the real-Poisson-distribution. The mathematical model seeks the achievement of the following restrictions:

- The sum of days of occurrence of each event is equal to the number of days in a month.
- The number of days of occurrence of each event is less than the days calculated in the upper limit and greater than the days calculated in the lower limit.
- The addition of the number of days that each event occurs per month, multiplied by the number of cement works, is equal to the total number of cementations registered in the scorecard.
- The addition of the number of days that each event occurs per month, multiplied by the number of pumping jobs is equal to the number of total pumping jobs registered in the scorecard.

Considering the pairs (*i*,*j*), the model to obtain the demand by day and by month is the following,

$$\max z = \sum_{i=0}^{6}\sum_{j=0}^{6} p_{ij} d_{ij}^{(m)} , \qquad (1)$$

subject to

$$\sum_{i=0}^{6}\sum_{j=0}^{6} d_{ij} = D_m \qquad (2)$$

$$d_{ij} \geq d_{ij}^{+} \qquad (3)$$

$$d_{ij} \leq d_{ij}^{-} \qquad (4)$$

$$\sum_{j=0}^{6}\sum_{i=0}^{6} i d_{ij} = T_{cmt} \qquad (5)$$

$$\sum_{i=0}^{6}\sum_{j=0}^{6} j d_{ij} = T_{pump} \qquad (6)$$

$$p_{ij}, d_{ij} \geq 0 \qquad (7)$$

where,

i = Number of Cementing jobs

j = Number of Pumping jobs

p_{ij} = Probability of each event

d_{ij} = Number of days

D_m = Total number of days that occur the pair (*i*,*j*)

T_{cmc} = Total number of Cementing jobs in a month (scorecard)

T_{pmp} = Total number of Pumping jobs (scorecard)

The mathematical model can be solved through MS-Excel Solver to obtain data of the combination of cementing and pumping jobs by day and by month.

3.2. Validation of the Model.

In order to validate the mathematical model, a simulation in SIMIO platform was made.

3.2.1. Simulation

The simulation models a generation of cementing jobs *(i)* and pumping jobs *(j)* month by month.

Figure 1 shows how the simulation works. In each run the number of jobs of each kind (*i* and *j*) per day are stored. At the end of the month it is determined the number of days that happened *i*, cementing and *j* pumping jobs. Results are then compared with those predicted by the mathematical model. The processes were set into SIMIO following the logic shown in Figure 2.

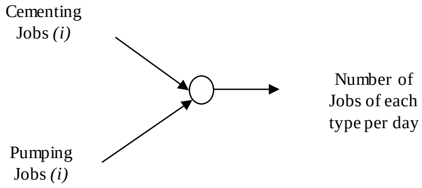


Figure 1: Conceptual Simulation Model

4. ANALYSIS OF THE RESULTS

4.1. Mathematical Model results

The search of information about a system can be done with different approaches by the managers. Some are based on the statistical mode per month and observe possibilities around it. Others, analyze all the different scenarios which can happen in one day. It depends on the approach of the managers what data are chosen to make decisions.

The model proposed in this paper, seeks the most feasible scenarios that may be presented based on scorecard data. Mathematical model was solved in MS-Excel Solver allowing to obtain the expected events per month. Maximization is performed for each month separately based on their corresponding scorecard data. Figure 3 shows the results, in color scale, of the mathematical model for each situation per month. Color scale goes from cero to seven, which can be interpreted as the number of times the event was obtained during month.

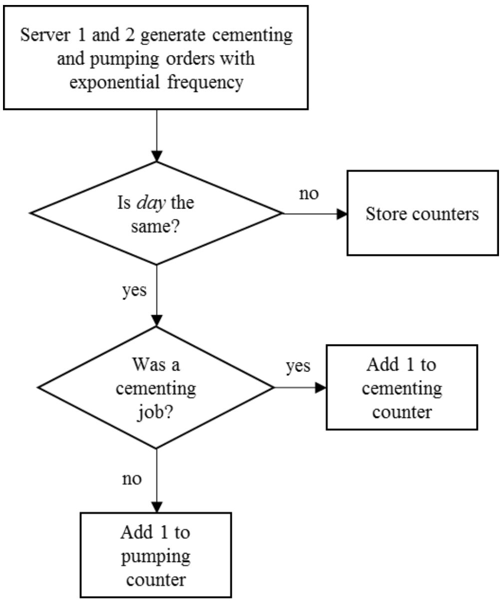


Figure 2: Logic structure of the simulation model

All these data results demonstrate a plethora of different possible scenarios or events that the organization could present during a year from the scorecard historical data. We can see in Figure 3 the results month by month, presenting different demand behavior between them. The most representative results are explained below.

In January, the expected events that can happen oscillates in the combination of $i = (0,1,...,3)$ and $j = (0,...,4)$. According to the numerical results, the events that are most likely to occur during the month are around 1 pumping job and 1 cementing job. However, an event which could demand a great quantity of resources is the pair $i, j = (3,3)$

The enterprise policy will influence resources management strategy. Will the organization be prepared for the hardest days of the month? Or for the most frequent days? In the first case, the maximum capacity of their resources will not be occupied in several days, which generates costs per day of non-occupation. In the second case, the company may require external resources causing variable costs.

Events obtained for June and July are different from January. June events are widespread combination of *i* and *j* indicating a high number of days in which there are several events that require more resources, this means that organization may have more cementing and pumping jobs in a day during several days. Both in June and in July, the variation within the same month is high which generates more management and scheduling uncertainty.

In September opposite happens because is the month with lowest reported jobs (see Table 1 scorecard data) in comparison to other months due to a few number of required pumping jobs. However, the obtained high number of cementing jobs causes that possibilities with 0 to 5 cementing jobs become feasible.

Events resulting from the model may show the months with fewer down time. How does it help us to know this data? If there are months with high demand of services, activities as maintenance, capacitation or preparation of resources must be anticipated by the organization. If the information provided to the stakeholders includes those months with more jobs, then contingency plans may be determined in advance. Thus, considering acquisition of outsourcing maintenance, personal and resources. The previous prevent the company to possible risks ensuring to avoid the surprise factor.

4.2. Simulation Model Results

The results of simulation indicate that expected events may be only slightly deviated from those predicated by the mathematical model. Such deviation is represented by the area enclosed by the dashed line in Figure 3 for January. In fact, some events out of the area also occur but with very low frequency and are related to not including restriction in the number of jobs suggested by expert opinions. For example, feasible sets for *i* and *j* were not included. In general, situation results concords with the mathematical model in determining the most probable events.

For example, in Figure 3 the results of the simulation do not exceed the white dotted line that can be appreciated for the month of January. This validates that mathematical model represents in a correct way the most frequent scenarios and the combinations of these.

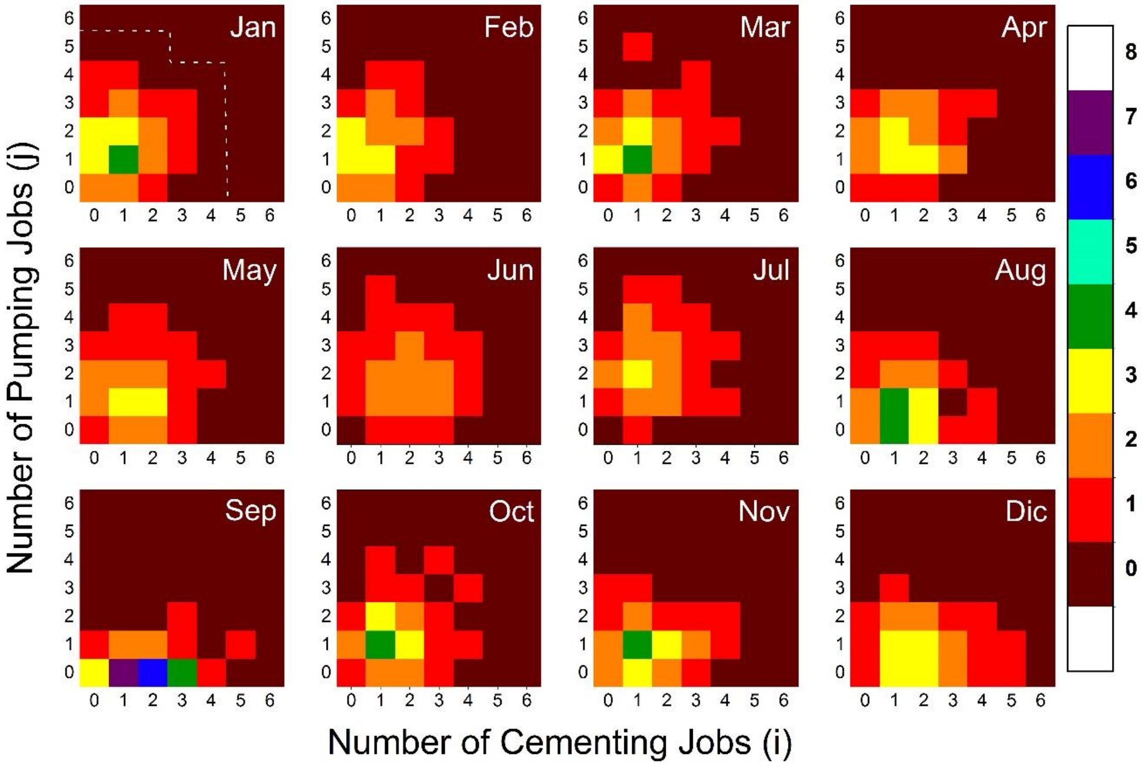


Figure 3: Mathematical model results per month

An important highlight is that both, mathematical and simulation modelling are a powerful tool to analyze the demand. But in this case, through simulation as a tool, a lot of runs must be done to determine the same feasible region obtained by the mathematical model. However, one of the advantage of simulation is that it invites to us to analyze more the system. For example, simulation gives explicit idea of worst scenarios, that may be considered as anomalous by the mathematical model. They occur rarely, but it does not mean that the company will never face those events.

In addition to the previous results, Table 1 shows a comparison between the average of number of jobs resulting from simulation, versus historical scorecard data. The averages were obtained in a period of 1000000 days (simulation time) for every month.

4.3. The worst scenarios

When the resources are planning, the scenarios or situations which are commonly considered are the most frequent. Determination of worst scenarios proposes to have a broader paradigm of demand behavior to make better decisions. It is always good to know which are the worst scenarios that organization could have before they happen. The resources must be managed to comply most jobs, but also should be useful to have a plan that allows a rapid response to a contingency. In this sense, the analysis of the worst scenarios is a powerful tool. A lot of risk appear when bad decisions are taken, and with high costs consequences.

Figure 3 shows the results obtained through the mathematical model. How can you accomplish a worst scenarios analysis with this type of data? In this framework, we could do through two options. The first one, is determining those events closers to the delimited area in Figure 3 (for each month). For example, for January the worst scenarios may be follow the pairs of *i,j*: (0,5), (1,5), (2,4), (3,4), (4,4), (4,3), (4,2), (4,1), (4,0) and (0,3). Those events, are very closer to the occurrence region and even the maximization results indicate that they do not occur, it does not that this will not happen in the real world.

If it is desired to explore even more the system, as a second option simulation can be used to find worst scenarios outside of the white dotted line in Figure 3. This results could present some ideas of the events which are considered as not feasible for the expert group but in real world have a tiny ocurrence probability.

With the results of this analysis, the managers plan what to do if the events with the color representing zero in Figure 3. Would the company refuse to provide the service? Or would they risk being penalized for lack of resources? Would request external resources with variable costs? How much they would this cost to the organization?

The point as a manager is to keep a curious thought regarding the situations that may occur. Different techniques used to analyse the demand allow to have more insights of the system and prevent organizations to incur in operational or other kind of risks.

5. CONCLUSIONS

Commonly, organizations deal with missing data for the planning of their systems and the lack of information generates high uncertainty. This method presented in this paper, can help to obtain more information about the probable demand situations that can occur beyond based on historical data.

In this work, a mathematical model was proposed and developed to analyse the demand from an historical scorecard (Table 1, scorecard data). The number of jobs per month generated by mathematical model, are the same as those in Table 1 scorecard data because it is a constriction of the model. The results were shown in feasible scenarios form with the corresponding frequency obtained (Figure 3) and validated through simulation model. Besides, the resulting events give us insights about the demand in systems.

Through mathematical model it was possible to obtain a broad picture of the demand behaviour. Also, the addition to the model own features permit to show data in a form useful to make decisions in management. The combination of techniques allows to discover the details of the resulting events.

A more careful analysis of the data available has the purpose of give more information of the system preventing delays, penalties and operational risks. This model helps to reduce lack of knowledge, which could decrease uncertainty and risks by helping prevention decisions.

ACKNOWLEDGMENTS

This research was supported by UNAM PAPIIT grant IT102117. The first author specially acknowledges the support from CONACYT scholarship program.

REFERENCES

Chen M., Lu J. and Shao Y., 2010. An application of Monte Carlo Simulation: Activity Based Urban Travel Demand Forecast Modeling. International Conference on Computer Design and Applications, 4, V4-251–V4-254.

Ma R., Wu R., Khanwala M., Li D. and Dang S., 2015. Power Demand Forecasting using Stochastic Model: Parameter Estimation. Modern Electric Power Systems.

Middendorf W.H., 1973. A modified Delphi method of solving business problems. IEEE Transactions on Engineering management, EM-20 (4), 130–133.

Niu L., ZhAO J. and Liu M., 2009. Application of Relevance Vector Regression Model based on Sparse Bayesian Learning to Long-term Electricity Demand Forecasting. Proceedings of the 2009 International Conference on Mechatronics and Automation. 2363–2367.

Qiu, X. and Suganthan P.N., 2016. Electricity Load Demand Time Series forecasting with empirical Mode Decomposition based Random Vector Functional Link Network. International Conference on systems, Man and Cybernetics, 001394 – 001399.

Wan X., Zhang Q. and Dai G., 2014. Research on Foresting method of natural gas demand based on GM, Model and Markov Chain. International Conference on Cognitive Informatics & Cognitive Computing, 436 – 441.

Zou A.Q., Huang G., Chen J.L. and Qu Y.Q., 2011. The Inventory Demand Forecasting Model of the Regional Logistics Network in Supply Chain. Industrial Engineering and Engineering Management, 1565 – 1569

AUTHORS BIOGRAPHY

Oroselfia Sánchez Sánchez studied Chemical Engineering at Faculty of Chemistry of Universidad Veracruzana (UV), then a Specialization in Quality Systems at Universidad Nacional Autónoma de México (UNAM) obtaining the Alfonso Caso Medal. She has worked on Petroleum Industry in Quality, Operation and Management areas. She received a Master in Operational Research with honors at UNAM. She was also a Professor of the Specialization in Quality Systems at UNAM. She is currently studying a PhD in Operations Research at UNAM.

Idalia Flores de la Mota 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 Postgraduate Program at UNAM and her research interests lie in simulation and optimization of production and service systems, based on this research she has been editor of two books edited by Springer.

SIMULATION OF BOARDING PEDESTRIAN IN MEXICAN SUBWAY: THE CASE OF PANTITLAN TERMINAL STATION

Yazmin Dillarza-Andrade[a], Aida Huerta-Barrientos[b], Geraldo Salazar-Diaz[c], Joali Evelyn Pérez-Bonilla[d]

[a],[b],[d]Department of System Engineering
[c]Department of Electronic Engineering
National Autonomous Mexico University

[a]dillarza_andrade@hotmail.com, [b]aida.huerta@comunidad.unam.mx, [c]geraldosalazar16@gmail.com,
[d]joali.perez@gmail.com

ABSTRACT
The Collective Transportation System (STC) of Mexico City transports more than 1.6 billion passengers annually and it is considered one of the cheapest in the world. The increasing demand has caused saturations that, together with the operational problems, affect the time of transfer of the passengers and therefore the quality of the service of the transport system. There are currently no measurements of the time passengers wait to board the train. The purpose of this paper is to simulate the operation of the Pantitlan terminal station using the Simulation Model-Based Methodology for Complex Systems Analysis, MoSASCoM, and AnyLogic™ simulation software, to measure the boarding time of the passengers and to generate a tool in which it can be varied: the arrival rate of passengers, number of passengers that approach, the opening time of train doors and the frequency of arrival of trains, in order to assist the decision making of the operational management.

Keywords: hybrid modeling, boarding time, urban mobility, Mexican subway.

1. INTRODUCTION

The accelerated and uncontrolled urbanization process that takes place every day in the Metropolitan Zone of the Valley of Mexico (ZMVM) has led, over the last decades, the implementation and the use of mass transit of passengers that are efficient, Facilitate the ability to move from one place to another. However, in addition to being a necessity, it is also a right, as stated in the Federal District Transport and Road Law, in its article 101 "Users have the right to the public transport service to be provided in a continuous, uniform manner, Permanent and uninterrupted and in the best conditions of safety, comfort, hygiene and efficiency. Faced with the exponential growth of the population in the ZMVM and the need for mobility, the project for the construction of a massive underground transport, proposed in 1965, was consolidated on April 29, 1967, through the creation of a public and decentralized body: of Collective Transport (STC). This system currently has 12 lines that cover 196,383 km and has an influx of just over 1.6 billion passengers annually. Based on the Origin-Destination 2007 survey, carried out by the National Institute of Statistics and Geography (INEGI), the majority of passengers use the STC as a means of connecting with other transfer options. In that same survey, it was identified that during the morning the passengers are added, at 06:00 a.m. start with 1.2 million passengers; Between 07:30 a.m. and 07:44 a.m., reaches the maximum of 1.9 million passengers, while at 09:00 a.m. in the morning reaches 1.3 million passengers.
According to the main aspects of the quality of the service of the STC, they are: the perception of the passenger with respect to the service, the capacity, the speed, the regularity, accessibility in stations and security. However, based on the results of the survey "one year of price increase to $ 5.00 M.N. By ticket "(closed at 9:00 p.m. on January 12, 2016) by the Political Animal news portal, it was observed that 79% of the people surveyed consider that the STC service is reflected in images related to agglomerations and over demand, as shown in Figure 1.

Figure: 1 Agglomerations and over demand of passenger in the STC

Together with a team of Department of System and the team of Engineering and New Projects Department of the STC, the various reasons that affected the operation were reviewed, which directly impacts the operation. Time of transfer of the passengers and therefore to the quality of the service.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The determination of the time of permanence to board the trains in mass transport systems like the STC, at a global level, has been analyzed from the point of view of the structure and design of the platform, studies have also been made of the systems that work on scheduling of the arrival of the train to the stations, under the assumptions of output frequency and capacity of the trains constant. The objective of this research is to simulate the operation of the Pantitlan terminal station using the software AnyLogic™, measure the boarding time of the passengers and generate a tool in which it can be varied: the arrival rate of passengers, the number of passengers that board the train, the opening time of train doors and the frequency of arrival of trains, to assist the decision making of the operating management.
This paper is prepared as follows: in section 2 the conceptual model of Pantitlan terminal station is presented, in section 3 the construction model is shown, in section 4 the simulation model of Pantitlan terminal station is implemented using AnyLogic™ software. Concluding remarks are drawn in Section 5.

2. THE CONCEPTUAL MODEL OF PANTITLAN STATION

Figure 2 shows the conceptual model of the Pantitlan terminal station, which analyzes the interaction of passengers and trains. Pantitlan terminal stations is a terminal and the most important transfer station because it has correspondence among lines one, five, nine and A. We focus on terminal Pantitlan terminal station line one, which was the first line opened, nowadays this line travels from the Pantitlan terminal station to the Observatorio terminal station. On the left side, the activities of the passengers to board the train are indicated and the right side describes the activities of the train to transport the passengers.

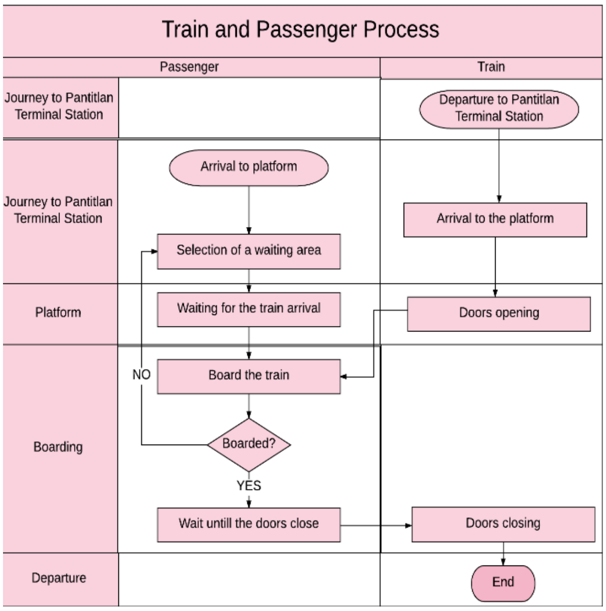


Figure: 2 Conceptual model of the Pantitlan terminal station

The entrance to the station by the passengers is done by four fixed stairs that are found throughout the station. The staircase leading to the first train cars is assigned for the exclusive use of women, children under 12 years and people with different capacities, and the rest of the train for use by the whole public, the procedure for boarding the train happens from the same way in both sections. The train, on the other hand, starts when it goes to the Pantitlan terminal station and up there, it remains in the platform of the station with the doors open so that the passengers can carry out the descent and ascent of passengers, if it comes from the Zaragoza station heading to the Pantitlan terminal station; or, only during the ascent of passengers when the train arrives from the train station to the Pantitlan terminal station.
In this station, the trains can start the race in two ways, one is by means of the trains that leave the train station and go to the Pantitlan terminal station, and the second is by means of the trains arriving from the Zaragoza station to the terminal station Pantitlan to finish and immediately start a new race, so passengers can board on both sides of the platform, as shown in Figure 3.

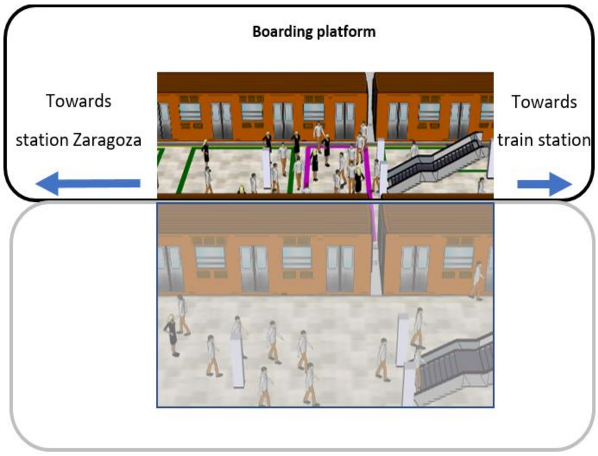


Figure: 3 Waiting area to board

The interaction between the trains and the passengers is on the platform. After the passenger arrives at the station, he goes to a waiting area and waits until the train arrives and opens the doors. Due to the capacity and demand of the system, users waiting on the platform may or may not board the train, in case they cannot climb, they repeat the process of choosing a waiting area, waiting for the train to arrive and attempting to board again to see it. It should also be noted that, as the first station, passengers choose to wait for the next train in the hope of reaching a seat. The boarding ends when the user leaves with the train to the next station. The measurement about the trains is made by cycles, composed by the time that is expected by the next train, plus, the train time on the platform. The time of arrival runs from the train on the platform closes the doors for the descent and ascent of passengers. One cycle includes both the time of arrival and the time on the platform.

3. DEVELOPMENT OF THE MODEL USING THE MOSASCOM METHODOLOGY

Passengers and trains are dynamic entities that move through the system, so they will be the agents of the system. The simulation model shows the dynamics of the door with the greatest number of passengers, located in the area where most men travel; however, women also travel, therefore the attributes used are men and women, being predominant that of men. The resource in our system is the capacity of passengers entering the section of the wagon analyzed by the door of greater influx. The decision variables are: the frequency of trains, the arrival rate of the passengers, the opening time of doors and the capacity of the train. For the characteristics described above, it was concluded in the construction of a hybrid model based on the Simulation Model-Based Methodology for Complex Systems Analysis, MoSASCoM, to guide the modeling and simulation process to analyze emerging properties to certain initial conditions. In this case, the problem is the complexity of the measurement of the time of boarding of the passengers of the STC, which cannot be determined by analytical methods, due to its process of self-organization among its elements; an important feature of the study of complex systems, Huerta (2016). The methodology is constituted by five defined stages: 1. Base question; 2. The development of the model (DEMO); 3. The simulation of the model (SIMO); 4. Analysis of the simulation model (AMSI); and 5. The documentation of the simulation model (DOMSI).

The first of five stages, is the base question, which for this study is What is the boarding time of passengers? The second stage is the development of the model, DEMO, whose objective is the design and development of the real system model through the recursive cycle of five phases: 1. Level of abstraction. 2. Perspective, 3. System operations, 4. Potential approaches to modeling and 5. Model communication (Huerta 2016). The analysis was about the door with a greater influx of both sides of the platform, which is located at the end of the first access stairs to the platform. In the software, the plane of the station was represented by placing the rails and the platform of passengers, whose length is 150 meters, along which are distributed the nine wagons of the train and four doors for each wagon. For the animation and representation of the doors we used the SketchUp 2016 animation software trial version. In order to obtain information, the area of interest separated from the rest of the animation was performed independently. For this purpose, the decomposition was carried out first and then the synthesis, as indicated by the MoSASCoM approach. The values of the parameters of the frequency between trains, the rate of arrival of passengers and the time of opening of doors raised in field, that were used for the simulation were realized through a data acquisition, during the fourth quarter of 2016, beginning, intermediate and weekend of business days, from 07:00 a.m. at 09:00 a.m. for the operations of the system, the operational flows of the arrival of trains to the platform and arrival of arrival of passengers were modeled.

The third stage is the simulation of the model, SIMO, which is the implementation of the conceptual model in a computer. For the choice of software, we reviewed and compared simulation software published with the name "Simulators Comparison V.1" of the year 2014 in Critical Manufacturing S.A. The tools that compare this publication are: Arena, FlexSim, AnyLogic, Simul8 and Internal Simulation, from which AnyLogic was selected, because it can integrate the simulation based on agents and processes, achieving the hybrid simulation through its Libraries and logical flow of processes. The implementation of the conceptual model, from the Pantitlan terminal station of Line 1 of the STC, was realized in the software AnyLogicTM by synthesis and by decomposition, using rail, process modeling analysis and pedestrian libraries. By means of the animation interface the platforms were built to scale; through the logical process, the objects were implemented, through which the simulation model was programmed; and the graphs show the result of the interest measurement with its respective mean square error (MSpE, Mean Square Pure Error), see Figure 4.

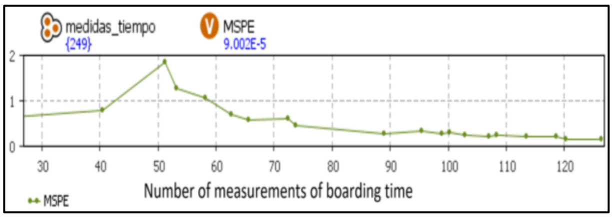


Figure: 4 MSPE Curve

4. THE SIMULATION MODEL

We implemented the simulation model of pedestrian and trains activities in the subway station using discrete-event simulation (DES) and Methodologic Based on Simulation Models the Analysis of Complex System (MoSASCoM). The first one uses the top-down perspective while the second one, uses the recursive perspective. So, in this simulation model we combine two modeling perspectives. Then we implemented the model using AnyLogic software. A tool was developed using the software Anylogic, which allows simulating the process of boarding passengers at the Pantitlan terminal station of Line 1, the measurement area is at the door of greater influx. In the right part of the window is a configuration area that allows establishing seven initial conditions of the simulation:

1. Passenger per minute measuring gate, which is the number of passengers arriving in one minute to the measuring gate. You can select a fixed rate or a Poisson distribution that generate random values with lambda equal to the value placed.
2. Duration of doors opening, in which it is possible to configure the time the train keeps the doors open waiting for boarding passengers. The values can follow the lognormal, exponential, uniform distribution; or, a fixed rate, depending on the option that is chosen.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

3. Time of arrival of trains Zaragoza Pantitlan: It allows to select between four distributions: Lognormal; Exponential; Poisson and Uniform. Depending on the distribution selected, the editable tables will be displayed to capture the parameters of the distribution. The selected distribution will determine the arrival values of the trains to the platform.
4. Arrival time of trains Zaragoza Pantitlan, is the time between trains arriving by the middle route, h2, and can be enabled or disabled. The options to choose are lognormal distribution, Exponential, Poisson, Uniform; Or fixed rate, depending on the selected option, the editable tables will be displayed that allow capturing the parameters of the same.
5. Arrival time of trains Pantitlan Zaragoza, is the time between trains arriving by the upper route, h1, and can be enabled or disabled. The options to choose are lognormal distribution, Exponential, Poisson, Uniform; or fixed rate, depending on the selected option, the editable tables will be displayed that allow capturing the parameters of the same.
6. Number of passengers that climb, is the number of people who decide to board the train at the measuring gate. In this case, you can choose the Poisson distribution, Uniform; or a fixed rate.
7. Number of passengers per minute, allows the simulation of the arrival of passengers along the platform. In this case the passengers arrive by four stairs of access to the platform of boarding, and in each one of them can be configured the amount of people that arrive per minute. The selected value will be used to generate random values that follow a Poisson distribution with mean equal to this value.

The Figure 5 shows the interface screen with the fields that can be edited by the users.

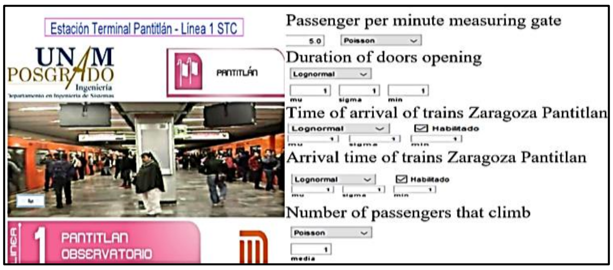


Figure: 5 Interface screen for users

The fourth stage of the MoSASCoM methodology is the analysis of the simulation model, AMSI. For verification, it is determined that the simulation model correctly reflects the conceptual model, and the context of the base question formulated, in addition to verification using the compiler included in AnyLogic. The execution time in the simulation is two hours, within which it is possible to observe in the graph of the MSpE, a square of pure error in the interface, the stabilization of the system response, see Figure 6.

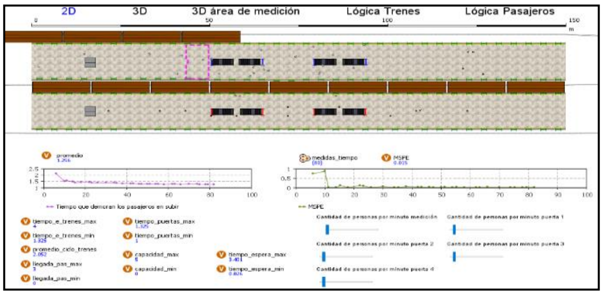


Figure: 6 Execution of the simulation model

4.1. Analysis of the simulation model

Through the design of experiments, the 2^k factor design was analyzed the impact of the variables at two levels, high and low. The factors of the simulation model are four: 1. Arrival rate of passengers, 2. Opening time of doors, 3. Time of arrival between trains and 4. Number of passengers that go up. Therefore, sixteen scenarios are required. Through the combinations of the high and low factors, the sixteen runs were executed with a simulation duration of two hours (120 minutes). The obtained results were: the time of boarding, the average time between trains and the average square of pure error. The latter increased considerably in the scenarios where the high level of passenger arrival rate was considered. In the graph of main effects for the result, it is possible to appreciate that the model presents greater sensitivity to the variable X_1, the rate of arrival of passengers, because depending on the value of this the time of boarding can be less than five minutes until greater to 30 minutes. The variable X_4, number of passengers that climb, also has a considerable impact on the response of the model, because as they board fewer passengers the time of boarding can reach the time it is reduced by approximately 20 minutes. The effect to the time of seam by X_3, time of frequency between trains, is approximately 10 minutes; And about five minutes, by the variable X_2, door opening time (see Figure 7).

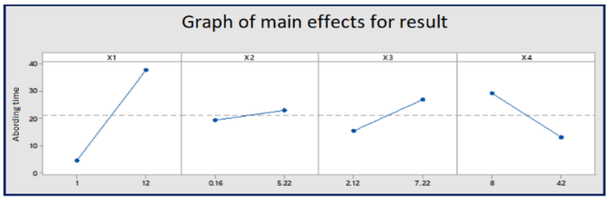


Figure: 7 Graph of main effects of boarding time

To obtain the results of the approach time of the terminal station Pantitlan, Line 1 of the STC, by principles of the *Central Limit Theorem*, it is known that while there is a considerable number of data, these will approximate a normal distribution. For this case, 50 executions were performed in the simulation model, the results were entered in Minitab software 17 and the Anderson-Darling test identified that the data follow a normal distribution. Besides the time of approach, the cycle time between trains was analyzed, in order to calculate in how many cycles the passengers manage to approach.

5. CONCLUDING REMARKS

The boarding time in the fourth quarter of 2016 was from 0.093 to 24.9893 minutes, indicating that passengers were up as soon as the first train arrived until the wait for six trains after their arrival.

The results for the month of October 2016 for the beginning of the week was on average boarding time of 1.3841 minutes, the passengers boarded in the first cycle of the arrival of trains. In the middle of the week, an average boarding time of 8.4181 minutes was obtained, so passengers waited for one cycle after having arrived and even up to five subsequent cycles. At the weekend of the month of October obtained an average time of 4.7604 minutes, reason why the passengers waited of one to two cycles at the most to be able to board.

For the month of November 2016, at the beginning of the week, an average boarding time of 3.3840 minutes was obtained, the passengers boarded in the first or maximum in the third cycle of arrival of trains. In the middle of the week an average value of boarding time of 12,014 minutes was obtained, so passengers expected an average of three cycles. At the weekend, an average boarding time of 6.0962 minutes was obtained, so the passengers waited for two train arrival cycles on average.

For the month of December 2016, for the beginning of the week, an average boarding time of 9.5636 minutes was obtained, the passengers boarded in the third cycle of arrival of trains. In the middle of the week an average boarding time value of 1.7771 minutes was obtained, reason why the passengers waited for a cycle after having arrived. For the weekend an average time of 4.4331 minutes was obtained, reason why the passengers waited of one to two cycles at most to be able to board.

With the above it is possible to say that to perform the simulation model in the Pantitlan terminal station of Line 1 of the STC by means of Methodology based on Simulation Models for the Analysis of Complex Systems, in this case, the time of boarding of the passengers, which under initial conditions, supported the development of the model with characteristics similar to the system under study; it was also possible to generate a tool that contributes to the decision-making of STC's operational management.

ACKNOWLEDGMENTS

The authors appreciate the partial support by National Council for Science and Technology of Mexico (CONACyT) (SNI number 65477). We are also grateful to Collective Transportation System (STC) for the facilities in the data collection, in particular to the team of Quality Assurance: Eng. Yolanda Carrillo Hernández, Eng. Agustín G. Patiño Acencio, Eng. Roberto M. Roman Mota and Ph.D. José Angel Bermejo Arenas.

REFERENCES

Auditoría Superior de la Ciudad de México, 2014. CXL. Informe final de auditoría, derivada de la revisión de la cuenta pública del Gobierno del Distrito Federal. Available from: http://www.ascm.gob.mx /IR/Informes/2732.pdf [accessed 02February 2017]

Huerta A., 2014. Metodología basada en Modelo de Simulación para el análisis de sistemas complejos (MoSASCoM), National Autonomous Mexico University. Available from: http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/6919/Metodolog%C3%ADa_basada_en_modelos_de_simulaci%C3%B3n_para_el_analisis_de_sistemas_complejos_.pdf?sequence=1 [accessed 12 April 2017]

Huerta A., 2016. Introducción a la Modelación y Simulación de Sistemas Complejos. 1st ed. Ciudad de México.

INEGI and GEM, 2007. Encuesta Origen-Destino 2007. Instituto Nacional de Estadística y Geografía – Gobierno del Estado de México. Available from: http://bicitekas.org/wp/wp-content/uploads/2013/07/2007_Encuesta_Origen_Destino_INEGI.pdf [accessed 07 May 2016]

STC, 2016. Sistema de Transporte Colectivo - Información pública. Available from: http://www.metro.cdmx.gob.mx/operacion/cifras-de-operacionmx [accessed 21 December 2016]

AUTHORS BIOGRAPHY

YAZMIN DILLARZA-ANDRADE currently is a master student of Systems Engineering, in the disciplinary field Industrial Engineering of the National Autonomous University of Mexico, Faculty of Engineering. She is interested in the optimization of industrial processes that impact on the financial resources and the management of the organizations, besides she is interested in the planning of policies and organizational actions that promote the development and growth of the same ones.

AIDA HUERTA-BARRIENTOS received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the School of Engineering, UNAM and she is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

GERALDO SALAZAR-DIAZ currently he is part of the team of the Electronic Instrumentation Laboratory for Space Systems at the UNAM, and specializes in the development of applications in micro controllers in diverse platforms like TIVA, DSP, PICs and Arduino. His areas of interest are the development of applications in SRAM FPGAs using Xilinx tools such as ISE and Vivado and the development of electronic applications for space systems in micro and nano satellites.

JOALI EVELYN PÉREZ - BONILLA Bachelor in Electric and Electronic Engineering, with major in energetic systems. Work experience as project engineer in solid residues degradation systems. Currently student of the master in system engineering at National Autonomous Mexico University (UNAM). Additionally, works as an external consultant in the enterprise wherein the current investigation took place. Her main interests are on business plans for small enterprises, her work experience and academic formation have led her to study the master in system engineering to obtain the necessary tools of a systemic thinking.

THE INTELLIGENT DECISION SUPPORT SYSTEM PROTOTYPE FOR PORT INFORMATIONAL INTEGRATION

Ana X. HALABI-ECHEVERRY[(a)], Sardar M. N. ISLAM[(b)]
Hugo L. NIÑO-VERGARA[(c)], Nelson OBREGON-NEIRA[(d)],

[(a),(c)] Escuela Internacional de Ciencias Económicas y Administrativas, Universidad de La Sabana, km 7 autopista norte de Bogotá, D.C., Chía (Cundinamarca), Colombia

[(b)] College of Business, Victoria University
City Flinders Campus, Level 13
PO Box 14428, Melbourne, Vic. 8001

[(d)] Instituto Geofísico, Universidad Javeriana,
Carrera 7 no. 42-27 Piso 7, Bogotá, Colombia

[(a, b)] ana.halabi, hugo.nino1@unisabana.edu.co, [(c)] nobregon@javeriana.edu.co [(d)] Sardar.Islam@vu.edu.au

ABSTRACT
This paper demonstrates the different set of values, queries and scenarios offered as a proof of concept of the Intelligent Decision Support System (*i*-DMSS) for port integration. In particular, we consider how an *i*-DMSS can support data and information integration across ports to deliver improved decision-making and outcomes. Each set of options can be saved to feed in the future a knowledge base with the choices made by the users. It constitutes an early prototype as a suitable visual schema for explaining in practical terms the number of scenarios that guide informational integration for ports.

Keywords: Intelligent Decision-making Support Systems, Early Prototype, Port Informational Integration, Port Strategic Decisions.

1. INTRODUCTION

Ports as any other organisation are facing challenging changes in their traditional ways of support decisions and the flexibility increasingly complex in their information systems. It is estimated that current information system must display environment strategies, norms, culture, behaviours and decisions that become increasingly difficult to be monitored, and are continuously affecting business processes and impacting operational strategic goals.

Information and Communication Technologies (ICT) in ports has traditionally focused on necessities at the operational level as a response to port-specific processes (Cetin & Cerit 2010; Mathew et al., 2005; Henesey, 2006). Electronic Data Interchange systems, Vessel Traffic Monitoring and Information systems, and Container Terminal systems are some examples. Vessel traffic monitoring and information systems (VTMIS) have evolved from website-based systems (Forward, 2003). In planning yard distribution and container layouts, ports generally use container terminal systems for managing the movement of cargo through terminals (Almotairi et al., 2011). Current Logistics and Transport Management and Collaboration systems mainly cover requirements of business-to-business (B2B) transactions.

Van Baalen & van Oosterhout (2009) discuss new necessities for IT in ports through information sharing, planning and execution in collaborative ways such as the called port community systems and the inter-organisational systems. By the use of advance ICT in ports, new technological dilemmas arise, such as the need for more intelligent support. The need to introduce intelligent support tools can cope with the complexity of global operations as pointed by Murty et al. (2005). Therefore, as these authors indicate, while current information systems may meet current needs, more intelligence is required to handle the growing complexity within the port domain. For example, information systems in the port domain rarely take advantage of Computational Intelligence technologies such as data-mining, knowledge-based systems and ontologies.

2. AN INTELLIGENT DECISION SUPPORT SYSTEM (*i*-DMSS) FOR PORT INFORMATIONAL INTEGRATION

Artificial Intelligence in its fusion with decision support systems (DSS) supports the prototype design for the (*i*-DMSS) port-to-port solution, that as to the best of the authors' knowledge, it is the first time for a solution of this type be offered. The proof-of-concept of the decision-aid tool, namely, the intelligent decision making support system for port informational integration (*i*-DMSS) was first presented in Halabi-Echeverry (2017). The aim of this paper is to propose the use of computational intelligence technologies to drive knowledge towards port informational integration. The port informational integration concept offered mainly refers to a higher perspective of port cooperation in which development of capabilities on sharing information, planning and execution allows two or more ports to advance and deliver benefits among the partners. The *i*-DMSS for port informational integration provide guidance to experts and decision makers. Cassaigne & Lorimier highlight that an important challenge for tactical/strategic or "non-programmable" decisions (in the words of Herbet Simon) places special emphasis in the DSS's future development. "Strategic decisions are mainly based on knowledge and gut feeling to answer a novel situation, in other words they are characterised by uncertainty and complexity (2006, p.402)". They propose interactions among the parts (human/technology) of an DSS to observe the complexity of the decision supported, i.e., decision maker and the expert knowledge (which sometimes do not reside in the decision maker) and the computational intelligent system.

The *i*-DMSS for port informational integration is meant to be used primarily by multilateral organisations involved in strategic global decision making. Public organisations such as IMO (International Maritime Organization), and private associations like IAPH (International Association of Ports and Harbors) demand a comprehensive integrated independent system to carry out follow-up and control of local developments that could have global influence in terms of economic, environmental, demographic or cultural performance. Major port operators such as HPH (Hutchison Port Holdings), PSA (Port of Singapore Authority); DP (Dubai Port World), APM Terminal (A.P. Moller) may find the *i*-DMSS for port informational integration useful to conduct data analysis based on a variety of aspects apart from the economical approach. The observance of law and regulations by this mega corporations will need a tool to objectively measure the impact of such acquisitions to support political decisions that could be influenced by private interests.

3. FUNCTIONALITIES

The *i*-DMSS for port informational integration promotes functionalities to respond to the next generation of intelligent decision support systems supporting data and information integration across ports to deliver improved decision-making and outcomes for the parties concerned.

The essential functionalities contended in the *i*-DMSS for port informational integration are:

- Integration of heterogeneous repositories,
- Exploitation of multiple learning algorithms,
- Metadata to enable future system automation and user support requirements
- Providing semantic interoperability

3.1 Integrating heterogeneous repositories,

Multiple data hierarchies are outlined as a proof-of-concept in this paper making special emphasis to the literature and the public available resources where they come from. Figure 1 shows various merged data hierarchies and performance indicators included in the *i*-DMSS for port informational integration. As the interrelated nature of these concepts may create complexity of computing, future efforts in this direction must be estimated. The available set of performance indicators is visualised by rectangles in yellow, estimated data-levels by rectangles in blue and measurements by rectangles in skintone. These relationships are not exhausted but an indication of the complexity of the data-driven approach dealt with in the system.

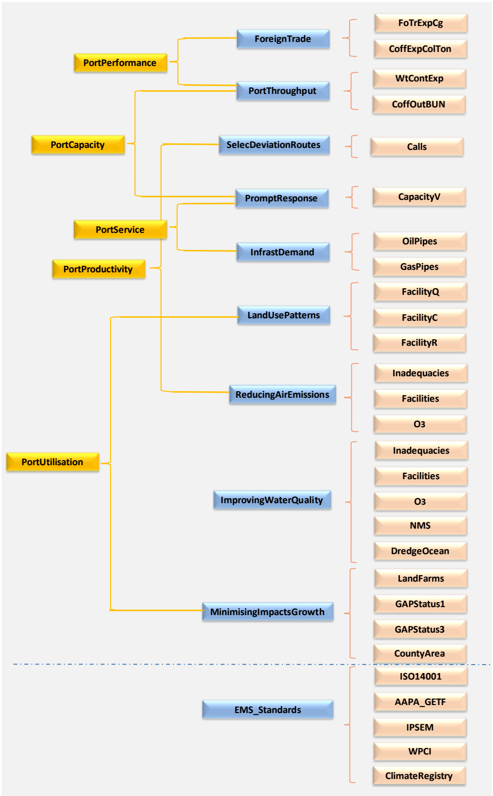


Fig. 1: Hierarchical diagram of concepts, factors and variables contented in the *i*-DMSS for port integration)

3.2 Exploiting multiple learning algorithms

This functionality refers to the general performance of a learner and its prediction capability. The learning space captures the relevant factors and measurement variables simplified during the data mining activity. The performance of learning algorithms is determined by dataset characteristics and algorithms.

Feature Selection (or variable selection) refers to a task that can be formulated as an optimization problem and used with learning algorithms of classification (or clustering). Kacprzyk & Pedrycz (2015) define feature selection in terms of three approaches: filter, wrapper or embedded (p.1216):

The concept of metafeature refers in general to "statistics describing the training dataset of the problem, such as number of training examples, number of attributes, correlation between attributes, class entropy, among others (Prudêncio et al. 2011, p.226)". Thus, metafeatures define the main properties of a specific dataset using a process known as a learning task, which should be computed in a data-driven way. Castiello & Fanelli (2011) state that Metafeatures must satisfy two basic conditions: "Firstly, they should prove to be useful in determining the relative performance of individual learning algorithms. Secondly, their computation should not be too difficult and burdensome (pp.163-164)".

An important issue in knowledge discovery regards to finding the finest classifier. Using the automatic system construction wizard in Rapid Miner 5.0®, the metalearning classification is a straightforward process. This wizard also aids evaluating each classifier and finding an optimal parameterisation for the dataset at hand. No single learning algorithm will construct hypothesis of high accuracy on all problems.

The automatic system process of Rapid Miner 5.0® better referred as the PaREn (Pattern Recognition Engineering) system, makes possible to obtain an overview of the performance of a classifier over different datasets. It also includes preprocessing when necessary, i.e., normalisation, discretisation, or missing value replenishment and parameter optimisation setups (Shafait et al., 2010). The success of an automatic pattern recognition is also due to the metafeatures of the datasets for metalearning.

PaREn evaluates the accuracy of the following classifiers on the datasets for cases included in the *i*-DMSS for port integration: 1) the learner supervised rules – OneR, 2) naïve bayes, 3) support vector machines, 4) knearest eighbor, 5) neural networks, and 6) random forest. The evaluation uses a crossvalidation technique along with the root mean squared errors (RMSE) for each classifier. Table 1 shows the corresponding evaluation.

Results of this process suggest Random Forest as the learning algorithm that better performs for the datasets at hand. It has the highest accuracy among the considered classifiers with an acceptable RMSE that brings confidence to the prediction. This special output of Rapid Miner 5.0® is based only on metafeatures.

Table 1: PaREn results for customised datasets on port informational integration and corresponding evaluation

Dataset	# Observ	#Attr+Class	OneR	Naïve Bayes	Support Vector Machines	K-Nearest Neighbors	Neural Networks	Random Forest
Accuracy (Cross-Validation)								
US West coast, the Gulf and Atlantic coasts Dataset (Case 1)	44	27	0.614	0.273	0.636	0.636	0.636	0.682
RMSE			0.083	0.083	0.068	0.056	0.078	0.084
US West coast, the Gulf and Atlantic coasts Dataset (Case 2)	44	26	0.727	0.932	0.932	0.932	0.955	1.000
RMSE			0.083	0.083	0.068	0.056	0.078	0.084
Rjin Schelde Delta Dataset (Case 3)	29	27	0.723	0.862	0.964	0.969	0.964	1.000
RMSE			0.083	0.083	0.068	0.056	0.078	0.084

3.3 Entail metadata to enable future system automation and user support requirements

As said, metadata serves as training and evaluation data for new learning processes (Hilario et al. 2011). This brings an advantage over black box systems giving the user the control and flexibility necessary to combine learning with experience. The metadata is organized in a hierarchy scheme using colours which demonstrate the relationships that may exist between the data elements. Further will be explained that although the *i*-DMSS for port informational integration prototype provides the fixed baseline for those hierarchies, in the future it would be desirable allow the user to interact with the hierarchies using his experience to redefine or confirm the baseline. Figure 2 shows the metadata relationships and Table 2 provides the metadata identification. Each metadata element is provided with a

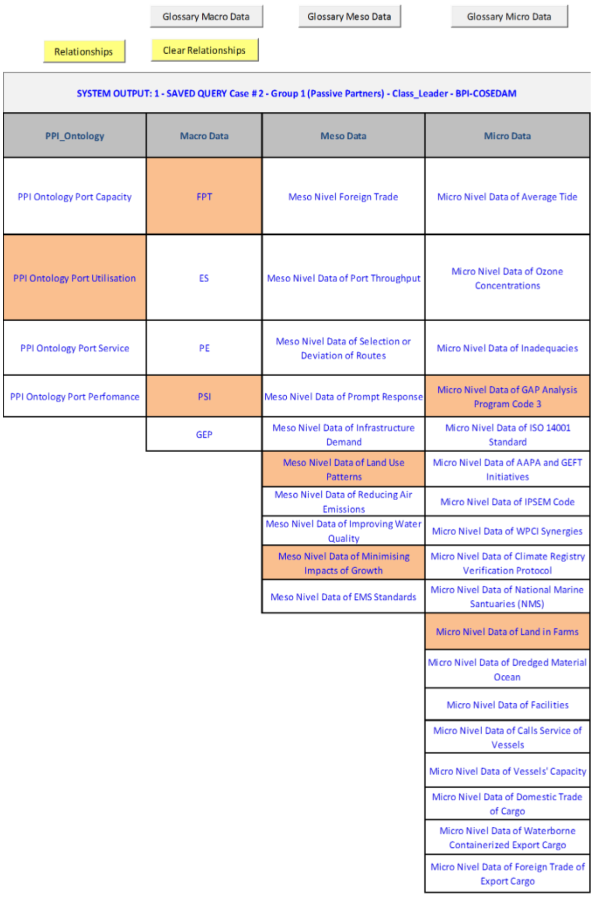


Fig. 2: System Output 2: Flexible Metadata Visualisation

mandatory or optional label which indicates if analyses are subordinate or not to that element. Users can offer other options to interpret the processes embedded in the system. This output available for the user is the flexible metadata produced in the different modelling steps.

Table. 2: Metadata identification

Identifier	Description
dc.title: PortName	*a name given to a port*
dc.title: Latitude	*valid values range between [-90.0,90.0]*
dc.title: Longitude	*valid values range between [-180.0,180.0]*
dc.title: Type	*refers to whether the port is seaport, river or deepwater*
dc.title: Size	*refers to whether the port is large, medium or small*
dc.title: Region	*refers to the maximum level at which data aggregation is useful for analyses*
dc.title: State Jurisdiction	*refers to the general legal competence of countries over their ports*
dc.title: LoCoast	*refers the coastal frontages over which ports are located*
dc.title: Context	*refers to the geographical scope for port integration*
dc.title: Port State Jurisdiction	*refers to the control port area endorsed and accredited in recent years at the international and local level*
dc.title: County	*refers to the narrower legal competence of counties over their ports*
dc.title: Time_Year_Scale_of_Interest	*refers to the time scale of interest (year) to run the queries*
dc.title: Time_Month_Scale_of_Interest	*refers to the time scale of interest (month) to run the queries*
dc.title: Important_events_Outliers	*refers to important events identified through the outliers'' analysis*
dc.title: ClusterStatus	*refers to the cluster status given to a port*
dc.title: EMSstatus	*refers to the EMS status given to a port*
dc.title: Str_Governance	*refers to strategy to integrate information with a port partner where normative and procedural pressures and actions take place*
dc.title: Str_TransportInterconnections	*refers to strategy to integrate information with a port partner belonging inter) organisational network where rational use of coastlines and their demands places special emphasis on.*
dc.title: Str_LogisticsFunctions_Operations	*refers to strategy to integrate information with a port partner where higher purchasing power and consumption levels tend to foster port development.*
dc.title: BPI_Environmental&EcologicalSustainability	*The BPI to promote port integration in cooperative decision-making on environmental and ecological sustainability*
dc.title: BPI_OrganisationNetworking	*The BPI to promote port integration in collaborative decision-making on transport or (inter) organisational networks*
dc.title: BPI_PortLogisticsPerf_Economics	*The BPI to promote port integration through value-added analyses on port performance in terms of economics*
dc.title: tide	*refers to the tide mean current rates (for vessels approaching and mooring)*

3.4 Providing semantic interoperability

The *i*-DMSS for port informational integration uses artificial intelligence to describe interoperability matters in heterogeneous repositories and data (or metadata), and the exchange/use of information such as content, format, semantics (ontologies) and defined standards. Semantic interoperability deals with meaningful and precise exchange and sharing of information. Technologies at this stage include metadata and ontologies.

There are a certain number of standards and technologies needed to achieve an enterprise integration and interoperability. It includes standards and technology for interoperability such as: the eXtended Mark-up Language (XML), Hypertext transfer protocols (HTTP/HTTPS), Web Services and Service-Oriented Architectures (SOAs), and in recent times, Predictive Modelling Mark-up Language (PMML)., XML is a widely used, standardised tagged language proposed and maintained by the World Wide Web Consortium (W3W). It has been proposed to be a universal format for structured content and data on the web but can indeed be used for any computer based exchanged. On then other hand, PMML is recently the most common approach to go towards XML-based formats.

The systems interoperability is a challenge posed for Inter-Organisational Systems (IOS) in ports. New technologies are meant to enable information exchange, planning at a higher level after the exchange of information, real-time chains and seamless communication between stakeholders.

The exchange/use of semantics (ontologies) is a component of the *i*-DMSS for port integration hat draws into conceptualisations on port performance indicators (PPIs) through the efficient use of data hierarchies. Differences between PPIs demand hard work for understanding the aggregation of the information in which they are based on; additionally, they are difficult due to the diverse number of methods for their calculation which is essential for decision-making. A single PPIs interpretation is almost impossible. No one measure will suffice, as the differences between ports and the interrelated nature of the metrics create multiple possible interpretations for single data elements.

For instance: In principle every port could be developed to its maximum capacity, reasoning about the description of port capacity comes to the relationship between vessels' capacity (CapacityV) and waterborne containerized export cargo (WtContExp) that can be measured with significant differences and variations per port. Illustration 1 presents one possible description for this relationship:

Illustration 1: Partial Ontology PPI: Port Capacity

OWL:
Class (PortCapacity partial
DataLevels
restriction (***hasA*** amongst other things some values From CapacityV)
restriction (***hasA*** amongst other things some values From WtContExp)
Paraphrase:
PortCapacity has *amongst other things, some* values from vessels' capacity in DWT (CapacityV) and also *some* values from waterborne containerised export cargo (WtContExp)

The ontological description given is provided as a first step to guide future development of a complete semantic model in the *i*-DMSS for port informational integration. Analyses rely on regional and aggregate statistical data to guide the decision maker on daily-basis.

4. PROTOTYPING

An early prototype has served for the purpose of showing some of the explained functionalities:

- **Visual Function: The State's Jurisdiction Choice**

The State's Jurisdiction visual function allows the user to make a choice on one or more territorial boundaries where ports exercise governance and managerial functions, moving from simple to complex outputs for analyses.

- **Visual Function: The Business Process Intelligence Choice**

Once the user has selected the State's Jurisdiction under analysis, options of Business Process Intelligence (BPI) are detected according to the embedded data mining workflows in the system.

- **Visual Function: Selecting Values out of each BPI**

The function on selecting values from each BPI allows the user to further drill down his analyses to a level in which classifications, groupings and/or forecasting are displayed. Although, the information accessed so far is static, i.e., cannot be replaced by the user, it indicates the logical sequence of the BPIs for analyses. It also indicates the targets to be accomplished, for instance, if the user is searching ports with leadership characteristics for competitive purposes or average behaviours to fulfill strategies on Corporate Social Responsibility (CSR).

- **Visual Function: Selecting the period of interest**

The function on selecting the period of interest allows the user to constrain the analyses to a particular period of time.

- **Visual Function: Selecting Possible Maps and or Schemas for Visualisation of Queries**

This function allows the user a simple visual identification of the query fields by pressing the option 'show maps'. In the future it is expected the user interacts with spatial and georeferenced information for each field. Two types of visual schemas are available: fixed maps and fixed forecasting reports.

- **System Ouput 1: Saved Query Report**

Once the user has completed his query, he received a confirmation for all choices. The output report allows the user to verify the selections made. It is possible to use glossaries at the upper part of the report.

- **System Output 2: Flexible Metadata Visualisation**

The second output available for the user is the flexible metadata produced in the different modelling steps. This brings an advantage over black box systems giving the user the control and flexibility necessary to combine learning with experience. The metadata is organised in a hierarchy scheme using colours which demonstrate the relationships that may exist between the data elements. Although the prototype provides the fixed baseline for those hierarchies, in the future it would be desirable allow the user to interact with the hierarchies using his experience to redefine or confirm the baseline.

- **System Output 3: Visualisation of What-IF Scenarios**

Finally, the user is provided with different rules originated from the data mining and analytical workflows. The rules are given names to ease the user's understanding of them.

5. SCENARIO ANALYSIS

Scenarios, demonstrations and examples have been developed to encourage the port authorities and other decision makers to utilise the tool. This subsection seeks to characterise one application case considered for port informational integration and show the effective knowledge in decision-making and the necessary assistance in understanding diverse and complex situations for port informational integration in the US West Coast.

A small concentration of ports, among are: Seattle, Oakland, Tacoma and Portland are showing mainly differences on the vessels' capacity (CapacityV) served by the port. Variables such as waterborne containerised export cargo in twenty-foot equivalent units (WtContExp) can be considered important for grouping 'mega-ports' and therefore, play a less important role for grouping medium-sized ports. The decision-making elements are concerned with the ability to integrate information with a port partner defining new port boundaries for the purpose of sustainability involving ecosystems, normative, systemic and procedural dimensions. The regulatory function of these ports has led port authorities to face high pressures to become accredited and internationally recognised. Moreover, a number of environmental measures produced by agencies and local administrative authorities, are difficult with respect to decision making, and as a result with defining strategies to understand the consequences of cooperation between ports.

Rule 06 in Figure 3 ratify that Portland and Seattle use benchmarks and standards becoming aware of the scope and impacts of their activities. They have done well in reducing air emissions, although there is a warning to Portland probably because of its activities near a water-base river basin. Oakland is in a monitoring stage for its air emissions and water quality conditions. All of them are of medium size belonging to the group of 'passive partners' (rules 11, 13 and 17). Certain rules' names may be duplicated but the outcome of the rule is slighty different.

SYSTEM OUTPUT: 2 - SAVED QUERY Case # 2 - Group 1 (Passive Partners) - Class_Leader - BPI-COSEDAM

PORT PARTNER (What-IF Scenarios)	Activated Rules				
Oakland	R05_Improving Water Quality Monitoring	R08_Reducing Air Emissions Monitoring	R11_Passive Partner_WtContExp	R13_Passive Partner_CapacityV	R17_Passive Partner_CapacityV
Portland	R06_Reducing Air Emissions Toll Free	R013_WholeVerificationWarning	R11_Passive Partner_WtContExp	R13_Passive Partner_CapacityV	R17_Passive Partner_CapacityV
Seattle	R06_Reducing Air Emissions Toll Free	R11_Passive Partner_WtContExp	R13_Passive Partner_CapacityV	R17_Passive Partner_CapacityV	

Fig. 3: System Output 2: Rules Visualisation

Four blocks of information about this case are obtained in Figure 4. The first block shows partners with their coordinates, coastal locations and role in an existing or potential cluster (i.e., an initiative or passive role for port informational integration). The second block shows the common interests for cluster formation. These refer to the identification of a partner port for sustainable development. For the pair of ports Seattle and Portland, a more positive outcome given the established actions (i.e., the reduction of air emissions in the area) indicate those ports can lead mutually advantageous actions in this direction. The third block shows cluster similarities (variables in common among the ports). Ports of Seattle and Portland are characterised by a lower infrastructure capacity and throughput. The cluster differences indicated in the fourth block are featuring the fact that in domestic trade there is a slight imbalance for the ports.

PORT	LATITUDE - N	LONGITUDE - W	LoCoast	ClusterRole
Portland OR	45.56845	-122.73899	Pacific	Passive
Seattle	47.6062095	-122.3320708	Pacific	Passive

PORT	Common interests for cluster formation		
Portland OR	R06: Toll free(keep going) on the reduction of air emissions		
Seattle	R06: Toll free(keep going) on the reduction of air emissions		

PORT	Cluster Similarities			
			waterborne containerised export cargo	*waterborne containerised export cargo*
Portland OR			R11: less than 706, 500 TEUs	R17: less than 131 million deadweight tons (dwt)
Seattle			R11: less than 706, 500 TEUs	R17: less than 131 million deadweight tons (dwt)

PORT	Cluster Differences			
	vessel calls	*domestic trade of cargo*		
Portland OR	R7: up to 2, 588 calls	R7: greater than 6.9bn U$ dollars		
Seattle	R8: up to 2, 588 calls	R8: greater than 5.5bn U$ dollars		

Fig. 4: System Output 2: Rules Visualisation

6. CONCLUSIONS

The role of this paper is not just to accept the port's technological status-quo, but also to identify what new tools may be required to support strategic decision-making of port managers/authorities. It demonstrates the conceptual i-DMSS through prototyping and adding some explanation of how it would support real port informational integration. At this stage the prototype is offered as a proof of concept.

The *i*-DMSS functionality is tested through a set of values, queries and scenarios that contribute to the identification of design choices under which the prototype for port informational integration may work. Each set of options can be saved to feed in the future a knowledge base with the choices made by the users. This knowledge-driven perspective offers to the community and practitioners the ability to learn from the metadata and metafeatures to build intelligent models for port informational integration that support the prototype design for a port-to-port solution, that to the best of the authors' knowledge, is the first time for a solution of this type to be offered.

This paper reveals inconsistencies in the terminology used in the port domain and suggests an accurate use of terms and links between attributes to allow efficient data mining and consequently decision support process. The aim is twofold: create an illustration of the data-level concept for port integration and describe semantically key data contended in the *i*-DMSS for port informational integration such as: port capacity,

7. FURTHER WORK

We are required to overcome current concerns about the *i*-DMSS for port informational integration update and data management as well as the limitations and complications that may rise adopting an easy-to-use platform available online.

In defining the *i*-DMSS modular development some considerations need to be made. The *i*-DMSS modules will require to provide a guide to describe the decisions and challenges simultaneously to decision makers and developers to incorporate the decision-making side and engineering requirements.

Looking to the future implications of this research, the author estimates a new view of these information systems will offer to the port decision makers an opportunity to integrate their information, and informing stakeholders on relevant issues.

LIST OF REFERENCES

Almotairi, B., Flodén, J., Stefansson, G., & Woxenius, J. (2011). Information flows supporting hinterland transportation by rail: Applications in Sweden. *Research in Transportation Economics*. doi: 10.1016/j.retrec.2011.08.003

Cassaigne, N., & Lorimier, L. (2006). A Challenging Future for i-DMSS. *Intelligent Decision-making Support Systems*, pp. 401-422. Springer London

Castiello, C., & Fanelli, A. (2011). Computational Intelligence for Meta-Learning: A Promising Avenue of Research. In N. Jankowski, W. Duch & K. Grąbczewski (Eds.), *Meta-Learning in Computational Intelligence* (Vol. 358, pp. 157-177): Springer Berlin Heidelberg.

Cetin, C. K., & Cerit, A. G. (2010). Organizational effectiveness at seaports: a systems approach. *Maritime Policy & Management*, 37(3), 195-219.

Forward, K. (2003). Recent Developments in Port Information Technology. In K. Jeffery (Ed.). London, U.K.

Halabi-Echeverry, A. X. (2017). Computational Intelligence for Development of Strategic Decision Making in Port Informational Integration. PhD Thesis. Macquarie University, Australia.

Henesey, L. E. (2006). Multi-Agent Systems for Container Terminal Management. Blekinge Institute of Technology, Sweden.

Hilario, M., Nguyen, P., Do, H., Woznica, A., & Kalousis, A. (2011). Ontology-Based Meta-Mining of Knowledge Discovery Workflows. In N. Jankowski, W. Duch & K. Grąbczewski (Eds.), *Meta-Learning in Computational Intelligence* (Vol. 358, pp. 273-315): Springer Berlin Heidelberg.

Kacprzyk, J., & Pedrycz, W. (2015). Introduction. In J. Kacprzyk & W. Pedrycz (Eds.), Springer *Handbook of Computational Intellig*ence, pp. 1-4. Springer Berlin Heidelberg.

Mathew, R., Leathrum, J. F., Mazumdar, S., Frith, T., & Joines, J. (2005). An Object-Oriented Architecture for the Simulation of Networks of Cargo Terminal Operations. *The Journal of Defense Modelling and Simulation: Applications, Methodology, Technolog*y, 2(2), pp. 101-116. doi: 10.1177/154851290500200205

Murty, K. G., Liu, J. Y., Wan, Y. W., & Linn, R. (2005). A decision support system for operations in a container terminal. *Decision Support Systems*, 39(3), pp. 309-332. doi: 10.1016/j.dss.2003.11.002

Prudêncio, R. C., Souto, M. P., & Ludermir, T. (2011). Selecting Machine Learning Algorithms Using the Ranking Meta-Learning Approach. In N. Jankowski, W. Duch & K. Grąbczewski (Eds.), *Meta-Learning in Computational Intelligence*, 358, pp. 225-243. Springer Berlin Heidelberg.

Shafait, F., Reif, M., Kofle, C., & Breuel, T. (2010). Pattern Recognition Engineering. In proceedings RapidMiner Community Meeting and Conference. Retrieve from http://madm.dfki.de/publication&pubid=4947

van Baalen, P.J. & van Oosterhout, M. (2009). IOS Adoption and Implementation in Global Port Communities. In R.A. Zuidwijk, P.J. van Baalen & J.A.E.E. van Nunen (Eds.), Port Inter-Organizational Information Systems: Capabilities to Service Global Supply Chains (pp. 57-72). Boston - Delft: Now Publishers.

Varga, J., Romero, O., Pedersen, T., Thomsen, C. (2014). Towards Next Generation BI Systems: The Analytical Metadata Challenge. In L. Bellatreche & M. Mohania (Eds.), *Data Ware-housing and Knowledge Discovery*, 8646, pp. 89-101. Springer International Publishing.

AUTHORS BIOGRAPHY

Halabi-Echeverry, Ana. X is the Director of the Master in Operations Management and a Professor in the School of Economics and Administrative Sciences at La Sabana University in Colombia. She is also an Associate Professor of the Decision Sciences and Modelling Program at Victoria University. She completed a PhD in Computing Sciences from Macquarie University in Australia. Her research interests include: Intelligent Systems and Information Science, Sustainable Supply Chains, Governance and Business Ethics. She has published in academic peer-review journals such as Springer Special Editions, Taylor and Francis, Elsevier and IEEE Publications.

Professor **Islam,** Sardar M. N. is Professor of Business, Economics and Finance; and Director, Decision Sciences & Modelling Program, Victoria University, Australia. One of the areas of Professor Islam's specialisations is Applied Managements Science/ applied quantitative modelling. He is currently undertaking research, teaching and doctorate supervision work in Applied Management Science addressing issues in a wide range of disciplines in accounting, economics, finance, business, and law. Many university libraries around the world including all top universities such as Harvard, Cambridge, etc. have many of his books. He has published many articles (total about 200) including a good number of journal articles in international journals.

HOLISTIC PLANNING OF PRODUCTION AND INTRALOGISTICS SYSTEMS THROUGH AUTOMATED MODELING WITHIN AND AMONG THE TOOLS OF THE DIGITAL FACTORY

David Weigert[(a)], Paul Aurich[(b)], Tobias Reggelin[(c)]

[(a),(b),(c)] Otto von Guericke University Magdeburg, Magdeburg (Germany)

[(a)]david.weigert@ovgu.de, [(b)]paul.aurich@ovgu.de, [(c)]tobias.reggelin@ovgu.de

ABSTRACT
The automated and semi-automated model generation has been discussed and developed for decades. With AutomationML (AML), an open, object-oriented, XML-based storage and exchange format is provided, which should also allow an exchange among visualization, construction and simulation tools. Separate cross-section functions and proprietary software solutions of the individual tools make it difficult to define a common transfer point. The presented concept and tool describes the development of an application-oriented and source-open middleware. The focus of the ongoing implementation phase is the development of these uniform, digital planning methods and tools by AML. The first prototype outlines the advantages but also the disadvantages of automatic model generation via AML. Up to now simple conveyor systems from straight, curved conveyor belts, rotary tables and single stations from the visualization tool Tarakos – taraVR can modelled automatically in the simulation tool Siemens – Plant Simulation and construction tool Autodesk – AutoCAD.

Keywords: automated and semi-automated model generation, AutomationML (AML)

1. INTRODUCTION
The versatility, speed and flexibility of the product creation process is increasing due to the ever-shorter product life cycles. This development has a direct impact on the participating logistics and production processes (Schenk 2014). The increasing digitization and automation accelerates the development of the continuity of available data models for the digital production and logistics. The use of digital tools in the areas of simulation, visualization and construction improve the quality of planning, increase efficiency and shorten the product development and launch (Schenk 2014; Daft 2016; Klepper 1996; Lüder and Schmidt, 2015). These benefits can only be fully used if it is enabled to translate all relevant and so far isolated digital methods and tools into an integrated planning system (Faltinski 2011; Schreiber and Zimmermann, 2011). Currently used tools cover only specific functional areas within the product lifecycle management (PLM). This focus on individual areas of application of PLM enables a high level of specialization. The tools include the possibility of simulation, visualization and construction. However, the significance of these functions is limited with respect to the specialization of the users (Figure 1). This limited solution spaces are created by a lack of interface integration. An exchange of data in a heterogeneous system environment is therefore limited (Faltinski et al. 2012; Rawolle et al. 2002). In practice, therefore, no holistic, neutral and IT-based approach has been developed for the integrated digital planning and control of intralogistics systems and production areas. It lacks a neutral exchange format for continuous availability and mutual availability of simulation data, geometric design points and visualization elements.

tool for: / ability to:	simulation	visualization	construction
simulate		◐	○
visualize	○		◐
construct	○	○	

○ low ◐ medium ● high

Figure 1: Ability of visualization, construction and simulation in the different tools

The ongoing implementation phase discusses the concept and tool and first results of establishing continuous modeling between simulation, visualization and construction tools. The further course discusses the concept and tool for continuous model building on the example of the exchange direction of visualization to simulation and visualization to construction.

2. STATE OF THE ART AND SCIENCE
The standardization of systems, processes and their components is an essential element in the mastering of complexity as well as control and structuring of future, digital challenges (Drath 2010; Eigner and Stelzer 2009). The previous focus was the combined planning phases on a shortening in the time-to-market of products through the integration of product, process and production system planning, entitlement to use today is

at an early stage of reliable data from the product development process (Schenk 2014; Schenk and Schumann 2008). The idea of combined planning phases is described for years by modern and powerful tools (Dangelmaier 2013). These special cases are made of individually designed solutions. It lacks the holistic planning approach to take part in processes to promote an integration of all. First approaches for generating layout-based models can be found at Lorenz and Schulze (1995). First approaches to use structural as well as system data from a production planning and control system (PPS) to create models can be found at Splanemann et al. (1995). A first classification of automatic model generation approaches was provided by Eckardt (2002). The author distinguishes between parametric, structural, and hybrid approaches. Another way to classify approaches of the model generation makes the classification of model generation approaches to Straßburger et al. (2010). Early developments through digital planning and control was designed by the enterprise application integration (EAI) and service-oriented architectures (SOA) (Aier 2006; Kaib 2004; Bieberstein 2008). The EAI represents integrated business processes along the value chain. Enterprise applications of different generations and system architectures can interact in this framework over a common network. The SOA describes a method that encapsulates from existing IT components and coordinates. This existing services will be consolidated and summarized to a higher service. The objectives of EAI as also SOA are reducing costs in the development of production processes and increase the flexibility of business processes in the long term. The reason for the low acceptance and development of the methods is the high demands on insecure systems, data security, continuity of the tool development and product development process (Fay 2006; Raupricht et al. 2002). The interaction of different digital planning tools within the product life cycle is summarized often under the term "Digital Factory". The term describes a comprehensive network of digital methods and models, including simulation and 3D visualization. Its purpose is the integrated planning, management and implementation, as well as a steady improvement in all key factory processes and resources (VDI 2008; Wenzel et al. 2003). Also here is a link of different planning tools. However, the use of continuous planning tools is missing. Weigert (2015) describes the first opportunity of an implemented middleware for the automatic data exchange among different tools of the digital factory by AML. The motivation for the use of a common concept and tool can be reducing costs for the planning, control and operation and maintenance of plants and factories. First options for the automatic generation of the model are described. However, no procedure is known to reach the current level which combines three essential tools for the digital factory by an open interface.

2.1. Storage and Exchange format

For the semi- and fully automated model generation is the origin and use of the data and information utmost importance (Bergmann 2014). The currently most popular standards for the automatic model generation are the data formats simulation data exchange format (SDX) (Sly and Moorthy 2001) and Core Manufacturing Simulation Data (CMDS) (Lee 2015; Bergmann et al. 2010). The hierarchically structured SDX format is used to exclusively provide layout information. With the open-source, XML-based format of CMDS can both layout - as also process-related information transmitted. The problem of implementation of comprehensive control and routing strategies and complex system behavior is not completely solvable but manageable with these data formats (Bergmann, 2014; Bergmann et al. 2010). By AutomationML (AML), an open source, free available, object-oriented, XML-based storage and exchange format is being developed. After initial evaluations of different exchange formats, Daimler AG initiated the development and standardization of AutomationML as an intermediate format of the digital factory together with ABB, KUKA, Rockwell Automation, Siemens, netAllied and Zühlke, as well as the University of Karlsruhe and Otto von Guericke University Magdeburg in October 2006. In 2009, the previously closed industrial consortium opened by establishing an association. The first new member was the Fraunhofer IOSB. It becomes clear that the efforts to use AML are driven by the industrial and scientific location of Germany. AML owns the technical requirements for the modeling of production, intralogistics and robotic systems that is used but so far mainly in the area of virtual commissioning. Fundamentally AML linking role profiles (Hoernicke et al. 2016; Hundt et al. 2009; Lüder and Schmidt 2015). So far, topology, geometry, kinematics and behavior of system components can be described with AML. The hierarchical picture of the topology of the subject of the planning is carried out by means of Computer Aided Engineering Exchange (CAEX). The CAEX library concept includes three types of library (Drath 2010):

- The **SystemUnitClass** library is a catalog of concrete physical or logical system objects or their combination. Attributes, interfaces and nested internal elements and their compounds are assigned to the elements
- The **RoleClass** library defined abstract physical or logical system objects, regardless of the actual technical realization. Roles describe the functioning of investment properties
- The **InterfaceClass** library describes the kind of interfaces between the system objects. The relations between investment objects are mapped

Geometry and kinematics can be associated with individual system components through COLLADA files. The control is defined by PLCopenXML and describes the system behavior. AML is adaptable and flexible, it offers the possibility to include other XML formats (Hundt et al. 2009). In addition, the AML format has an inherent distributed data structure. The information is instead of a monolithic XML document, saved as separate documents. The reusability of individual system components and the development of element libraries will be easier (Lüder and Schmidt 2015).

3. APPLICATION GOALS AND CONCEPTUAL DESIGN

A continuous model generation between simulation, visualization and construction tools for the integrated planning of production and intralogistics systems describes the linking of the three digital tools. The expertise in dealing with the tools than the substantive complexity of each tool is highly classified. The presented concept and tool describes the development of application-oriented and open source middleware. The goal is to develop of uniform, digital planning methods and tools for a consistent design. The lossless and accelerated conversion and modelling within the various tools is in the focus. On reason of the different core functions of each tool, as well as their proprietary interfaces, the goal is to develop an open source automated import and export solutions (Figure 2).

Following advantages arise from the use of an automatic exchange system:

- Existing simulation, visualization, and construction tools in the company remain in place, preventing costly new investments
- Productivity and cost reduction can be achieved by the use of development, combining the individual benefits of the tools
- Visualization, modeling, and simulation of real-world intralogistics systems accelerates, reducing the largely manual and costly effort in creating a new model

At present, many software solutions have only limited access to the ability to offer planning data in a heterogeneous system landscape. This fact is due to the lack of conformity of the interfaces and the resulting lack of integration of all necessary planning data. The reasons for these so-called insellations have been and are the efforts of machine manufacturers to exclude competing products by means of ever new proprietary interfaces in favor of their own "complete solutions". A circumstance that today is hardly accepted by customers. Furthermore, the lack of networking between the planners and partners involved in the process contributes to the existing problem. In order to support the definition of the mapping rules, a graphical interface is developed which contains the following features:

- Visualization of the data model of the source system in a tree structure
- Selection of pre-made rule templates for repetitive application forms:
- *1:1 mapping* (Due to a type attribute; On the basis of a checked attribute value)
- *m:1 mapping* (Illustration of a group with the same or similar attribute value; Illustration of a group by means of object relations (parent-child, Sibling)
- *0:n mapping* (Creating new node with no equivalent in the source model)
- Memory function for the created rule sets with additional information:
  - Author
  - Version
  - Description
  - Source-Tool
  - Target-Tool

The exchange system is not limited to the generation of exchange files but can also be used for exchanging data between current network instances if the tools involved permit such an integration. The concept and tool for the continuous and comprehensive model building is appropriate on the example of the simulation tool Plant Simulation – Siemens and construction tool AutoCAD – Autodesk. The visualization tool taraVR – tarakos serves as a starting point for the investigations.

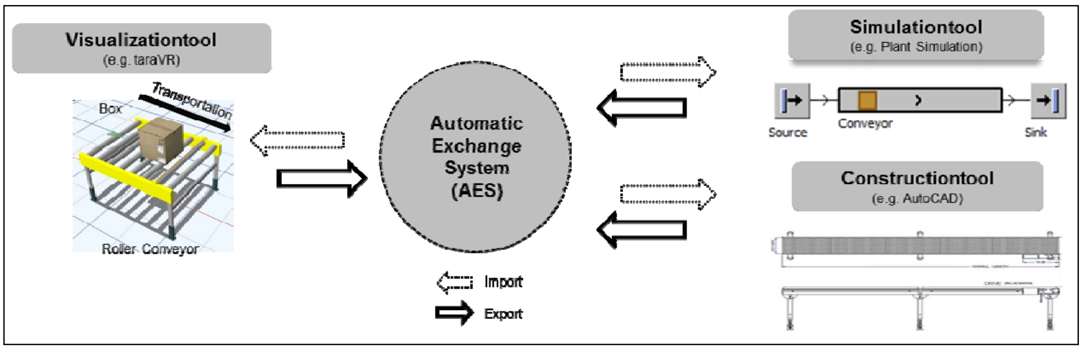


Figure 2: Overall concept of development

4. CONTINUOUS MODELING BETWEEN SIMULATION, VISUALIZATION AND CONSTRUCTION TOOLS

The Automatic Exchange System (AES) forms the basis for the common data and information exchange. The AES is defined by a system of mapping rules and an associated user interface. The model item get its role profiles from the developed AutomationML roll library within the Exchange System. In addition to the transmission of model elements it is important to transfer their attributes such as capable of taking over transport direction, speed, and more state descriptions. Furthermore, the relationships between the elements must be transferred. The approach of an open source standards about the AML function narrows the existing interface packages and fee-based libraries in the commercial product suites for the exchange of data. Goal is to realize independent modeling of closed software packages. Based on the classification ability of model generation approaches to Straßburger et al. (2010) the represented development can be classified as follows:

- Application: parallel (tactical) planning
- Focus: conveyor systems
- Degree of automation: semi-automated
- Approach: direct generic structure
- Support creation of model
- Interfaces: Text - and XML-based

Before models can be automatically generated it is necessary that the modules and elements of different tools are mapped. Data mapping is the process of mapping data elements between different data models and thus constitutes a fundamental step in the information and data integration (Alexe et al. 2008). A so-called rule interpreter is designed to enable this mapping. From the respective visualization, construction, simulation tools the libraries with all modules and their parameters are exported and mapped. This manner defines rules for the transfer. If current rules exist, the import file for the target tool can be generated by the rule interpreter emanating from an export file of source tool automatically.

The import file is loaded using the created interfaces or corresponding data interpretation in the target tool and interpreted. To integrate the relevant elements of the model and to interpret is a systematization of elements of importance for further processing. After the reduction for the respective tools it is important to define the requirements for the interfaces. The needs and situations of the interfaces between the designed EAS, the simulation tools as well as design tool have been analyzed and defined.

4.1. Automated model generation – Construction

The Autodesk AutoCAD tool has considerably more modification options in the area of the interface creation. Through the import of other CAD tools with the formats 3D Studio (.3ds), Autodesk Inventor (.ipt; .iam), SolidWorks (.prt; .sldprt; .asm; .sldasm) and step (.stp; .ste; .step) is there quite a broad base. For the development and integration of and into the EAS is the variety but not of importance. The neutral exchange format AML forms the basis of the development. For this purpose, separate commands are developed for import and export. The goal of consistency to the visualization tool ensures that the interface in the format programmed .net and used. Determined at a minimum for the required parameters of the construction tool for the communication with the AES include:

- Object information: name, geometry
- Layout information: location (connection are not necessary)
- Detailed information on geometry are not needed, can be optional attached in an AML file through the COLLADA format

A direct import AML files is not possible and is not supported by the Autodesk software suite. AutoCAD is used within the development as a two-dimensional representation of intralogistics planning and construction data.

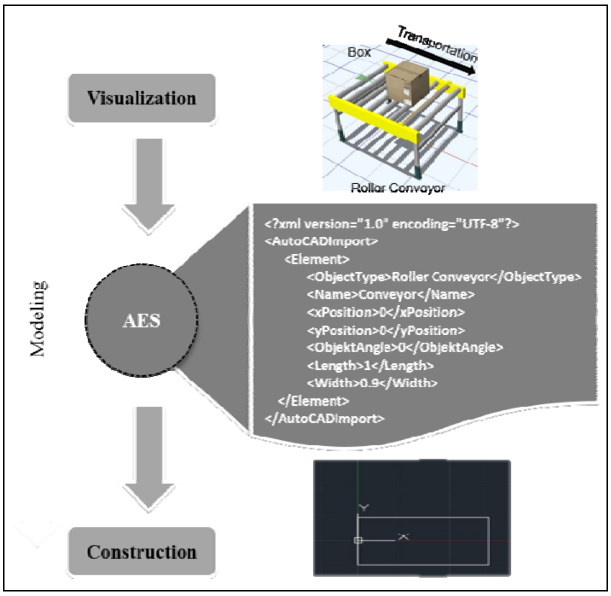


Figure 3: Model generation between visualization and construction tools via AES

Not only static blocks are addressed within the tool. The blocks are predefined two-dimensional body by the user. They are stored in a block table. Hereby the purpose a dynamic adaptation of the body within the construction tool to get.

A pure one-to-one relationship of modules is no longer necessary. Blocks to dynamically build and parameterize the interface with c#.NET was programmed and encapsulated in a Dynamic Link Library (.dll) file. The enclosed program environment is performed as AES.dll.

Algorithm 1: Pseudocode for Modeling – Construction

```
define fcn = rootnode.childnodes and
        scn = FCN.childnodes
open xmlfile
while rootnode has fcn
    read SCN.objecttyp, SCN.x–pos, SCN.y–pos
    create object from blocktable.objecttyp at
            x–pos.,y–pos.
    if object is dynamic
        for i = 1 to SCN.parameter.count
            read SCN.parameter(i) and
                parameter(i).value
            set object.parameter(i)=parameter(i).value
        next
    end if
end while
close xmlfile
```

After the call of the encapsulated .dll file it is possible to run various commands. The pseudo code describes the reading and processing of the XML file in the construction tool (Algorithm 1). The XML file is created from the AES and read in the target tool. In contrast to the simulation tool, the structure can be an XML file because the tool can work with XML and .dll file to process the models (Figure 3). The file is built in a common XML format. The linking of several objects is ensured by the updating of the source code. By working with dynamic blocks, the parameters can be arbitrarily varied according to the read-in to the construction tool. This is the processing of dynamic blocks. The blocks are user-defined two-dimensional bodies. They are stored in a block table. The aim of the invention is to obtain a dynamic adaptation of the bodies within the construction tool. For a static block, the dimensions are invariable. It is therefore stored with fixed dimensions in the table and can only be generated with these dimensions again. In a dynamic block, the user defines individual elements of the block as variable.

4.2. Automated model generation – Simulation

For the development of the simulation tool, event-oriented simulation approach from Plant Simulation by Siemens has been used. “Plant Simulation” has a variety types of license and fee-based libraries. Costly licenses such as “Professional” in combination with the “Interface Package” have been avoided for the development of a common data interface. The selection would not meet the main goal of an open-source and neutral communication interface. The license type of "Standard" forms the basis for the interface between the EAS and the simulation tool. An advantage is the upward compatibility of the simple data structure (.txt file). A disadvantage is the complex data processing in the AES before import and export in the simulation tool. The strings must correspond to a given form can be accurately encoded and decoded. As a result is a text file as opposed to a binary file without the use of special programs to read and can be viewed with a simple text editor and edited. Determined at a minimum for the required parameters of the simulation tool for the communication with the AES include:

- Object information: name, type, geometry
- Layout information: location and connection to other elements
- Material flow parameters: time usage, routing

A direct import of AML files is not possible and is not supported by the selected simulation tool. A master file was developed for the simulation tool, which contains all requirements needed for the model creation. The system works according to a hierarchical management of access to a common resource as a master/slave system. The master file allows only the backup model derivatives and contains all the required methods to model creation, export model, library export and import of the modules. A new generation section of source code is generated for each previously mapped module. Sections for identical statements, for example the combination of elements, can be applied. The text file is read from the AES into the simulation tool.

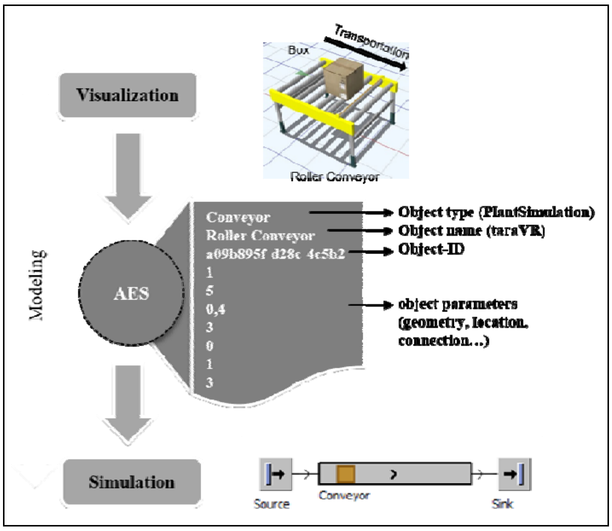


Figure 4 Model generation between visualization and simulation tools via AES

The file contains a hierarchical structure, which allows to process all objects equally. Changes to entire model elements or parts thereof can thus be easily incorporated. The file is integrated by the integrated data interface. This has resulted from the use of general interfaces.

The content of the text file is deliberately simple. The file is divided as follows. The object type is determined by the target tool and is part of the library in the simulation tool. The object name describes the name of the object in the source tool. The identification number of the object is a one-to-one number sequence that identifies the object in the model. Further details describe object parameters such as geometry, location, connection and speed. For multiple objects, the ID also describes the auto-linking points in the model. After conversion from the visualization tool into the AES, a text file is created (Figure 4). This is read in the simulation tool by a simple data interface. The necessary intelligence of the simulation tool is generated by the programmed interpretation of the source code (Algorithm 2). The programmed method reads the text file and checks whether the object exists in the simulation tool library. If the response is positive, the method begins to link the object to the element from the simulation library. After this, the method parameterizes the created object according to the specifications of the text file. Because the entire intralogistics system is present in a text file, the objects are separated by a separator. The unique ID knows all objects predecessors and successors. If the predecessor or successor is missing, the source or sink is created.

Algorithm 2: Pseudocode for Modeling – Simulation

```
open textfile
while not textfile.end
read bibliothek, objecttyp, x-pos., y-pos., name
create object from bibliothek.objecttyp at x-pos.,y-pos.
    if next textfile.line ≠ seperator
        do
        read parameter and parameter.value
        case parameter of
        case "input-id"
            if object.exists(input-id)
            connet object with input-object
            end if
        case "output-id"
            if object.exists(output-id)
                connet object with output-object
            end if
        else
            set object.parameter = parameter.value
        end case
        until next textfile.line ≠ seperator
    end if
endwhile
close textfile
```

The flexibility of the model creation is given by scanning the entire simulation library in the master file at start. This long term causes that even individual building blocks of the end user can be read out and dynamically addressed.

5. SUMMARY AND OUTLOOK

The described concept and tool represents a comprehensive approach for the automatic and open-source modeling and data transmission between different tools. The development describes a possibility to repeatedly use individual AML libraries for simulation, visualization and construction. This can be ensured by an AML role library. First results show that the implementation is successful. Similarly, no proprietary software solutions are used. With the possibility to combine modeling within the different tools it can be possible to produce precise and fast models from a single source. As a result, costly, error-prone and time-intensive re-modeling is already avoided. The development level reached the basis for the transfer of the models into the corresponding target tool, described by simulation and construction. Further steps are the increasing detailing and standardization of the automated exchange system. In the future, the question about the implementation and transmission of the work plans and the accompanying control and the routing of the elements in the simulation tool will be explored. Up to now simple conveyor systems from straight, curve conveyor belts, rotary tables and individual stations from the visualization tool Tarakos - taraVR can be modeled automatically in the simulation tool Siemens Plant Simulation and construction tool Autodesk - AutoCAD. The use of cost-intensive and optional interface libraries can be dispensed. Due to the simplicity of the development it is possible to connect arbitrary tools. To this end, investigations are currently being carried out to prepare the interface to other vendors' tools. Similar advantages are promised. It was necessary to use only existing licenses throughout the development. Only the internal specifications of the tool were used. No license rights of the providers were infringed.

The cooperation with end-users from the practice makes clear what additional requirements on the range of functions the AES must solve in future. The question about the implementation and transmission of the work scheduling and the accompanying control and the routing of the elements in the simulation tool are also to be investigated. The repatriation of information and models of the simulation and construction tool in the visualization tool be made manually in the AES at the current time. Target is to process the data from the construction through renewed data mapping in the AES. Here, the AAS repeatedly assumes the role of the rule interpreter and plausibility auditor. The integration of the event log obtained by the simulation runs is conceived for the simulation. The described and planned scope and procedures are to be simplified and made more dynamic in the future. At the same time, a significant increase in the user-friendliness in handling the tool is planned.

REFERENCES

Aier, S. (ed.): Enterprise application integration: Serviceorientierung und nachhaltige Architekturen. Berlin: GITO-Verl. 2006.

Alexe, B.; Chiticariu, L.; Miller, R.J.; Tan, W.-C.: Muse: Mapping Understanding and deSign by Example. In: IEEE 24th International Conference on Data Engineering, 2008, Cancun, Mexico, 7/4/2008 - 12/4/2008, 2008, p. 10–19.

Bergmann, S.: Automatische Generierung adaptiver Modelle zur Simulation von Produktionssystemen. Ilmenau: Univ.-Verl. 2014.

Bergmann, S.; Fiedler, A.; Straßburger, S.: Generierung und Integration von Simulationsmodellen unter Verwendung des Core Manufacturing Simulation Data (CMSD) Information Model. In: Integrationsaspekte der Simulation: Technik, Organisation und Personal: Generation and integration of simulation models using the Core Manufacturing Simulation Data (CMSD) information model. Karlsruhe: KIT Scientific Publ 2010, p. 461–468.

Bieberstein, N.: Executing SOA: A practical guide for the service-oriented architect. Upper Saddle River, NJ: IBM Press/Pearson plc 2008.

Daft, R.L.: Organization theory & design. Boston, MA: Cengage Learning 2016.

Dangelmaier, W.: Fertigungsplanung: Planung von Aufbau und Ablauf der Fertigung ; Grundlagen, Algorithmen und Beispiele. Berlin [u.a.]: Springer 2013.

Drath, R. (ed.): Datenaustausch in der Anlagenplanung mit AutomationML: Integration von CAEX, PLCopen XML und COLLADA. Berlin: Springer 2010.

Eckardt, F.: Ein Beitrag zu Theorie und Praxis datengetriebener Modellgeneratoren zur Simulation von Produktionssystemen. Aachen: Shaker 2002.

Eigner, M.; Stelzer, R.: Product Lifecycle Management: Ein Leitfaden für Product Development und Life Cycle Management. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg 2009.

Faltinski, S.: A dynamic middleware for real-time automation systems. Lemgo, Hannover: InIT; Technische Informationsbibliothek u. Universitätsbibliothek 2011.

Faltinski, S.; Niggemann, O.; Moriz, N.; Mankowski, A.: AutomationML: From data exchange to system planning and simulation. In: Industrial Technology (ICIT), Athen, 2012, p. 378–383.

Fay, A.: Reduzierung der Engineering-Kosten für Automatisierungssysteme. Industrie Management (2006) 22, p. 29–32.

Hoernicke, M.; Messinger, C.; Arroyo, E.; Fay, A.: Topologiemodelle in AutomationML. Atp-Edition : automatisierungstechnische Praxis ; Organ der GMA (VDI-VDE-Gesellschaft Meß- und Automatisierungstechnik) und der NAMUR (Interessengemeinschaft Automatisierungstechnik der Prozessindustrie) 58 (2016) 5/2, p. 28–41.

Hundt, L.; Lüder, A.; Barth, H.: Anforderungen an das Engineering durch die Verwendung von mechatronischen Einheiten und AutomationML. SPS/IPC/DRIVES 2009 (2009), p. 341–349.

Kaib, M.: Enterprise Application Integration: Grundlagen, Integrationsprodukte, Anwendungsbeispiele. Wiesbaden: Dt. Univ.-Verl. 2004.

Klepper, S.: Entry, exit, growth and innovation over the product life cycle. Estados Unidos: American Economic Review 1996.

Lee, Y.-T.T.: A Journey in Standard Development: The Core Manufacturing Simulation Data (CMSD) Information Model. Journal of research of the National Institute of Standards and Technology 120 (2015), p. 270–279.

Lorenz, P.; Schulze, T.: Layout based model generation. In: Lilegdon, W.R. (ed.): Proceedings of the 27th conference on Winter simulation, Arlington, VA, USA, 3-6 Dec. 1995, 1995, p. 728–735.

Lüder, A.; Schmidt, N.: AutomationML in a Nutshell. In: Handbuch Industrie 4.0 : Produktion, Automatisierung und Logistik. Wiesbaden: Springer Fachmedien Wiesbaden 2015, p. 1–46.

Raupricht, G.; Haus, C.; Ahrens, W.: PLT-CAE-Integration in gewerkeübergreifendes Engineering und PlantMaintenance. atp – Automatisierungstechnische Praxis (2002) 44 (2), p. 50–62.

Rawolle, J.; Ade, J.; Schumann, M.: XML als Integrationstechnologie bei Informationsanbietern im Internet. Wirtschaftsinformatik 44 (2002) 1, p. 19–28.

Schenk, M.: Fabrikplanung und Fabrikbetrieb: Methoden für die wandlungsfähige: Springer Berlin Heidelberg 2014.

Schenk, M.; Schumann, M.: Interoperable Testumgebung für verteilte domänenübergreifende Anwendungen. In: Scholz-Reiter, B. (ed.): Technologiegetriebene Veränderungen der Arbeitswelt. Berlin: GITO-Verl. 2008, p. 155–169.

Schreiber, W.; Zimmermann, P.: Virtuelle Techniken im industriellen Umfeld: Das AVILUS-Projekt - Technologien und Anwendungen: Springer Berlin Heidelberg 2011.

Sly, D.; Moorthy, S.: Simulation data exchange (SDX) implementation and use. In: Peters, B.A. (ed.): Proceedings of the 2001 Winter Simulation Conference, Arlington, VA, USA, 9-12 Dec. 2001, 2001, p. 1473–1477.

Splanemann, R.; Roth, M.; Soravia, S.: Einsatz der Materialflußsimulation zur Planung, Analyse und Optimierung von verfahrenstechnischen Produktionsanlagen. Chemie Ingenieur Technik 67 (1995) 9, p. 1107–1108.

Straßburger, S.; Bergmann, S.; Müller-Sommer, H.: Modellgenerierung im Kontext der Digitalen Fabrik-Stand der Technik und Herausforderungen. Proceedings der 14 ASIM-Fachtagung Simulation in Produktion und Logistik (2010), p. 37–44.

VDI: Digitale Fabrik: Grundlagen ; VDI-Richtlinien ; VDI 4499, Blatt 1. Berlin: Beuth 2008.

Weigert, D.: Automated exchange system between simulation, visualization and construction tools: The 15th International Conference Modeling and Applied Simulation, MAS 2016: p. 112-120 .

Wenzel, S.; Hellmann, A.; Jessen, U.: e-Services - a part of the "Digital Factory". In: Bley, H. (ed.): Proceedings / 36th CIRP International Seminar on Manufacturing Systems, 2003, p. 199–203.

AUTHORS BIOGRAPHY

DAVID WEIGERT studied Industrial Engineering with specialization in Logistics at the Otto-von-Guericke-University Magdeburg. He became a research associate at the Chair Logistical Systems at the Otto-von-Guericke-University Magdeburg and scientific project assistant at the Fraunhofer Institut for Factory Operation and Automation IFF Magdeburg. His areas of competence are the analysis and optimization of logistics processes, as well as modelling, simulation and optimization of logistics systems.

PAUL AURICH holds a bachelor's degree in Industrial Engineering in Logistics. Currently he is studying mechanical engineering at the Master's degree. His research interests include the modeling, simulation and mathematical optimization of logistics systems.

TOBIAS REGGELIN is a research and project manager at Otto von Guericke University Magdeburg and the Fraunhofer Institute for Factory Operation and Automation IFF. He received a doctoral degree in engineering from Otto von Guericke University Magdeburg. His research interests include modeling and simulation of production and logistics systems and the development and application of new modeling and simulation methodologies.

A SIMULATION TOOL FOR MODELLING AND OPTIMIZATION OF A JOB-SHOP PRODUCTION SYSTEM

Eleonora BOTTANI[(a)], Marta RINALDI [(b)], Roberto MONTANARI [(c)], Massimo BERTOLINI [(d)], Francesco ZAMMORI [(e)]

[(a),(b),(c)]Department of Engineering and Architecture, University of Parma, viale delle Scienze 181/A, 43124 Parma (Italy)

[(a)]eleonora.bottani@unipr.it, [(b)]marta.rinaldi@unipr.it, [(c)]roberto.montanari@unipr.it; [(d)]massimo.bertolini@unipr.it; [(e)]francesco.zammori@unipr.it

ABSTRACT
This study carries out an analysis of the machinery department of a main Italian company, which operates as a manufacturer of plants for the food industry. The analysis targets expressively the improvement of the total shop floor time of the jobs in the department and is supported by a simulation model developed under SIMUL8™. Thanks to the model, the dispatching rule currently used by the company to schedule the jobs to be manufactured in the department was compared to additional 6 scheduling strategies, to evaluate potential improvements.
The results obtained have shown that the most remaining operations (MROP) rule returns the most interesting results in terms of total shop floor time for the jobs examined. The outcomes of this paper could be useful to the targeted company to evaluate the implementation of alternative dispatching rules to schedule the jobs.

Keywords: job shop; dispatching rule; simulation model; food machinery industry.

1. INTRODUCTION

"Job shop" production is a manufacturing process where products are designed and produced as per the specification of customers, within prefixed time and cost. The distinguishing features of this kind of production system are low volumes and high variety of products. Also, from an operational point of view, a job shop system comprises of general purpose machines arranged into different departments. Each job demands unique technological requirements and need being processed on machines in a specific sequence. In general, jobs are manufactured following a sequence of orders from the customers, with fixed routes for each job through the plant (Anil Kumar and Suresh, 2008).
Among others, job shop systems can be found among industries described as make-to-order (MTO) production units, engineer-to-order (ETO) production units, high-variety production units or order-driven production units (Velaga, 2016).
As known, to remain competitive on the market, companies should increase the throughput rate, reduce the costs and improve the customer satisfaction. At the production management level, a production manager has to improve the performance of the manufacturing system, by, e.g., reducing the process time, mean set-up time, mean time between failures, mean time to repair and demand variability (Renna, 2017). Due to the complexity and the number of processes typical of job shop systems, production optimisation is a challenging tasks and, frequently, job shop systems relies on experienced workers (e.g. when manufacturing a special product) and production managers in the production planning and management (Supsomboon and Vajasuvimon, 2016).
In this regard, a particularly crucial issue is the scheduling problem that, if not properly faced, may lead to relevant idle times (i.e. a non-value added time when a machine is not being used despite the fact that it is available to be used) and delays. Unfortunately, in many cases, scheduling is a non-deterministic polynomial-time (NP) hard problem, and cannot be optimally solved in a suitable computational time. This is the reason why, managers often require specific tools to manage it., which, in many cases,
Among these tools, simulation-based optimization methods, is generally used as a robust way to find an effective solution within fairly shorter time for larger sized problems (Xie and Allen, 2015). More in general, simulation is one out of several tools that can be used to investigate a real system, its dynamics and logic, as well as its evolution in time or as a function of the input set (Carley, 2003). Also, discrete-event simulation is one of the most powerful methods that allows to recreate the dynamics of a real system in a controlled environment, thus enabling to analyse the interdependencies between its elements, to monitor its main control parameters and evaluate its performance. Nowadays, simulation is widely used in industrial factories, as it allows to vary some parameters or factors to investigate their impact on the system's performance, without affecting the whole system and before implementing any solution in the real situation.
This paper describes the development of a simulation model reproducing a real job shop system and the use of

this model to analyse and improve the performance of the system, with a particular attention to scheduling improvement. . The context where this study was carried out is the mechanical department of a milling plant of a manufacturing company, located in northern Italy.

The remainder of the paper is organized as follows. Section 2 lists the steps followed in the research. Sections 3-5 provide the details of the steps previously listed. Section 6 summarizes the results obtained in the analysis, describes the implications and limitations of the study and outlines future research directions.

2. THE RESEARCH METHODOLOGY

The research methodology adopted in this study is detailed in the steps listed below.

1. *Analysis of the machinery department*. The first step of the study was the analysis of the machinery department and of the related criticalities, with the aim to identify the potential for improvement, as detailed in section 3;
2. *Data collection*. The main figures of the system (e.g. set-up times, processing times, jobs routings, jobs families, etc.) were collected, either from the company's information system or by direct observation. These data were used as input in the simulation model, as detailed in section 4;
3. *Model development and validation*. A simulation model was built to reproduce the mechanical department. The simulation model reproduces the targeted department and provides, as output, some key performance indicators (KPIs). Details related to the development of the simulation model, its validation and the performance level measured for the current scenario are given in section 5;
4. *Analysis and performance improvement*. Starting from the simulation outcomes, new operating policies were proposed, focusing expressively at the scheduling problem. In this regard, seven common dispatching rules were tested and their performance were compared to that of the current scheduling policy. Results of this step are proposed in section 6.

3. ANALYSIS OF THE MACHINERY DEPARTMENT

The company considered in this study (referred to as "Company A" for confidentiality) is an Italian manufacturer of plants for the milling industry. Company A exports its products to several countries worldwide. The company operates on an Engineer-To-Order (ETO) basis, meaning that any new product starts with the design of the plant and includes assembly, installation, civil works, test of the equipment at the customer's site and staff training. In Italy, Company A is leader in its market segment while, on a global basis, it ranks second among the top producers of milling plants.

The department where the research has been carried out (i.e. the mechanical department) covers an area of approximately 2,400 square meters and includes 18 machines. The production process starts from the raw materials (e.g. bars and rods in steel, cast iron, aluminium, brass or other alloy metal alloys), which are processed using mechanical machineries, to realize specific components according to the 2-D or 3-D drawing prepared by the technical division. The mechanical processes performed in the targeted department consists mainly in chip removal operations, to obtain a final component of desired shape. Such processes are carried out by means of turning, grinding and milling machines. These machines are organised according to a job-shop layout, where machines are aggregated into specialized departments. Jobs are scheduled on the different machines using an Earliest Due Date (EDD) rule, i.e. the first job to be processed is the one with the closest delivery date.

Each employee of the department typically controls two machines, with only one exception of an employee who controls three machines. The manufacturing activities are organized in two work shifts of 7 hour each; only one machine works on one work shift of 8 hours.

The main criticalities can be summarised as follows:

- the fact that each employee should control more than one machine simultaneously could lead to some inefficiencies. For instance, the employee could be forced to stop the processing of one production lot to introduce a more urgent one;
- because an employee could supervise two or three machines, it often happens that if he/she is operating on a machine (for instance, for set-up or manual operations), the second machine cannot work until the employee has finished working on the first one. This causes delays in the overall processing time of the department;
- as a consequence of the situations described above, the jobs processed are often sent to the next production department late. The delay is typically caused by a series of inefficiencies and waste, which are often not known to the company;
- the company's top management is aware that the set-up times recorded in the company's information system do not reflect the real values. Indeed, from an administrative point of view, the set-up time is used only to derive an estimate of the manpower cost in the targeted department. All set-up activities that do not strictly require the presence of an employee have a zero-set-up time by default. Therefore, the real efficiency of the machines of the targeted department is not known.

The analysis carried out in this paper aims at solving some of the criticalities above, and in particular at developing a simulation model that evaluates the shop floor throughput time and tries to minimize it.

4. DATA COLLECTION

The correct management of a work cycle in the department depends on many factors, including, among others:

1. Number, shape and size of the pieces to be manufactured;

2. The amount of items to be manufactured in a job (batch size);
3. Delivery date of the order;
4. Manufacturing and set-up time.

A data collection phase was carried out in December 2016 and January 2017 to gather data listed above. To be more precise, the delivery date of each job is recorded in the company's database, together with the order date (thus allowing to estimate the total throughput time of a job as recorded by the company). Similarly, the number, shape and size of the pieces to be manufactured, as well as the batch size, were collected from the company's data base, over a time horizon of 2 months (November-December 2016). These data were also confirmed through direct observation of the department's operating conditions.
The data extracted from the company's database showed that overall, the company has manufactured 1002 jobs in the targeted department, and 213 during the observation period.
As far as the type of equipment and tools used are concerned, jobs can be processed by more than one machine of the targeted department; the analysis of the data retrieved from the company' database showed that approximately 73% of the jobs manufactured are processed by only one machine, while approximately 24% should be processed by two machines; the remaining jobs need to be processed by either three (2.75%) or four (0.52%) machines.
The manufacturing and set-up time are both recorded in the company's database. More precisely, the processing time is recorded in the company's database for 968 out of the 1002 jobs manufactured, while the set-up time is available for 853 jobs. However, as already mentioned, the company's top management was aware that the set-up time recorded could differ from the real values in a significant way. Therefore, the activity of the targeted department was directly observed, approximatively, for a month (2 weeks in December 2016 and 2 weeks in January 2017) to collect both set-up times and the processing times of the jobs manufactured by the company. The batch size was also recorded in the observation period.
Some of the results are shown in Figure 1, which clearly shows that most of the products manufactured (94.3% approximately) exhibit a processing time of less than 294 min. More specifically, the processing time ranges from 0 to 6 minutes for 19.1% of these jobs and from 6 to 12 minutes for 14.89% of these jobs.

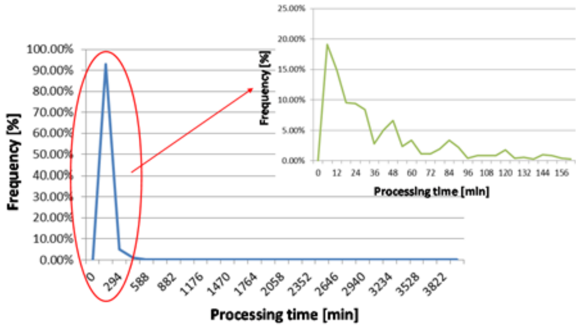


Figure 1: Processing time (company's data)

A similar analysis was carried out for the set-up times; this was made with respect to 756 jobs out of the 1002 manufactured by the company. Table 1 provides an extract of obtained results, which have been organized into classes of 15 minutes for visualization purpose.
From Table 1 it is easy to see that more than 85.4% of the jobs have a total set-up time of less than 60 minutes, corresponding to the sum of the set-up time on the different machines of the department under examination. For approximately 8.8% of the jobs, the total set-up time ranges from 60 to 90 minutes.
The data recorded in the company's database were compared to those obtained from the direct observation. To this aim, the ratio "time recorded in the company's database" over "time observed" was computed. The corresponding results are proposed in Table 2. From Table 2 it can be seen that for only 16.13% of the jobs observed the ratio between the time recorded in the company's database and that observed is in the range 0.8-1, which shows correspondence between these data. Instead, for the remaining jobs the correspondence is significantly weaker. Also, for approximately 41% of the jobs observed, the ratio is less than 1, indicating that the time recorded in the company's database is lower than the real value.
The same analysis, carried out on the set-up time, leads to the results proposed in Table 3. Results are similar. Only for less than 10% of the jobs examined, the observed set-up time corresponds to that recorded in the company's database; also, for more than 78% of the jobs, the set-up time recorded in the company's database is lower than that observed in the real functioning of the department.
The dissimilarities between the data recorded and that observed are one out of several possible causes of jobs' tardiness at the departmental level. Overall, in the light of the observed dissimilarities, subsequent analyses were carried out using the real data of manufacturing and set-up time, instead of those recorded in the company's database.

Table 1: set-up time (company's data)

Class	Set-up time (min)	Absolute frequency	Relative frequency	
1	0	93	12.3%	
2	15	217	28.7%	
3	30	200	26.5%	
4	45	46	6.1%	
5	60	90	11.9%	85.4%
6	75	20	2.6%	
7	90	37	4.9%	
8	105	9	1.2%	
9	120	8	1.1%	
10	135	0	0.0%	
11	150	7	0.9%	
12	165	1	0.1%	
13	180	12	1.6%	
14	195	3	0.4%	
15	210	2	0.3%	
16	225	0	0.0%	
17	240	1	0.1%	
18	255	1	0.1%	
19	270	1	0.1%	
21	300	3	0.4%	
22	315	0	0.0%	
23	330	0	0.0%	
24	345	0	0.0%	
25	360	0	0.0%	
26	375	4	0.5%	
27	405	1	0.1%	

Table 2: comparison between the processing time observed and that recorded in the company' database

Class	Ratio	Absolute frequency	Relative frequency	
1	0	0	0.00%	
2	0.2	3	2.42%	
3	0.4	16	12.90%	
4	0.6	15	12.10%	
5	0.8	17	13.71%	41.13%
6	1	20	16.13%	16.13%
7	1.2	12	9.68%	
8	1.4	11	8.87%	
9	1.6	8	6.45%	
10	1.8	5	4.03%	
11	2	3	2.42%	
12	2.2	4	3.23%	
13	2.4	3	2.42%	
14	2.6	2	1.61%	
15	2.8	0	0.00%	
16	3	5	4.03%	42.74%

Table 3: comparison between the set-up time observed and that recorded in the company' database

Class	Ratio	Absolute frequency	Relative frequency	
1	0	0	0.00%	
2	0.2	18	19.78%	
3	0.4	24	26.37%	
4	0.6	17	18.68%	
5	0.8	12	13.19%	78.02%
6	1	9	9.89%	9.89%
7	1.2	2	2.20%	
8	1.4	2	2.20%	
9	1.6	3	3.30%	
10	1.8	0	0.00%	
11	2	0	0.00%	
12	2.2	2	2.20%	
13	2.4	1	1.10%	
14	2.6	0	0.00%	
15	2.8	0	0.00%	
16	3	0	0.00%	
17	3.2	1	1.10%	
18	3.4	0	0.00%	
19	3.6	0	0.00%	
20	3.8	0	0.00%	12.09%

5. MODEL DEVELOPMENT AND VALIDATION

A simulation model of the machinery department was built using the commercial software SIMUL8^{TM} for Windows. SIMUL8^{TM} uses dynamic discrete simulation and is commonly exploited to simulate systems that involve processing of discrete entities at discrete times. Examples of those systems are production, manufacturing, logistic or service provision systems. As output, it generates statistics of performance parameters and metrics of the production system examined (Concannon et al., 2007). Also, the software embodies statistical tools, able to process the input data and output of the model, set the number of replicates, and elaborate the output. In real situations, there are many other factors that can involve the criticalities in this department.

The scheme of the simulation model, as it has been developed in SIMUL8^{TM} is proposed in Figure 2. From Figure 2 it can be seen that the model structure reflects the layout of the machinery department. Each machine is reproduced using a set of 4-5 work centres and a buffer (or queue). Each work centre represents one of the activities required to process a job. These include, for each job: set-up time; processing time; job material handling and waiting time of the job, once processed. Specific rules have been set, so as to allow these activities to be executed in the correct order, avoiding overlapping. Also, each machine has a defined availability, which was estimated starting from the direct observation of the machinery department.

The buffer is used to represent the physical area available in front of each machine, where jobs can be temporarily stored.
As input, the model uses data collected in the Data collection step; these data include, among others, the number, shape and size of the pieces to be manufactured; the sequence of machinery operations to process each job; the manufacturing and set-up time.

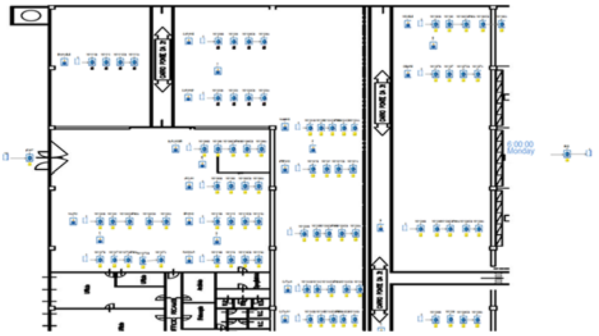


Figure 2: Scheme of the simulation model.

The model was run with the data collected and validated by comparing the number of jobs manufactured per day with real data. Comparing the simulated and real data is one of the possible ways to validate a simulation model (Kleijnen, 1995). To this end, the number of replicates was set at 10, which should ensure sufficient reliability of the results obtained, and the simulation duration was set at one week. The comparison showed a good correspondence between the simulated data and the real one, in terms of the number of jobs processed per day and over the whole week.

6. ANALYSIS AND PERFORMANCE IMPROVEMENT

Some key performance indicators (KPIs) were quantified using the output of the model, in order to evaluate the effectiveness of the targeted department; these KPIs include:

- The total shop floor time of the department, i.e. the time required to complete the set of activities required to process the full set of jobs scheduled to be manufactured in a week in the targeted department;
- the machine effectiveness, i.e. its effective use in the day (excluding, e.g., stops, waiting, loading/unloading operations).

A modified scenario was then designed and tested by means of the simulation model, to evaluate possible improvements in the above KPIs. To be more precise, the new scenario makes use of different dispatching rules, i.e. of specific rules to assign the jobs to the different machined of the department (Sculli and Tsang, 1990). A total of 6 common and easy to use rules was derived from the literature and tested on the set of jobs processed, and the relating results were compared to that returned by the scheduling rule currently used by the company (i.e. the EDD). The rules chosen for testing are:

- Total Work (TWORK): the selected job owns the lowest total processing time (this latter obtained as the sum of the working time of the operations already performed and those still to be performed);
- Minimum set-up time (MSUT): The selected job involves the minimum setup time on the machine being considered;
- Operational Expiration Date (OPNDD): the chosen job has the shorter "expiration date". This latter is obtained as the ratio of the difference between the delivery date and the starting time of the job and the number of operations to be carried out on the job;
- Few remaining operations (FROP): the chosen job owns the lowest number of remaining operations to complete;
- Most remaining operations (MROP): the chosen job has the greatest number of remaining operations to be performed;
- First in the system first served (FISFS): The allocated job is the one that entered in the system first.

The simulation model was used to reproduce 104 (random) weeks of functioning of the department, to test the rules listed above and compare the results with those of the current scheduling policy, obtaining (6 + 1) x 104 = 728 scenarios in total. The simulation was stopped once all the jobs have been processed; hence, the simulation duration reflects the total shop floor time of the department.
A summary of the performance of the dispatching rules, in terms of the total shop floor time, is proposed in Table 3. The table shows that the best rule to prioritize the jobs in the targeted system is the MROP, which returned a lower total shop floor time compared to the current scenario. More precisely, this rule returns the lowest shop floor time for the system considered in 22.3% of the scenarios simulated. Further rules that provided interesting findings are OPNDD and FISFS. The rule currently used by the company to prioritize the jobs, i.e. the EDD, returned optimal results in terms of total cycle time only in 11.18% of the scenarios simulated.

Table 3: comparison between the shop floor time observed and that recorded in the company' database

Scheduling rule	Number of optimal scenarios	Percentage of optimal scenarios
TWORK	16	10.53%
MSUT	17	11.18%
MROP	34	22.37%
FROP	20	13.16%
OPNDD	25	16.45%
EDD	17	11.18%
FISFS	23	15.13%

In the light of the fact that MROP returned the most effective results, the performance of this rule was investigated in further details. To this end, the total processing time of the MROP was compared to that of the EDD, to quantify the savings achievable when using the new rule. It was found that the average difference in the total shop floor time (across the whole sample of 104 weeks simulated) accounts for 400 minutes,

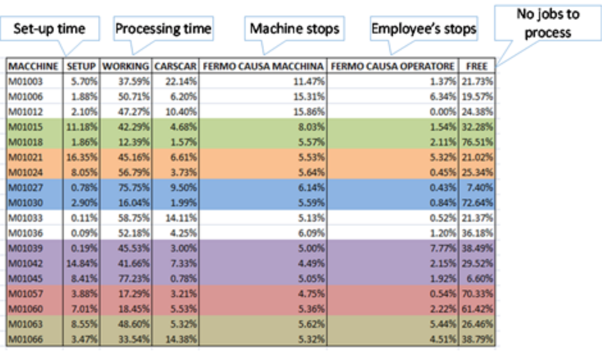
ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

corresponding to the average saving achievable using the MROP rule; a peak of 2877 minutes saved was also observed (Table 4).

Table 4: savings in the processing time (EDD vs. MROP)

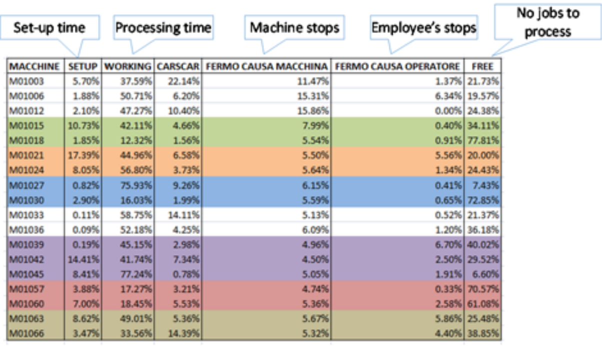
Class	Absolute frequency	Relative frequency
0	0	0.00%
411	42	84.00%
822	5	10.00%
1233	2	4.00%
1644	0	0.00%
2055	0	0.00%
2466	0	0.00%
2877	1	2.00%

As a further analysis, the simulation model was used to reproduce the operating conditions of the machinery department over 3 (real) weeks observed between December 2016 and January 2017, with the use of the MROP rule. This analysis aims at evaluating the possible savings in the total shop floor time the company could have achieved if it had applied of the MROP rule in the observation period. For all weeks simulated, the MROP rule returned better performance than the EDD one in terms of the total shop floor time. Also, the efficiency of the machines was found to improve considerably when using the MROP rule. Results, for a sample week (week no.2) are proposed in Figure 3 and 4).



MACCHINE	SETUP	WORKING	CARSCAR	FERMO CAUSA MACCHINA	FERMO CAUSA OPERATORE	FREE
M01003	5.70%	37.59%	22.14%	11.47%	1.37%	21.73%
M01006	1.88%	50.71%	6.20%	15.31%	6.34%	19.57%
M01012	2.10%	47.27%	10.40%	15.86%	0.00%	24.38%
M01015	11.18%	42.29%	4.68%	8.03%	1.54%	32.28%
M01018	1.86%	12.39%	1.57%	5.57%	2.11%	76.51%
M01021	16.35%	45.16%	6.61%	5.53%	5.32%	21.02%
M01024	8.05%	56.79%	3.73%	5.64%	0.45%	25.34%
M01027	0.78%	75.75%	9.50%	6.14%	0.43%	7.40%
M01030	2.90%	16.04%	1.99%	5.59%	0.84%	72.64%
M01033	0.11%	58.75%	14.11%	5.13%	0.52%	21.37%
M01036	0.09%	52.18%	4.25%	6.09%	1.20%	36.18%
M01039	0.19%	45.53%	3.00%	5.00%	7.77%	38.49%
M01042	14.84%	41.66%	7.33%	4.49%	2.15%	29.52%
M01045	8.41%	77.23%	0.78%	5.05%	1.92%	6.60%
M01057	3.88%	17.29%	3.21%	4.75%	0.54%	70.33%
M01060	7.01%	18.45%	5.53%	5.36%	2.22%	61.42%
M01063	8.55%	48.60%	5.32%	5.62%	5.44%	26.46%
M01066	3.47%	33.54%	14.38%	5.32%	4.51%	38.79%

Figure 3: machine's efficiency in week no.2 – MROP rule.



MACCHINE	SETUP	WORKING	CARSCAR	FERMO CAUSA MACCHINA	FERMO CAUSA OPERATORE	FREE
M01003	5.70%	37.59%	22.14%	11.47%	1.37%	21.73%
M01006	1.88%	50.71%	6.20%	15.31%	6.34%	19.57%
M01012	2.10%	47.27%	10.40%	15.86%	0.00%	24.38%
M01015	10.73%	42.11%	4.66%	7.99%	0.40%	34.11%
M01018	1.85%	12.32%	1.56%	5.54%	0.91%	77.81%
M01021	17.39%	44.96%	6.58%	5.50%	5.56%	20.00%
M01024	8.05%	56.80%	3.73%	5.64%	1.34%	24.43%
M01027	0.82%	75.93%	9.26%	6.15%	0.41%	7.43%
M01030	2.90%	16.03%	1.99%	5.59%	0.65%	72.85%
M01033	0.11%	58.75%	14.11%	5.13%	0.52%	21.37%
M01036	0.09%	52.18%	4.25%	6.09%	1.20%	36.18%
M01039	0.19%	45.15%	2.98%	4.96%	6.70%	40.02%
M01042	14.41%	41.74%	7.34%	4.50%	2.50%	29.52%
M01045	8.41%	77.24%	0.78%	5.05%	1.91%	6.60%
M01057	3.88%	17.27%	3.21%	4.74%	0.33%	70.57%
M01060	7.00%	18.45%	5.53%	5.36%	2.58%	61.08%
M01063	8.62%	49.01%	5.36%	5.67%	5.86%	25.48%
M01066	3.47%	33.56%	14.39%	5.32%	4.40%	38.85%

Figure 4: machine's efficiency in week no.2 – EDD rule.

7. CONCLUSIONS

This study has proposed the analysis of the machinery department of a main Italian company manufacturing plants for the food industry. The analysis, supported by a simulation model developed under SIMUL8^{TM}, has targeted in particular the revision of the dispatching rule currently used by the company to schedule the jobs to be manufactured in the department. A total of 6 alternative rules was tested using simulation and the relating results were compared in terms of the total shop floor time generated by each rule. The choice of this specific KPI was motivated by the fact that one of the main criticalities of the department analysed refers to the delay experienced by the jobs processed.

The results obtained have shown that the MROP rule returns the most interesting results in terms of total shop floor time for the jobs examined. From a practical point of view, this outcome could be useful to the targeted company to evaluate the implementation of alternative dispatching rules to schedule the jobs. At the same time, however, it is not expected that these results are optimal in general and, probably, they do not provide a definitive solution to the targeted production department. In this respect, the natural future step of the research will be to implement the alternative scheduling rule in the manufacturing department in a pilot trial, to provide a quantitative evaluation of the resulting benefits.

REFERENCES

Carley, K.M. (2003). Computational organizational science and organizational engineering. Simulation Modelling Practice and Theory, 10, 253–269

Concannon, K., Elder, M., Hindle, K., Tremble, J. and Tse, S. (2007), Simulation Modelling with Simul8^{TM} (ISBN: 0973428503), Visual Thinking International, available at: www.simtech.hu/_data/VFS_6084183539ff1c826da47589a021838c.pdf (accessed 2 August 2014).

Kleijnen, J.P.C. (1995). Verification and validation of simulation models. European Journal of Operational Research, 82, 145-162

Renna, P. (2017). Allocation improvement policies to reduce process time based on workload evaluation in job shop manufacturing systems. International Journal of Industrial Engineering Computations 8, 373-384

Sculli, D. and Tsang, K.K. (1990). Priority dispatching rules in a fabrication/assembly shop. Mathematical and Computer Modelling, 13(3), 73-79

Supsomboon, S., Vajasuvimon, A. (2016). Simulation model for job shop production process improvement in machine parts manufacturing. International Journal of Simulation modelling, 15, 611-622. Retrieved July 2017 from http://www.ijsimm.com/Full_Papers/Fulltext2016/text15-4_611-622.pdf

Xie, C. and Allen, T.T. (2015). Simulation and experimental design methods for job shop scheduling with material handling: a survey. International Journal of Advanced Manufacturing

Technology, 80, 233–243. DOI 10.1007/s00170-015-6981-x

AUTHORS BIOGRAPHY

Marta RINALDI is research fellow at the Department of Engineering and Architecture of the University of Parma. She graduated (with distinction) in Industrial Engineering and Management in 2011 and got a Ph.D. in Industrial Engineering in 2015. From 2011 to 2014 she was scholarship holder at the Interdepartmental Center CIPACK. She currently works on discrete event simulation and its application to industrial plants, manufacturing systems and business processes, and is author of more than 20 papers in this field.

Eleonora BOTTANI is Associate professor in Mechanical Industrial Plants at the Department of Engineering and Architecture of the University of Parma. She graduated (with distinction) in Industrial Engineering and Management in 2002, and got her Ph.D. in Industrial Engineering in 2006, both at the University of Parma. Her research activities concern logistics and supply chain management issues. She is author (or co-author) of more than 130 scientific papers, referee for more than 60 international journals, editorial board member of five scientific journals, Associate Editor for one of those journals, and editor-in-chief of a scientific journal.

Roberto MONTANARI is Full professor of Mechanical Plants at the University of Parma. He graduated (with distinction) in 1999 in Mechanical Engineering at the University of Parma. His research activities mainly concern equipment maintenance, power plants, food plants, logistics, supply chain management, supply chain modelling and simulation, inventory management. He has published his research in approx. 70 papers, which appear in qualified international journals and conferences. He acts as a referee for several scientific journals, is editorial board member of 2 international scientific journals and editor of a scientific journal.

Massimo BERTOLINI graduated in 1999 in Mechanical Engineering at the University of Parma and received his Ph.D. in Production Systems and Industrial Plants from the same university in 2004. Since 2002 he has been a Lecturer in Industrial Plants at the Department of Engineering and Architecture of the University of Parma. He is currently Associate professor of Operations management at the same university. His research activities mainly concern equipment maintenance, reliability, production planning, logistics and food safety. He has published more than 60 papers in international journals and conferences.

Francesco ZAMMORI graduated with distinction in 2004 in Management Engineering and completed his post-graduate studies in 2009, when he received a Ph.D. in Industrial Engineering from the University of Pisa. From 2012 he is Assistant Professor at the University of Parma where he teaches Management Accounting Systems, Project Management, Information Science, Data Bases and Information Systems. His research interests mainly concern: *(i)* Lean Thinking, *(ii)* Hybrid Production Planning and Control Systems, *(iii)* Modeling and Simulation and *(iv)* Project Management. His research activities led to the publication of more than 40 works, most of which accepted on prestigious international journals.

MODEL-BASED REMOTE RUNWAY SOLUTION FOR A CONGESTED AIRPORT: MEXICO CITY AIRPORT CASE

Miguel Mujica Mota [(a)], Paolo Scala[(b)]

[(a),(b)]Aviation Academy, Amsterdam University of Applied Sciences,
1097 DZ Weesperzijde 190, Amsterdam, The Netherlands

[(a)] m.mujica.mota@hva.nl, [(b)] p.m.scala@hva.nl

ABSTRACT
The airport of Mexico City is suffering from congestion problems during part of the day. For that reason the government announced the construction of a complete new one which is supposed to be operative in 2020 in its first phase. However the economic downturn in the country jeopardize the complete project besides other technical issues; for that reason it is important to have intelligent alternatives that allow to invest in a progressive fashion so that the investments are not lost or end up in useless infrastructure. The current work explores the option of using a remote runway in the location of the new airport so in case the original project is delayed or even cancelled the objective of absorbing more traffic is still achieved. The results indicate that the proposed infrastructure allows the growth of the airport while the operational indicators are not in bad shape as taxi times are similar to those in airports with remote runways such as Schiphol in The Netherlands.

Keywords: simulation, airport, congestion, taxi time, airport failure, downturn, over dimension, NAICM

1. INTRODUCTION

Mexico City Airport (ICAO:MMMX) is the main airport of the metropolitan region. It is an airport that has been evolving since the early 20^{th} century. Due to the economic evolution of the country and the main city (Mexico City), the airport has been evolving in parallel. Its evolution at the beginning was due to the local carriers mainly Mexicana and Aeromexico, some international mainly from the US have been also part of the traffic flying from and to Mexico. Mexico has no Open skies policy however recently Mexico signed an agreement with the US that allows American airlines to fly from anyplace in the US to any airport in Mexico (Transportation 2014). Due to this and the expectancy of a trend towards Open skies, the demand is expected to grow in the coming years (around +5% in aircraft movement and +12.2% in number of passengers).
In 2015, the airport network of Mexico as a country transported more than 73 million passengers and about 655 500 tons of cargo. The 63% were domestic passengers and 37% internationals. Mexico City International Airport moved 38.43 million passengers, more than a third of the total traffic of the country (SCT, 2015), which makes it the major airport in Mexico and one of the most important airports in Latin America. The domestic general aviation sector accounted for the 8.5% of the total movements. For the coming years, 41 million passengers have been forecasted by 2020.

1.1. The Situation of Mexico City Airport

Mexico City Airport is considered key for the development of the metropolitan region in Mexico and also for the development of the country. The development of the new airport in Mexico City is under progress however due to different causes the project has suffered delays and it is at risk of not being on time (De Jung 2017). The airport once finished will have a final capacity of 120 mill pax/yr. However this airport will not be operative until 2020 (only the first phase). As expected, Mexico City as a destination is still growing and the country has also gained importance as a tourist and business destination.
According to OCDE, the MMMX's slot allocation process is one of the major hurdle for airline competition, the claim by the government is that airport is fully used and thus competition by new entrants is restricted. In 2013, the Ministry of Communications and Transportation (SCT) issued a new Saturation Declaratory for the Airport (DOF, 2013). According to Mexican authorities, the maximum capacity is limited to 61 operations per hour (with hard restriction of 40 landings per hour) under optimal conditions. However, during adverse weather conditions the limit will drop to 30 or even 20 landings per hour.
There are several consequences of MMMX being constrained:

- Traffic at regional airports bound for AICM suffers delays (ground delays) due to ATC problems at MMMX.
- The result is increasing delays in the domestic aviation system and underutilization of aircraft by the airlines.
- Different queues appear at different locations in the airport making the ATC a real challenge for the controllers

- ATC apparently gives priority to intercontinental flights hence domestic flights will suffer the most the lack of slots.

On the other hand, low cost carriers and international airlines put pressure on the airport to let them enter to operate in the airport. In consequence with the agreement with the US, recently JetBlue, SouthWest started operations from the US to Mexico City Airport putting more competition to the Low Cost Carriers currently operating at the airport, so the expectation of continued growth will be maintained. Fig. 1 illustrates the growth trend of recent years in which the impact of the LCC is revealed.

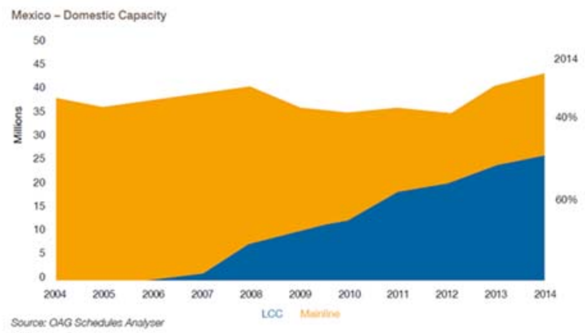


Figure 1. The growth trends in MMMX

Besides that not only the narrow body aircrafts want to operate in the airport, also the biggest commercial aircraft have claimed their right to operate there. Last year AirFrance started with the operation of the A380 three times a week to Mexico City and Lufthansa and Emirates have stated that they have intentions to start operating with the A380 from Frankfurt and Dubai to Mexico City respectively (CAPA 2014).

For all these reasons and with the risk of severe delays in the delivery of the new airport that will absorb all the expected growth, it is critical to evaluate the current and alternative solutions to the airport saturation. The required tool needs to be able to incorporate the different key elements that determine the capacity such as infrastructure configuration, taxiways, runways, restrictions, airlines business models, weather, and the uncertainty inherent to those elements for understanding the emergent dynamics that appear once all the elements are put together in place.

Due to the level of investment of the project (between US$3.7bn to US$ 10.3bn), the risk involved in the project is quite high, on the one hand the current airport is under congestion levels that hinder its growth hence it will be necessary to come with a solution such as the construction of a new airport; on the other hand if the levels of demand that justify such a huge investment are not fulfilled then the risk of having a half-empty system is also high. For this reason it is necessary that these types of investments are made in such a flexible way that the previous risks are minimized as much as possible. It has been mentioned that there are diverse factors whose uncertainty might have an adverse effect on the original plans such as economic downturns, strikes, government change, technical problems, among others. For this reasons the authors have emphasized in previous works (Mujica et al. 2017) that the inclusion of simulation in several planning steps are the only way for evaluating the effect of uncertainty in these types of investments. In particular for this type of project, it is proposed intermediate states of the project in which for each of them the risk is inherently minimized. Figure 2 shows the approach proposed.

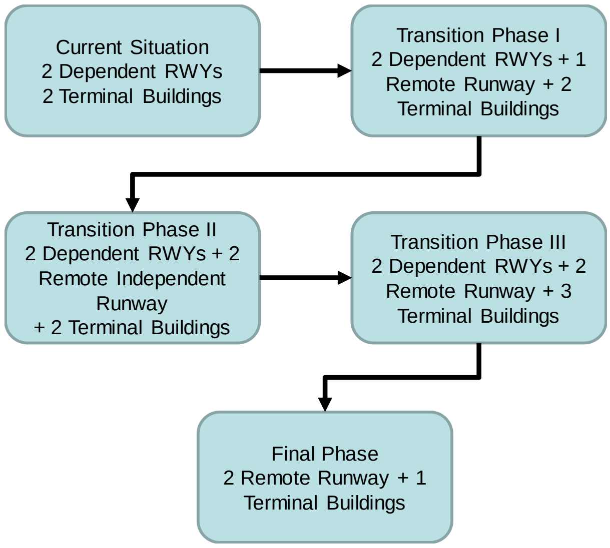


Figure 2. The flexible approach for the implementation of a new airport in Mexico

As it can be appreciated from the figure, the development is progressive and at each phase it is possible to evaluate the risk of failure or anomalous conditions which were not in line with the original plan, and theoretically wise it would be possible to stop the development in each phase if it identified that the original assumptions (demand, operational conditions etc.) are not valid in reality.

In this article we present the analysis of an alternative operational option (Transition Phase I) which is a remote runway in the location of the future airport. The option has been demonstrated as feasible in other important airports in the world. The most familiar one is Schiphol Airport in Amsterdam which has a runway that is almost 6 km away from the passenger terminal building. The developed model allows to evaluate the feasibility of the operation and the identification of potential operative problems as well as the future performance indicators that the new system might have. This configuration can be used as a transition between the current airport to the new one or as a hybrid model if due to unexpected reasons the new airport cannot be finished.

2. SIMULATION MODELS

MMMX runway system consists of two runways with dimensions of 3963x45m and 3985x45m, respectively. Runways operate in a segregated mode which means that one is used for arrivals and the other one for departures. According to Herrera (2012), the arrivals are performed on the runway 05R/23L and departure on runway 05L/23R. The current airport facilities have a total surface area of 790 ha. It has two passenger terminals, Terminal 1 (T1) and Terminal 2 (T2), it includes 96 parking positions, 36 direct boarding gates in T1 and 34 in T2, and a total of 74 operative gates. T1 is used for both national and international flights, whereas T2 is used mainly operated by Aeromexico. Figure 3 depicts an aerial view of the airport and the location of the assigned for the new airport.

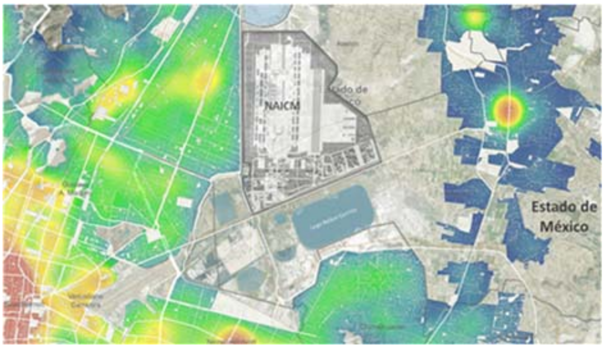


Figure 3. The future location of the new airport of Mexico City

In this work, we used a validated model of the current Airport of Mexico City as the base case and then we implemented the remote runway taking into consideration the logistical challenges of the new system.

2.1. Current System Description

The validated model is the simulation model of the current airport; Fig. 4 illustrates the base model.

Figure 4. Base Case: Current Model of Mexico City Airport

For the current system, the following table presents the values of the different elements in the system.

Table 2. Current Airport Model Description

Element	Value	Comment
Terminal 1 Gates	36	Direct boarding gates
Terminal 2 Gates	34	Direct boarding gates
Turnaround Times	-	Based on different probability distributions
Taxiing speeds (Km/h)	9.26 - 18.5 - 27.7 - 37	Depending on which part of the taxiway A/C are crossing
Runway Occupancy Time Arrival (sec.)	39	Time consumed by the A/C in order to cross the runway at landings
Runway Occupancy Time Departure (sec.)	46	Time consumed by the A/C in order to cross the runway at departures

2.2. Remote Runway System Description

The proposed configuration is presented in Fig.5 where it can be seen the location of the new airport of Mexico city (NAICM). This will be the base for making the model of the proposed configuration.

Figure 5. Remote Runway Model of Mexico City Airport

Table 3 describes the main characteristics of the remote runway model.

One of the main disadvantages of the use of a remote runway is the taxi time from terminals to the runway, however the objective of evaluating this configuration is also for verifying if the impact of taxi time degrades the performance of the system as a whole.

Table 3. Remote Runway Airport Model Description

Element	Value	Comment
Runway Length (m)	3,985	It can accommodates all different types of aircraft
Runway Occupancy Time Arrival (sec.)	39	Time consumed by the A/C in order to cross the runway at landings
Speed runway exit (Km/h)	50	Speed of the aircraft after a landing at the runway exit
Speed Taxing (Km/h)	9.26 - 18.5 - 27.7 - 37	Depending on which part of the taxiway A/C are crossing
Distance to Terminal 1 (m)	13,000	Approx. distance from the remotest runway exit to Terminal 1 area
Distance to Terminal 2 (m)	12,655	Approx. distance from the runway exit to Terminal 1 area

3. EXPERIMENTAL DESIGN

For evaluating the future performance of the airport there were some assumptions taken:

- It is possible to construct a taxiway system that connects the current airport with the new runway
- The analysis of the airspace topology has been taken out of the study
- The effect of the wind direction has been neglected
- The traffic mix is kept similar
- The increase of traffic is only in the available slots (morning and night)

The experimental design takes the current model as the base case and then a second scenario with the increase of 30% is evaluated. Then it will be compared to the operation of the scenario with the remote runway, keeping some variables controlled. Table 1 presents the characteristics of the four Scenarios.

Table 1. Scenario definition

Base Scenario	**Scenario 1**	**Scenario 2**	**Scenario 3**
Current Traffic	+ 30% traffic	Current Traffic	+ 30% Traffic
2 Dependent Runways	2 Dependent Runways	Remote Runway	Remote Runway

With the development of these scenarios we established the feasibility and potential of the proposed solution. The first two scenarios (base case + Scenario1) put focus on what is the impact of the increase of traffic in the current system. As it has been presented by previous studies (Mujica et al. 2016) the limiting element of the airport is the runway system, in the first scenario we present the performance indicators of the operation and the ones if we increase the traffic by 30%. In the following two scenarios we implement the remote runway under the current situation and then the increase of traffic in 30% so that we are able to identify the potential and performance indicators of the future airport with the remote runway.

4. RESULTS

After running the scenarios with the proposed variations, it was identified what the current parameters were and with those values we were able to determine the difference and potential for improvement of the current system. Table 4 presents the main performance indicators of the four scenarios evaluated.

Table 4. Performance Indicators of the Different Scenarios

	Base scenario	Scenario 1	Scenario 2	Scenario 3
ATMs	941	1175	939	1181
Max hourly ATMs	61	62	70	73
Avg time in queue before departing (min.)	22.84	24.06	3.3	3.24
Max A/C in queue from T1	10	12	3	3
Max hourly A/C in queue from T1	5	5	3	3
A/C in queue from T2	28	33	4	4
Max hourly A/C in queue from T2	9	9	4	4
Taxi time from runway to gate (min.)	1.87	1.73	25.81	25.73

As it can be appreciated in the base case, the queues of A/C are high while with the new scenario not only the number of ATMs are increased drastically but also the queues are reduced dramatically. In addition, it is noticeable that in the new configuration if the A/C make the taxi from the remote runway to the terminal building at approximately 37 km/hr it might take approximately 25 minutes time to get to the Terminals, however, the taxi speed can be increased which might reduce the time to get to the terminals. Hourly ATMs are showed in figure 6 and 7 for the Base scenario and Scenario 2, respectively. In these two graphs it is possible to see how the hourly ATMs evolve during the day and the difference between the two scenarios in terms of number of hourly movements. The figures reveal the level of improvement that can be achieved with the new implementation which is in line with the objective of the NAICM.

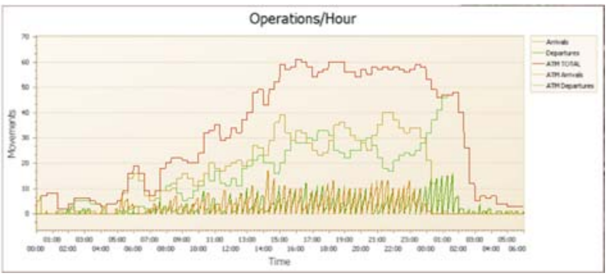


Figure 6: Hourly ATMs Base scenario

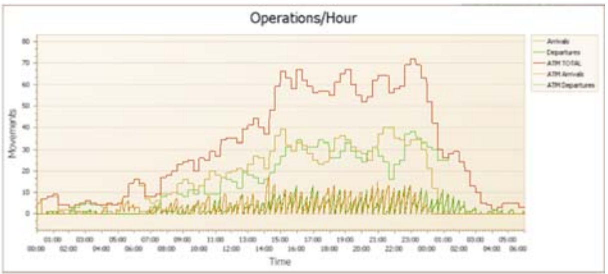


Figure 7: Hourly ATMs Scenario 2 (with remote runway)

5. CONCLUSIONS

It has been presented a simulation model of a proposed operational configuration of a runway which will be part of the new airport of Mexico City. The proposed model is used to investigate the performance indicators that the airport with a remote runway might have in case it is implemented as a transition to the future airport of Mexico City. This approach will enable the decision-makers to reduce the risk of an economic catastrophe if the assumptions regarding demand and operational capacity are not fulfilled. The concept of a remote runway was inspired by Schiphol which is one of the most important hub airports in Europe and has a runway which is approximately 7 km away from the terminal building. The results confirm that the operational values are satisfactory and that with the implementation of the remote runway it is possible to absorb approximately 470,000 ATM movements per year assuming an operation of 17 hrs/day, while reducing highly the levels of congestion. In other words it is expected that with this configuration the airport is able to transport 60-70 mill passengers per year which is approximately the double of the current values. This approach structured in an intelligent way will enable governments and decision makers reduce the risk of failed investments by a very high degree. For this reason the authors foster the use of simulation as a tool for addressing the inherent uncertainty that these systems show. As future work other types of scenarios will be studied such as the ones in which the impact of new aircraft like A380s, B787 or A350 might have in the capacity or how a smart allocation based on the business model of the airlines might improve the efficiency of the operation of the airport.

Acknowledgments

The authors would like to thank the Aviation Academy of the University of Applied Sciences for supporting this study and the Dutch Benelux Simulation Society (www.dutchbss.org) and EUROSIM for the dissemination of the findings of this study.

REFERENCES

CAPA,2014,<http://centreforaviation.com/news/mexico-dgac-air-france-lufthansa-emirates-and-turkish-airlines-interested-in-a380-to-mexico-city-339019>

DOF (2014). Declaratory: saturation at the Mexico City International Airport. Official Newspaper of the Federation (DOF), Mexico.

Secretaria de Comunicaciones y Transportes, (SCT) 2015. Estadística histórica (1992-2014) / Historical statistics (1992-2014).

Feike De Jung, 2017.The billionaire and the airport: could his last act in Mexico City ruin Carlos Slim?, The guardian, June 28th 2017

Mujica Mota M., Boosten G., De Bock N.,Jimenez E.,Pinho de Souza J., 2017. Simulation-Based Turnaround Evaluation for Lelystad Airport, Journal of Air Transport Management

Mujica Mota M., Boosten G., Zuniga C., 2016. MAKE SPACE!:Simulating the A380 Operations in Mexico City Airport, Model-Based Disruption Analysis, in Proc. of EUROSIM 2016, Oulu Finland

Transportation,2014<www.transportation.gov/briefing-room/united-states-mexico-reach-modernized-aviation-agreement>

COMBINED SYNTHESIS/SIMULATION APPROACH TO DESIGN AND VERIFICATION OF PRODUCTION CONTROL STRATEGIES

Juliana Keiko Sagawa[(a)], Gašper Mušič[(b)]

[(a)]Production Engineering Department, Federal University of São Carlos,
Rodovia Washington Luís - Km 235, São Carlos, SP 13565-905, Brazil
[(b)]University of Ljubljana, Faculty of Electrical Engineering,, Tržaška 25, 1000 Ljubljana, Slovenia

[(a)]juliana@dep.ufscar.br, [(b)]gasper.music@fe.uni-lj.si

ABSTRACT
Model based approaches to production control have been discussed in a variety of sources. Several attempts to derive production models by data mining and machine learning techniques have been reported recently. A more classical approach is to develop a discrete simulation model and explore various control scenarios in simulation setting. Another line of research explores the usability of analytical models in combination with standard control design techniques. More specifically, an application of bond graph modelling to a multi-product manufacturing system is reported in the literature. During the modelling process the discrete production dynamics is "fluidised" and analytical model is derived in matrix form. This paper extends the previous work in the sense that analytical model is used to tune the PI production controller and, for verification purposes, the discrete event simulation is applied, which renders a more realistic representation of discrete production dynamics. Among others, various disturbances in the planned production and their effect on the performance of the developed control strategy are more easily evaluated this way.

Keywords: Manufacturing systems, modelling, simulation, production control

1. INTRODUCTION

Contemporary manufacturing systems are characterized by flexibility and are dynamic multiproduct manufacturing systems operating in volatile environment. Related management methods are complex and can significantly benefit from computer support techniques including modelling and simulation.

Model based approaches to production control have been discussed in a variety of sources. Related modelling approaches are developing in line with the evolution from static deterministic to dynamic stochastic analysis of these systems. Viswanadham and Narahari (1992) state that mathematical modelling plays a vital role in the design, planning and operation of FMSs. They give a review of performance evaluation techniques of FMSs using stochastic modelling by Markov chains, queueing networks, and stochastic Petri nets. Further review on applications of queuing theory based techniques in manufacturing is given by Papadopoulos and Heavey (1996). Numerous works use discrete-event simulation for analysis and optimization of manufacturing systems. Longo (2013) proposes a simulation-based tool for solving short period production planning problem within a real manufacturing system. Jain et al. (2017) discuss manufacturing data analytics based on a virtual factory representation, which includes multi-resolution modelling capabilities.

A large number of works is applying Petri nets (PNs) for modelling various aspects of FMSs. In Recalde et al. (2004) a set of examples is collected illustrating the use of PNs for performance evaluation, performance control in terms of schedule optimization, modelling production management strategies, such as Pull Control and Kanban and also deadlock prevention techniques. Among others, hybrid models are discussed where discrete event view is combined and/or substituted by continuous view. Discrete state is relaxed into continuous variables, which opens the way to dynamic performance analysis of manufacturing systems as well as application of control design techniques.

Glavan et al. (2013) present an approach where a manufacturing performance model is derived from historical data on key performance indicators (KPI). Data mining techniques are used to derive a prediction model that is used in model predictive control setting for selected target KPI values.

A more classical approach is to develop production control strategies using analytical models in combination with standard control design techniques. Wiendahl and Breithaupt (2000) use a continuous flow model of a single work centre and develop a backlog controller and a work-in-process controller. Ortega and Lin (2004) present a review of control theoretic methods applied to production-inventory systems with classification of the used models and discuss their integration into production management hierarchy. A recent review of control theoretic approaches to production control is given by Duffie et al. (2014).

Another approach is to develop a dynamic simulation model and explore various control scenarios in a simu-

lation setting. Sagawa et al. (2017) apply bond graph modelling to a multiproduct manufacturing system. During the modelling process the discrete production dynamics is "fluidised" and an analytical model is derived in matrix form. In Sagawa and Freitag (2016) the model is used to explore PI-controller-type production control strategies by simulating a closed loop model.
This paper extends the previous work of Sagawa et al. (2017) in the sense that the analytical model is used to tune the PI production controller and, for verification purposes, the discrete event simulation is applied, which renders a more realistic representation of discrete production dynamics. Among others, various disturbances in the planned production and their effect on the performance of the developed control strategy are more easily evaluated this way.

2. MODELLING OF THE MANUFACTURING SYSTEM

The modelling of manufacturing system in this paper follows the approach of Sagawa and Nagano (2015), Sagawa et al. (2017). The model depicts the dynamics of multi-product manufacturing system and was developed based on bond graph methodology. This brings the advantage of modularity, which allows the integration of different systems, subsystems and components. A dynamic model is developed representing the actual work in process (WIP) and actual production quantity over time. The objective of this dynamic modelling is to keep the WIP levels of the production system under control, even when the system faces disturbances. Bond graphs modelling allows an easy implementation of closed loop models. Hence, different automatic control strategies may be implemented and evaluated, grounded in a systematic dynamic analysis.
In order to develop the bond graph multi-product model, the steps of the general methodology for dynamic modelling (Doebelin 1998) were followed. To apply these to a manufacturing system, however, some issues had to be resolved, such as: how to represent multiple products in the bond graph model; and how to define the constraints of the problem, i.e., how to incorporate the data related to the multiple products in the model. The related solutions are described in Sagawa et al. (2017). Here, only an illustrative manufacturing system will be presented with a corresponding matrix model derived by the described methodology.
The model is a simplified version of the multi-station and multi-product model used in the aforementioned works. The example is based on real manufacturing system and can be considered as a stage in the manufacturing process of polymer bags and fabric in a textile company (Figure 1).
The process starts with the extrusion of the polymer. The molten mix is forced out through the die head into a cooling tank, in the form of a film. After quenching, the film is slit into tapes by a row of equally spaced blades. The resulting tapes are stretched and annealed by passing them over and under a set of rolls. The finished tape yarns are then wound onto packages (bobbins). The weaving of

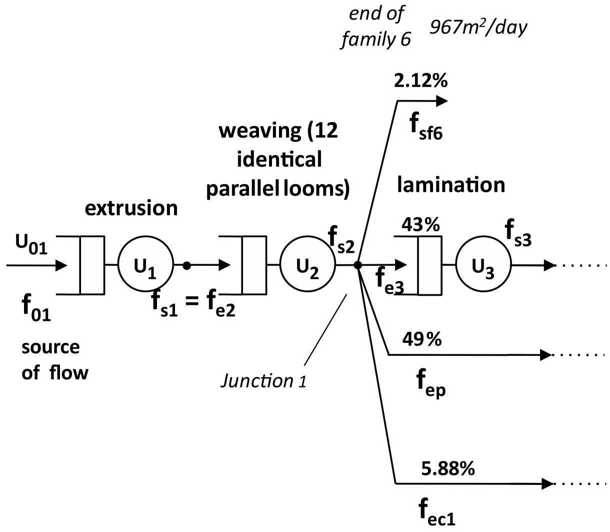


Figure 1: A stage in manufacturing system

the polypropylene tapes is done in circular looms, which can produce tubular cloth of various widths. Depending on the specifications of the products, this cloth goes through different finishing processes, such as printing or lamination.
The model of the described manufacturing stages in matrix form is shown in equation (1)

$$\begin{bmatrix} q_1' \\ q_2' \\ q_3' \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0.43 & -1 \end{bmatrix} \cdot \begin{bmatrix} U_1 min(1,q_1) \\ U_2 min(1,q_2) \\ U_3 min(1,q_3) \end{bmatrix} + \begin{bmatrix} U_{01} \\ 0 \\ 0 \end{bmatrix} \quad (1)$$

where q_1 to q_3 is the instantaneous production volume stored in the three buffers of the system, U_1 to U_3 are the processing frequencies of the workstations, and U_{01} is the processing frequency of the source of raw material that feeds the system. The term $min(1, q_i)$ is the result of fluidization of material flow, which is inherently discontinuous. The expression enables to properly handle the material flow in case a buffer is empty, see Sagawa and Nagano (2015) for details.

3. PI PRODUCTION CONTROL STRATEGY

The control strategy under investigation aims at maintaining the required buffer levels to meet the demanded daily production quantities of different product families. Figure 1 shows the target for one product family ($967m^2/day$), the other targets are defined further along the stages of manufacturing process. Nevertheless, the shown flow percentages in *Junction 1* indicate steady state material flow requirements. E.g. the source flow must correspond to 45548 m^2/day, which is also the steady state processing frequency of workstation 1 and workstation 2. The steady state processing frequency of workstation 3 corresponds to 43% of the total flow.
In order to achieve the desired WIP levels in the plant, the processing frequencies can be varied around steady state values. A suitable control law must be applied, and this was studied in Sagawa and Nagano (2015), Sagawa et al. (2017) and Sagawa and Freitag (2016). Among others,

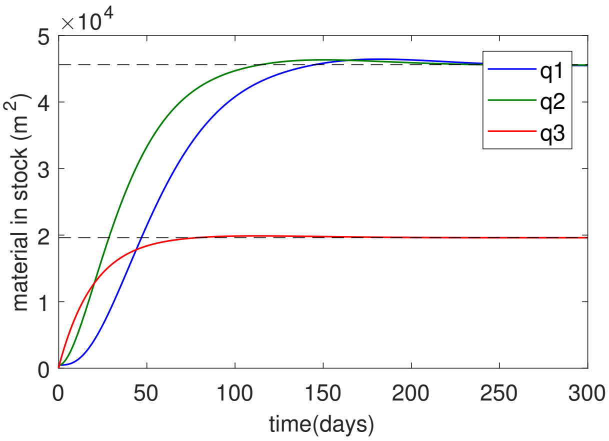


Figure 2: Building of WIP when starting with empty buffers

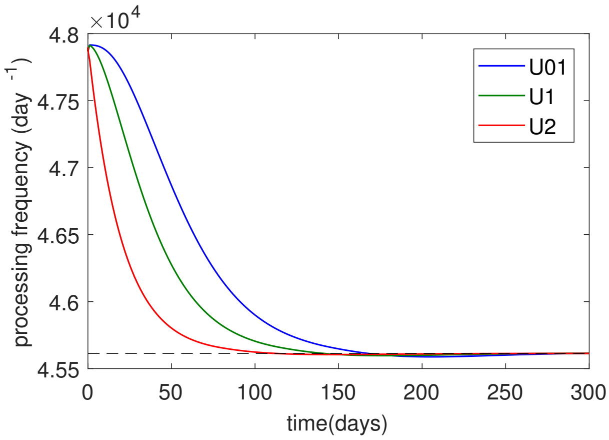


Figure 3: Adjustments of processing frequencies and material input flow

PI controllers were investigated in Sagawa and Freitag (2016). The controllers operate based on the normalized error

$$e_j = \frac{q_{jc} - q_j}{q_{jc}} \quad (2)$$

where q_{jc} is the desired buffer level, and q_j is the actual buffer level at position j. The control output is interpreted as a relative deviation from the steady state processing frequency. The actual processing frequency is determined by

$$U_i = U_{ip}(1 + u_{jPI}) \quad (3)$$

where U_{ip} is the steady state processing frequency of workstation i, and u_{jPI} is the control output calculated on the basis of level error in buffer j. In the case of linear material flow the processing frequency U_i is influenced by the level of first buffer downstream of workstation i. In case of parallel branches, the minimal control output of the corresponding controllers is considered.

Figure 2 shows the initial building of WIP when starting with empty buffers, until the processing frequencies stabilize at required material flows. Corresponding processing frequencies of workstations 1 and 2 as well as input flow U_{01} are shown in Figure 3. Note that processing frequency of workstation 3 is kept constant at steady state value U_{3p} and is not shown in the figure.

Figure 4 shows the WIP levels when the process operation is disturbed by temporary breakdown of workstation 2. The shown responses are obtained by continuous system simulation in Simulink where equation (1) is implemented and corresponding PI controllers are added.

In accordance with Sagawa and Freitag (2016) the controller parameters are set to $K_p = 0.05$ and $K_I = 0.001$. The integrator state was limited at ± 1.

4. DISCRETE EVENT SIMULATION BASED VERIFICATION

Continuous simulation is well suited to testing control strategies and controller tuning, yet the operation of the production system involves discrete operations. The question is how well the abstract continuous dynamic

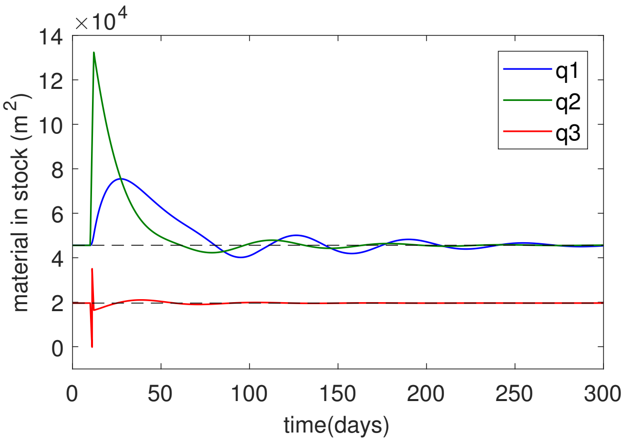


Figure 4: WIP levels at temporary workstation 2 breakdown

model represents the actual production dynamics. A partial answer to that can be obtained by verifying the developed production control strategy in discrete-event simulation (DES) setting.

A DES model was developed in SimEvents, a discrete-event simulation add-on to Simulink. The production system is represented as a network of queues and servers, and SimEvents permits to link the entity generation rates and server service rates to external signals, which can originate from continuous simulation. This way the derived PI control strategy can be tested also in DES setting. Still, some adjustments of the PI controllers are required to obtain comparable results. Firstly, a pure continuous implementation of the controllers is not feasible, since the buffer levels do not change continuously. This would also not be realistic considering potential industrial implementation, where on-going adjustments of the workstation processing frequencies would be impractical. Instead it was decided to determine the frequencies on a daily basis, which corresponds to discrete-time implementation of the PI controller with sampling time of 1 day. Secondly, the discrete-time implementation required an alternative implementation of the integrator saturation, which was introduced in continuous version to prevent integral controller windup.

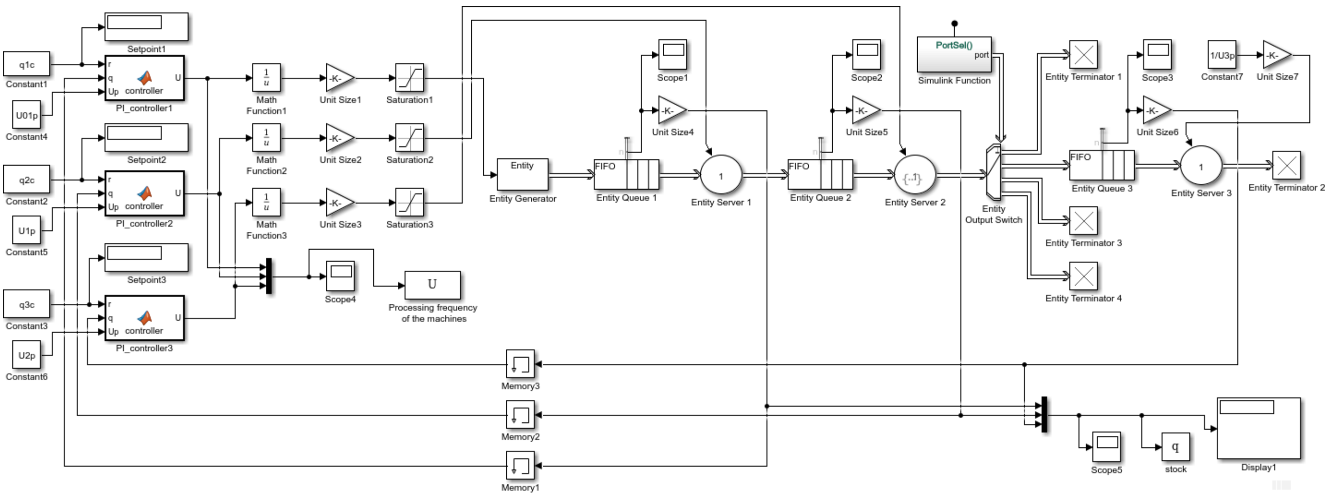


Figure 5: Simulation scheme - DES simulation

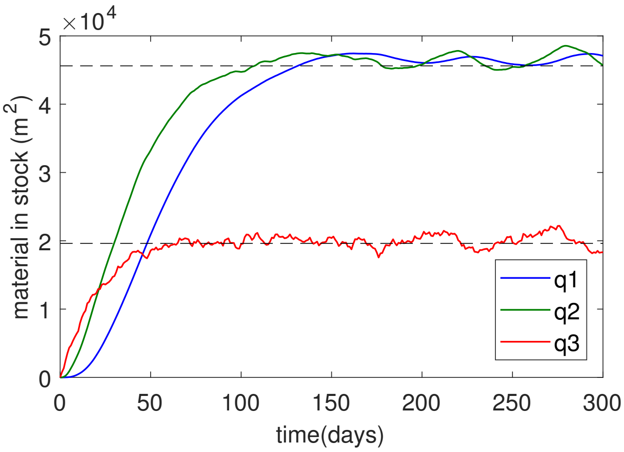


Figure 6: Initial building of WIP - DES simulation

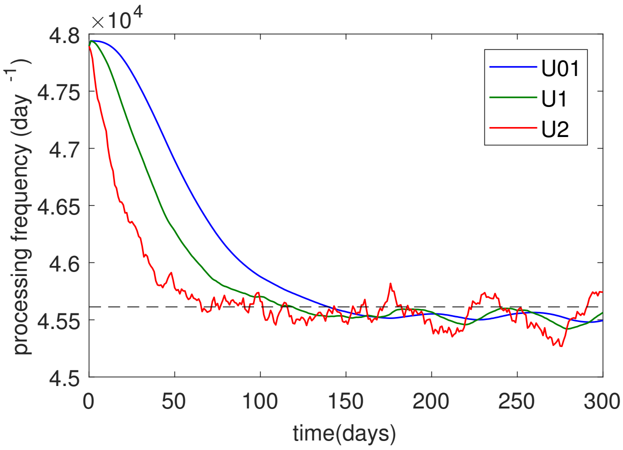


Figure 7: Adjustments of processing frequencies and material input flow - DES simulation

Figure 5 shows the main part of the simulation scheme. The manufacturing system is modelled as a series of three queue-server pairs, each representing a workstation in the system with the accompanying input buffer. Entities in the model represent a fixed quantity of product (*unit_size* parameter) expressed in m^2. A path splitting element is inserted in between workstations 2 and 3, which models the division of material flow. Entities are routed through the split based on result of the *PortSel()* function, which is programmed to act randomly but also to maintain prescribed values of average material flow through the split output ports. In particular, the average flow through port 2 must correspond to 43% of the total flow.
WIP levels in the system directly correspond to the number of waiting entities in queues. These are used as feedback signals to the controllers. Controller outputs U_i represent input material flow to workstation 1, and processing frequencies of workstations 1 and 2, respectively. Input material flow requirement is translated into the intergeneration time of the *Entity Generator*, while processing frequencies are transformed into service time of the related *Entity Servers*. This way the production control loop is closed within the simulation model. *Memory* blocks are inserted in feedback paths to break the algebraic loops.
Figures 6 and 7 illustrate how the controller builds the required WIP levels when starting from empty buffers and using DES model. The stochastic behaviour in the shown result is mainly induced by the implementation of the material flow splitting in DES, which is based on specified discrete distribution of random values.
This is closer to operation of the actual system, where the material flow is not necessarily continuous in between production orders. Each product family includes different products, and machine setup is needed in between production of different products, i.e. before executing the production orders. The actual material flow in a given moment depends on the actual production orders prescribed by master production schedule (MPS), detailed operation scheduling, standstills and other disturbances etc.
Therefore the product mix percentage only represents average values, the actual daily quantities of product families vary during the manufacturing operation.

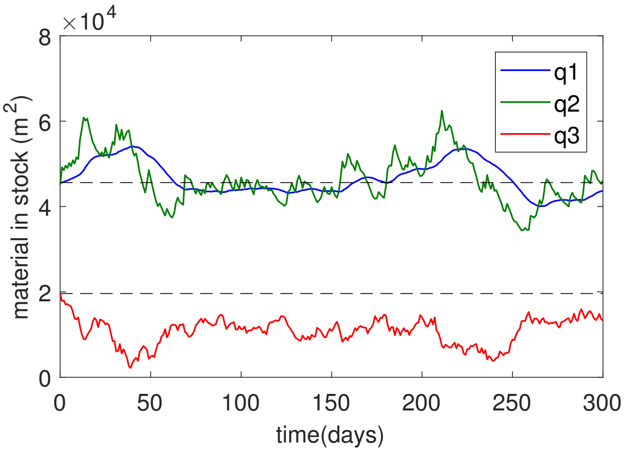


Figure 8: WIP levels in the presence of disturbances - machine breakdowns

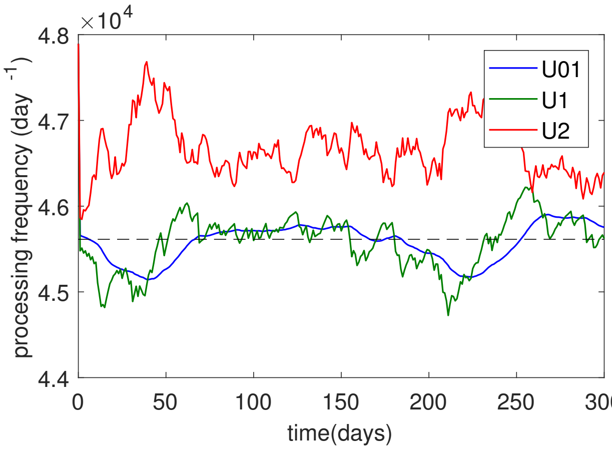


Figure 9: Adjustments of processing frequencies and material input flow in the presence of disturbances - machine breakdowns

Similarly, other random effects could be explored, such as nondeterministic processing frequencies of the workstations etc. E.g., instead of simulating complete workstation 2 breakdown, Figures 8 and 9 illustrate WIP levels and related adjustments of processing frequencies when occasional machine breakdowns occur within workstation 2. Note that workstation 2 consists of 12 parallel looms and simulated scenario assumes exponentially distributed breakdown times of a single loom with average rate of 1/2 day and uniformly distributed repair times in the interval (0,1). Figure 10 shows machine breakdown times where value 1 indicates single machine breakdown and value 2 simultaneous breakdown of 2 machines.
Comparing Figure 6 and Figure 8, it is possible to see that the WIP levels of buffers 1 and 2 present oscillations with a little bit higher amplitude (and present also high-frequency oscillations, in the case of buffer 2). Nonetheless, the results show that, even in the presence of these disturbances, the WIP levels of buffers 1 and 2 remain around the reference values. This indicates that the defined control strategy for the source of flow and for the workstation 1 remains effective in the presence of disturbances. For the buffer 3, an expected offset is observed, since the workstation 2 had its capacity reduced. In addition, it may be noted that the controller tries to compensate the machine breakdowns by increasing the processing frequency of workstation 2 (Figure 9). This means that the remaining parallel machines of this workstation would have to work faster or for longer times (i.e. to do overtime) in order to compensate the downtimes of the temporarily broken machines. These expected results shown that the continuous model and the discrete model are coherent.
The continuous formulation is suitable for this manufacturing system presented in previous works since the manufactured products are continuous until they reach the cutting operation (i.e. the final products are based on a continuous tubular fabric, in this case). However, the representation of the system as a discrete model is a relevant step towards its application in a much broader range of cases, where the products are measured in discrete units. Even if the product quantity is measured continuously,

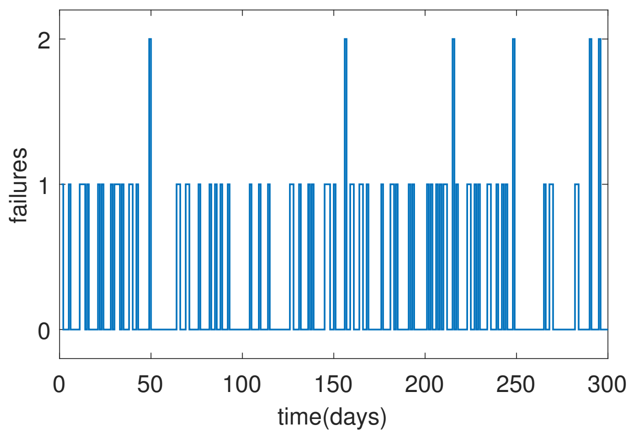


Figure 10: Machine breakdowns

such as in the given case, it makes sense to verify how the controller designed for the aggregated continuous model deals with non-ideal and discontinuous material flow situations that are present in reality. On the other hand, building exact simulation model that would exhibit all the specifics of the given manufacturing system is often too difficult.
In addition, the capacity adjustments prescribed by the controller do not correspond, in practice, to a direct increase or decrease in the production rate of a given machine, since this rate is subjected to technological constraints and can only vary within a specific range, as known. Rather, these capacity adjustments may be converted into an indication of downtime or overtime of the machines. In other words, if the controller prescribes that the capacity must be increased in 20%, it does not mean increasing the production rate in 20%, because this is usually infeasible. In practical terms, this adjustment has to be implemented by means of overtime or an extra shift, for instance, and the representation of the system as a discrete model allows to include and test these more realistic conditions.
Unfortunately, it can be shown that chosen control strategy does not cope well with larger disturbances in the

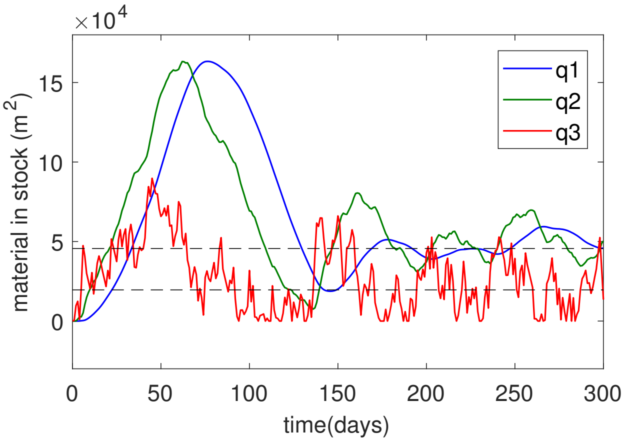


Figure 11: DES simulation considering lot sizes

material flow. Figure 11 shows the results of WIP level control when material flow splitting is not smooth but considers lot sizes. A product dependent lot size of material goes through the same path over the manufacturing system. In case of larger lots this substantially disturbs the flow balance and represents a challenge for the WIP level controller. As a result, large overshoot is present in the initial WIP building phase and oscillations around required WIP levels remain present after the required buffer levels are reached. Additionally, integrator state limitations had to be relaxed in order to avoid steady state error (offset) of WIP levels. The improvement of the control strategy remains one of the issues for the further research work, especially when considering lot sizes.

6. CONCLUSIONS

The verification of the production control strategies indicates that the developed continuous model adequately represents the production system dynamics and is useful for development of production control strategies. In comparison to the controlled DES model, the simulation with continuous model is computationally more efficient and therefore much more appropriate for controller tuning. Verification of the control strategy in DES setting is advantageous in the sense of more realistic testing scenarios that can be easily implemented. It therefore presents a useful stage in verification of production control strategies.

REFERENCES

Doebelin, E. O., 1998. *System dynamics : modeling, analysis, simulation, design*, New York : Marcel Dekker.

Duffie, N., Chehade, A. and Athavale, A., 2014. Control theoretical modeling of transient behavior of production planning and control: A review, *Procedia CIRP*, 17, 20–25.

Glavan, M., Gradišar, D., Strmčnik, S. and Mušič, G., 2013. Production modelling for holistic production control, *Simulation Modelling Practice and Theory*, 30, 1–20.

Jain, S., Shao, G. and Shin, S.-J., 2017. Manufacturing data analytics using a virtual factory representation, *International Journal of Production Research*, 55 (18), 5450–5464.

Longo, F., 2013. On the short period production planning in industrial plants: a real case study, *Int. J. Simulation and Process Modelling*, 8 (1), 17–28.

Ortega, M. and Lin, L., 2004. Control theory applications to the production-inventory problem: a review, *International Journal of Production Research*, 42 (11), 2303–2322.

Papadopoulos, H. and Heavey, C., 1996. Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines, *European Journal of Operational Research*, 92 (1), 1–27.

Recalde, L., Silva, M., Ezpeleta, J. and Teruel, E., (2004). *Petri Nets and Manufacturing Systems: An Examples-Driven Tour*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 742–788.

Sagawa, J. K. and Freitag, M., 2016. A simulation model for the closed-loop control of a multi-workstation production system, *9th Eurosim Congress on Modelling and Simulation*.

Sagawa, J. K. and Nagano, M. S., 2015. Modeling the dynamics of a multi-product manufacturing system: A real case application, *European Journal of Operational Research*, 244 (2), 624–636.

Sagawa, J. K., Nagano, M. S. and Neto, M. S., 2017. A closed-loop model of a multi-station and multi-product manufacturing system using bond graphs and hybrid controllers, *European Journal of Operational Research*, 258 (2), 677–691.

Viswanadham, N. and Narahari, Y., 1992. *Performance Modeling of Automated Manufacturing Systems*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Wiendahl, H.-P. and Breithaupt, J.-W., 2000. Automatic production control applying control theory, *International Journal of Production Economics*, 63 (1), 33–46.

AUTHOR BIOGRAPHY

JULIANA KEIKO SAGAWA received B.Sc., M.Sc. and Ph.D. degrees in production engineering from the University of São Paulo, Brazil in 2006, 2008, and 2013, respectively. She is Adjunct Professor at the Faculty of Production Engineering of the Federal University of São Carlos. Her research interests include dynamic modelling of production systems, planning, scheduling and some organizational topics such as integration and servitization. More information can be found at: `http://lattes.cnpq.br/6549481623896786`.

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 Full 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`.

DOWNSIZING THE AIR FREIGHTER: SIMULATION-BASED STUDY OF KLM CARGO IN SCHIPHOL AIRPORT

Miguel Mujica Mota [a], Abdel El Makhloufi[b], Pedro González Hernández[c]

[a]Aviation Academy, Amsterdam University of Applied Sciences,
[b]Research group Smart Mobility & Logistics, Amsterdam University of Applied Sciences
1097 DZ Weesperzijde 190, Amsterdam, The Netherlands
[c] National Autonomous University of Mexico, Mexico

[a] m.mujica.mota@hva.nl, [b] a.el.makhloufi@hva.nl, [c] phetergh@gmail.com

ABSTRACT

Recently KLM has revealed the plan to downsize the full-freight cargo fleet in Schiphol Airport, for that reason it is important for the company and the airport to explore the consequences of moving the cargo transported by the full freighters into the bellies of the passenger flights. The consequences of this action in terms of capacity and requirements are still unknown for the stakeholders. The current study illustrates that once the freighters are phased out, the commercial traffic needs to adjust mainly their load factors in order to absorb the cargo that was previously transported by the full freighters. The current model is a version that includes the airside operation of the airport and also the vehicle movement which allows addressing the uncertainties of the operation as well as the limitations and potential problems of the phasing-out action.

Keywords: air cargo, airport operations, transhipment, full freighter, belly cargo, uncertainty, warehouses, KLM cargo

1. INTRODUCTION

On a worldwide level, the transport of commercial cargo is a key economic indicator of international trade as well as a thermometer for the state of the global economy. The supply chain and logistics industry exists to connect manufacturers with suppliers and middlemen shippers with the end customer (Feng, 2015) and for some valuable goods aviation plays an important role.

In terms of the business model between cargo operations and passenger operations, there exist many similarities as well as differences. According to some authors, the key differences between cargo and passenger operations are uncertainty, complexity and flexibility (Bartodziej et al., 2009; Leung et al., 2009; Li et al., 2009; Wang and Kao, 2008).

Air cargo transport is more complex than passenger transport because the former involves more players, more sophisticated processes, a combination of weight and volume, varied priority services, integration and consolidation strategies, and multiple itineraries of a network than the latter.

Key similarities include:

- Similar revenue management tools and concepts such as demand forecasting, overbooking, capacity forecasting, route generation and bid prices.
- A growing movement towards network capacity as opposed to leg capacity, in a similar fashion for passenger airlines. However, cargo shipments do not care about the quality of the service (nonstop vs. connecting) as much as passengers, disregarding time.

Major differences include:

- Cargo focuses on building customer, supplier, and retailer relationships because of a limited number of customers.
- Cargo focuses on profitability rather than load factor. Average load factors for passenger flights hovered around 80% in 2015 while cargo load factors in the passenger flights are around 40% and 70% in full freighters.
- Cargo requires medium to long term allotment management - optimal assignment of space to customers.
- Different optimization factors such as freight mix based on density of payload and revenue.
- Optimization of schedule based on alternate constraints on noise and airport utilization (night-time flying).
- Cargo ground transport in Europe is at least approximately 700 km radius from airport. For passengers is mostly less than 3 hr. travel.
- Air cargo transport has higher uncertainty than passenger transport in terms of capacity availability. In passenger transport, passengers may cancel reservations, and a small number of passengers may not show up. However, in capacity booking for air cargo, freight forwarders have to pledge the use of the cargo capacity on specific flights twelve or six months ahead (Amaruchkul et al., 2011).
- Because freight forwarders do not need to pay for unused capacity, they may book more than the actual

needed capacity to cut risks or to compete with others forwarders.

- Many bookings in air cargo can be cancelled and/or rebooked because airlines typically do not charge for changing reservations. Therefore, the booking process is subject to considerable volatility (Petersen, 2007).
- Forecasting cargo capacity is more complex than forecasting passenger capacity. While the capacity of a passenger aircraft is fixed by its number of seats, cargo capacity depends on the type and dimensions of the container used (called unit load devices, ULDs), and specified according to pivot weight, pivot volume, type, and centre of gravity (Leung et al., 2009).
- Transhipment itineraries between an origin and destination (OD) pair for cargo transport benefit the airline more than they benefit passenger transport. In general, all major airlines operate within so-called hub-and-spoke networks. Both passengers and cargo are transported from many different origins to a small number of hubs, where passengers and cargo are consolidated and then transported further to other hubs by wide-body aircrafts. The total number of transits are limited for passenger transport, and much larger for air cargo transport i.e. air cargo can be transhipped via several intermediate airports from the origin to the destination to meet the delivery time (Amaruchkul et al., 2011). The airline only needs to declare the origin, stopover (transit) airports, and destination to the forwarders and can decide on the optimal use of transhipment itineraries of air networks.

Common items shipped by air include perishables, pharmaceutical products, high-tech and electronics, clothing, animals and high-value products such as diamond, art cars among others.

Because of these differences, air freight accounts for less than 1% of total freight carried by all transport modalities (air, sea, water, and road) in terms of both volume and weight. However, air freight accounts for about 40% of its value (Damme et al, 2014) and almost 1% of global GDP is spent on air transport (IATA 2016).

1.1 The Situation of KLM in Schiphol

The cargo operations at Schiphol airport faces major challenges from macro- developments (such as rapid changes in aviation sector and cargo market, technology/ICT revolution, transitions to green and circular economy, e-commerce and 3D-printing), increasing volumes and volatility and uncertainty of airfreight and logistics, and last but not least changes in freight strategy of the hub carrier KLM cargo. The last one consists in reducing its full-freighter and increasing cargo transport in the bellies of the passenger's aircrafts. Schiphol is important for the logistics sector and economic growth of the Amsterdam metropolitan region economy. Airfreight operations are fully concentrated at Schiphol airport, with minimal cargo activities in other regional airports such as Maastricht and Eindhoven airports which makes it a key node of some important and valuable products such as flowers or pharmaceuticals.

Schiphol is ranked third in Europe in term of airfreight aggregated volumes (2 million tons in 2015), behind Charles-de-Gaulle and Frankfurt airports.

The main carrier in Schiphol is KLM whose passenger operation accounts for more than 80% of the revenues of KLM-Air-France group, however, an important part of airfreight volumes are transported in combined (belly) aircrafts. In this way, revenues generated from airfreight operations are complementary to passenger operations, especially on intercontinental networks that are difficult to maintain financially.

During the financial year, the Group transported nine billion Revenue Ton-Kilometres of which 75% in the bellies of passenger aircraft and 25% in the dedicated full-freighter fleet, to a network of 316 destinations in 115 countries.

In 2010, confronted with the crisis in the sector, the Group adopted a new "priority to bellies and combis" strategy aimed at optimizing the economics of passenger aircraft bellies. Therefore, the full- freighter fleet is used as a complement to cover the routes not served with passenger flights or for products that cannot be carried in bellies or markets in which belly capacity is not adapted to demand.

Roughly speaking, 30% of the total cargo capacity of Schiphol is handled by KLM and Martinair (a cargo subsidiary of KLM). At December 31, 2015, the KLM fleet comprised 113 aircraft, of which 65 long-haul aircraft and 48 medium-haul aircraft. KLM reduced Martinair cargo fleet from 10 Full Freighters (FF) to only 4 (KLM 2016). The reason for this is that cargo transported in the Bellies is less costly and improve the profit margin of airlines. Especially the popular destinations such as New York and Shanghai, where more passengers are flying to, bellies offer new possibilities to improve operational costs and improve competitiveness of airlines. However, the transition from full freighters to more bellies has tremendous implications on the organization of airport operations (time slots, schedule and punctuality). As a result, cargo operations have to be alienated with the passenger's operations, where both processes should be combined in a limited and common airports areas.

The current study is a model-based approach for understanding the initial consequences of reducing the KLM fleet of full freighters at Schiphol and the knock on effects that this decision might have. As it has been mentioned, the downsizing of the fleet will continue due to the situation previously mentioned. For KLM and Schiphol group the understanding of the consequences of the phasing out of the remaining full freighters is key for improving the operational management of the airport systems which besides all the limitations its growth is still on with consequences in congestion, delays and capacity.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

1.2 Similar studies

In the literature there are some studies that deal with some specific problems of the aviation industry where uncertainty is addressed in a particular fashion. For instance Chandran et al. (2007) proposed a dynamic programming algorithm for robust runway and their work considered uncertainties in the aircraft arrival times, ending up with a trade-off between runway throughput and the probability of violating separation minimum on the runway as objective. Kim and Feron (2011) focused their efforts in the robust gate assignment problem when stochastic delays are introduced. Arias et al. (2013) studied the stochastic aircraft recovery problem by employing constraint programming in combination with simulation techniques. In the work of Lee (2014) the airport surface congestion problem is studied combining optimization with simulation. The work of Mujica Mota et al. (2017) presents a high detailed model in which for the first time include the analysis of an airport operation with the ground handling operation. Other specific material for different problems in aviation can be found in literature however to the knowledge of the authors no similar problem that deals with the simulation of vehicles, and bellies can be found in literature. For this reason we proposed the use of simulation as the ultimate tool for addressing the situation of the presented problem.

2. METHODOLOGICAL APPROACH

Schiphol airport is one of the biggest hubs in Europe, it has 6 runways, however for the commercial traffic only 5 are under operation. The five runways have different dimensions and orientations. Table 1 shows the information of the runways and orientation.

Table 1. Schiphol's Runway Description

Number	Runway Direction/ code	Common Name	Length (m)
1	18R/36L	Polderbaan	3800
2	06/24	Kaagbaan	3500
3	09/27	Buitenveldertbaan	3453
4	18L/36R	Aalsmeerbaan	3400
5	18C/36C	Zwanenburgbaan	3300
6	04/22	Oostbaan	2014

The orientation of the different runways together with the taxiway system make the airport very complex. Figure 1 illustrates the runway system in which also the taxi ways can be appreciated.

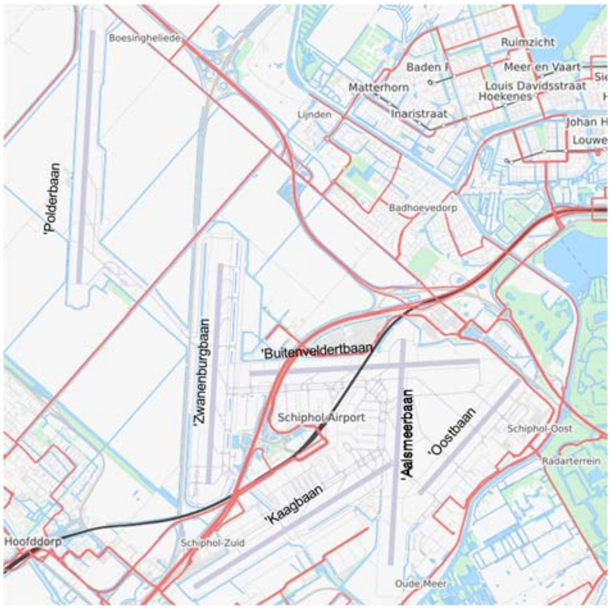


Figure 1. The location of runways in Schiphol

The runway configuration is continuously changed, in general terms it changes 16 times during the day due to the noise restrictions imposed to the system. The configuration is adjusted according to the noise quota and also to the peak hour of the day, for the departure peak sometimes 2 runways are used for departing and 1 runway for arriving while in arrival peaks 2 can be used for arriving and one for departing in an independent fashion. Just in exceptional cases 2+2 configuration can be used but that is not common. The Polderbaan is used very frequently due to its remote location from the city therefore the annoyances of noise are minimized by putting priority in the use of this runway.

Regarding the cargo operation, several companies are located in Schiphol, and KLM has its own warehouse in between the Kaagbaan and terminal building as Figure 2 illustrates. In that location the transhipment of goods is performed and also exporting products arrive to the airport via cargo trucks.

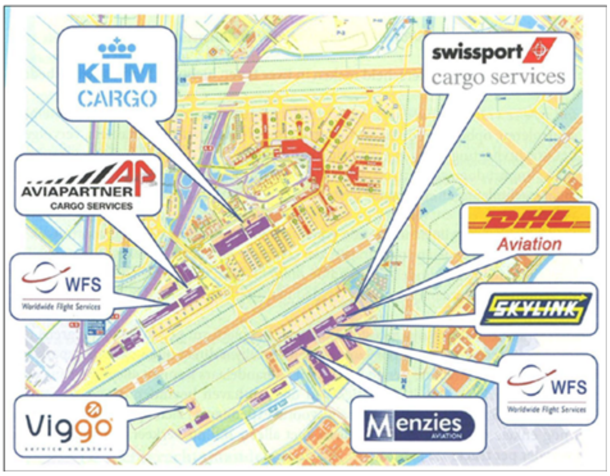


Figure 2. The most important warehouses in Schiphol

2.1 Assumptions of the model

For the evaluation of the transition from Full freighter to the transport of cargo in the bellies of the aircraft (A/C) we took the public information from KLM and Martinair together with the traffic numbers from OAG database for developing and analysing four different scenarios. We performed some assumptions based on different studies and reports and we came up with the initial numbers. These numbers together with the configuration of the airport allowed us to build the initial model in which the operation of vehicles+A/C can be simulated and the uncertainty of it addressed. The following table summarizes the main assumptions used for the model.

Table 2. Assumptions for the model

Parameter	Domain	Description
Traffic	Piers B- M	The traffic is generated only for KLM and Martin Air, based on a day of traffic. The allocation is performed on a probabilistic way based on the amount of available gates. Thus no attention is paid to the type of aircraft.
Runway usage	The scenario is based on a peak arrival mode: 2 RWY for arrival and 1 for departure	The arrival are Polderbaan and Zwanenburgbaan, departure is Kaagbaan
Load factor of Aircraft Passengers	0.4	The load factor for the cargo bellies is assumed 40%
Aircraft Output/Input Ratio	0.96	Based on public information the out/in ratio is 96%
Load factor of full freighters	0.7	The load factor of the full freighters is 70%
Truck Vehicles	5.2 tons	This corresponds to the capacity of the cargo trucks
Interarrival time of trucks	Exponential (1.4) hr.	The inter arrival time of trucks is 1.4 hr. average
Truck Ratio Output/Input Ratio	0	Due to the lack of information we assume that no cargo is taken by the trucks from the warehouse.
Runway and Taxi layout		The layout is based on the real configuration of taxiway system and runway system

The main assumption for estimating the amount of cargo that enters and leaves the airport is based on a mass balance of the total input/output reported by Schiphol (Schiphol Group 2016). Figure 3 illustrates the flow of gross tons in which roughly 30% corresponds to KLM+Martinair.

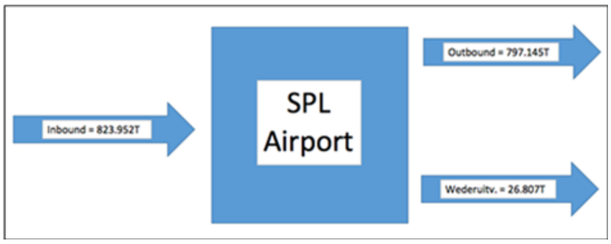


Figure 3. Simplified model of Schiphol Operation

Roughly speaking, from the inbound value, approximately 600 tons get into Schiphol via full freighters, bellies and trucks. As it can be appreciated, the outbound is less than the inbound so that is how the ratios out/in are estimated. The third flow is the amount of cargo that stays for being exported some days after it has arrived to Schiphol. The information was also used for developing the operational model which was developed using a discrete-event-oriented program called SIMIO (SIMIO 2017) which is very flexible for the types of operations performed in the airport. Figure 4 illustrates the simulation model developed which is composed by different elements such as the taxi way network, gates, runways, road for vehicles and the warehouse of KLM.

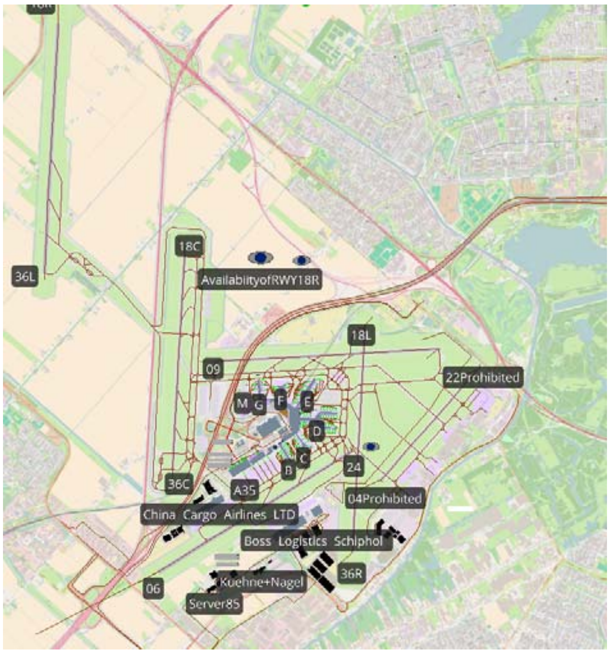


Figure 4. Discrete-event-oriented model of Schiphol

The program combines the object-oriented approach together with a process-based logic which makes it very flexible for making modular implementations and specify processes where the precedence is clearly identified.

3. EXPERIMENTAL DESIGN

Based on the previous assumptions+layout+ traffic we make the initial model of the cargo operation at Schiphol including some basic restrictions such as the limit of runway use by only 1 A/C at the time, runway occupancy times of approximately 50 seconds and the speeds of the taxi operations in approximately 40 km/hr.

With the initial model we were able to develop the base case in which the numbers are evaluated. It is worth to mention that the initial traffic includes the 4 Full freighters and the following scenarios are used for evaluating what would be the impact of phasing out the freighters:

- Scenario 1: reduction to 3 Full freighters and the increase of load factor of passenger A/C to 0.6
- Scenario 2: Reduction to 2 Full freighters at load factor of passenger A/C of 0.6
- Scenario 3: Reduction to 1 Full freighter and load factor of passenger A/C of 0.7
- Scenario 4: Reduction to 0 Full freighter with a load factor of 0.7

The scenarios were run with 15 replications for addressing the initial values of the operation. Table 3 presents the results of the different scenarios.

Table 3. Scenario Results

		Input (KG)	STD. Dev.	Output (KG)	STD. Dev.
Base Case	Bellies	309,875	10,400	298,909	12,923
	Full Freighters	277,900	0	268,729	0
	Total	587,775		567,638	
Scenario 1	Bellies	449,532	52,894	434,504	51,264
	Full Freighters	199,500	0	192,917	0
	Total	649,032		646,285	
Scenario 2	Bellies	468,411	1,134	450,467	1,096
	Full Freighters	133,000	0	128,611	0
	Total	601,411		581,565	
Scenario 3	Bellies	523,985	72,226	506,152	71,555
	Full Freighters	66,500	0	64,306	0
	Total	590,485		570,457	
Scenario 4	Bellies	546,980	0	528,930	0
	Full Freighters	0	0	0	0
	Total	546,980		528,930	

The previous scenarios assume that the passenger bellies are able to get to the load factor of the full freighters (0.7) but in reality this numbers need to be verified with the ground handlers.

As it can be appreciated from the results, as long as we reduce the amount of full freighters, we can counterbalance the lost in capacity of the full freighters by increasing the load factors of the passenger traffic. It is worth to note that until the scenario 3 we are able to counterbalance the lost in capacity by increasing to 0.7 the load factor of the passenger flights. However when we get to the final scenario it is not possible anymore to absorb the lost in capacity with the assumed limit for the load factors of 0.7. This results can be clearly appreciated in the Figure 5 where it can be appreciated the reduction of the operation of the full freighters while the amount of cargo transported by the bellies increases and they are able to transport the same quantity as the base scenario except in the final one.

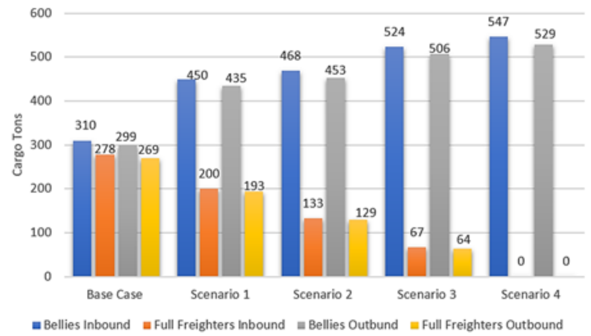


Figure 5. Inbound-Outbound cargo analysis

From the figure one can note that the outgoing cargo in the base case is approximately 570 tons; most of the scenarios are able to transport that amount of cargo except the final one (scenario 4) where the outbound is only 529 tons thus approximately 50 tons of cargo cannot be transported via the bellies of the cargo passengers.

The previous result is quite interesting since it suggests that if KLM is not able to increase their load factors to a higher value than 0.7 (which is the current value for the full freighters) the company will have to implement another strategy if they are not willing to lose the revenue that comes from the cargo operation.

4. CONCLUSIONS AND FUTURE WORK

It has been presented a simulation model of the airside of Schiphol airport which integrates the truck operation and the main elements of the airside operation for making an initial evaluation of what the impact will be once the full freighters of KLM are phased out. The results indicate that the company is able to counterbalance the lost in capacity by increasing the load factor of the passenger flights in the cargo compartments until reaching the level of the current full freighters. The analysis also suggests that if the assumption of the limit for increasing the load factor to the maximum value of 0.7 is correct then the company will have to take a different approach once the last full freighter is removed from the fleet. The strategy could be the increasing of the flight frequency or start new routes with new aircraft. Furthermore, if the company is not evaluating in the short term to take the suggested actions, they can follow a hybrid strategy in which part of the cargo can be transported in the passenger bellies increasing the load factors and at the same time keeping at least one full freighter.

The previous results are initial ones that give light into the potential outcome of the policy of phasing out the full freighters of KLM. In the future the authors will evaluate other important effects such as the uncertainty in the load factors at the bellies and by using real information it will be possible to evaluate what is the correct limitation for the load factors once the other restrictions that play an important role are included in the model such as turnaround times, capacity of ground handlers and on time performance limitations.

Acknowledgments

The authors would like to thank the Aviation Academy of the University of Applied Sciences for supporting this study and the Dutch Benelux Simulation Society (www.dutchbss.org) and EUROSIM for the dissemination of the findings of this study.

REFERENCES

Amaruchkul, K., Cooper,W.L., & Gupta, D. 2011. A note on air-cargo capacity contracts, Production and Operations Management, 20(1), 152-162

Arias, P., D. Guimarans, M. Mujica, and G. Boosten. 2013.A Methodology Combining Optimization and Simulation for Real Applications of the Stochastic Aircraft Recovery Problem. In Proc. of the 2013 EUROSIM Congress on Modelling and Simulation, pp. 265–270. Cardiff, UK.

Bartodziej, P., Derigs, U., Malcherek, D., & Vogel, U. 2009. Models and algorithms for solving combined vehicle and crew scheduling problems with rest constraints: an application to road feeder service planning in air cargo transportation. Operation research-spektrum, 31(2), 405-429

Chandran, B., and H. Balakrishnan. 2007. A Dynamic Programming Algorithm for Robust Runway Scheduling. In Proc. of the 2007 American Control Conference, 1161–1166. NewYork City,USA.

Damme, van.D., Radstaak, B., & Santbulte, W. 2014. Luchtvrachtlogistriek een dynamische keten in perspectief. SDU, Den Haag.

Feng, B. et al. 2015. Air cargo operations: Literature review and comparison with practices

IATA, 2016. Another Strong Year for Airline Profits in 2017. press release 76

Kim, S. H., and E. Feron. 2011. Robust Gate Assignment. In AIAA Guidance, Navigation, and Control Conference, edited by 1-12, 1161–1166. Portland,USA.

KLM.2016.OurFleet<https://afklcargo.com/NL/en/common/about_us/fleet.jsp>

Lee, H. 2014. Airport Surface Traffic Optimization and Simulation in the Presence of Uncertainties. Ph.D. thesis, Department of Aeronautics and Astronautics at Massachusetts Institute of Technology

Leung, L.C., van Hui, Y., Wang, Y., Chen, G., 2009. A 0–1 LP model for the integration and consolidation of air cargo shipments. Operational Research. 57 (2), 402–412

Li, Y., Tao, Y., & Wang, F. 2009. A Compromised large-scale neighbourhood search heuristic for capacitated air cargo loading planning. European Journal of Operational Research, 199(2), 533-560

Mujica Mota M., Boosten G., De Bock N.,Jimenez E.,Pinho de Souza J., 2017. Simulation-Based Turnaround Evaluation for Lelystad Airport, Journal of Air Transport Management

Petersen, J. 2007. Air Freight Industry—White Paper. Georgia Institute of Technology, Supply Chain and Logistics Institute, H. Milton Stewart School of Industrial and Systems Engineering. April 1, 2007. www.scl.gatech.edu/industry/industry-studies/AirFreight.pdf. Accessed December 12, 2016

SIMIO 2017. <www.simio.com>

Schiphol Group. 2016. Transport and Traffic Statistics https://www.schiphol.nl/en/schiphol-group/page/transport-and-traffic-statistics/

Wang, Y.J., & Kao, C.S. 2008. An Application of a fuzzy knowledge system for air cargo overbooking under uncertain capacity. Computer & Mathematics with Applications, 56(10), 2666-2675

MATHEMATICAL MODELS TO SIMULATE THE PROCESS INDUCED STRESS IN SILICON NITRIDE THIN FILM COATINGS AND ITS POSITIVE EFFECTS ON SOLAR CELLS

Abderrazzak El Boukili

Al Akhawayn University in Ifrane, Morocco

Email: a.elboukili@aui.ma

ABSTRACT

In this paper, we are developing advanced mathematical models to simulate the material mismatch induced stress in silicon solar cells. This stress is caused by the intrinsic stress created in the top coating layers during their growth. This investigation can't be done using experiments. Because, we can't separate the stress due to thermal mismatch and the one due to material mismatch in experiments. Our investigation will help designers and manufacturers of solar cells choose the right coating materials and optimize the growth parameters as the thicknesses of coating materials, thicknesses of emitter and base layers, growth temperatures, growth pressures, growth times and more. Another important and original work, in this paper, is the investigation of the positive effects of this calculated stress on the improvement of the optical efficiency of silicon solar cells. 2D numerical results will be presented, analyzed, and validated for a sample 2D solar cell.

Keywords: Mathematical models, simulation, material mismatch induced stress, solar cells.

1. INTRODUCTION

The application of process induced stress to enhance the performance of solar cells is really a new and promising area of research. Besides our previous paper (El Boukili, 2016), we found only one paper, in literature, talking about this subject (Santhosh et al., 2014).

On top of that, the values of stress used in (Santhosh et al., 2014), have not been calculated using a mathematical model as the one we are proposing here or the one we have proposed in (El Boukili, 2016). Then, they do not represent the real values of stress that could exist in a real silicon solar cell. In this paper, we are attempting to use an accurate mathematical model to calculate the real values of stress that could exist in a real silicon solar cell. And, in this paper, we are focusing only on the modeling of the stress that due to the material mismatch between the coating layers and the N doped emitter layer of the solar cell.

In our previous paper (El Boukili, 2016), we have investigated the effects of the stress, due to the thermal mismatch and not material mismatch, on the performance of a silicon solar cell. And, in (El Boukili, 2016), we have considered the effects a *compressive* stress. In this paper, we are considering the effects of a *tensile* stress on the efficiency of a silicon solar cell.

The reason for which we are investigating separately the material mismatch stress and the thermal mismatch stress is to quantify the amount of stress we get from each source. Like that, designers and manufacturers will have a clear idea on how to choose the right growth conditions for the anti-reflecting coating layers. And, they will have a clear idea on how to choose the right materials for the ant-reflecting coating layers. This quantitative analysis that separates between material mismatch and thermal mismatch stresses can't be carried out using experiments. Because, the values of stress we get from experiments (when we can do an experiment) include both

stresses. On top of that, it is not always possible to measure stress in real solar cells, specially, in nano solar cells. Since the light (X-rays) used in the experiments will disturb the atomic structure of the nano solar cell.
The investigation we are carrying out to model and calculate separately these two types of stress is new and original.

The authors in (Santhosh et al., 2014), have used values of stress in the range of -2 GPa and +2 Gpa. And, they have used a commercial TCAD software for device simulation to investigate the effects of these amount of stress on the main characteristics of a 2D silicon solar cell. The results they got are extremely encouraging and promising for the future of solar cells. They found that the application of a tensile stress (σ) enhances significantly the power output of a sample silicon solar cell.
The idea to use stress to enhance the performance of nano transistors is not new. It started in 2010 (Moroz, 2004, 2005, 2012, 2013; El Boukili, 2010, 2015; Chui et al., 2007; Jian Li et al., 2014; Ghani et al., 2003, 2010). And, today most manufacturers of nano CMOS transistors use stress to enhance dramatically their electrical performances.

However, using stress to enhance the electrical and optical performances of the solar cells is not the fashion yet in the scientific community as far as we know.
But, solar cells are the one that suffer from low performances. Only 20% or less of the sun energy is converted into electricity by most commercial solar cells. Then, 80% of the sun energy is still lost out there. So, it is time to search for all the possible ideas and methods to reduce this loss from 80% to 20% or less. The day in which we will achieve this, will be similar to the day when we discovered the wheel. Anyway, we believe that this day will come.

In this paper, we are going to calculate the real values of stress in the same sample silicon solar cell used in (Santhosh et al., 2014).
And, we are going to use these real values of stress, the Figure 15 in (Santhosh et al., 2014), and Lagrange polynomials to interpolate accurately the obtained values of the power output in the studied solar cell.

In the most commercial silicon solar cells used today, silicon nitride (SiNx) is the material of choice for the top anti-reflecting layer. In this paper, we are using SiNx as the top anti-reflecting layer for the silicon solar cell sample we are studying. Here, x represents the Nitrogen content.
We are going to investigate and analyze the effects of the:

i) thickness of the top SiNx film,
ii) Nitrogen content (x) in SiNx,
iii) lattice parameter of SiNx,

on the intrinsic stress, σ_i, in the SiNx films, the extrinsic stress, σ, in the base and emitter regions of the solar cell, and on the power output, and the short circuit current.

This paper is organized as follows. Section 2 presents the theoretical models we have developed to calculate accurately the intrinsic stress σ_i due to the material mismatch between the top coating SiNx film and the silicon material of the solar cell. Section 3 presents the theoretical models we have developed to calculate accurately the extrinsic stress σ in the whole structure of the sample silicon solar cell. Section 4 presents the 2D numerical results and their analysis for both the intrinsic stress σ_i and the extrinsic stress σ. This section will also present a quantitative analysis of the impact of the extrinsic stress on both short circuit current and power density of the silicon solar cell studied. Section 5 presents the concluding thoughts and future work.

2. MODELING OF INTRINSIC STRESS IN SILICON NITIRIDE ANTI-REFELECTIVE COATING

Silicon nitride (SiNx) thin films have enjoyed considerable industrial applications. They are used as insulators, dielectric material, barrier coatings, surface and bulk passivation, thin film transistors, support structures for Fabry-Perot filters, and as antireflective coatings for silicon solar cells. The wide industrial applications of SiNx are attributed to its molecular composition, structural, mechanical, electrical, thermal, and optical high quality properties. All of these SiNx properties can be tuned and controlled by varying the growth conditions as:

- Growth temperatures.
- Growth pressures
- Nitrogen content
- Growth rates
- Growth techniques and more.

Plasma enhanced chemical vapor deposition (PECVD) is the industrial workhorse used to fabricate SiNx thin films. Other alternative techniques known as hot-wire chemical vapor deposition (HWCVD) offers superior growth rates and lower cost.

Generally, the Nitrogen content x of SiNx thin films is the most important factor governing the optoelectronic properties of SiNx. The refractive index of SiNx depends strongly on x which depends on the growth technique. To use SiNx as an anti-reflective coating for the top surface of a solar cell, its refractive index must be tuned properly by optimizing the growth conditions.
There exist three crystallographic phases of silicon nitride, designated as α-SiNx, β-SiNx and c-SiNx. The α-SiNx and β-SiNx phases are the most common forms of silicon nitride. They are produced under ambient pressure and temperatures below 2000K. The c-SiNx phase has been discovered in 1999 (Zerr et al., 1999). It has been produced at temperatures higher than 1927°C and at pressures higher than 15 GPa. The structure of c-SiNx has been determined as a cubic spinel structure at ambient temperatures. It possesses a hardiness of 35 GPa which makes c-SiNx an excellent top coating layer for solar cells. The SiNx thin film we are using as the anti-reflective coating of our sample silicon solar cell corresponds to the cubic phase: c-SiNx.
The cubic structure of c-SiNx will make it easy for us to calculate its lattice constant. In fact, the lattice constant of cubic materials is the same in all the directions of the crystal.

Let afx be the lattice constant of SiNx. And, let as be the lattice constant of the silicon substrate of the solar cell.

SiNx thin film is grown at elevated temperatures on the top of the silicon (Si) substrate to fabricate the anti-reflective coating for the solar cell. When the SiNx thin film cools down an intrinsic stress σ_i will develop in the SiNx thin film. And, when the growth process is over, the intrinsic stress σ_i will create an extrinsic stress σ in the whole solar cell according to the first Newton's law of action and reaction. The amount of intrinsic stress σ_i generated in the SiNx thin film will depend on the growth conditions. We can tune up and optimize the growth conditions to generate the desired amount of stress that will improve the efficiency of the solar cell.

The intrinsic stress σ_i is due to both:

i) the material mismatch between SiNx and Si substrate, and
ii) the thermal mismatch between SiNx and Si substrate.

In this paper, we are focusing only on the modeling and simulation of the intrinsic and extrinsic stresses due to material mismatch.

We assume that silicon and SiNx are elastic materials. Then, we use Hooke's law to calculate the material mismatch induced intrinsic stress σ_i from the material mismatch induced intrinsic strain ε_i by:

$$\sigma_i = D\varepsilon_i \, . \tag{1}$$

D represents the stiffness matrix calculated from the elastic constants c11, c12, and c44 as in (EL Boukili, 2015). The stress σ_i and the strain ε_i are tensors given by:

$$\sigma_i = (\sigma_i^{xx}, \sigma_i^{yy} \sigma_i^{xy}).$$

$$\varepsilon_i = (\varepsilon_i^{xx}, \varepsilon_i^{yy}, \varepsilon_i^{xy}).$$

The terms with xx, and yy indices represent the normal intrinsic stress or strain components. The terms with xy represent the shear intrinsic stress or strain components which are assumed to be zero in this paper.

To calculate the normal components of ε_i, we propose the following accurate model:

$$\begin{aligned} \varepsilon_i^{xx} &= a / b \\ \varepsilon_i^{yy} &= -\varepsilon_i^{xx} \end{aligned} \quad . \qquad (2)$$

Where a = afx – as, b = afx . To calculate accurately afx, we propose the following accurate and nonlinear model which takes into account the Nitrogen content x in the SiNx thin film.

$$afx = 0.02733x^2 + 0.1992x + 5.431. \qquad (3)$$

The unit of afx and as is Angstrom.

To calculate accurately the components of the material mismatch induced intrinsic stress σ_i in the *SiNx* thin film, we are proposing the following accurate model based on (1)-(3):

$$\begin{aligned} \sigma_i^{xx} &= \alpha\varepsilon_i^{xx} + \beta\varepsilon_i^{yy} \\ \sigma_i^{yy} &= \beta\varepsilon_i^{xx} + \alpha\varepsilon_i^{yy} \\ \sigma_i^{xy} &= \sigma_i^{yx} = 0 \end{aligned} \quad . \qquad (4)$$

Here, α and β are the elastic constants c11 and c12 of *SiNx* thin film at room temperatures. The models we are proposing to calculate accurately the resulting extrinsic stress σ in the whole solar cell structure are presented in Section 3.

3. MATHEMATICAL MODELS OF EXTRINSIC STRESS IN THE SOLAR CELL

To calculate the extrinsic stress σ from the extrinsic strain tensor ε, intrinsic stress σ_i, intrinsic strain ε_i, and the stiffness matrix D, we use again Hooke's law to get the constitutive equation:

$$\sigma = D\varepsilon + \sigma_i \qquad (5)$$

And, we use the relation between the extrinsic strain $\varepsilon = (\varepsilon^{xx}, \varepsilon^{yy}, \varepsilon^{xy})$ and the displacements $u(x, y), v(x, y)$ given by:

$$\begin{aligned} &\varepsilon^{xx} = \frac{\partial u}{\partial x}; \varepsilon^{yy} = \frac{\partial v}{\partial y}; \\ &\varepsilon^{xy} = \varepsilon^{yx} = \frac{1}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) \end{aligned} \qquad (6)$$

Then, we solve for the displacements $u(x, y), v(x, y)$, the following system of second order partial differential equations:

$$\begin{aligned} &\alpha \frac{\partial^2 u}{\partial x^2} + \beta(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 v}{\partial x \partial y}) + \gamma \frac{\partial^2 v}{\partial x \partial y} = u_0 \\ &\alpha \frac{\partial^2 v}{\partial y^2} + \beta(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 u}{\partial x \partial y}) + \gamma \frac{\partial^2 u}{\partial x \partial y} = v_0 \end{aligned} \qquad (7)$$

Here, α, β, γ represent the elastic constants c11, c12, and c44 of silicon and SiNx at room temperature. The term u_0, v_0 depend on the first partial derivatives of the intrinsic stress σ_i. The model (7) we are using to calculate the extrinsic stress σ is accurate and physically based. Since the model (4) we are using to calculate σ_i is physically based and accurate. For example, the model (4) is based on Hook's law and it does take into account, accurately, the effects of the Nitrogen content on the lattice constant of SiNx.

The system (7) is approximated accurately using Finite Element Method within the framework of the standard process simulator Suprem4 (Hansen et al., 1993).

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

4. NUMERICAL RESULTS AND DISCUSSION

The proposed models (2), (3), (5), and (7) have been implemented within the framework of Suprem4 to enhance the old stress models used in Suprem4. Suprem4 has been used to build and simulate the stress in the studied solar cell. The total width of the solar cell is 1 um. The length of the solar cell is 0.4 um. The thickness of the grown SiNx is in the range 80 nm to 200 nm. The Nitrogen content x is in the range 20% to 59%. This structure is similar to the one used in (Santhosh et al., 2014). Figure 2 shows the extrinsic stress distribution in the whole solar cell when the Nitrogen content x is 59% and the thickness of SiNx is 80 nm. This figure shows that the calculated values of the extrinsic stress σ^{xx} in x direction are in the range -3.5 GPa and +1.8 GPa in the Solar Cell. The enhancement of the solar cell efficiency corresponding to this amount of stress is shown in Figure 3. The Figure 4 shows the extrinsic stress distribution in the whole solar cell when the Nitrogen content x is 20% and the thickness of SiNx is 80 nm. Figure 4 shows that the calculated extrinsic stress σ^{xx} is in the range -0.137 GPa and 0.4 GPa. This shows that the stress in the solar cell increases with increasing Nitrogen content x. Figure 5 shows the short circuit current enhancement as function of extrinsic stress σ^{xx}.

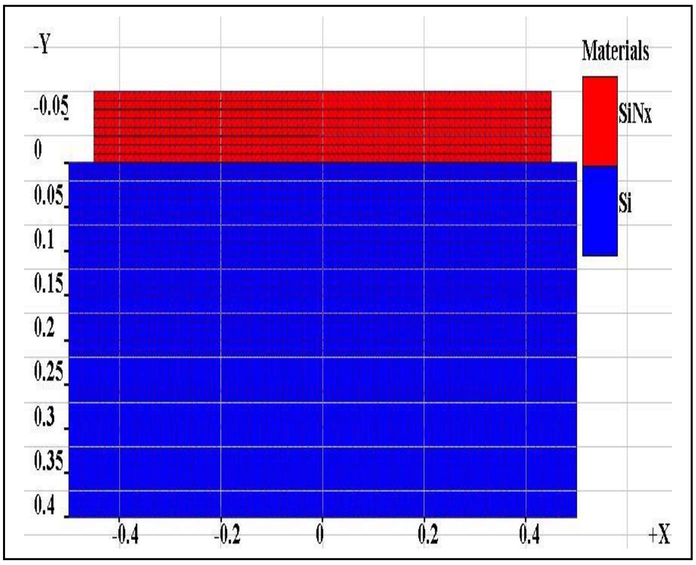


Figure 1: Structure and Mesh of the Solar Cell.

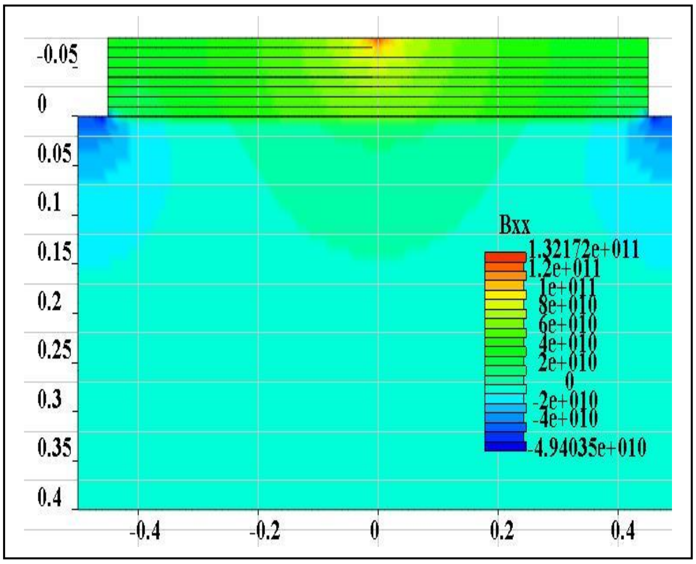


Figure 2: Extrinsic stress in x direction when x=59%.

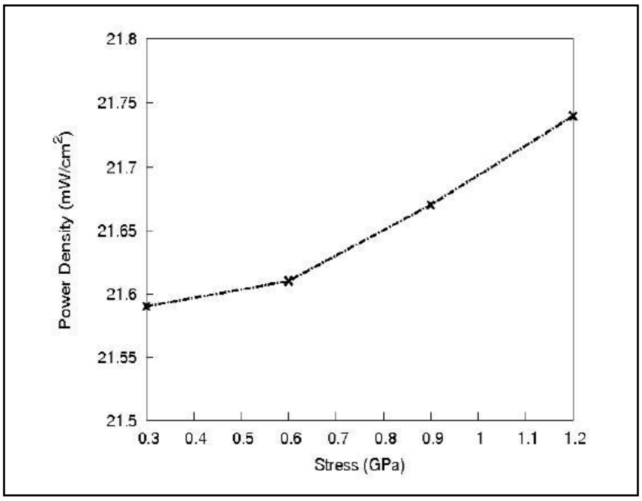


Figure 3: Stress induced enhancement of power density.

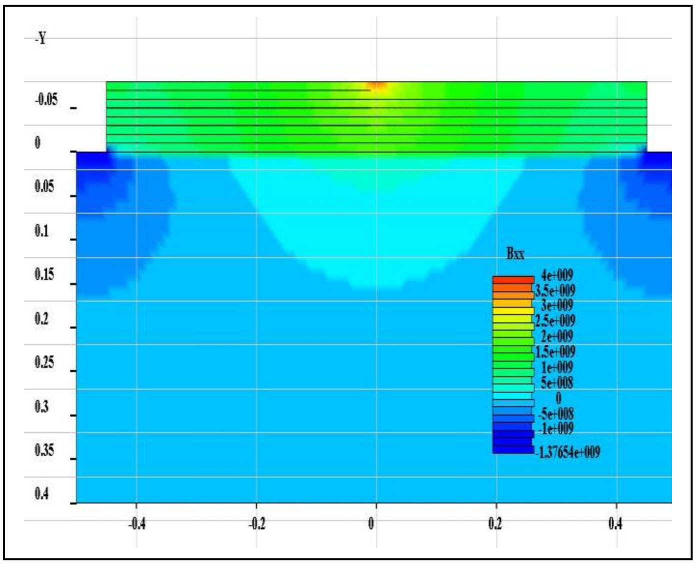


Figure 4: Extrinsic stress in x direction when x=20%.

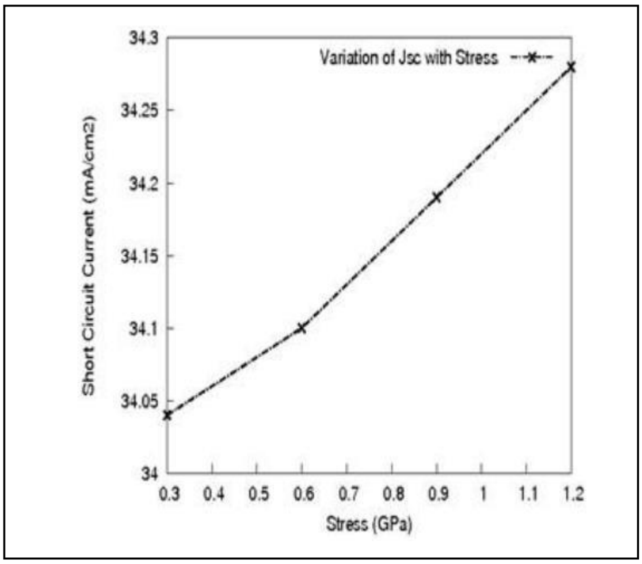


Figure 5: Stress induced enhancement of short circuit current.

5. CONCLUSIONS

The application of the right mechanical stress will enhance dramatically the optical and electrical performances of the modern solar cells. The models we have proposed in this paper have been used to calculate accurately the amount of the extrinsic stress we get from the material mismatch between the SiNx thin films used as anti-reflecting coatings and the silicon substrate of the solar cell. In the future work, we will investigate the amount of stress in silicon solar cells that is due to both thermal and material mismatch.

REFERENCES

Chui, K.J., 2007. n-MOSFET with silicon-carbon source/drain for enhancement of carrier transport. IEEE Trans. Electron Devices, Vol. 54, 249-256.

El Boukili, A., 2010. 3D Stress simulations of nano transistors. In: Gunther, M., Bartel, A., Brunk, M., Schops, S., Striebel, M., eds. Progress in Industrial Mathematics at ECMI. Berlin Heildelberg: Springer, 85-91.

El Boukili, A., 2015. New models for thermal stress simulation in 3D nano CMOS transistors. Proceedings of the first international conference on electrical and information technologies. pp. 25-27, March, 25-27, Marrakech (Morocco).

El Boukili, A., 2016. Quantitative analysis of the stress effects on the performance enhancement of silicon solar cells. Proceeding of 4th International Renewable and Sustainable Energy Conference, Nov. 14-17, Marrakech, Morocco.

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. Dec. 8-10, Washington DC (USA).

Ghani, T., 2010. Challenges and innovations in nano-CMOS transistor scaling. IEEE workshop on microelectronics and electron device. April 16, Boise (Idaho USA).

Hansen, S., Deal, M., (1993). User manual for Suprem 4. Stanford University, USA.

JianLi, M., Fu, Z., Liu, P., Zhang, H. (2014). Hole mobility enhancement in uniaxial stressed Ge dependence on stress and transport direction. Science China, Phys., Mech. & Astr., Vol. 57, 1860-1865.

Moroz, V., 2004. Analyzing strained-silicon options for stress-engineering transistors. Solid State Technology Magazine. July, Vol. 47, issue 7, p49.

Moroz, V., Washington, L, Farah, N., Sunderraj, T., Eneman, G., Smith, L., Screntelkamp, R., (2006). pMOS with 200% mobility enhancement induced by multiple stressors. IEEE Electron Dveice Letter, Vol. 27, No. 6, 511-513

Moroz, V., et al., 2013. A group IV solution for 7nm FinFET CMOS: Stress engineering using Si,Ge, and Sn. Proceeding of IEDM, pp. 26.3.1-26.3.4 Dec. 9-11, Washington DC (USA).

Santhosh, K., K., Srinivasan, R. (2014). Effects of stress on the performance of silicon solar cells, IJIRSET, Vol. 3, Issue 3, 725-730.

Zerr., A., Miehe, G., Serghou, G., (1999). Synthesis of cubic silicon nitride. Nature 400, 340-342.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

SOCIOTECHNICAL SYSTEMS ACCEPTANCE AND SUSTAINABILITY ASSESSMENT USING DYNAMIC SIMULATION MODEL

Dace Aizstrauta[(a)], Egils Ginters [(a)]

[(a)] Sociotechnical Systems Engineering Group, Department of Modelling and Simulation, Faculty of Computer Science and Information Technology, Riga Technical University, 2 Daugavgrivas Street, Riga LV-1048, Latvia

[(a)]egils.ginters@rtu.lv

ABSTRACT
Since the creation of sociotechnical systems Integrated Acceptance and Sustainability Assessment Methodology (IASAM) in 2013 it has experienced several levels of upgrade. Each next step of improvements has mainly added new dimension and increased the potential of applications of the methodology. In addition, a web-tool was created for the first version of IASAM and this was also improved along the changes of methodology.

Keywords: IASAM3, acceptance and sustainability assessment, sociotechnical systems, system dynamics

1. INTRODUCTION

The prevalence of technologies in our everyday lives keeps increasing and so does the number of devices and gadgets around us. Technologies are being embedded in most processes underlying our work, education, travel, entertainment, etc. However social pressure of society resulting in technologies changes and initiates the new and combined solutions related with RFID, VR/AR, Future Internet, OCR applications, etc. This has not only led to development of complex sociotechnical systems, but also to the evolution of acceptance and sustainability research. The complexity, in turn, makes the development, requirements definition, testing and the analysis of eventual adoption and diffusion of sociotechnical systems very difficult.

The complexity makes also the sociotechnical systems research field highly multi-disciplinary. According to Morris (2009) there are at least four differing viewpoints in the literature about such systems and they represent the interdisciplinary research conducted in social sciences, organizational sciences, engineering, and complex systems.

Complexity can be explained by the logical and physical structures of sociotechnical systems, where first one involves ideas, concepts, guidelines, rules, algorithms, but second - environment (technical ones and persons) for implementation and running of logical structure (Ginters et al 2017). Typical features of sociotechnical systems are their emergent properties (some of the properties of the system emerge after it has gone into use and cannot be predicted in advance), their non-deterministic nature and the fact that they are influenced by the organisations culture, rules and objectives (Somerville 2014). The term sociotechnical systems is widely used to describe many complex systems, but there are five key characteristics of open sociotechnical systems (Badham et al 2000):

- Systems should have interdependent parts;
- Systems should adapt to and pursue goals in external environments;
- Systems have an internal environment comprising separate but interdependent technical and social subsystems;
- Systems have equifinality. In other words, systems goals can be achieved by more than one means. This implies that there are design choices to be made during system development;
- System performance relies on the joint tuning of the technical and social subsystems. Focusing on one of these systems to the exclusion of the other is likely to lead to degraded system performance and utility.

Sociotechnical systems involve complex interaction between people, machines, and the environmental aspects of the work system (Baxter and Sommerville 2011). The involvement and participation of people significantly affect the architecture and design of those systems (Reymondet 2016). These systems usually contain technology subsystems and components central to its performance and have societal, political, and economic relevance and impact (Susmann 2013).

Therefore, it is necessary to understand and assess emerging sociotechnical systems. Especially nowadays, when technologies develop in a rapid manner and technological changes can often disrupt a market or industry's established rules, living conditions, orders, beliefs, and values. Organizations will succeed only if they are able to embrace the change (Jonathan and Chung-Shing 2015). But organizations and their management often does not have the capacity to cover the main insights of several disciplines and make an informed decision with limited time and knowledge resources and without comprehensive methodology. The decision makers need tools to evaluate sociotechnical systems and become aware of the new reality.

Furthermore technological change and innovation is itself at the center of many research papers (Jonathan and Chung-Shing 2015, etc). One of the issues that the developers of technologies, researchers and organizations are most interested in is the potential adoption (and acceptance) of the technology and its diffusion within society and the consequent success in the market. Technology acceptances are information service theories that model how users come to accept and use a specific technology. These theories suggest that when users are presented with a new technology, a number of factors influence their decision about how and when they will use it (Samson and Jongsu 2015).

Many authors have studied different aspects of new technology acceptance from a variety of theoretical perspectives explaining the relationship between user beliefs, attitudes, and intentions and analyzing different factors that influence information technology acceptance – individual, organizational aspects, cultural, gender and professional differences. These studies focus on behavioral aspects of technology acceptance or adoption, for example Theory of Reasoned action (Fishbein and Ajzen 2010). One of prominent models to be mentioned is Technology Acceptance Model (TAM) (Davis, Bagozzi, and Warshaw 1989). It has been criticized for focusing on initial adoption and not on continuous use. There are also other approaches, for example Expectation-Confirmation Theory (ECT) (Bhattacherjee 2001) that initially originated in marketing sphere and Unified Theory of Acceptance and Use of Technology (UTAUT) that tries to consolidate eight approaches into one (Venkatesh et al 2003). These theories question the factors behind the intentions and behaviors of users from psychological perspective. Different variations of TAM, UTAUT model, ECT and some others are mentioned in the discussion of technology acceptance and adoption research.

But none of them gives full and combined understanding about the human and technological factors that are influencing acceptance and sustainability of sociotechnical systems.

The aim of the research underlying this paper is to create a comprehensive tool to help interested stakeholders to embrace the change brought by technology innovations and to analyze the acceptance and sustainability of emerging technologies.

The authors propose the concept of sustainability for evaluation of the set of factors that let the technology to be developed, implemented, maintained properly (i.e. according to the needs of all stakeholders) and attract long-term users and create positive output and/or outcome according to the purpose of the technology and initial intentions of its developers (financial, social, etc) (Aizstrauta, Ginters and Piera Eroles 2014). Technology sustainability thus combines the different ways for looking at technology development, acceptance and diffusion, and fills the gaps in the literature. Because the understanding of technology development and exploitation is increased by combining individual factors with both internal (connected with ICTE development management and quality of technology) and external (connected with domain development) sociotechnical factors.

For measurement of potential acceptance and sustainability an IASAM was created in 2013. It is a methodology that helps to evaluate technologies from four perspectives, called flows – management, quality of technology, acceptance and domain development. This evaluation approach is based on a viewpoint that technology acceptance research should not be divided apart from the technological, economic and social evaluation. In other words it introduces a new approach for evaluation of new technologies by combining socio-economic aspects and technical characteristics of technology development and exploitation (Aizstrauta and Ginters 2015).

The following sections of this article are organized as follows. The next section describes the development of IASAM3 and the underlying reasons from several upgrades making the IASAM more experienced. It also describes the fundamentals of using IASAM3 methodology. The third section then explains the role of system dynamics simulation to build an interactive IASAM3 tool. Finally, the conclusion contains a summary of the main ideas of the paper.

2. DEVELOPMENT OF IASAM3 AND THE APPLICATIONS OF METHODOLOGY

In terms of content and themes to be covered the research on technology sustainability and acceptance evaluation was driven by the need to understand, how technologies, people, organizations, policies, and societal values are intertwined. It was clear that to satisfy this need a tool had to be created that would include technical and human factors assessment. But sociotechnical systems are governed by organizational policies and rules and may be affected by external constraints such as national laws and regulatory policies (Sommerville 2014). Also Reymondet (2016) points out that in sociotechnical systems research, the dichotomy between the technical on one side, and social and human on the other side is salient. Therefore the research had to look further and widen the scope of factors under assessment to create a methodology that identifies measures and arranges the criteria that impact technology's sustainability and cover a wide range of issues. All these dimensions are built the IASAM3 index.

In terms of structure the tool had to meet certain prerequisites. Namely, the integrated acceptance and sustainability assessment model had to:

- Be easy accessible to use and available to different stakeholders with different experience and knowledge, from inventors and programmers to potential investors;

- Be usable for initial evaluation as well as within later stages of technology development;
- Offer an opportunity to analyze and assess the anticipated dynamics of the evaluation results. Thus, the assessment model can be used also as guidelines for technology development.

2.1. IASAM Development from the First Version to IASAM3 - Reasons and Benefits

Since the creation of IASAM methodology in 2013 it has experienced several levels of upgrade. Each next step of improvements has added new dimension and increased the potential of applications of the methodology. In addition, a web-tool was created for the first version of IASAM and this was also improved along the changes of methodology.

At the first level, the methodology was just a set of 49 questions that were grouped according to the model's criteria. The initial IASAM version was validated against Skype as it met the indications of a successfully implemented technology (Aizstrauta, Celmina and Ginters et al 2013). Then IASAM was used for socio-technical evaluation of actual products within FP7-ICT-2009-5 CHOReOS project No. 257178 (2010-2013) *"Large Scale Choreographies for the Future Internet (IP)"*. Within this project eight products were evaluated. The results gave some understanding of the technology and its development process, but the result lacked dynamics and its interpretation was rather narrow.

But the most important flaw of this version was that the model included a survey of potential users that was based on UTAUT model and that was a burden for the users of the methodology. The methodology is planned to be applicable at any stage of technology development and potential user surveys might be very challenging during early stages of technology development where no prototype is available, in cases were potential users are hard to reach or in situations when there is no time. Such a survey would need either excessive time, human or financial resources.

Therefore IASAM2 replaced the potential user survey with additional criteria that are based on attributes of innovations based on Rogers (2003) Theory of innovations. This resulted in a model that consisted of 61 questions that remain grouped under 17 variables that together constitute four main groups, called flows, that make up the result. Additional validation measures were taken, again using Skype.

This version was applied within another international project "Future Policy Modelling" (Ginters et al 2013), where four products were assessed using IASAM2 methodology.

The IASAM2 version had several significant benefits, including that the assessment could be carried out by the interested party oneself, there was no need for resource-consuming potential user survey. Besides, the model itself became more comprehensible, as the calculus, analysis and reporting could be done within one methodological framework. This version of IASAM2 met the initial goals of this methodology to make it better – it became easier to use and more universal in its applications.

Later on the methodology was improved once more and the functionality and respective advantages of system dynamics simulation were incorporated within the model. With the help of mathematically reprocessed life-cycle data of other technologies the static IASAM2 was reshaped into a dynamic analytical tool that helps not only to evaluate the current condition of the technology, but also to make judgements on potential life cycle parameters at any step of development the technology under assessment. It was done using Skype as a "role model". That means that the assessed technology IASAM3 reference trendline is being compared with Skype's life-cycle function and therefore, the measurements done with IASAM3 methodology became comparable with reference trendline and the sustainability and acceptance of Skype. The technology assessment results can now be viewed and interpreted in contrast with the results of Skype. Such approach makes the results of IASAM3 more comprehensible, enable less complicated interpretations, and be more user-friendly (Aizstrauta and Ginters 2017).

2.2. IASAM3 Methodology Fundamentals

From the very beginning the use of IASAM3 methodology was intended to be relatively easy. The basis for the assessment is a self-assessment questionnaire. The developer or any other stakeholder or interested party can do the assessment, using available information and his or her subjective attitude. To reduce the subjectivity of the assessment, it is advisable to use information from internal procedures, employee evaluation, client surveys, etc.

And the sources of information that is needed for the evaluation are omnifarious – the information can be:

- Described or embedded in the documents of the organization;
- Embedded in processes connected with the development or running of the technology;
- Publicly available sources about the company or the overall market situation;
- Ascertained with the help of people involved in the technology development.

IASAM3 index is made of four basic flows – management, quality of technology, technology acceptance and domain development. Each of them consists of certain criteria – all together eighteen criteria. Each criterion is evaluated with the help of specially formulated criteria descriptions/statements. That means the user evaluates 61 pre-defined criteria descriptions using a simple 7 point Likert scale. The evaluation of criteria is undeniably subjective, but it relies on assumption that every evaluator, whether a technology developer or potential investor, will be concerned to receive the most reliable evaluation for decision-making. After the criteria have been evaluated the sum of all values from the questionnaire is calculated and divided by maximum possible value of questions answered. The result gives a numerical value of integrated technology sustainability index

$$IASAM3_{index} = \frac{\sum_{n=1}^{12} F_n + \sum_{i=1}^{49} B_i}{(N-C)*7} \quad (1)$$

, where F – additional IASAM2 survey response values; B– initial IASAM survey response values, N – total number of questions; C – number of questions marked with “N/A”.

In addition, the procedure calculates the consistency of the result. This value E is called IASAM3 credibility and it looks at the number of questions left when C values (those marked with N/A) are excluded and decreases the „internal credibility” of the index calculated in the previous step. The more questions marked with N/A the less consistent is the result. The E is calculated as:

$$E = C / N \quad (2)$$

, where E – credibility, N – total number of questions, C – number of questions marked with N/A.

Thus the measurement gives two results – the sustainability index of the assessed technology and the credibility for the calculated index. Besides it is worth mentioning, that model allows making evaluation multiple times during the life-cycle, gathering these results and comparing the evolution or development of the technology from IASAM3 perspective.

The final value of the IASAM3 index indicates the percentage of the potential maximum that could be reached in comparison with the Skype. Therefore, the next step is to construct a trendline of evaluated technology and use it to compare the potential dynamics of IASAM3 index in comparison to Skype reference trendline. This trendline is constructed using a polynomial function (3):

$$IASAM3_{reference-trendline} = 2.5327x^3 - \\ -7.1021x^3 + 4.6355x - 0.0485 \quad (3)$$

The polynomial function is a result of mathematically reprocessed life-cycle data of other successive and wide used technologies to predict continuation of the Skype life-cycle in conformity with IASAM3 rules. Its aim is to reshape the static IASAM2 into a dynamic analytical tool that helps not only to evaluate the current condition of the technology, but also to make judgements on potential life cycle parameters of the technology under assessment. Finally by the aid of interactive visualizations applicant can make many experiments using different what-if scenarios.

3. SYSTEMS DYNAMIC USE FOR TECHNOLOGY ASSESSMENT USING IASAM3

System dynamics discipline is an attempt to address the changes of an ever increasing complexity of everyday life and long-term policy problems. Applications of system dynamics cover a very wide spectrum, including national economic problems, supply chains, project management, educational problems, energy systems, sustainable development, politics, psychology, medical sciences, health care, and many other areas (Barlas 2002; Nielsen and Nielsen 2015).

But dynamic complexity and the interdisciplinary policy problems are often embedded in or rely on sociotechnical systems, therefore system dynamics approach is suitable to address problems connected with sociotechnical systems research.

In addition, as Nielsen and Nielsen (2015) put it, the system dynamics does not offer one “grand theory” instead each model is a theory by itself. And therefore IASAM3 is a model that is a theory of acceptance and sustainability assessment of technologies.

The concepts of the system dynamic philosophy are primarily centered on a certain understanding of causality in a system’s setting. System dynamics is concerned with aggregate social phenomena and not individual actions (Nielsen and Nielsen 2015). The simulation approach provides the opportunity to analyse a time-varying system with multiple feedback links and analyze quantitative and qualitative factors (Ginters, Barkane and Vincent 2010)

The simulation and modeling tool Insight Maker has been designed to make modeling and simulation accessible to a wider audience of users and integrates three general modeling approaches – System Dynamics, Agent-Based Modeling, and imperative programming – in a unified modeling framework. The environment provides a GUI that is implemented purely in client-side code that runs on users’ machines (Fortmann-Roe 2014). Therefore Insight Maker was selected as simulation environment of IASAM3.

3.1. IASAM3 Tool and Case Studies

For the purposes of IASAM model application, a custom made web-based evaluation tool was created for the initial IASAM version. It’s approbation is described in the documentation of CHOReOS project “Large Scale Choreographies for the Future Internet” (2010-2013). Later the tool was improved following the model advancements and now offers the users not only to carry out evaluations, but also use Insight Maker as a system dynamics modeling platform. This tool enables the users to:

- Create evaluations in a user-friendly, intuitive environment;
- View the results of IASAM3 index of evaluated project or technology;
- Visually compare the results with Skype life-cycle;
- Change the values of indicators in an interactive environment to run different what-if scenarios.

Two case-studies were used to test this new approach and the web tool – one of improvements of technology that calculates the plays of poker game (project "Poker Calculator") and another of implants that simplify the communication with computer with the help of user's hands (project "Implants").

First, the documents, procedures and responsible persons that could provide information for evaluation purposes for each case study were identified and the information was gathered. Second, the self-assessment questionnaire was filled out and the IASAM3 index of the evaluated project was calculated according to the answers of the questionnaire. Then, according to the new perspective added by IASAM3 compared with IASAM2, Skype reference trendline is constructed (3). IASAM3 index indicates how the function of the technology under assessment looks pro rata. The function for the technology under assessment is constructed by scaling the function on both axis.

Further the web-tool lets to change the values of four main flows – acceptance, quality of technology, management, and domain development – and run InsightMaker based simulations to test different scenarios. The process is visually described in BPMN2 notation in Figure 1.

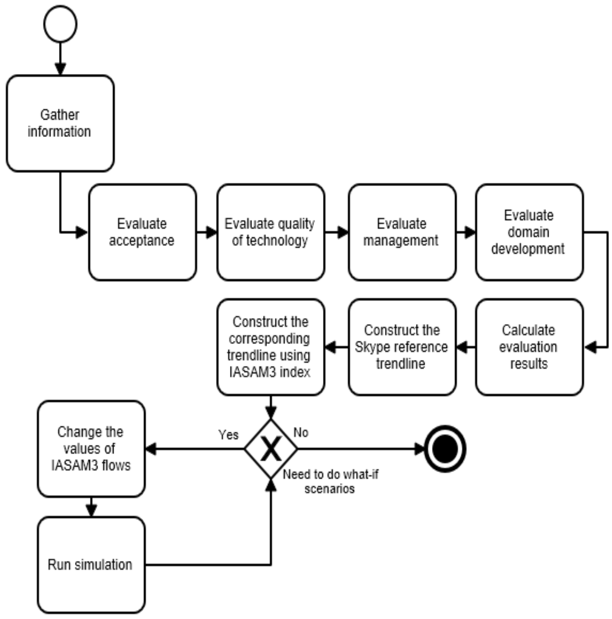


Figure 1: IASAM3 Implementation

Graphical results of simulation one case studies can be seen in the Figure 2 and Figure 3.

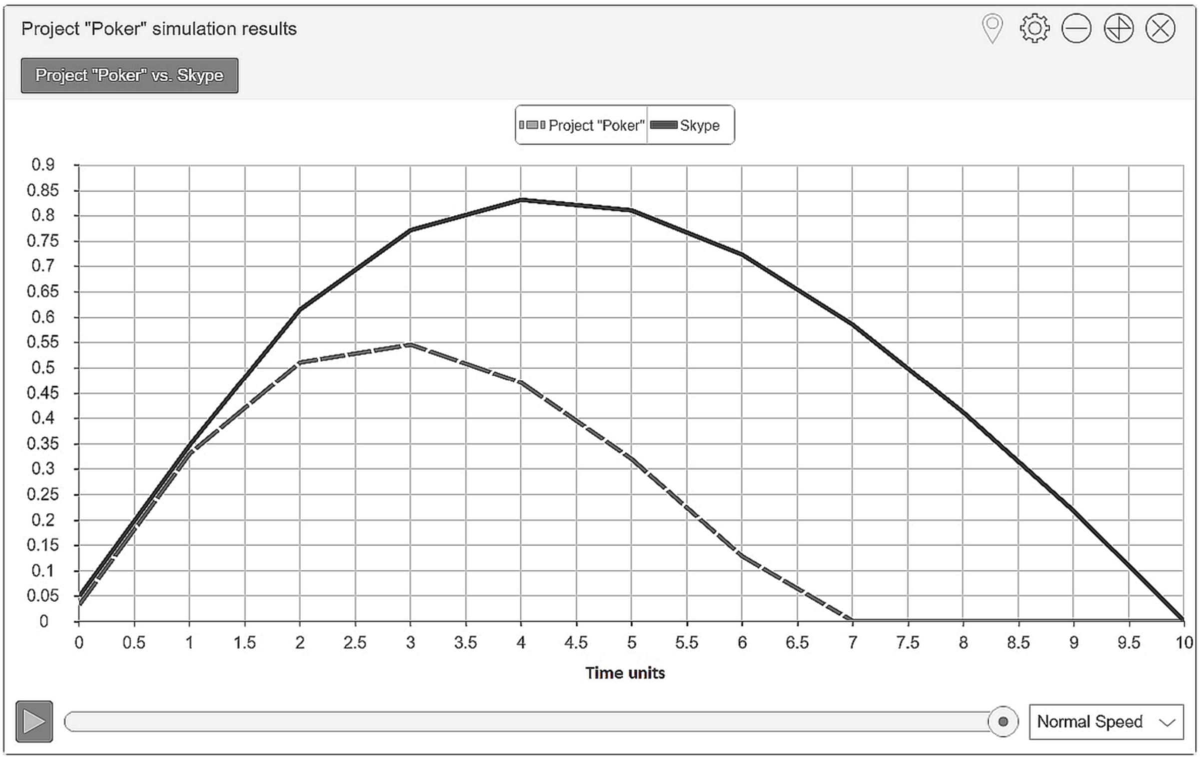


Figure 2: Project "Poker Calculator" Results Generated by Insight Maker Engine within the IASAM3 Web-tool

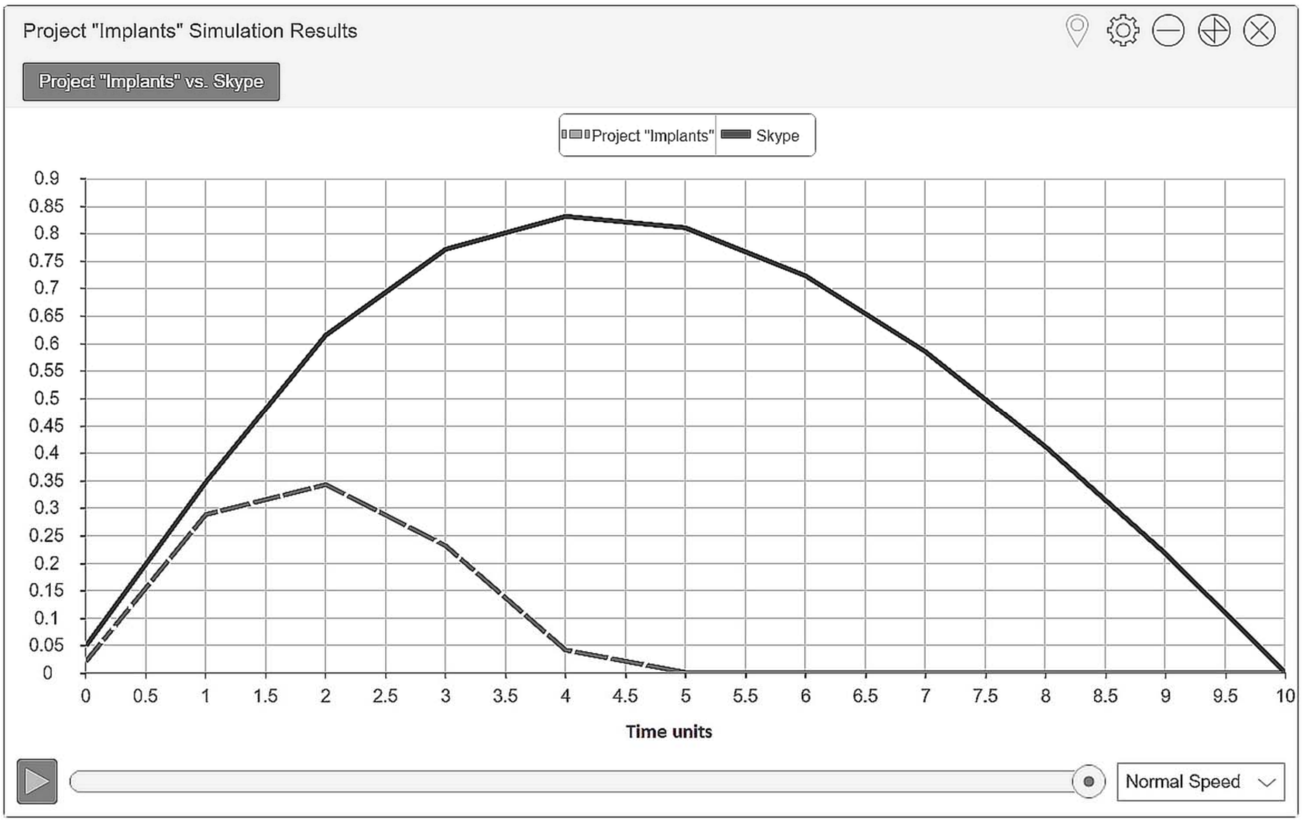


Figure 3: Project "Implants" Results Generated by Insight Maker Engine within the IASAM3 Web-tool

The result of the poker calculator shows that the technology complies with the development cycle of Skype with IASAM3 index value 0.67. That means that the technology will be able to reach 67 % of Skype's theoretical maximum. The implants on the other hand reach 47 %.

The tool enables the user to manually change the influencing values of IASAM3 index (1), by changing the value four flows – key impact factors – management, technology acceptance, quality of technology and domain development (see Figure 4). The main elements of the interface are:

- InsightMaker panel which shows the IASAM3 model's graphical visualization;
- Visualization the results of simulation within the same InsightMaker panel;
- The numerical result of the simulation, namely, IASAM3 index and Credibility values on the right side;
- Custom made sliders for adjusting the value of four flows for the simulation of what-if scenarios also on the right side;
- Web-tool management menu on the upper side of the panel.

By repeatedly running the simulation the predicted life-cycle trendline of the evaluated project is calculated according to the new values and shown in a graph, also including the Skype life cycle trendline as a reference.

4. CONCLUSIONS

The increasing amount of technologies that define more and more aspects of our everyday lives has led to a massive spread of different technologies, including complex sociotechnical systems. The human factor involved in every sociotechnical system makes these systems and their outcome both more valuable (these systems cannot yet be substituted by technical solutions) and more complex (they are harder to build, predict and exploit).

Sociotechnical systems are interdisciplinary in their nature and therefore need an interdisciplinary approach in their research. IASAM3 methodology proposed by the authors offers an interdisciplinary assessment methodology and a web-based tool to evaluate the potential acceptance and sustainability of technology. The methodology combines the technical criteria of technology and its quality itself, the human factors of technology adoption, the socio-economic aspects. During the development of IASAM3 it has been through three development stages and has resulted in a tool that can help the interested stakeholders to get a well-founded evaluation that shows also the prospective success of the technology.

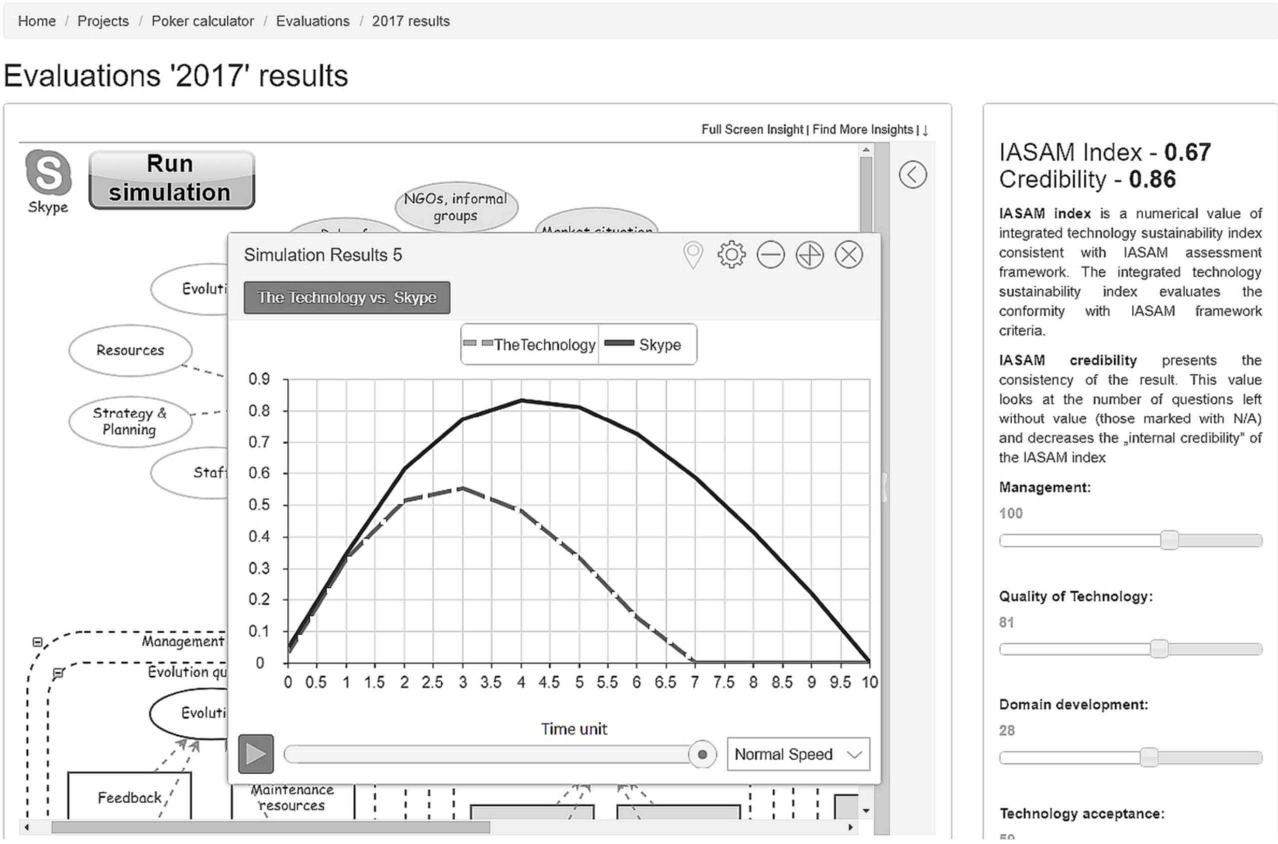


Figure 4: Screenshot of the IASAM3 Web-tool's GUI

IASAM3 adds to the evaluation results perspective of time that is displayed in the context of Skype life-cycle path. By showing visualizations of two functions, this tool makes the result more intuitive and easier to comprehend. System dynamics and InsightMaker simulation platform enable the user to carry out deeper analysis of the results.

Additional benefit of IASAM3 is that it provides users with a ready-to-use and user friendly tool that does not require additional specific knowledge in the field of modeling, programming, or statistics.

Therefore, it can be used as widely available open source tool across specialists with different backgrounds.

To ensure the availability of this tool to as many interested parties as possible, it is easy accessible and is provided as a ready-to-use software. Even more, it could become used as the "software on-demand". The Software as a the Service (SaaS) lets the companies to avoid traditional software installation, maintenance and management steps in favor of delivering cloud-based applications via the Internet. However the aspects of IASAM3 tool to become a SaaS needs additional deliberation.

ACKNOWLEDGEMENTS

The article presentation in I3M 2017 conference is supported by FLAG-ERA JCT 2016 project "Large scale experiments and simulations for the second generation of FuturICT" (FuturICT 2.0).

REFERENCES

Aizstrauta D, Celmina A, Ginters E, Mazza R, 2013. Validation of integrated acceptance and sustainability assessment methodology. Procedia Computer Science 26:33-40

Aizstrauta D, Ginters E, 2017. Using Market Data of Technologies to Build a Dynamic Integrated Acceptance and Sustainability Assessment Model. Procedia Computer Science 104: 501-508

Aizstrauta D, Ginters E, 2015. Integrated Acceptance and Sustainability Assessment Model Transformations into Executable System Dynamics Model. Procedia Computer Science 77: 92-97

Aizstrauta D, Ginters E, Piera Eroles M.A, 2014. Applying theory of diffusion of innovations to evaluate technology acceptance and sustainability. ICTE in Regional Development 43:69-78

Badham R, Clegg C, Wall T, 2000. Socio-technical theory. In: Karwowski, W, eds Handbook of Ergonomics. New York: John Wiley

Barlas Y, 2002. System Dynamics: Systemic Feedback Modeling for Policy Analysis, in Knowledge for Sustainable Development: An Insight into the Encyclopedia of Life Support Systems. Paris,

France; Oxford, UK: UNESCO Publishing – Eolss Publishers pp. 1131-1175.
Baxter G, Sommerville I, 2011. Socio-technical systems: From design methods to systems engineering. Interacting with Computers 23:4-17
Bhattacherjee A, 2001. Understanding Information Systems Continuance: and Expectation-Confirmation Model. MIS Quarterly 25(3): 351-370
Davis F.D, Bagozzi R.P, Warshaw P.R, 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science,35(8): 982—1003
Fishbein M, Ajzen I, 2010. Predicting and changing behavior: The reasoned action approach. New York: Psychology Press
Fortmann-Roe S, 2014. Insight Maker: A General-Purpose Tool for Web-based Modeling &Simulation.
Ginters E, Merkuryev Y, Grabis J, Straujums U, Bicevskis J, 2017. Requirements Model of Sociotechnical Systems Simulator Architecture. Adavances in Intelligent Systems and Computing. Vol. 569. Recent Advances in Information Systems and Technologies. Vol. 1. (eds) A.Rocha, A.M. Correia, H. Adeli, L.P. Reis, S. Costanzo. ISSN 2194-5357, ISBN 978-3-319-56534-7. Springer International Publishing AG, 797-807.
Ginters E, Aizstrauts A, Aizstrauta D, Lauberte I, Piera Eroles M.-A, Buil R, Sonntagbauer P, Sonntagbauer S, 2013. FP7 FUPOL Project – Innovation in Policy Science. Proceedings of the CBU International Conference IISE, pp. 231-237. 4-17April, Prague. Simulation Modelling practice and Theory 47:28-45
Ginters E, Barkane Z, Vincent H, 2010. System Dynamics Use for Technology Assessment. Proceedings of the 22th European Modeling & Simulation Symposium (EMSS). October 13-15 2010, Fes, Morocco.
Jonathan C.H, Chung-Shing L.A, 2015. Typology of technological change: Technological paradigm theory with validation and generalization from case studies. Technological Forecasting & Social Change 97:128–139.
Morris A, 2009. Socio-technical systems in ICT: a comprehensive survey. University of Trento, Italy. Available from: http://eprints.biblio.unitn.it/1670/1/054.pdf [accessed 2 March, 2017]
Nielsen S, Nielsen E.H, 2015. The Balanced Scorecard and the Strategic Learning Process: A System Dynamics Modeling Approach. Advances in Decision Sciences. Volume 2015: 20 pages
Reymondet L, 2016. A Framework for Sense-Making of Complex Sociotechnical Systems. Thesis (M.Sc). Massachusetts Institute of Technology.
Rogers E.M, 2003. Diffusion of Innovations, 5th Edition. New York: New Press
Samson Y.D, Jongsu L, 2015. Technology adoption: A conjoint analysis of consumers' preference on future online banking services. Information Systems 53:1-15
Sommerville I, 2014. Software Engineering. Essex: Pearson Education, 810.
Susmann J.M, 2013. Understanding and Designing Complex Sociotechnical Systems. Massachusetts Institute of Technology SDM Systems thinking Webinar Series. Available from: https://sdm.mit.edu/news/news_articles/webinar_040813/sussman_040813.pdf (accessed November, 2016]
Venkatesh V, Morris M, Davis G, Davis F, 2003. User acceptance of information technology: toward unified view. MIS Quarterly 27:425-478

AUTHORS BIOGRAPHY

Dace Aizstrauta finalized the doctoral studies in Vidzeme University of Applied Sciences in IT program Sociotechnical Systems Modeling and is working on technologies/projects evaluation tool – Integrated Acceptance and Sustainability Assessment Model (IASAM). Dace holds a master's degree in Computer Sciences in the field of Sociotechnical Systems Modeling and a bachelor's degree in Political science and Public administration. She worked in Sociotechnical Systems Engineering Institute and participated as researcher in FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2014). Her work on IASAM versions resembles the multi-disciplinary education and her previous professional experience in research. Her interests are related with sociotechnical phenomena in technologies development and introduction. She has more than 10 scientific articles related with the research field.

Egils Ginters is the leader of Sociotechnical Systems Engineering Group at Department of Modelling and Simulation of faculty of Computer Science and Information Technologies. He is full time professor of Riga Technical University. Graduated with excellence from Riga Technical University Department of Automation and Telemechanics in 1984 and holds Dr.Sc.Ing. in 1996. He is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE), European Social Simulation Association (ESSA) and Latvian Simulation Society. He participated and/or coordinated EC funded research and academic projects: FP7 FLAG-ERA JCT 2016 FuturICT 2.0 (2017-2020), FP7 FUPOL project No. 287119 (2011-2015), FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2014) and other. His interests in science are related with sociotechnical systems modeling and simulation, logistics information systems and systems sustainability research. He has more than 165 scientific articles related with the research fields.

ANALYTICAL REVIEW AND THE ANALYSIS OF EDUCATIONAL PROGRAMS WITH COMPUTER MODELLING AND SIMULATION ENGINEERING CONTENT

Borut Zupančič[(a)], Yuri Senichenkov[(b)], Gašper Mušič[(c)]

[(a),(c)]University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, 1000 Ljubljana, Slovenia
[(b)] Peter the Great Saint-Petersburg Polytechnic University, Polytechnicheskaya 29, 195251 St. Petersburg, Russia

[(a)]borut.zupancic@fe.uni-lj.si, [(b)]sen@dcn.icc.spbstu.ru, [(c)]gasper.music@fe.uni-lj.si

ABSTRACT

The paper deals with outcomes of a working package in the InMotion ERASMUS+ project, in which European partners should help to establish improved programs with modelling and simulation content in Malaysia and Russian Federation. By the aid of a survey sent to partner universities and worldwide, the analytical review and the analysis of educational programs with computer modelling and simulation engineering content was done. The answers were analysed for bachelor, master and PhD programs with regard to the basic information (duration, contact hours and individual work, final work, practical orientation of the study and elective courses), with regard to curricula and competencies. The final part is devoted to the investigation whether there are some bachelor, master and PhD programs that are completely in the area of modelling and simulation. We were able to find only one example of such institution in US.

Keywords: education, curriculum, syllabus, computer modelling and simulation engineering

1. INTRODUCTION

Most of engineering higher education programs have Computer Modelling and Simulation Engineering (CMSE) content as important part inside several courses. InMotion (Innovative teaching and learning strategies in open modelling and simulation environment for student-centered engineering education) (Inmotion, 2017) is an ERASMUS+ project under Key Action 2 – Capacity Building in the field of higher education (EACEA, 2017) with the general aim to continue the reform of the system of engineering higher education in in partner countries (PC) Malaysia (MY) and Russian Federation (RU), to improve quality of education and teaching according to the priorities established in the Bucharest (Bucharest, 2012) and Yerevan Communiqués (Yerevan, 2015), and to meet the demands of Strategic Framework for European Cooperation in Education and Training (European Commission/EACEA/Eurydice, 2013). This has to be done by the aid of European partners (EU universities): University Bremen (UniHB), National Distance Education University, Madrid (UNED) and University of Ljubljana (UL).

The following aims were defined:

- to improve the level of competences and skills in CMSE by developing new and innovative education approaches and learning modules,
- to provide relevant learning activities in appropriate contexts for different types of learners, including lifelong learning,
- to ensure a quality higher education system in CMSE and enhance its relevance for the labour market and society,
- to promote a European dimension in higher education for the modernisation, accessibility and internationalisation of the higher education in CMSE in MY and RU and
- to contribute to the cooperation between the EU and PC universities.

The main objectives are:

- Updated Curricula in CMSE with new Syllabi and educational content as fundamental educational program for three level educational model and development of guidelines for Long Life Learning (LLL).
- Development of a common approach for student-centred learning in the use of modern computer simulation packages and tools for solving innovative engineering problems for various application areas.
- Introduction of eScience approach and research-based learning; development of eLearning modules based on innovative teaching strategies and creative learning approaches using workflow modelling tools and blended learning approaches based on the best information-communication technologies (ICT).
- Elaboration and implementation of Open Modelling and Simulation Environment platform (OMSE), and Massive Open Online Courses of the new generation (MOOC) for qualitative improvement of the engineering education process and academic workflow

support among universities and stakeholders across the PC and EU Member States.

When implemented it is supposed the project will change the situation in the following ways:

- Student-centred learning will make the educational process more flexible and more efficient by the choice of the desired studying areas.
- Graduates from MY and RU universities will obtain appropriate competences from the CMSE field.
- With OMSE a new paradigm with respect to integration, harmonization and aggregation of various types of quality-controlled eLearning components derived from internationally operated learning and research facilities will be created.
- The stakeholders will get access to the MOOCs for the LLL training of their professionals.
- Prospectively, other faculties of partner universities and universities outside the consortium may adopt the learning environment (OMSE) and use it for the teaching.

As much as possible the results from a previous TEMPUS project eMaris will be used (Wishnewsky et al., 2013, Gordeeva et al., 2014).

The focus of this paper is one working package - WP1.2, with the goal to make an analytical review of educational programs with CMSE content in EU and PC universities. The analysis should include bachelor, master and PhD level. The results will be used as a basis for new or updated curricula and Syllabi with CMSE content. University of Ljubljana, Faculty of Electrical Engineering (UL) was responsible for this action and for the report.

2. INITIAL ACTIVITIES

The first idea was to collect the current curricula and syllabi of all project partners. However some beginning activities showed that the huge volume of materials would be collected in quite different forms, so it would be later very difficult to extract any usable information. After some meetings within UL group and with some consultations with our partners, we decided to develop a survey with which each partner would be forced to develop a document which already analyses their programs in a way that synthetical results can efficiently be used for further developments in working package WP1 and further. So we decided to collect surveys for the programs, which have most modelling and simulation courses, but simultaneously also to collect curricula. All materials should be in English. One program means one survey. We expected from each partner several surveys, if possible for bachelor, master and PhD cycle.

Although the emphasis was given to CMSE it was rather clear that there are at least according to our knowledge no CMSE programs. We have in mind more general engineering programs, which hopefully contain several CMSE courses (e.g. electrical, computer, mechanical engineering …).

Simultaneously we also asked partners to send curricula.

3. SURVEY

3.1. Description of the survey

The first part of the survey collects general program information: institution, name of the program, duration, number of credit points (CP), information about the actual amount of 1 CP load for a student, the number of contact hours (CH), the number of hours of individual work (IW), then the information in CP for final work (diploma), practical work, and at the end the total amount of CP of compulsory and elective courses. Of course we did not know the situation in Russia and Malaysia with regard to the credit system. Therefore we explained in the instructions the European credit system and asked partners to recalculate their own credits into European in order to be able to better and easier compare programs.

The remaining part of the survey is more dedicated to the CMSE area. It consists of PART I and PART II.

In PART 1 we analyse the curriculum with regard to CMSE: three types of courses should be listed:

- Basic courses in engineering programs without direct CMSE contents but very important (essential) for CMSE (e.g. Mathematics).
- Courses, which parts are also important parts of CMSE courses (e.g. Numerical methods).
- Pure CMSE courses (e.g. Continuous systems modelling and simulation).

In this part the survey asks for syllabus outline. We wanted that important topics mostly from all (pure) CMSE courses are itemised. This means that the syllabi of several courses should be analysed and more important items included. The survey asks also for proposals for additional topics in case of reforms, possibilities for new courses, … and for some interesting CMSE applications.

PART II was included at the request of a Russian partner and deals with competencies. We ask to indicate three types of competencies:

- General (general outcomes that students must obtain in higher education programs, e.g. critical thinking on the basis of analysis and synthesis, …)
- Professional-general (these are competencies related to the particular engineering program, e.g. optimal use of ICT.
- Professional-specific (these are competencies devoted to modelling and simulation in

engineering programs, e.g. experimental modelling, …).

The PART II is followed by formation of competencies distribution. Namely we ask for the numbers of courses (among the listed ones) that give three types of competencies. We expected that basic engineering courses give mostly general competencies, the courses with CMSE content more professional-general competencies, and the pure CMSE courses mostly professional specific competencies. So we expected to obtain a matrix with bigger numbers at the diagonal.

The survey is concluded with appropriate web links, where more information about the programs can be found.

A detailed instructions for survey completion were also included.

3.2. Distribution of the survey
The survey was sent to 10 EU and PC partners. As we wanted to obtain more results we sent the survey also to 19 other European partners who do not participate in the project. We did not obtain responses from eight partners. All partners with appropriate acronyms, which are later used in the analysis, are presented in Table 1.

Table1: All partners to which the survey was sent.

	Partners in the project – EU and PC	Country	Acronym
1	University of Bremen	Germany	UniHB
2	St.Petersburg State Marine Technical University	Russia	SMTU
3	St.Petersburg State Politechnical University	Russia	SPBPU
4	Novosibirsk State Technical University	Russia	NSTU
5	Universiti Kuala Lumpur	Malaysia	UniKL
6	Universiti Teknologi Malaysia	Malaysia	UTM
7	Universidad Nacional de Educacion Adistancia (The National Distance Education University)	Spain	UNED
8	University of Ljubljana	Slovenia	UL
9	St. Petersburg Institute for Information of RAS	Russia	SPIIRAS
10	Universiti Teknologi PETRONAS	Malaysia	UTP
	Other universities which responded		
1	University of Ljubljana, FRI	Slovenia	UL FRI
2	Amsterdam University of Applied Sciences	Netherland	AUAS
3	Faculty of Information Studies in Novo mesto	Slovenia	FIS
4	Technical University Riga	Latvia	TUR
5	University of Maribor	Slovenia	UM
6	Wismar University of Applied Sciences	Germany	WU-M
7	University of Glasgow	Scotland	UG
8	Politecnico di Milano	Italy	PoliMi
9	Vienna University of Technology	Austria	TUW
10	University of La Rioja	Spain	UR
11	University of Zagreb	Croatia	UZG

3.3. Reception of surveys and curricula
From project partners we obtained surveys for 10 bachelor studies, for 7 master studies and for 4 PhD studies (Zupančič et. all. 2016, see Table I). From other European institutions we obtained surveys for 8 bachelor programs, 7 master programs and 1 PhD program. All together 37 surveys were completed.
From project partners we obtained curricula for 7 bachelor studies, for 5 master studies and for 4 PhD studies. We did not collect curricula from other European partners.

4. ANALYTICAL REVIEW

4.1. General program information
It is well known that European credit system specifies 25-30 hours (usually 25) of student work (CH+IW) for 1 CP. 1 semester has 30 CP. So 3 years program has 180 CP and 4 years program 240 CP. 1 semester normally contains 15 weeks. According to Slovenian rules the number of CH/week must be between 20 and 30. Russian system also operates with CP, which are even entitled ECTS. However 1 CP means 36 working hours. The max. no. of CH/week is 32. As one semester contains 17 weeks, it results in much higher number of contact hours in the program. In Malaysian system the credit system is also used. However they operate with 1 CP_{Mal}=40 hours of student work. As 4 years programs have app. 140 CP_{Mal}, then 1 semester means 17.5 CP_{Mal}. To compare programs more easily we asked partners for appropriate recalculations to European system. It seems that hours for 1 CP_{Mal} should be divided by 1.6 and Malaysian credits for courses and programs must be multiplied with 1.6 to obtain European credits.

4.1.1. Analysis of general information for bachelor programs
General program information is analysed in Zupančič et. al. 2016, see Table II.

Duration
The duration in Europe is usually 3 years (180 CP), sometimes 4 years (240 CP). In Russia and Malaysia all programs have 4 years (240 CP).

Contact hours and individual work
It was already commented that Russian programs have much more contact hours. The ratio CH/IW is very

different: from 0.37 for UNED (what can be explained as this is an e-learning institution), app. 0.8 for Russian universities (2.6 for SPbPU is probably a misunderstanding) and 0.6-1 for EU universities.

Final work
The final work, which includes preparation, thesis, defence … has usually 5-15 CP with some exceptions (AUAS 30 CP, UL 0 CP). The difference between EU and PC partners is not observed.

Practical orientation of the study
It is very important for engineering studies to have a strong practical component. Therefore we introduced two questions in the survey: practical work, which includes lab. exercises, seminars, tutorials, … and practical work, which includes field internship (typically in industry). The percentage ratio between the sum of these two data and the CP of the program shows the value from 24% (UM) to 65% (UniKL).

Elective courses
Traditional European programs were based on compulsory units. Bologna reform required more elective courses. The percentage of elective CP against program CP shows for most programs the value 10-20%. According to Slovenian rules the minimal value is 10% (5 % of professional courses, 5% of general courses, also from any other institutions).

Although the instructions clearly explained that elective courses must be counted from a student point of view (i.e. how many can a student select) some partners included the sum of all elective courses credits.

4.1.2. Analysis of general information for master programs
General program information is analysed in Zupančič et. al. 2016, see Table III.

Duration
The duration is in most cases 2 years (120 CP). There are also shorter programs: UNED (Spain) with 2 e-learning programs - 1 and 1.5 years, WU-EM (Wismar) and UG (Glasgow) 1 year. UG (Glasgow) has also one integrated master program with duration of 5 years. The number of CP sometimes differs from expected values, as some 2 year programs are actually 3 semester programs and some programs have more intensive teaching – also the work during vacation period.

Contact hours and individual work
The ratio CH/IW is very different: from 0.2 for SPbPU to 0.9 for UG. Typical value is app. 0.5.

Final work
Final work which, includes preparation, thesis, defence … has very different amount of CP: from 6 CP (on SPbPU) to 30 CP (on UNIHB, GU, UL). In general the number of CP is higher compared to the bachelor level.

Practical orientation of the study
The percentage ratio between the sum of the practical work, which includes lab. exercises, seminars, tutorials, … and practical work, which includes field internship (typically in industry) and the CP of the program shows the value between 20% (UL, PoliMi) and 70% (SPbPU, NSTU, UNED1, UR).

Elective courses
The percentage of elective CP against program CP shows for most studies the value 10-80%. Typical value is app. 30%, which is more than on the bachelor level.

4.1.3. Analysis of general information for PhD programs
Unfortunately we received only 5 surveys for PhD programs (SMTU, NSTU, UL, SPIIRAS, UR). Although most institutions, which were included into investigation, answered that they have PhD programs but as there are no CMSE courses, they did not complete the survey. General program information is analysed in Zupančič et al. 2016, see Table IV.

Duration
The duration is in two European programs and SPIIRAS 3 years (as proposed by Bologna rules) and in two Russian programs (SMTU, NSTU) 4 years.

Contact hours and individual work
Of course in all programs there are much more individual (research) work as contact hours. The ratio CH/IW is 0.03-0.09.

Final work
Final work which includes preparation, thesis, defence … has on SMTU and NSTU 9 CP, on UL 30 CP and on SPIIRAS 4 CP.

Practical orientation of the study
The percentage ratio between the sum of the practical work, which includes lab. exercises, seminars, tutorials, … and practical work, which includes field internship (typically in industry) and the CP of the program shows the value between 3% (SMTU) and 8% (NSTU, SPIIRAS). Such small numbers are expected due to the fact that PhD study is based mainly on individual research work.

Elective courses
The percentage of elective CP against program CP shows for most studies the value 3-8%. This is also expected as the majority of CP is devoted to individual research work.

4.2. PART I. Analysis of the curriculum with regard to CMSE
In this part we collected three types of courses: basic courses in engineering programs without direct CMSE

contents but very important (essential) for CMSE, courses, which parts are also important parts of CMSE courses and pure CMSE courses (see Section 3.1.). We collected the titles of courses, appropriate CP, CH and IW and the information whether courses are compulsory or elective. Later we learnt that according to educational standard (named 3+) there are even more categories in Russia:

- compulsory basic - units, that are obligatory to take place in the curriculum,
- compulsory variable - units a department should choose from some certain quantity and include them in the curriculum; all practices (after the 1, 2, 3 years),
- elective - units a student should choose from some certain quantity and include them in his own individual educational plan,
- facultative – units, a student may additionally choose; such disciplines are not marked in CP.

The analysis shows, how much a particular program is oriented into modelling and simulation. As expected usually engineering programs have only few courses, which can be treated as pure CMSE courses.

As partners completed this part with very different understanding, the results are rather questionable. It appeared that there are very different interpretations about the particular course types.

4.2.1. Analysis of bachelor programs

The results are analysed by Zupančič et al. 2016, Table V. We see that programs have 0-4 pure CMSE courses with 0-11 CP. In average there are 2-3 courses with app. 10 CP. There is no big difference between EU, Russian and Malaysian programs.

4.2.2. Analysis of master programs

The results are analysed by Zupančič et al. 2016, Table VI. We see that programs have 0-6 pure CMSE courses with 0-30 CP. In average there are 3 courses with app. 18 CP. There is no big difference between EU and Russian programs. UTM as the only Malaysian representative declared only 1 course with 5 CP.

4.2.3. Analysis of PhD programs

The results are analysed by Zupančič et al. 2016, Table VII. In this part the analysis is difficult as there are only 5 surveys. We see that programs have 0-2 pure CMSE courses with 0-20 CP. It seems that Russian partners have more CMSE contents as European partners. NSTU declared 2 courses with 21 CP. SMTU has even Mathematical modelling included in the name of the specialisation within the PhD program, but actually only 1 pure CMSE course with 5 CP.

4.3. Syllabus outline

As mentioned we asked partners and others to write in itemised form the most important topics, which are in their opinion typical for modelling and simulation. We expected that this items are mostly from pure CMSE courses, but can be also from other courses. The completed surveys show, that the thinking, what is actually important for modelling and simulation is very different. Some partners listed contents that are in our understanding important but not in the real focus of CMSE. Some surveys were in this part empty. We understand that such partners meant that they do not have real CMSE contents in the program.

The survey asked also for proposals for additional topics in case of future program updates or reforms.

Based on all the responses we made a selection of more important items for the current situation and for the future plans separately for bachelor, master and PhD level. As the programs are very different one can find same contents on different cycles.

Syllabus outline for bachelor programs
Current status

– Conventional mathematical modelling of dynamical systems.
– Theoretical, experimental and combined modelling.
– Simulation methods: from differential equations, transfer functions, state space description to simulation program.
– Multi-components models.
– Unified and universal modelling.
– Object-Oriented modelling.
– Tools: UML, Matlab, Simulink, Stateflow, Modelica, Maple, Mathematica, Rand Model Designer.
– Simulation with general purpose programming languages.
– Models based on partial differential equations.
– Numerical methods and problems: integration methods, numerical stability, the problem of discontinuities, the problem of algebraic loops.
– Analysis of simulation results.
– Experiment design and optimization.
– Verification and Validation.
– Experimental modelling – Identification.
– Finite element methods.
– Modelling and simulation of discrete-event systems (DEVS).
– Tools for DEVS: Matlab, SimEvents, Enterprise Dynamics, AnyLogic.
– Petri nets, coloured Petri nets.
– Agent-based modelling.
– Analysis of bottlenecks.
– Modelling, simulation and optimization of production systems.
– Monitoring and supervision of processes units.
– Logistics: Ports, airports, shopping centres...

- Operational Research.
- Queuing theory.
- Monte Carlo method.
- Probabilistic models, modelling of random inputs.
- Hybrid systems.

Future plans
- Numerical libraries.
- Planning and carrying out computer experiments.
- Real time, hardware in the loop simulation.
- Visualisation and animation.
- Artificial intelligence in modelling and simulation.
- Virtual-reality based simulation.
- Simulation of complex and distributed control systems.
- Modelling and simulation of hybrid systems.
- Agent-based modelling.
- Web and cloud computing based simulation.
- Industry 4.0 in modelling and simulation.

Syllabus outline for master programs
Current status
- Simulation of complex systems (discontinuous, variable structure, …).
- Hybrid systems. Event detection, software tools.
- Component models (variable structure, agent based modelling).
- Model simplification.
- Bond graphs.
- Evolutionary computation for modelling and simulation.
- Modelling and simulation with PDE.
- Dynamical model parameter estimation.
- Identification of non-parametric models.
- Multivariable and non-linear system identification.
- Paradigm of physical modelling.
- Object oriented modelling. Modelica. Rand designer.
- Computational causality. Overdetermined and underdetermined systems. DAE index. Index reduction. System initialization. Algebraic loops. Symbolic manipulations. Tearing.
- Finite automata and state charts.
- Modelling with partial differential equations. Initial and boundary conditions. Numerical methods for solving PDE.
- Fundamentals of solving partial differential equations using finite element method.
- Real time simulation, hardware in the loop, software in the loop, rapid prototyping.
- Discrete-event models, cellular automata, agent-based models.
- The basics of cellular automata and Monte Carlo methods.

Future plans
- The supply chain modelling.
- Parallel computing.
- Planning and carrying out computer experiments.
- Real time simulation.
- Web-based simulation.
- Co-simulation.
- Modelling with partial differential equations.
- Virtual reality based simulation.
- Integrated marine logistics optimization.
- Multi-body systems.
- Visual analysis and animation.

Syllabus outline for PhD programs
Current status
- Hybrid systems: discontinuous, variable structure, events and accurate detection, stiff systems, numerical integration methods.
- Hybrid automata. State diagrams. Block diagrams. Block-textual diagrams.
- Software for modelling and simulation of hybrid systems.
- Graphical modelling languages and visual computer models. Formal languages. Syntactic and semantic analysis.

Future plans
- Multi-agent models. Use of software for developing and analysis of agent-based models.
- Hybrid system simulation.
- Visual interactive simulation.

4.4. PART II. Competencies

As already mentioned, we introduced this part in the survey at the request of a Russian partner. It is still not clear whether the future analysis of this part can give some practical results. Many partners were unsatisfied to complete this part as they did not feel competent for it. This was also the reason why some surveys were not sent back or were empty in this part. The fact is that proper fulfilment of this part is complicated and time consuming.

Some countries (also Russia) have special catalogues for all types of competencies and then it is easier to fill out appropriate data. However in other countries they do not use catalogues and then one has to invent many answers, which are than quite different and cannot be compared. But it is clear that one should fill out this part from accredited program. Unfortunately competencies of accredited programs usually (at least in Ljubljana) do not include modelling and simulation items. These items can be found only in a document, which precisely specify the competencies and outcomes of particular courses. So some partners developed huge lists of competencies which are usually rather self-understanding but still very difficult for comparisons.

We know that we should develop programs starting with competencies. This is a systematic approach. However we do not plan to build new programs but to upgrade the existing ones. So we do not need to think about some general competencies but about very specific ones for the CMSE area. Going through all surveys we can find a useful information.

To conclude, many surveys came back also with competencies distribution table fulfilled. What we expected that basic engineering courses give mostly general competencies, the courses with CMSE content more professional-general competencies, and the pure CMSE courses mostly professional specific competencies, was proved.

5. OTHER MODELLING AND SIMULATION PROGRAMS

As mentioned several times all engineering programs have usually a small amount of modelling and simulation content. It would be really interesting to make an investigation whether there are some bachelor, master and PhD programs that are completely in the area of modelling and simulation. We found an example on Old Dominion University in Norfolk, US (Old Dominion University, 2017) which offers an undergraduate four-year degree program leading to the Bachelor of Science in Modelling and Simulation Engineering. The department also offers programs of graduate study leading to the degrees Master of Engineering, Master of Science, Doctor of Engineering, and Doctor of Philosophy with a major in Modelling and Simulation.
The institution offers many small courses usually for 3 CP. Details can be found on the WEB page
http://catalog.odu.edu/courses/msim/
This is the list with some courses titles:

- Introduction to Modeling and Simulation Engineering.
- Discrete Event Simulation.
- Continuous Simulation.
- Simulation Software Design.
- Continuous Simulation Laboratory.
- Simulation Software Design Laboratory.
- Topics in Modeling and Simulation Engineering.
- Introduction to Distributed Simulation.
- Introduction to Game Development.
- Secure and Trusted Operating Systems.
- Computer Graphics and Visualization.
- Introduction to Medical Image Analysis.
- Design and Modelling of Autonomous Robotic Systems.
- Introduction to Game Development.
- Machine Learning.
- Optimization Methods.
- Finite Element Analysis.
- High Performance Computing and Simulations.
- Cluster Parallel Computing.
- Advanced Analysis for Modelling and Simulation.
- Modelling Global Events.
- Computational Methods for Transportation Systems.
- Internship.
- Practicum.
- Doctor of Engineering Project.

6. CONCLUSION

In the report we briefly summarize some important facts, which were obtained from the surveys. We are aware that some results and comparisons are questionable also due to some misunderstandings, which also occur due to time limitations all partners had for completing surveys.

More information can be found in the surveys (Zupančič et al. 2016 – Appendix). The surveys are divided into classes for bachelor, master and PhD programs. In each class there are surveys of partner institutions (all partners in the project) and of other European universities.

ACKNOWLEDGMENTS

The work described in this paper was conducted within InMotion project, co-funded by the Erasmus+ Programme of the European Union, No. 573751-EPP-1-2016-1-DE-EPPKA2-CBHE-JP.

REFERENCES

Bucharest 2012, Ministerial Conference. Available from: https://www.ehea.info/cid101043/ministerial-conference-bucharest-2012.html [accessed 30 May 2017]

EACEA 2017, Education, Audiovisual and Culture Executive Agency. Available from: https://eacea.ec.europa.eu/homepage_en [accessed 30 May 2017]

European Commision/EACEA/Eurydice, 2013. Education and Training in Europe 2020: responses from the EU Member states. Eurydice Report. Brussels: Eurydice, ISBN 978-92-9201-512-1, doi: 10.2797/49490. Available from: http://eacea.ec.europa.eu/education/eurydice http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/163EN.pdf [accessed 30 May 2017]

Gordeeva S., Novopashenny I., Ogurol Y., Ryzhov V., Zhao J. (2014). Sakai CLE for blended learning model. Proc. of International Academic Conference on Education Teaching and e-Learning, Prague, pp. 157 – 158.

InMotion, Innovative teaching and learning strategies in open modelling and simulation environment for student-centered engineering education, 2017.

Available from: http://www.inmotion-project.net [accessed 30 May, 2017]
Old Dominion University, 2017. Available from: http://www.odu.edu/msve [accessed 30 May 2017]
Wishnewsky M., Novopashenny I., Zhao J., Ogurol Y., Podgaiskii E. et al. (2013). Experience with Sakai CLE in the framework of international educational TEMPUS project eMaris in the field of Applied Marine Sciences. CALMet X Conference and Eumetcal Workshop, Toulouse, France, pp. 64 – 65.
Yerevan, 2015, Ministerial Conference and Fourth Bologna Policy Forum. Available from: http://bologna-yerevan2015.ehea.info/ [accessed 30 May 2017]
Zupančič B. Mušič G., Simčič M., 2016. Analytical review on the analysis of EU & PC educational programs, Report for the working package WP 1.2 for ERASMUS+ project InMotion, University of Ljubljana, Faculty of Electrical Engineering. Available from: http://msc.fe.uni-lj.si/Download//Zupancic/Report_WP1.2.pdf [accessed 30 May 2017]

AUTHORS BIOGRAPHY

BORUT ZUPANČIČ received his Ph.D. and became a full professor at the Faculty of Electrical Engineering, University of Ljubljana in 2000. His major research interests are: control systems, multi-domain and object oriented modelling and simulation, continuous and hybrid control systems design, harmonization of thermal and flows in buildings. He is the author of more than 200 conference papers and 50 papers in scientific journals, co-author of one international book (published by Prentice Hall Inc.) and author or co-author of several books in Slovene language. He was the president of EUROSIM - the Federation of European Simulation Societies 2004-07 and the Secretary 2010-16 and the president of SLOSIM - the Slovene Society for modelling and simulation. Currently he is the Head of the Laboratory for Modelling, Simulation and Control at the University of Ljubljana, Faculty of Electrical Engineering.
His Web page can be found at
http://msc.fe.uni-lj.si/Staff.asp?person=2

YURI SENICHENKOV is a professor of the Higher School «Software engineering» at Institute of Computer Science and Technology of Peter the Great Saint-Petersburg Polytechnic University (dcn.icc.spbstu.ru).
He received his degree "Candidate of Science" in the field of Numerical Mathematics from St. Petersburg State University in 1984, degree "Doctor of Science" in the field of Numerical Software from St. Petersburg Polytechnic University (2004). His fields of scientific interest are: Mathematical and Computer Modeling, Numerical software and Numerical Analysis. Yuri Senichenkov is is the author of more than 150 conference papers and papers in scientific journals, co-author of Model Vision Studium (www.MvStudium.com), Rand Model Designer (www.rand-service.com/ru) - tools for modeling and simulation of complex dynamical systems, co-author of books (in Russian): "Visual Modeling" (2001), "Practical modeling of complex dynamical systems" (2003)(www.bhv.ru),"Numerical modeling" (2004)(http://gpupress.ru/), "Modeling of systems. Dynamical and hybrid systems", "Modeling of systems. Object-Oriented approach"(2006) (www.bhv.ru); «Mathematical modeling. Component technologies» (2013) (http://gpupress.ru/), «Mathematical modeling of hybrid dynamical system» (http://gpupress.ru/) (2014); «Object-Oriented Modeling in Rand Model Designer 7» (2016), chairman of Polytechnic University Annual Computer Modeling Conference -COMOD (http://dcn.icc.spbstu.ru), a member of editorial boards of Computer Instruments in Education Journal (www.ipo.spb.ru), Simulation News Europe Journal (www.sne-journal.org), Humanities and Science University Journal (http://en.unipress.pro), a member of board Russian Federation the National Simulation Society (http://simulation.su/en.html).

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 Full 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.

USING HIGH LEVEL ARCHITECTURE TO COMBINE SIMULATIONS IN A COMPANY CONTEXT: MOBILE FACTOR

[a]Simon GORECKI, [b]Gregory ZACHAREWICZ, [c]Nicolas PERRY

[a], [b]Integration from Material to Systems lab, University of Bordeaux, France
[c]Arts et métiers, ParisTech, ENSAM of Bordeaux, France

[a]simon.gorecki@u-bordeaux.fr, [b]gregory.zacharewicz@u-bordeaux.fr
[c]nicolas.perry@ensam.eu

ABSTRACT
Modeling and simulating play an important role in the industry, and even more in the innovative domain, for testing scenarios by anticipating different eventualities. Those simulations are often very complex and difficult to modify, that is why they could be built separately first and then run together. One of the most popular standards for distributed simulations is the IEEE 1516-2010-Evolved of the High Level Architecture (HLA) that supports implementation of distributed simulations.
In our context, a company has launched a project of mobile factory to set up solar panel field in several countries. That implies various study domains and expertise (foundation definition, study of solar transmitters structure, risk analysis, project management, factory modeling). The goal of this paper is to demonstrate the interconnections proposed between all these entities. It will be implemented in the frame of a global simulation managed by HLA.

Keywords: High Level Architecture; Mobile Factory

1. INTRODUCTION

In industrial design domain, the modeling and simulation tools are very important. They allow the representation of any complex system, to study their behavior and their interactions with the environment. Once the modeling phase is completed, the simulation of these systems will allow us to virtually design our subject to anticipate and avoid problems during the building phase.
In our case, we will use various modeling and simulation technologies in a semi-academic, semi-professional context: a French company has launched an innovative project for setting up a solar power plant. This project deals with different domains including risks. University of Bordeaux supports research in M&S and specifically in DS (Distributed Simulation). Most of these research works have created specific domain simulators. Each of these autonomous simulations is capable of representing a fragment of the global project. One of the last phases of this project is to assemble all these simulations to obtain a global simulation of the problem. However, all these simulations use different technologies and manipulate heterogeneous data, which complicates the assembly. To solve these problems, we will use the HLA standard, and especially the Pitch Technology. (Möller, 2012)

2. DISTRIBUTED SIMULATION & HIGH LEVEL ARCHITECTURE

In the computer simulation domain, distributed simulations are one of the most useful and powerful applications. Indeed, they consist of several components (often associated with one or more functions) that can be processed by different processes. All these components are part of a single execution which can be relocated to a different computer / server, hence the term "distributed". This concept of functions relocation makes the loads distribution possible on different machines, and thus increase the efficiency of a program.
One of the advantages of distributed simulation is to solve some interoperability problems. Interoperability is the interactions ability between systems. This problematic appears when several highly dissimilar systems (by their internal structure, exchanged data format, or semantic data) have to communicate. The interoperability problematic must be considered if interactions are at data level, service level or process level (Zacharewicz, Labarthe, Chen, & Vallespir, 2009). Those problematics involve at least two entities which try to communicate. Consequently, establishing interoperability consist in relating two systems together and remove incompatibilities. Incompatibility is the fundamental concept of interoperability. The concept of 'incompatibility' has a broad sense and is not only limited to 'technical' aspect - as usually considered in software engineering, - but also 'information' and 'organisation' (Zacharewicz, Chen, & Vallespir, 2009).
Indeed, in distributed simulations, the components are modular. They can have a heterogeneous architecture and exchange different structured messages. This enables the solution of interoperability problems.
In our application case, this notion of distributed simulation will be used with the High Level Architecture standard (HLA) (IEEE Computer Society, 2010). It is a specification of software architecture. It defines a framework which allows the creation of global execution. This framework defines how to create a

"global" simulation, which is made of several distributed simulations. This distributed simulations can communicate with one another. It was originally created by the Office of Defense Modeling and Simulation (DMSO) of US Department of Defense (DoD) to facilitate the assembly of stand-alone simulations with a different architecture. The original goal was the reuse and interoperability of military applications, simulations and sensor. This standard is designed to resolve interoperability and reusability issues between software components. Another interesting aspect of this technology is the synchronization. It allows to dynamically manage interoperability issues with simulations exchange messages: it must be ensured that messages are sent at the right time, in the right order, and that they do not violate causal constraints. In order to do this, various systems for synchronization of processes and time management are proposed by HLA.

According to the HLA standards, each simulation participating to the application is called "federate". A federate interacts with other federates. There are forming a group named HLA federation. All of these entities can communicate with each other through a Run-Time Infrastructure (RTI). It is the RTI that will manage the federation, authorize federates to communicate or not, and provide various services such as time management, file or data exchange, etc.

3. RUN-TIME INFRASTRUCTURE (RTI)

A federation is composed of a set of federates and a Run-Time Infrastructure (Falcone et al., 2015). This RTI provides to federates all functionalities which are described by the specification. Federates can only interact through the RTI. They can "Publish" to inform the RTI and the other federates about an intention to send information. They also can "Subscribe" to reflect some information created and updated by other federates. This is the basic communication mechanic. The data flow exchanged between all the federates is represented in the same form of classical object-oriented programming. There are two kind of objects which are exchanged in HLA standard: Object Class and Interaction Class. (Möller, 2012)

Object Class are time persistent during the simulation. They have attributes that can be updated. For instance, in the case of a simulation where the Object Class would be a car, his attributes would be a position, a speed, a name. Interaction are not persistent over time and can have parameters. For example, in this case, "Start car", "Stop car", "Accident" would be possible Interaction.

This two kind of objects are described by a XML file named Federation Object Model (FOM) attached to a federate. That is an important point: the FOM describes all the information that will be exchanged between federates. It is the only thing which will be shared during the simulations. This has an impact on the safety of this technology. The federate are totally autonomous, the only exchange between them is described by the XML file.

Each federate are single. Originally, HLA was created for reusability, all the components can be executed separately, or with others through a federation.

4. BUSINESS CONTEXT

The company behind this work designs solar power plants. This project consists in installing solar panels fields in several countries in order to provide electricity in world areas which are not powered so far. However, the transport of solar panels fields is extremely expensive. To reduce these blows, we aim to design a mobile factory which would manufacture the solar panels on site, and thus reduce the transport cost. Rather than transporting finished products, only the mobile plant and raw materials will be carried out. This project's main challenges are the miniaturization of this factory in order to get it in the least transport containers (around 20), and the factory designing taking into account risks and low knowledge on this project.

In terms of implementation and difficulties caused by the project complexity, this innovation gave rise to several works which subjects deal with the following problems:

- Optimization and decision helping for defining the structure foundations of the solar panels field, depending on the ground structure. Simulation executed with Mathlab and Excel. (Piegay & Breysse, 2015)
- Study of the concept maturity integration in decision making process: applied for designing the solar transmiter supporting structure. Simulation executed with Mathlab and Excel. (El Amine, 2016)
- Study and dimensioning of the mobile factory dimensions, cost, etc, according to the demand. Simulation executed with Excel. (Benama, 2014)
- Study of project management method integrating risks. Calculating risks probabilities into project management Simulation executed with Excel. (Rodney, 2014)
- Tool to concatenate all the mentionned works to make them run under Papyrus engine: BPMN for modeling and simulate systems. (Posse, 2015)

Most of these works have led to a simulator creation which solve a specific domain problem submitted to it. It is now necessary to allow them to communicate for obtaining an overall simulation representing the whole solar panel field.

5. PROPOSITON

In order to overcome the interoperability problems, we will link the different components represented in the next scheme thanks to the HLA specification.

This allows the orchestration of our simulations set communicating within "global simulation". With HLA, allowing us to create and manage independent simulated

components, the global solar panels project will be set as a federation, composed of each simulator that we just describe (as federate).
In our federation, the Run-Time Infrastructure (RTI) provides different mechanisms to manage federate and federation. It will administer data exchange between federates, allow connection authorization to federation, time management and simulations synchronization.

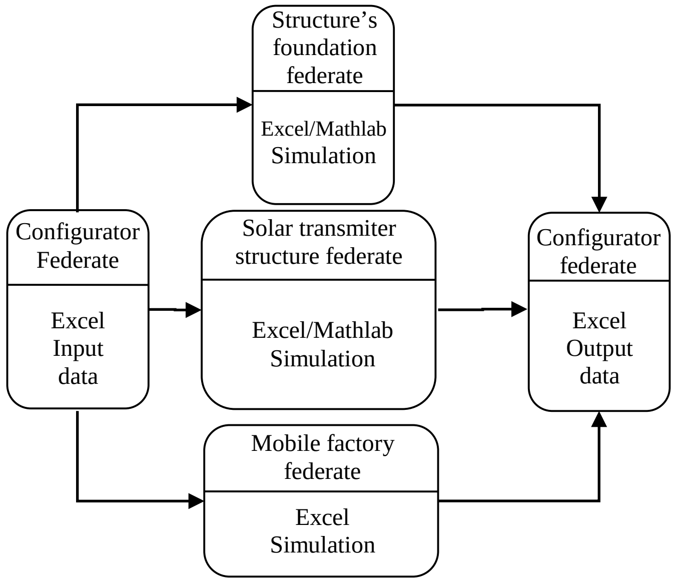


Figure 1: Business Process Model – Choreographies Diagram Mobile factory

We can see on the diagram above all the simulators organized in functional blocks connected according to the flow of the necessary flows. For the project management simulator with risk taking, it must be linked to each function in order to be able to generate untimely events that correspond to technical problems during the simulation.
In this context, the interoperability issues are strong on the one hand according to the data handled by each of the simulators, and on the other hand according to the various technologies used (Excel and Mathlab).

6. CONFIGURATOR

The starting point of all this simulations is the 'configurator'. It is used as a central point where all process will draw data, perform calculations, and store back their results. It acts as a data base module where each simulation will read and write values. No calculation are made in the part. There are 3 data modes managed by the configurator:

- Input data: will be filled by the user who defines the simulation's parameters.
- Intermediate variables: will be filled and used by several simulations in order to communicate between them.
- Output data: will contains all simulation's results.

Among the "input data" parameters, there are three different data types:

- Design variables: are parameters which must be optimized. This means we can act on those values in order to research the best compromise between all the goals. Each design variables are associated to a range of values which will vary until we found the right value.
- Project variables: are imposed by the project specifications. They depend on the system definition and strategic decision taken by the decision maker. Contrary to the design variables, we cannot act on these values. During the preliminary conception phase, the decision maker can be imprecise on some strategic choices, so, all the project variables are not every time constant, they can evolve.
- Environmental variables: describe natural properties (climatic, air, water, ground…) of the system environment, and also the social, economic and political environment.

The "intermediate variables" part has the role of link between simulations. It allows them to communicate. Any simulation can write in this field a value, a name and a unit which will be available for others programs.
Finally, the "output data" part is the global simulation final step. The user gave information in "input data", and after the run, he takes the results back in "output data". There are several output data types allocated, one for each simulator:

- Mobile factory
- Structure's foundation
- Solar transmitter

The figure below represents the data flow exchanged between the configurator and the three other simulations according to the (Yourdon & DeMarco, 1989) notation.

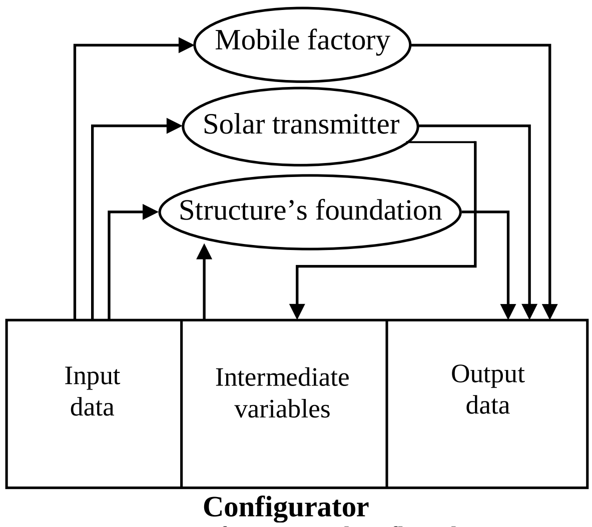


Figure 2. Configurator's data flow diagram

We can see that the “structure’s foundation” simulation need in input data from, “input data” and “intermediate variables”. This means that data from “intermediate variables” must be completed before the “structure’s foundation” runs. In order to fill this “intermediate variable” database, that we can see on the figure 2, we must run the “Solar transmitter” simulation first. This execution will complete the missing field. After that, The Structure’s foundation simulation followed by the mobile factory simulation can be executed. Those three simulations will complete the “output data” part and give the user the simulation’s result.
The figure 3 describes the simulation launch scenario that we have just described.

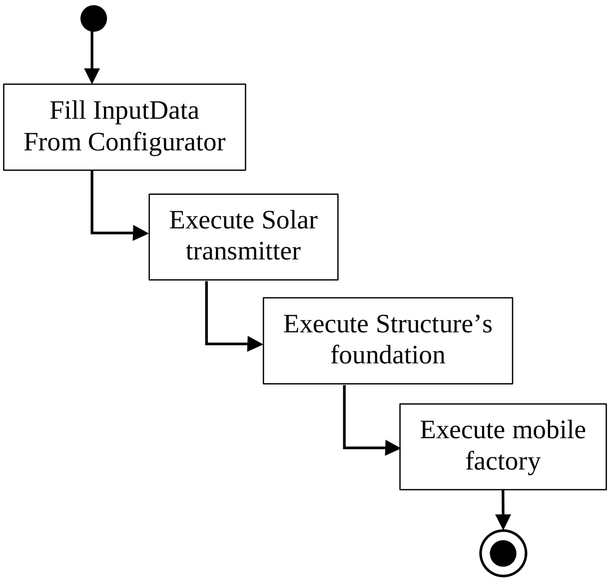


Figure 3. Simulation launch scenario

From an HLA point of view, the configurator could be a federate that:

- Contains input, intermediate, and output data we have just described.
- Distributes this data to each federate when they need them.
- Verifies data values.
- Manages the simulation launch scenario.

7. SOLAR TRANSMITTER

This simulation has for objective to determinate which parameters are the most efficient for the solar transmitter system. The presence of immature concepts makes decision-making difficult in this development phase. Furthermore decisions taken during this conceptual design phases have a critical impact on the product life cycle cost. To deal with issue, a simulation has been developed in order to anticipate conceptual problems (El Amine, 2016). As pictured in the figure 4, this simulation is composed of several Mathlab functions that use configurator’s data in order to define the best solar transmitter’s parameters considering the constraints imposed by specifications. This simulation also determinates intermediate variables that are required for the structure’s foundation simulation. That is why this simulation must be executed first.

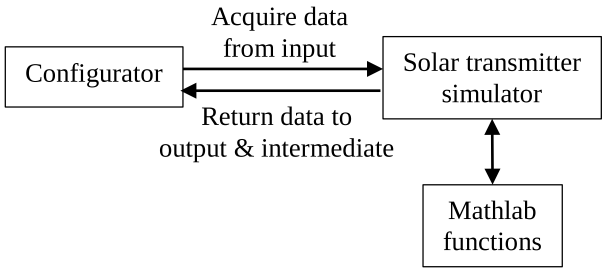


Figure 4. solar transmitter simulation architecture

8. STRUCTURE FOUNDATIONS SIMULATION

This simulation is about defining structure foundation of the solar field, and defining steel frame according to the specifications (Piegay & Breysse, 2015). The development of those structures is complex because it must respect contradictory instructions from the decision-maker. Thus, there is a need to find the best compromise between safety requirements, and economic stakes. Variability of environmental parameters and uncertain acceptance threshold are incorporated in a decision process simulation. The responses of the system can then be characterized by a performance measurement and by a dispersion measurement, making possible the judgment of the solution robustness while respecting the probabilistic character of the input data.
The effects of the interaction between ground and structure are also considered for achieving an overall structure optimization. When the simulation is over, we will have several parameters in output that give information about the most optimized structure foundation and steel frame.
For this simulation, we deal with two data sources. The first is the configurator described before. It will draw data from “input data” and "Intermediate variables” (which has just been filled by the Solar transmitter simulation) to store them in its own Excel file. The second data source is another Excel file that contains technical information about different sorts of metal beams (it is named “profile”). Those two databases will be used by two calculators: one in Excel file, it calculates structure’s foundations parameters, stock and share all simulation results to the configurator. The second calculator is a set of Mathlab functions which use the two databases for calculating the solar field steel frame. At the end of the simulation, data are brought back in the first calculator as you can see on the figure 5.

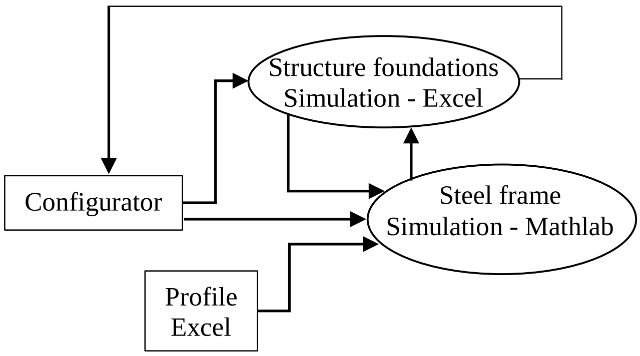


Figure 5. Structure foundations + steel frame simulation architecture – dataflow diagram

In the solar panel field system, the "Profile" database that we have just mentionned is not used by any other simulation. Therefore, it can be added to the structure foundations block. This set of two simulations plus one database could form an HLA federate. This federate has the role of simulate and defines the solar field structure foundation and defines the steed frame specification.

9. MOBILE FACTORY SIMULATION

This mobile factory simulation concerns strategic decision making between making or buying materials. In France, industrial production is coming to grief: the country exports few manufactured products. In addition, some companies have undertaken the relocation of their production and getting closer from the final customers in order to increase their margin. When a customer needs a production in a limited time, the mobile production system (MPS) can be the solution. The concept of MPS consists in mobilizing the same production system to successively satisfy several orders from geographically dispersed customers, directly on the final customer's site. (Benama, 2014)

This simulation is executed on a single Excel file. It is a system which takes in input data from the configurator, and have also its own data input. Those other data can be divided in two categories:

- Workforce variables: employees numbers available on site, teams numbers, managers numbers, break time of teams, etc…
- Technical variables: mirrors numbers per palette, palettes number per container, field surface, reflector number, etc…

As for the previous simulation, this one has its own private database which contains information about each technical operation of the solar field factory. For every construction's step of the field, we have got, on this database, information about costs in time, human, and material resources. This Excel sheet will be used as reference for mobile factory simulation.

With all this data, the simulation will execute step by step several calculations, as represented in the figure 6. The first step consists in calculating some of the mobile factory costs. Then, we also calculate the quantities and some of the raw materials buying costs. After this, comes the transport phase (comparison of the different costs depending on the type of transport carried out, storage, etc…). Later, we appraise the raw materials supply. This step permits the organization in time of the raw materials supply, according to the duration of the production cycles. And finally, the last stepis about the calculation of the plant assembly and dismantling phases, which will be done to define how much time should be expected, if qualified staff would be requiered and how much, and the cost of this project.

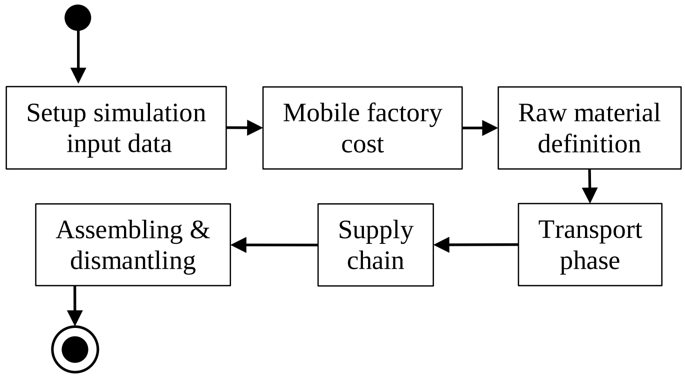


Figure 6. Mobile factory simulation scenario

In output of this simulation we will have data about solar field fabrication cost (in €/m²), information about the factory production rate, solar field installation time, the operators number required to install the plant, and the operators number required to operate the plant.

From an HLA point of view, this simulation and its associated database will be in a federate. The only communication will be for storing data input and output with the configurator.

10. ASSEMBLING SIMULATION WITH HLA

As we said earlier, each of those simulations will be defined as a federate in HLA architecture (Figure 7). The configurator will also be a federate that contains data necessary for all simulations. It will manage simulation scenario (order of execution), manage permission access to databases for each federate, and verify data values.

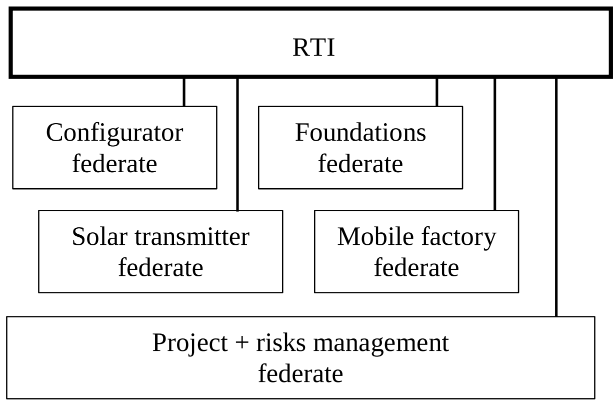


Figure 7. Global simulation as HLA view

Another federate which will be added to the federation is the project management simulation integrating risks (Rodney, 2014). Its function has the role of generating random events in the system to simulate risks, unexpected events, and defaults. It will act in the database system to generate understaffed situation, supply chain breaks, and so on. But it will also act in every simulation for generating problems, downturns, etc.

This federate has a special role because it will not only interact with the configurator, but also with all federates present in the federation.

In HLA standard, to guarantee a great communication, each federate must declare a Federation Object Model (FOM) that will describe the data exchange. In our case, each simulation previously described must declare in a file (the FOM) data that they must exchange with any other federate. For instance, the solar transmitter federate will communicate with the configurator federate, and the Project + risks management federate. So, its FOM will describe, data stored in input and output of the configurator as Object Class, and it will also describe as Interaction Class the communication with the risk generator. According to the configurator description, it will manage federate execution order, and will be described as an Interaction Class.

11. CONCLUSION

The industrial problem has been previously tackled domain by domain independently. Here the state of the art revealed that distributed simulation and the HLA standard can give an interesting answer to couple these heterogeneous works. One of the advantages of using this technology is the concept of interoperability and adaptability. The federation has permitted to define a set of entities representing different functions of the enterprise. Each function is (or must be) autonomous and have its own architecture. This allows the possibilities simply adding or removing federates to the simulation federation in order to revoke (or invoke) a functionality of the global simulation. The future step will consist in defining, according to HLA Federation Development Process (FEDEP), the behavior of each federate to preserve a global synchronization of information exchange over time.

REFERENCES

Benama, Y. (2014). Supporting make or buy decision for reconfigurable manufacturing system, in multi-site context. Ajaccio, France, 150–158.

El Amine, M. (2016). Integration of concept maturity in decision-making for engineering design: An application to a solar collector development. Springer-Verlag, London, pp 235-250.

Falcone, A., Garro, A., Anagnostou, A., Chaudhry, N. R., Salah, O.-A., & Taylor, S. J. . (2015). Easing the Development of HLA Federates The HLA Development Kit and its exploitation in the SEE Project. 2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT), Chengdu, China.

IEEE Computer Society. (2010). IEEE Standard 1516-2010 for M&S - HLA - Framework and Rules. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5553440

Möller, B. (2012). The HLA Tutorial. Pitch Technologies, Sweden. Retrieved from http://www.pitchtechnologies.com

Piegay, N., & Breysse, D. (2015). Multi-Objective Optimization and Decision Aid for Spread footing design in Uncertain environment. Geotechnical Safety and Risk 5, Rotterdam, the Netherlands, pp 419-424.

Rodney, E. (2014). Integrating Risks in Project Management. 16th International Dependency and Structure Modelling, Paris, pp 419-424.

Yourdon, E., & DeMarco, T. (1989). Mordern Structured Analysis. Englewood Cliffs, NJ, Yourdon Press.

Zacharewicz, G., Chen, D., & Vallespir, B. (2009). Short-Lived Ontology Approach for Agent/HLA Federated Entreprise Interoperability. International Conference on Interoperability for Enterprise Software and Applications, Beijing, China.

Zacharewicz, G., Labarthe, O., Chen, D., & Vallespir, B. (2009). MULTI AGENT HLA ENTERPRISE INTEROPERABILITY (SHORT-LIVED ONTOLOGY BASED). The International Workshop on Modelling & Applied Simulation, Puerto de La Cruz, Spain, pp 187-196.

AUTHORS BIOGRAPHY

Simon GORECKI: Ph.D Student at IMS lab (Integration from Material to Systems laboratory), University of Bordeaux. Domain research is about simulating process with distributed simulations and HLA (High Level Architecture).

Gregory ZACHAREWICZ: Associate professor HDR at IMS lab, University of Bordeaux. Domain research is about M&S and enterprise engineering.

Nicolas PERRY: Full Professor at ParisTech ENSAM of Bordeaux. Domain research is about system engineering, product process integration and green manufacturing.

CONTROLLED TRIANGULAR BATCHES PETRI NETS: A HIGHWAY CASE STUDY

Leonardo Brenner[a], **Isabel Demongodin**[b]

[a],[b]Aix Marseille Université, CNRS, ENSAM, Université de Toulon
LSIS UMR 7296, 13397, Marseille, France

[a]Leonardo.Brenner@lsis.org, [b]Isabel.Demongodin@lsis.org

ABSTRACT

In the discrete event and hybrid systems theory, Triangular Batches Petri Nets (TBPN) have been defined as an extension of Batches Petri Nets (BPN) to represent conges- tion/decongestion phenomena. Applied to mesoscopic modeling of traffic road networks, TBPN define a triangular flow-density relation, that allow to modeling these phenomena. An other extension, Controlled Triangular Batches Petri Nets (CTBPN) define controlled events in the formalism that allow to represent variation speed limit and flow when a control is applied on the traffic road networks. In this paper, we present the hybrid formalism defined in CTBPN, i.e., the hybrid behavior of a batch in free, congestion or decongestion behavior. A simulation method is introduced and a case study of a real highway is modeled. A comparison of the collected data and simulation results shows the accuracy of the model.

keywords: Petri Nets, Discrete Event Systems, Hybrid systems, Flow, Traffic road

1. INTRODUCTION

Traffic congestion is a growing problem in many metropolitan areas . Congestion increases travel time, air pollution, carbon dioxide (CO2) emissions and fuel use because cars cannot run efficiently. In fact, when the traffic demand is greater than the traffic capacity of the road, a phenomenon of traffic congestion appears. Mathematical models can simulate the real world scenario to predict the behavior of traffic in proper planning and design of the road network. These same models can be used to test different control strategies to reduce the congestion. In these contexts, mathematical description of traffic flow has been and is always a lively subject of research (Lighthill and Whitham 1955).

In this paper, we model and simulate a real highway road section using the Controlled Triangular Batches Petri Nets (CTBPN) formalism (Gaddouri, Brenner, and Demongodin 2014; Gaddouri, Brenner, and Demongodin 2016). This extension of the Generalized Batches Petri Nets (BTN) formalism (Demongodin 2001, Demongodin 2009) integrates the flow-density relation observed on the traffic flow.

The section 2. introduces the concepts of the CTBPN formalism. The section 3. presents the simulation method that compute the dynamic evolution of the model. A real highway section is simulated in the section 4. and the results are compared with collected data.

2. CONTROLLED TRIANGULAR BATCHES PETRI NETS

We present in this section the main concepts of the Controlled Triangular Batches Petri Nets (CTBPN) formalism. This formalism extends the flow-density relation in a batch place to represent the triangular fundamental diagram used in traffic road models (Daganzo 1994) to describe the flow-density relation.

2.1. Triangular Batches Petri Nets

A Triangular Batches Petri Net (TBPN) is an extension of a Generalized Batches Petri Net (see (Demongodin 2001, Demongodin 2009, Demongodin and Giua 2014) for more details on this formalism), where some new characteristics related to the batch place have been added.

TBPN is an hybrid Petri Nets formalism defined by six kind of nodes (places and transitions) (Figure 1).

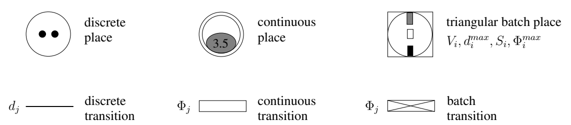


Figure 1: Nodes of Triangular Batches Petri Nets

Definition 1 *A* Triangular Batches Petri Net *(TBPN) is a 6-tuple* $N = (P, T, Pre, Post, \gamma, Time)$ *where:*

- $P = P^D \cup P^C \cup P^{TB}$ *is a finite set of places partitioned into the three classes of discrete, continuous and triangular batch places.*
- $T = T^D \cup T^C \cup T^B$ *is a finite set of transitions partitioned into the three classes of discrete, continuous and batch transitions.*
- $Pre, Post : (P^D \times T \rightarrow \mathbb{N}) \cup ((P^C \cup P^B) \times T \rightarrow \mathbb{R}_{\geq 0})$ *are, respectively, the pre-incidence and post-incidence matrices, denoting the weight of the arcs from places to transitions and from transitions to places.*
- $\gamma : P^{TB} \rightarrow \mathbb{R}^4_{\geq 0}$ *is the* triangular batch place function. *It associates with each triangular batch place* $p_i \in P^{TB}$ *the quadruple* $\gamma(p_i) = (V_i, d_i^{max}, S_i, \Phi_i^{max})$ *that represents, respectively, the maximum speed, the maximum density, the length and the maximum flow.*
- $Time : T \rightarrow \mathbb{R}_{\geq 0}$ *associates a non negative number with every transition:*
  - *if* $t_j \in T^D$*, then* $Time(t_j) = d_j$ *denotes the* firing delay *associated with the discrete transition;*
  - *if* $t_j \in T^C \cup T^{TB}$*, then* $Time(t_j) = \Phi_j$ *denotes the* maximal firing flow *associated with the continuous or batch transition.* ■

2.2. Triangular Batch Place

A triangular batch place (TB-place) $\gamma(p_i) = (V_i, d_i^{max}, S_i, \Phi_i^{max})$ is a transfer zone and is defined by four characteristics: *maximum speed*, *maximum density*, *length* and *maximum flow*.

Definition 2 *Let a triangular batch place* p_i*, with* $\gamma(p_i) = (V_i, d_i^{max}, S_i, \Phi_i^{max})$*. A* propagation speed of congestion*, denoted* W_i*, and a* critical density d_i^{cri}*, are associated with* p_i*, defined respectively by:*

$$W_i = \frac{\Phi_i^{max} \cdot V_i}{d_i^{max} \cdot V_i - \Phi_i^{max}} \tag{1}$$

$$d_i^{cri} = \frac{W_i . d_i^{max}}{V_i + W_i} \tag{2}$$

The flow-density relation *that governs the dynamics of TB-place* p_i *is defined as follows:*

$$\phi = \begin{cases} d.V_i & \text{if} \quad 0 \leq d \leq d_i^{cri} \\ W_i.(d_i^{max} - d) & \text{if} \quad d_i^{cri} < d \leq d_i^{max} \end{cases} \tag{3}$$

where d *denotes density and* ϕ *denotes flow.* ■

Figure 2 represents these definitions.

Let us now introduced some definitions needed for the rest of this paper. Let a Triangular Batch place p_i, with $\gamma(p_i) = (V_i, d_i^{max}, S_i, \Phi_i^{max})$. An *input flow* $\phi_i^{in}(\tau)$ and an *output flow* $\phi_i^{out}(\tau)$ of place p_i are respectively: $\phi_i^{\text{in}}(\tau) = Post(p_i, \cdot) \cdot \boldsymbol{\varphi}(\tau)$ and $\phi_i^{\text{out}}(\tau) = Pre(p_i, \cdot) \cdot \boldsymbol{\varphi}(\tau)$ where $\varphi(\tau)$ is the instantaneous firing vector of continuous and batch transitions (see (Demongodin and Giua 2010) for more details).

2.3. Controlled Triangular Batches Petri Nets

A Controlled Triangular Batches Petri Net (CTBPN) has the same syntax than TBPN. However we associate with CTBPN a different semantics, assuming that the maximal firing flow of continuous and batch transitions and, the maximal transfer speed of triangular batch places are control inputs.

Definition 3 *A Controlled Triangular Batches Petri Net (CTBPN) is a TBPN where the maximal transfer speed of TB-place* $p_i \in P^{TB}$ *and, the maximal firing flow associated with a continuous or batch transition* $t_j \in T^C \cup T^B$*, can varied. We denote respectively these variables:* $v_i(\tau)$*, with* $0 \leq v_i(\tau) \leq V_i$*, and* $\phi_j(\tau)$*, with* $0 \leq \phi_j(\tau) \leq \Phi_j$*.* ■

It should be noted that the variation of the speed of TB-places imposes a variation of the critical density and of the maximum flow of TB-place while the propagation speed of congestion, W_i stays constant as shown in Figure 2.

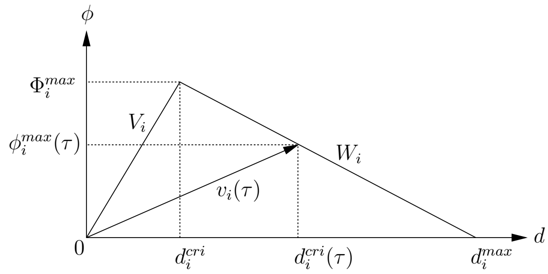


Figure 2: Flow-density relation of a TB-place

Definition 4 *Let TB-place* p_i *with* $\gamma(p_i) = (V_i, d_i^{max}, S_i, \Phi_i^{max})$ *with a maximal transfer speed* $v_i(\tau)$ *such that* $0 \leq v_i(\tau) \leq V_i$*. At time* τ*, the* controlled critical density $d_i^{cri}(\tau)$ *and the* controlled maximum flow $\phi_i^{max}(\tau)$ *are respectively defined by:*

$$d_i^{cri}(\tau) = \frac{W_i . d_i^{max}}{v_i(\tau) + W_i}, \tag{4}$$

$$\phi_i^{max}(\tau) = v_i(\tau) . d_i^{cri}(\tau) \tag{5}$$

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

with $0 \leq \phi_i^{max}(\tau) \leq \Phi_i^{max}$ *and* $\frac{\Phi_i^{max}}{V_i} \leq d_i^{cri}(\tau) \leq d_i^{max}$.

2.3.1. Controllable Batches

Each TB-place contains a series of controllable batches ordered by their head positions.

A controllable batch, i.e., a group of discrete entities characterized by continuous variables, is a four continuous variables defined as follows (Demongodin 2009).

Definition 5 *A* controllable batch $C\beta_r(\tau)$ *at time* τ*, is defined by a quadruple,* $C\beta_r(\tau) = (l_r(\tau), d_r(\tau), x_r(\tau), v_r(\tau))$ *where* $l_r(\tau) \in \mathbb{R}_{\geq 0}$ *is the* length*,* $d_r(\tau) \in \mathbb{R}_{\geq 0}$ *is the* density*,* $x_r(\tau) \in \mathbb{R}_{\geq 0}$ *is the* head position *and* $v_r(\tau) \in \mathbb{R}_{\geq 0}$ *is the* speed*. An* instantaneous batch flow *of* $C\beta_r(\tau)$ *is defined by:* $\varphi_r(\tau) = v_r(\tau).d_r(\tau)$. ■

Definition 6 *The* marking of a TB-place *at time* τ *is a series of controllable batches. If* $p_i \in P^{TB}$ *then* $m_i = \{C\beta_h, \cdots, C\beta_r\}$.

Definition 7 *A controllable batch* $C\beta_r(\tau) = (l_r(\tau), d_r(\tau), x_r(\tau), v_r(\tau))$ *of TB-place* p_i *with* $\gamma(p_i) = (V_i, d_i^{max}, S_i, \Phi_i^{max})$*, where its head position equals to the length of* p_i*, i.e.,* $x_r(\tau) = S_i$*, is called an* output controllable batch*, denoted* $OC\beta_r(\tau)$*. The* output density*,* $d_i^{out}(\tau)$*, of a TB-place is defined as follows. If at time* τ*, TB-place* p_i *has an output controllable batch* $OC\beta_r(\tau)$*, then* $d_i^{out}(\tau) = d_r(\tau)$*, else* $d_i^{out}(\tau) = 0$. ■

All controllable batches composing the marking of a TB-place must respect the triangular flow-density relation (see eq.3). This condition allows us to define states of controllable batches.

Definition 8 *(States of batches) Let* $C\beta_r(\tau) = (l_r(\tau), d_r(\tau), x_r(\tau), v_r(\tau))$ *be a controllable batch of TB-place* p_i*, with* $v_i(\tau)$ *variable speed and* V_i *maximum speed of* p_i *(*$v_i(\tau) \leq V_i$*).*

- $C\beta_r$ *is in a* free state *if its density is lower or equal to the critical density of* p_i*:* $d_r(\tau) \leq d_i^{cri}(\tau)$*;*
- $C\beta_r$ *is in a* congested state *if its density is greater to the critical density of* p_i*:* $d_r(\tau) > d_i^{cri}(\tau)$. ■

More details concerning the continuous-time dynamics of controllable batches inside a TB-place are explained in (Gaddouri, Brenner, and Demongodin 2016).

3. SIMULATION

The simulation of a CTBPN model is based on a discrete event approach with linear or constant continuous evolutions between timed events. Between two timed events, the state of the model has an invariant behavior state (*IB-state*). The state of the system is calculated only when it undergoes discontinuity. Three kind of events change the *IB-state*.

- Internal events (timed events inside a TB-place)
  - i.1 a batch becomes an output batch $C\beta_r = OC\beta_r$;
  - i.2 two batches meet;
  - i.3 an output batch is destroyed $OC\beta_r = \mathbf{0}$.
- External events
  - e.1 a discrete transition is fired: t_j;
  - e.2 a continuous place becomes empty: $m_i^n = 0$;
  - e.3 a discrete transition becomes enabled $m_i^n = a$;
  - e.4 a batch becomes an output batch (i.e. event i.1 above);
  - e.5 an output batch is destroyed (i.e. event i.3 above);
- Controlled events
  - c.1 the flow of a batch transitions is modified: $\phi_j(\tau) = \phi'_j(\tau)$;
  - c.2 the speed of a BB-place is modified: $v_i(\tau) = v'_i(\tau)$.

The dynamic evolution of the model is computed by the algorithm in Figure 3.

4. CASE STUDY: HIGHWAY

We consider in our case study the highway section between Marseille and Aix-en-Provence of the A51 highway in France.

4.1. Physical Model

The highway section studied has $1.207km$ long. The section start at the kilometer 2.3 and end at 3.507. The section has no input or output ramps, maximum speed is limited to $90km/h$ and only one direction traffic from Aix-en-Provence to Marseille will be modeled here.

Figure 4 represent this road section.

The data are collected by inductive loops at three points: at the begin of the section (B1), $0.9km$ away (B2) and at the end of the section (B3) for $1:36$ hours starting October, 15, 2015 at $15:58$. The inductive loops in each points send the flow (vehicles/hours), density, average speed and the state (congested or free) each six minutes.

The initial state of the section is a congestion between the point B1 and B2 and free traffic between B2 and B3.

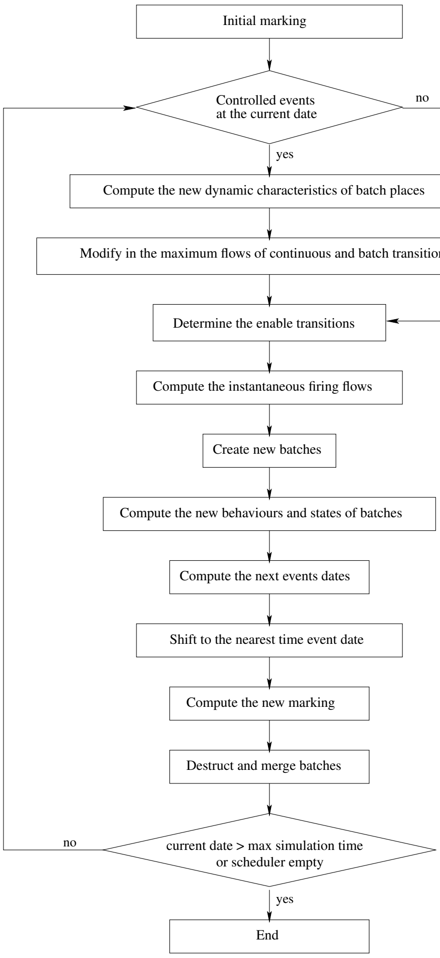


Figure 3: Dynamics of a triangular Batches Petri Nets

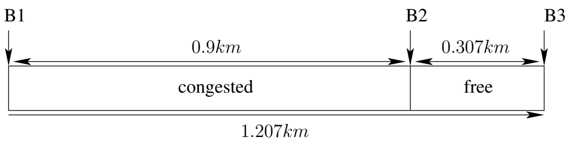


Figure 4: Road section physical model

4.2. CTBPN Model

The CTBPN model (Figure 5f this highway section is composed of two places : continuous place p_1 used to model the section capacity and triangular batch place p_2 to model the section. The characteristics of the place p_2 are computed based on observed flow, speed and number of ways of the road section.

The critical density of this road section is $\frac{4450}{90} = 49.44$ and we can define the initial marking of the place p_2 with two batches. The first batch $OC\beta_1 = (0.307, 34.33, 1.207, 90)$ is in free state. The second batch $C\beta_2 = (0.9, 122.58, 0.9, 31)$ is in congested state.

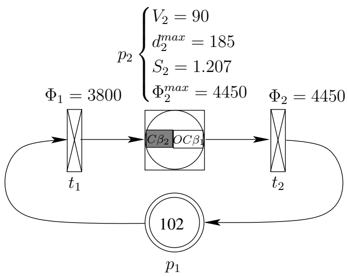


Figure 5: CTBPN model

4.3. Simulation Results

Ours interest here is to compare the output flow measured at the point B3 and the simulation results (flow at the transition t_2). All long the simulation, the variations of the input flow (measured by B1) is applied also in our model. We use controlled events to increase and decrease the maximum flow of the transition t_1 .

The simulation results are presented in Figure 6.

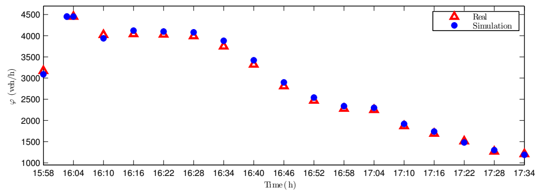


Figure 6: Output flow in decongestion behaviour

We can see at the beginning of the simulation an important increase of the output flow. This flow increasing (from 3090 to 4450 veh/h) is due the decongestion of the section between the points B1 and B2. We can also determine the exact instant when the congestion is finished. This instant is represented by the second round point (simulation) at $16:03$ hours in the graph.

The maximum error measured between the simulation results and the collected data are 3.20% and can be justified by the no respect of the maximum speed that was observed at the collected data.

5. CONCLUSION

Firstly, we presented the CTBPN formalism that was used to model the highway section. A schematic algorithm based on discrete events approach was shown in section 3.

The main contribution of the paper is the case study of a highway section. We proposed a simple CTBPN model that describe a physical highway section. The parameters of the model used in the simulation were extracted from a set of collected data.

The comparison between the simulation results and collected data show a good accuracy between our model and behaviour observed in the traffic road.

REFERENCES

Daganzo C., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological, 28(4):269-287.

Demongodin I., 2001. Generalised Batches Petri Net: Hybrid Model For High Speed Systems With Variable Delays. Discrete Event Dynamic Systems, 11(1-2):137-162.

Demongodin I., 2009. Modeling and Analysis of Transportation Networks Using Batches Petri Nets with Controllable Batch Speed. Proceedings of the Conference on Applications and Theory of Petri nets LNCS, 5606:204-222.

Demongodin I. and Giua A., 2014. Dynamics and steady state analysis of controlled Generalized Batches Petri Nets. Nonlinear Analysis: Hybrid Systems, 12:33-34,

Gaddouri R., Brenner L., and Demongodin I., 2014. Extension of batches petri nets by bi-parts batch places. Proceedings of the Workshop on Petri Nets for Adaptive Discrete-Event Control Systems (ADECS), pp. 83-102. June 24, Tunis (Tunisia).

Gaddouri R., Brenner L., and Demongodin I., 2016. Controlled triangular batches petri nets for hybrid mesoscopic modeling of traffic road networks under VSL control. Proceedings of the Conference on Automation Science and Engineering (CASE), pp. 427-432. August 21-24, Fort Worth (Texas, USA).

Lighthill M. and Whitham G., 1955. On kinematic waves. ii. a theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London, 229(1178):317-345.

SMART MAINTENANCE LIFECYCLE MANAGEMENT: A DESIGN PROPOSAL

Jan Zenisek [(a)], Letizia Nicoletti [(b)], Francesco Longo [(c)], Gabriele Traugott [(d)], Antonio Padovano [(e)], Michael Affenzeller [(f)]

[(a),(d),(f)] University of Applied Sciences Upper Austria, Austria
[(c),(e)] DIMEG, University of Calabria, Italy
[(b)]CAL-TEK SRL, Italy

[(a)] jan.zenisek@fh-hagenberg.at , [(b)] l.nicoletti@cal-tek.eu , [(c)] f.longo@unical.it ,
[(d)] gabriele.traugott@fh-hagenberg.at , [(e)] antonio.padovano@unical.it , [(f)] michael.affenzeller@fh-hagenberg.at

ABSTRACT

Given the latest trends toward the implementation of Industry 4.0 principles, the proposed research work combines human assistance technologies such as Virtual and Augmented Reality with advanced data analysis techniques and tools to develop a comprehensive strategy for Predictive Maintenance planning and execution. On one hand, using Augmented and Virtual Reality technologies, workers are effectively assisted during maintenance operations towards better performances and lower error rates. On the other hand, Predictive Maintenance entails strategies for maintenance planning based on machines' current conditions to avoid unnecessary overhauls.
Thus a seamless integration of Virtual and Augmented Reality with Predictive Maintenance is envisaged to bring substantial advantages in terms of productivity and competitiveness enhancement for manufacturing systems and represents a step ahead toward the real implementation of the Industry 4.0 vision.

Keywords: Industry 4.0, Virtual / Augmented Reality, Data Stream Analysis, Predictive Maintenance, Symbolic Regression

1. INTRODUCTION

Industry 4.0 is characterized by an unprecedented interconnection of things over the Internet, which brings the physical and the virtual world together. Cyber-Physical Systems (CPS) integrate physical devices (i.e. sensors) with software components, thus providing the availability of an immense amount of sensor data describing the condition of products, machines etc. The challenge for manufacturing companies is to obtain substantial benefits from analysis of the collected data in various application areas such as production planning, quality management, maintenance and so on.
Predictive Maintenance (PdM) means strategies to plan maintenance actions based upon a machine's current condition instead of reacting to breakdowns (corrective maintenance) or following fixed empirically intervals (preventive maintenance) as stated by Li, Wang and He (2016). Therefore, sensors, attached to a machine, keep track of its behavior over the time and a subsequent analysis component evaluates the resulting data stream in order to prognose factual maintenance needs and determine an optimal moment to trigger correcting actions. Hence, PdM aims at preventing machine breakdowns without performing unnecessary overhauls, leading to higher productivity and increased predictability.
Using Augmented Reality (AR) / Virtual Reality (VR) technologies, workers are effectively assisted during maintenance work – all the more so as using a common data base for PdM and assistance systems in maintenance. Wang et al. (2017), for example, propose a cloud based approach for PdM. Closing the loop of the maintenance life cycle leads to increased productive time of machines and thus contributes positively to productivity and competitiveness of manufacturing companies.
This work proposes a blueprint of how such a life cycle could look alike for real-world implementations, by combining cutting-edge data analysis and assistance technologies.

2. VIRTUAL / AUGMENTED REALITY TECHNOLOGIES FOR SMART OPERATORS

In the context of smart manufacturing systems, operators play a crucial role for the optimal integration of virtual and real assets. It is mainly due to the fact that, operators are expected to be fit from the knowledge generated by and within CPS while providing valuable inputs/feedbacks that are likely to drive CPS toward higher levels of intelligence.
In this perspective the smart operator concept is considered an enabling factor for a practical implementation of the Industry 4.0 vision. Smart operators are endowed with a superior knowledge of the working environment deriving not only from the abilities they acquire while executing daily tasks but

also from the interaction with high value-added contents/tools that contribute to enhancing his ability to perceive, understand and act in the workplace. Such contents/tools are mainly attributable to virtual and augmented reality. As a matter of facts, virtual and augmented reality contents are able to provide different levels of immersion and therefore to engage operators in unique experiences where real and digital objects are intertwined. As well known, augmented reality technologies allow keeping the actual view of real objects/systems while adding further contents levels on top of them whereas virtual reality allows recreating digital twins and highly interactive objects. These basic features make AR and VR well suited to achieve an optimal integration between real words objects and cyber resources in Industry 4.0 environments. Hence, within the scope of this research work, a substantial effort has been done toward the definition of an application methodology/approach that leverages on AR and VR as enabling factors for smart operators in smart factories with a special focus on maintenance operations. In particular, the basic idea behind the proposed approach/methodology is to support operators in complex, real time man-machine interactions that usually occur during maintenance operations providing visual, self-explanatory information on how to execute a specific task/procedure as well as demonstrating the best interaction patterns. The main building blocks of the proposed approach include:

1) Develop geometric models recreating real objects such as machines and equipments in the working environment;
2) Build Virtual environments recreating a typical or particular manufacturing system;
3) Dynamical integration of geometric models within the relevant virtual environment;
4) Identify and recreate each object dynamical behavior;
5) Set up interconnections with real-time data sources (such as sensors networks) to recreate within the virtual word the operational conditions of the real word system;
6) Procedures mapping, analysis and 3D reconstruction.
7) Knowledge resources digitalization and organization.
8) Dynamic binding of knowledge digital assets to real objects in the workplace for AR / VR contents delivery.

The practical implementation of the proposed methodology turns out into a working tool whose main features/functionalities can be summarized as follows:

- let operators be immersed and interact with a cyber space where they can gain meaningful insights, based on virtual and augmented reality, about man-machine interaction procedures for maintenance operations compliant with safety standards and principles;
- exploit virtual and AR resources for operators' preliminary training on high risk tasks in maintenance operations;
- support operators providing information that is usually not available in the workplace (i.e. expected maintenance operations, warning on unexpected dangers, risks that are likely to occur, suggestions on how to increase productivity, etc) as well as operator's training;
- send warning messages about the outcomes of improper operations (i.e. what happens if a maintenance operation is not performed, if the operator fails, etc).

Keeping in mind the need to preserve operational efficiency without hindering operators' workability, the many possibilities offered from wearable technologies have been investigated to detect the most suitable system configurations. As a consequence it can be deployed and integrated with different fruition technologies such as: tablets, headsets, interactive whiteboards, monocular eyewear, smart glasses, armbands and gesture recognition technologies as shown in Figure 1 and Figure 2.

Figure 1: Augmented reality contents delivered through mobile devices


Figure 2: Augmented Reality contents delivered through monocular eyewear

3. DATA BASED PREDICTIVE MAINTENANCE

Real world data series, collected in a PdM scenario as depicted in Section 1 are presumably quite complex to analyze. Some of these generated features might have correlations to others, which leads to redundancies and could bias analysis results. Moreover, it is most likely that not all relevant system aspects are covered by sensors, which also complicates a reasonable analysis. Another challenge refers to the high volume of a data stream, which is updated e.g. every second to represent a system's inner workings correctly (Gubbi et al. 2013), but therefore, computationally demanding to analyze. Furthermore, sensor values can go missing or get skewed while they're transferred from one to another component in a PdM setup as proposed in Section 4. In addition, fault system behavior might not be manifestable in single sensor series, but in their changed interplay (Saxena et al. 2008). Hence, the transgression of sensor thresholds, might not suit as a trigger for maintenance in a robust decision support system, since these thresholds only deal with one system dimension at a moment.

In order to address these considerations, first a preprocessing component, consisting of filters and routines for missing value handling, signal smoothing and feature extraction has to be setup. Based on the preprocessed series the actual stream analysis is performed in a second component downstream, which models the inner workings and output of the monitored system. Recent works on smart maintenance solutions, propose machine learning methods like Random Forests (Scheibelhofer et al. 2016) and Hidden Markov Models (Cartella et al. 2015) to build models in order to identify a tracked system's behavior as close as possible.

In this work we focus on Symbolic Regression models, developed with Genetic Programming (GP), a method applicable to machine learning. Symbolic Regression models suit well to describe systems which incorporate biological or physical processes, like the industrial production plants in a PdM implementation. The models are mathematical functions, combining terminals like variables and constants with a broad palette symbols, such as arithmetic, trigonometric or conditional operators. The GP algorithm develops a population of these functions, which are usually represented as syntax trees, following evolutionary concepts like parental selection, crossover and mutation, in order to find a good estimator for the observed system. One major advantage of Symbolic Regression models compared to other representations is that they are interpretable for domain experts, which might enable them to gain deeper insights (Affenzeller et al. 2009).

4. SMART MAINTENANCE

Based on current research for innovation of maintenance in industrial applications, using human assistance technology on one end and data analysis on the other, the idea of joining these two approaches towards a smart PdM adaption arose.

4.1. Related Work

In the current movement of Industry 4.0, several recently presented projects deal with modernizing maintenance towards PdM. In the work of Li, Wang and He (2016) the interdependencies of *IoT*, *Cyber-Physical Systems* and *Big Data Analytics* in the context of Industry 4.0 are highlighted and a PdM setup, using these technologies and methods, is briefly outlined. Sayed, Lohse and Madsen (2015) present a component based reference architecture for PdM implementations, using a very similar mix of IT, in greater detail. The work presents the overall architectural approach and describes all employed components on a software level, down to activity diagrams. Furthermore, Wang et al. (2017) present a mobile agent based paradigm for PdM, including implementation details and an experimental study on induction motors.

The potentials of AR-based knowledge sharing in industrial maintenance are explored in Aromaa et al. (2018). Here a field study has shown positive users' experiences as well as a good level of acceptance of AR systems. Similarly Uva et al. (2017) evaluate the effectiveness of augmented reality for conveying technical instructions. A seven-task maintenance procedure on a motorbike engine has been considered as case study and the outcomes of the study confirmed a performance improvement and lower error rates especially for difficult tasks.
Interesting applications can be found in Alam et al. (2017) where a AR/VR IoT prototype for maintenance tasks in complex working environments is introduced; in Kranzer et al. (2017) where in Intelligent Maintenance planner to provide augmented reality information are designed; in Jayaweera et al. (2017) that propose an AR wearable system to enhance the capabilities of machine operators and repairmen; in Akbarinasaji and Homayounvala (2017) that present an optimized framework which combines context-awareness and AR for training and assisting technicians in maintaining equipment to improve field workers effectiveness in an industrial context; and many others.

A state of the art on AR in maintenance can be found in Palmarini et al. (2018) highlighting also the areas where AR technology still lacks of maturity that are mainly related to reliability and robustness.

Although data analysis for PdM as well as AR/VR have been addressed in several projects in the context of industrial maintenance, to the best of our knowledge there is none proposing an architecture to combine them. Even if an interesting attempt to merge PdM and AR can be found in Mourtzis et al. (2017) that consider condition-based preventive maintenance supported by augmented reality and smart algorithms. However, unlike the approach we propose in this research work, little attention is paid to VR contents and the system is devised to work real-time with no possibility to be used for off-line training.

4.2. Maintenance Lifecycle

This work aims at proposing a general concept of how data analysis methods and AR/VR technologies could be employed beneficially to develop a comprehensive smart maintenance strategy. Starting with a monitored industrial production plant, Figure 3 illustrates how information, initially in form of sensor data, is transferred and transformed by cloud services to analysis software on a high performance cluster. Based on this system's output, concrete tasks are derived and sent to tablets and head mounted virtual/augmented reality glasses used by maintenance operators. These people close the depicted cycle when performing the generated instructions on the real-world plant. Thus, the figure presents how the pathway of data from acquisition, over digital transformation to its physical reintegration can be designed. We refer to this proposed design as *Smart Maintenance Lifecycle*.

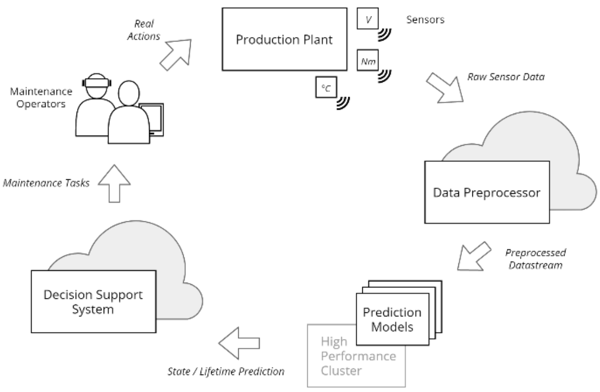


Figure 3: Smart Maintenance Lifecycle

Before the workflow incorporated by the presented lifecycle can be performed, all components have to be setup on the base of qualified test data series, each representing known states – i.e. normal and defective behavior – of the monitored system. During a second phase the installed filter rules, trained prognosis models and developed decision trees are evaluated on unseen sensor data stream. In the following, purpose and tasks of all components during these phases are described:

4.2.1. Production Plant

The production plant represents the maintenance lifecycle's start and end point. In the initial setup phase, it is equipped with condition monitoring sensors, which are generating data series in high frequency. As part of an IoT implementation, each sensor is individually connected to the cloud and continuously transferring the sensed values to the subsequent service. On the other end, the plant receives maintenance actions, performed by human operators, eventually on the base of its self-produced data. Hence, the production plant is source of digital information and sink for physical actions.

4.2.2. Data Preprocessor Service

This cloud service aggregates sensor data to a dense, equidistance stream, which makes it reasonably analyzable. The component filters outliers and signals with too many missing values, interpolates missing values, levels out the impact of noise and conclusively transfers the preprocessed data stream to the prediction models. Therefore a pipeline of rules and filters is developed in the initial phase.

4.2.3. Prediction Models

As outlined in Section 3 the model creation process is guided by the GP algorithm so that domain experts only have to provide qualified data sets. In our design approach we propose to employ a High Performance Cluster (HPC) for the initial training and the subsequent evaluation of Symbolic Regression ensemble models. These prediction models enable to analyze the preprocessed data stream considering two different aspects:

- Prediction of Remaining Useful Lifecycles (RUL) for the production plant's components (Saxena et al. 2008)
- Detection of abnormal or unknown system behavior and therefore, the need for checks and possibly maintenance

The emerging estimations are subsequently transferred to the cloud based decision support system for final reasoning. The necessity to deploy the model training algorithms on a HPC infrastructure origins from the complex task to analyze a great amount of high dimensional test data series. Furthermore, the infrastructure also speeds up the real-time evaluation of the created models in the second phase, especially when employing not only one, but large ensembles of models.

4.2.4. Decision Support System

Based on the previously made estimations regarding RUL and anomalies, this component reasons if, what and when actions should be triggered in order to prevent or correct an allegedly encountered defect. Hence, this component holds accountable for maintenance scheduling, which might not be a computational demanding task, but one, which is predestinated to run at the easy accessible and resilient position in the cloud. The decision support system consists of a set of hierarchically connected rules, formulated by domain experts. These *decision trees* evaluate the previous system's output so that the most proper concrete maintenance task is generated and textual and graphical content from a knowledge database to the subsequent system is forward.

4.2.5. AR/VR Technology and Operators

The maintenance operators, equipped with tablets and head mounted virtual/augmented reality glasses receive the decisions in form of instruction material from the connected DSS cloud service. First, the operator is acoustically and visually informed about the issue and subsequently guided to the respective machine. The necessary maintenance actions, construction plans and (video-) tutorials etc. are displayed in order to support the service process optimally. The physical actions, which the operator performs, close the smart maintenance lifecycle. In greater detail, after contents are displayed, the operator can freely interact with them i.e. zoom in and zoom out of 3D representations, read text info, interact with virtual reality and/or augmented reality contents, etc.

Moreover, as mentioned above, operators have access to 3D animations explaining safety and maintenances procedures required along the period of operation. Each procedure is broken down into steps and each step is visually shown through an augmented 3D virtual animation and vocally explained (as shown in Figure 4). Along the explanation, visual and vocal messages draw the attention on potential dangers or risks the operator may incur.

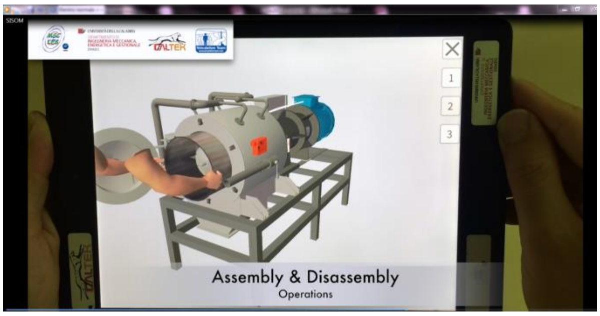


Figure 4: A sample of 3D animation for maintenance procedures

Besides, the operator can even search for additional contents typing keywords in a text search box.

This approach can be applied both real-time and offline for a variety of purposes:

- make available knowledge resources that usually may not be directly available in the workplace;
- be a consultation mean for operators' performances improvement thanks to easy and fast access to information contents;
- be an immersive and absorbing environment for preliminary training on new and/or complex procedures for maintenance operations;
- be an instrument for safety and security enhancement.

4.3. Continuous Adaption

In order to adapt current trends and to validate the applied maintenance actions and their effect, we propose to repeat the initial setup phase occasionally. Optimally, the phases work interleaved, so that single components are updated according to their performance. We propose to trigger an update if the quality of a component's output starts to decrease significantly. For instance, maintenance decision rules should be altered if the system triggered wrong or not necessary actions, or the adopted actions did not generated the expected results. On the other end, the preprocessing rules as well as the prognosis models should be updated after some cycles in order to react on changing conditions.

To detect the necessity for updates, a comprehensive maintenance validation workflow shall be employed: If the actions proposed by the decision support system do not fit the actual needs of the monitored plant, according to a hands-on analysis by a human operator, this information shall flow back and update the rule base automatically. We propose that the operator uses the assistance technology – i.e. tablets or AR/VR glasses – to respond her own impression by entering an estimation for the remaining useful lifecycles into a digital form or by using a voice recognition module. The modification of rules could be automated by penalizing poor decisions so that they are executed less likely. A gradual or more abrupt increasing frequency of such wrongful decisions might indicate, that the monitored system is under change. Hence, the decision support system should trigger the training of new prediction models on the base of more recently collected data series, if a certain threshold of a "poor-decision-counter" is exceeded. A similar routine could be applied to the preprocessor component. This continuous adaption process manages updating the components so that the maintenance lifecycle remains smart.

5. CONCLUSION

Within the scope of this research work, a substantial effort has been done toward the definition of an application methodology/approach that combines AR and VR technologies with data analysis methods as enabling factors for smart operators in smart factories to develop a comprehensive smart maintenance strategy. The practical implementation of such strategy turns out into a working approach that can be deployed at operational levels in order to:

- plan maintenance actions based upon a machines current conditions;
- deliver virtual and augmented reality contents on how and when procedures for maintenance operations are to be executed;
- exploit virtual and AR contents for operators' preliminary training and performance improvement with respect to high risk tasks in maintenance operations;
- support operators with information that is usually not available in the workplace (i.e. expected maintenance operations, warning on unexpected dangers, risks that are likely to occur, suggestions on how to increase productivity, etc)

ACKNOWLEDGMENTS

The work described in this paper was done within the projects:

“Smart Factory Lab” which is funded by the European Fund for Regional Development (EFRE) and the state of Upper Austria as part of the program “Investing in Growth and Jobs 2014-2020”.

“SISOM” funded by the Italian National Institute for Insurance against Accidents at Work (INAIL) under the BRIC 2015 program.

“SG-ICT: Serious Game at Increased impact on Culture and Tourism” funded by Ministry of Education, Universities and Research (MIUR) under the StartUP program.

REFERENCES

Affenzeller M., Wagner S., Winkler S., Beham A., 2009. Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications. Chapman & Hall / CRC Press.

Akbarinasaji, S., Homayounvala, E. (2017) A novel context-aware augmented reality framework for maintenance systems (2017) *Journal of Ambient Intelligence and Smart Environments*, 9 (3), pp. 315-327.

Alam, M.F., Katsikas, S., Beltramello, O., Hadjiefthymiades, S. (2017)Augmented and virtual reality based monitoring and safety system: A prototype IoT platform (2017) *Journal of Network and Computer Applications*, 89, pp. 109-119.

Cartella F., Lemeire J., Dimiccoli L., Sahli H., 2015. Hidden Semi-Markov Models for Predictive Maintenance. Mathematical Problems in Engineering, 2015, 1-23.

Gubbi J., Buyya R., Marusic S., Palaniswami M., 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29 (7), 1645-1660.

Jayaweera, M., Wijesooriya, I., Wijewardana, D., De Silva, T., Gamage, C. (2017) Demo abstract: Enhanced real-time machine inspection with mobile augmented reality for maintenance and repair (2017) *Proceedings - 2017 IEEE/ACM 2nd International Conference on Internet-of-Things Design and implementation*, IoTDI 2017 (part of CPS Week), pp. 287-288.

Kranzer, S., Prill, D., Aghajanpour, D., Merz, R., Strasser, R., Mayr, R., Zoerrer, H., Plasch, M., Steringer, R. (2017) An intelligent maintenance planning framework prototype for production systems (2017) *Proceedings of the IEEE International Conference on Industrial Technology*, art. no. 7915520, pp. 1124-1129.

Li Z., Wang K., He Y., 2016. Industry 4.0–Potentials for Predictive Maintenance. Atlantis Press. Available from: http://www.atlantis-press.com [accessed 22 May 2017]

Mourtzis, D., Vlachou, A., Zogopoulos, V. (2017) Cloud-based augmented reality remote maintenance through shop-floor monitoring: A product-service system approach (2017) Journal of Manufacturing Science and Engineering, Transactions of the ASME, 139 (6), art. no. 061011, .

Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.(2018) A systematic review of augmented reality applications in maintenance (2018) *Robotics and Computer-Integrated Manufacturing*, 49, pp. 215-228.

Saxena A., Goebel K., Simon D., Eklund N., 2008. Damage Propagation Modeling for Aircraft Engine Run-To-Failure Simulation. Prognostics and Health Management, 2008. Proceedings of 2008 IEEE International Conference on PHM, pp. 1-9.

Sayed M. S., Lohse N., Madsen A. L., 2015. SelSus: Towards a reference architecture for diagnostics and predictive maintenance using smart manufacturing devices. Proceedings of 2015 13th IEEE International Conference on Industrial Informatics (INDIN), pp. 1700-1705. July 22-24, Cambridge (UK).

Scheibelhofer P., Gleispach D., Hayderer G., Stadlober E., 2016. A Methodology for Predictive Maintenance in Semiconductor Manufacturing. Austrian Journal of Statistics, 41 (3), 161-173.

Wang J., Zhang L., Duan L., Gao R. X., 2017. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. J. Intelligent Manufacturing, 28 (5), 1125-1137.

HOW HUMAN FACTORS AFFECT THE OUTCOMES OF DISASTERS MANAGEMENT ON INDUSTRIAL SITES

Francesco Longo[(a)], Letizia Nicoletti[(b)], Antonio Padovano[(c)], Jean Cazorla[(d)], Marco Vetrano[(e)], Alessandro Chiurco[(f)], Caterina Fusto[(g)], Luigi Bruno[(h)], Agostino Bruzzone[(i)], Marina Massei[(l)]

[(a) (c) (d) (f) (g) (h)]DIMEG, University of Calabria, Italy
[(b) (e)]CAL-TEK Srl
[(i) (l)]DIME, University of Genoa

[(a)]f.longo@unical.it, [(b)]l.nicoletti@cal-tek.eu, [(c)]antonio.padovano@unical.it, [(d)]jean.cazorla@msc-les.org, [(e)]m.vetrano@cal-tek.eu, [(f)]a.chiurco@unical.it, [(g)]c.fusto@msc-les.org, [(h)]l.bruno@msc-les.org, [(i)]agostino@itim.unige.it, [(l)]massei@itim.unige.it

ABSTRACT
Modeling & Simulation plays a crucial role in the field of Industrial Disaster Management, especially because its outcomes are profoundly affected by human factors. The present research intends to show quantitatively how human behavior affect the outcomes of disasters management on industrial sites. A case study of a fire incident in an Ecuadorian public company engaged in the exploration and extraction of hydrocarbons is proposed. Pareto analysis combined to main effect and interactions effects plots clearly show to the reader how different factors (mostly related to the personnel's behavior and skills, interactions among workers and with the environment) affect the disaster evolutions and its outcomes in terms of loss of human lives, number of injured workers and evacuation times.

Keywords: Simulation, Human Behavior, Human Factors, Industrial Disaster, Emergency Management

1. INTRODUCTION

Most important disasters happened worldwide in industrial plants are characterized by common aspects (e.g. rescue entities involved, emergency management procedures used, etc.). However, the analysis of the disaster type, entities involved and emergency procedures is not enough to understand how a disaster usually evolves over the time and its main effects on the people involved. To this end, the human behavior and the human error must be taken into account. As mentioned by Rosenthal et al. (2001), when a disaster occurs, humans introduce randomness and unpredictability. Randomness and unpredictability introduce, in turn, emergent situations (due to the interactions among the people involved in the disaster and between the people and the external environment) that cannot be predicted and explained a-priori. Such emergent situations may strongly affect the evolution of the disaster over the time in terms of number of injured or death people or evacuation time.

As clearly demonstrated by the current literature, simulation (and specifically Multi Agent Systems, MASs) plays a critical role as tool to recreate the inner complexity of an industrial plant (and workers involved) before and after an emergency situation (e.g. a disaster caused by an explosion and/or fire). Bessis et al. (2011) underlines that MASs have to be regarded (above all when integrated with other methodologies) as next generation technologies for disaster management. By moving in this direction, Hashemipour et al. (2017) propose a framework, based on a Multi-Agent Coordination Simulation System, to be used as a decision-support system to help response manager and operations both for man-made and natural disasters. In addition to general multi agent frameworks and review guidelines about how to use MASs, there are a number of research works proposing specific applications and case studies. For instance, Mat et al. (2017) use a multi agent 3D simulation for flood evacuation, Bruzzone (2013) uses intelligent agent-based simulation for supporting operational planning in country reconstruction after a disaster, while Shi et al. (2009) proposes an agent-based evacuation model of large public buildings under fire conditions.

Therefore, the research effort is moving ahead trying to improve our capacity to understand how emergency scenarios evolve in complex systems (e.g. industrial plants) and to improve our preparedness and responsiveness, even coupling (as already mentioned by Bessis et al., 2011) multiple technologies. To this end, Sugie et al. (2018) develop a disaster prevention system where multi agent simulation is jointly used with robots for evacuation guidance. However, in such an evolving context, it is also worth mentioning that a lot has been done to reduce the probability to have disasters and emergencies. In the industry sectors, modern plants are usually high-automated: this drastically increased productivity whilst reduced the risk of accidents (Woods et al., 2010). The risk of accidents is also reduced by an ad-hoc training of the personnel; from this point of view, the Industry 4.0 and the digital revolution are now giving more emphasis to the importance and effectiveness of simulation based solutions for training (already proved in the past, see for instance Bruzzone and Massei 2010; Beroggi et al., 1995). Currently researches show the Virtual and Immersive Reality and Serious Games can be effectively used to support advanced training in different sectors and application areas (Cohen et al., 2013; Crichton, 2009; Davis et al., 2017).

1.1 Contribution of this article

A survey of the industrial accidents along the last years reveals that, despite the effort to tackle the problem from both sides (from one side better understanding of the emergency evolution and management if a disaster occurs and personnel training, from the other side reduction of the risk of accidents), the ideal condition of zero accidents in the industrial sector is still far to be reached (Twaalfhoven and Kortleven, 2016). Therefore, it is imperative to continue the research efforts in all the directions identified above. To this end, this article propose a multi agent simulation model to investigate the after disaster evolution in an industrial plant. In particular, the authors observe how different factors (mostly related to the workers behaviors and skills, interactions among workers and with the environment) affect the disaster evolutions and its outcomes in terms of loss of human lives, number of injured workers and evacuation times.

The remaining is organized as follows. After a brief description of the industrial site scenario (Section 2), section 3 introduces the multi agent system that has been developed with a focus on the human behavior modeling including human interactions with the external environment. Then, section 4 presents the factors and the performance measures taken into account for the simulation experiments and discusses the simulation results explaining how the chosen factors affect the evolution of the disaster and its outcomes.

2. INDUSTRIAL PLANT SCENARIO DESCRIPTION

Petroamazonas EP is an Ecuadorian public company dedicated to the exploration and production of hydrocarbons. It operates 21 plants, 18 located in the eastern region of Ecuador and three in the coastal zone as shown in Figure 1.

Figure 1. Petroamazonas EP plants

As part of the production process, the oil dispatched at the pumping stations initially passes through a filtering process so that solids contained in the fluid do not affect the different types of equipment it will encounter throughout its journey. Afterwards, if necessary, the oil is heated through heat exchangers (furnaces) in order to reduce its viscosity. Finally, the oil is delivered into centrifugal pumps, which provide the necessary energy so the fluid may be delivered to the next pumping station. These centrifugal pumps operate with internal combustion engines that use crude oil as fuel. For the main engines and pumps to operate properly, it is necessary to have backup systems that perform various functions:

- Air compressors for all instruments;
- Treated fuel (filtered and heated);
- Water for engine cooling;
- Power generators;
- Oil metering systems;
- Oily water drainage and treatment systems, and others.

The plant considered in this paper is the Sansahuari pumping station. The Sansahuari station (Figure 2) is part of the Cuyabeno production system and operates 24/365. Among the main operating systems, we find:

- Oil receiving system;
- Water, gas and oil separation system;
- Oil storage system;
- Water storage system training;
- Re-injection water system;
- Oil pumping system;
- Overheating system;
- Fire system.

Figure 2. Sansahuari Pumping Station

Figure 3 shows the main working points in the plant area.

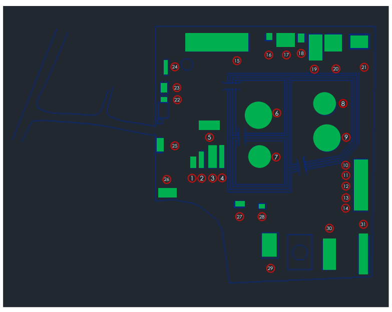


Figure 3. Plant Main Working Points

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

The main professional figures involved with their duties and responsibilities are summarized below:

- Plant Operators;
- Assistant Plant Operators;
- Chemical Engineers;
- Assistant Chemical Engineers;
- Plant Inspectors;
- Security Operator;
- Maintenance Operator;
- Automation Operator;
- Integrity Operator.

The **Plant Operator** performs all the basic operations and control of the various processes including reception, storage and transfer of petroleum to the central station. He is also in charge of operations synchronization according to the needs of recipients. Therefore, he acts on start-up and shutdown of machines, equipment and facilities while maintaining high levels for safety, quality and environmental conditions. Finally plant operators take care of review of oil production and transfer reports for the preparation of general reports.

The **Assistant Plant Operator** supports the work performed by the plant operator; in particular, control and monitoring of machineries and equipment, production reports and petroleum transfer.

The **Chemical Engineer** takes under control chemical processes during reception, storage and transportation of oil. The Chemical Engineer is also in charge of performing chemical analysis to keep under control the quantities of chemical substances to be injected into the different processes.

The **Assistant Chemical Engineer** supports the work done by the Chemical Engineer, in particular process control and sampling. He also conducts tests under the supervision of the Chemical Engineer.

The **Security Operator** is in charge of controlling public goods for their correct use; he carries out surveillance activities including control of work and safety standards.

The **Maintenance Operator** carries out predictive maintenance on machines and equipment in order to detect anomalies before failures occurences. He coordinates, plans and manages the maintenance activities also executing fault maintenance when needed.

The **Automation Operator** executes preventive and corrective maintenance on instruments including the control of links, interconnections and control loops for rooms and stations.

The **Integrity Operator** executes productive maintenance (non-destructive tests) in facilities and static equipment; he takes under control the distribution of fluids and pressures to ensure the safe operation of the petroleum transfer.

2.1 Emergency Procedures in case of incident

In order to combat unwanted events such as fires, in the operational areas of the Sansahuari station, Fire Systems are available to effectively combat events that may occur within the locations. Figure 4 shows the main coverage of the Plant Fire System.

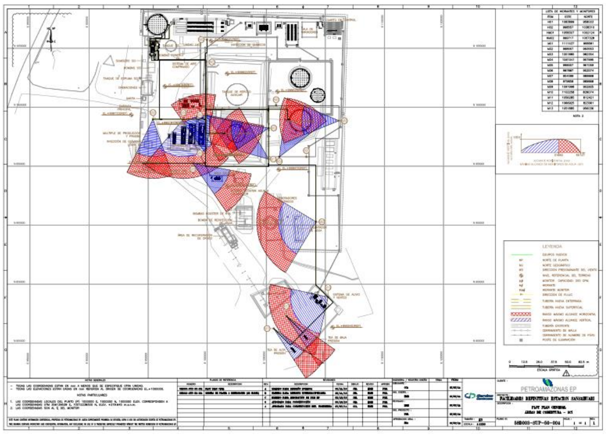


Figure 4. Plant Fire System Coverage

The purpose of the fire emergency response plan is to establish a structured fire control organization at the Sansahuari Station Production Facilities, as well as define roles and functions of firefighters.

The procedure used at the Sansahuari Station for detecting a fire at production facilities is summarized below:

1. The person who detects a fire should immediately notify the Facility Control Room and try to fight the fire only if it has the appropriate resources and if there are minimum safety conditions (fireman's suit, fire extinguisher, PPE, etc.), otherwise it will wait for the arrival of the internal emergency team.
2. The Control Room Operator shall assess the need to activate the Emergency Shutdown System and immediately notify the Operations Supervisor and the Monitoring Center. The person must provide the following information:
    - What is happening;
    - Fire Location;
    - Number of injured (if any).
3. The Plant Supervisor shall immediately notify (in the order given below) the following people to activate the general emergency alarm if necessary:
    - Field Manager;
    - Operations Superintendent;
    - Safety and Security Superintendent;
    - Maintenance Superintendent;
    - Constructions Superintendent.
4. When the general emergency alarm is activated, employees should go to the safety zones while the Rescue and First Aid Team and the internal firefighter team should get ready for starting their emergency management activities. In the meanwhile, the Emergency Manager Coordinator is informed on the magnitude and conditions of the fire, whether or not there is an associated spill and the need to rescue injured personnel.
5. According to the information received, appropriate response activities will start; materials, containment and control equipment will be directed to the

specific site designated by the Emergency Coordinator to start controlling the emergency.

The actions to be performed in the injection area are summarized below:

- Suspend in a safe way the operations that are still running at the time of the event.
- Re-direct the people to the designated meeting points.
- Secure the area.
- Identify the cause of the fire
- Do not enter the fire area without proper personal protective equipment, including self-contained breathing apparatus.
- Disconnect power supplies (close pipe valves, disconnect switch from electrical appliances) near the fire point.
- Identify the area where the fire starts.
- Locate the nearest extinguishers (do not use directly water jets) or fire systems and activate them
- Quickly isolate the area.
- If there are tanks exposed to the fire, use water spray to cool them.

3. HUMAN BEHAVIOR SIMULATION

Multi-agent based systems are particularly suitable for simulating human individual cognitive processes and behaviors in order to explore emergent macro phenomena such as social or collective behaviors. Therefore the authors have developed a multi-agent simulation model to recreate human behavior within an industrial plant in case of emergency.

As part of the simulation model, each human individual is modeled as an autonomous agent who interacts with a 3D Virtual Environment and other agents according to an Individual Behavior Model and some global rules and crowd dynamics rules that derived at the levels of interactions among individuals and group.

3.1 The 3D Virtual Model of the Sansahuari Plant

The purpose of this part is to produce 2D and 3D geometries representing the physical environment of the Sansahuari Plant. Starting from the Sansahuari Plant lay-out, the 2D representation (see Figure 6) and 3D representation (see Figure 6) of the plant has been derived. Elaborating more on the geometries, the 3D models of the most important equipment, machines and components of the plant have been developed.

The 3D Virtual Environment has been also equipped with multiple points of view and the possibility to move around, fly, and observe people behavior during the simulation. A 3D representation of the zone interested by a fire has been also created (see Figure 7) by using Fire Severity Zones circles.

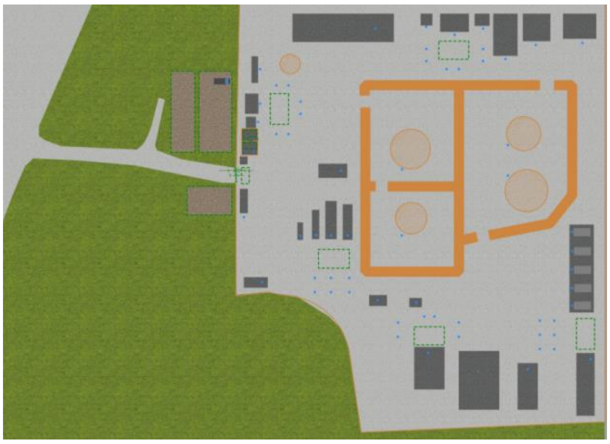


Figure 5. 2D representation of the Sansahuari Plant

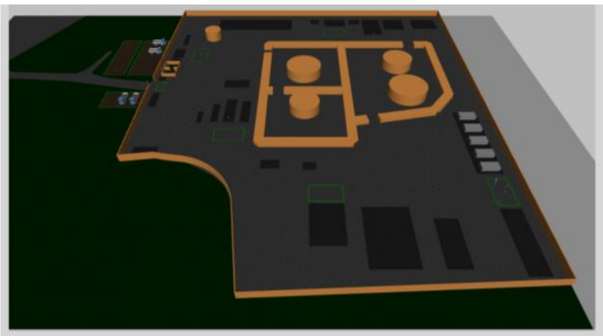


Figure 6. 3D representation of the Sansahuari Plant

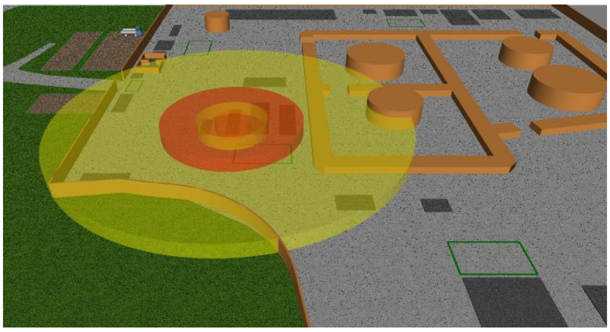


Figure 7. 3D View of the plant and representation of the area interested by the fire

3.2 Personnel Behavior Modeling

According to the information reported in section 2, the personnel working in the plant have their own duties (in the case of normal operations) and should follow specific procedures and action plans in emergency situations. As we have 12 different professional figures working in the plant, 12 different agents have been implemented in the simulation model, each one with its own behavior and designed tasks. Each agent has been programmed individually in order to set-up correctly tasks starting times, executions and durations. In addition, the use of the state charts allow simulating the behavior of the personnel during the emergency, the actions to take, interaction with the environment and the interaction with other people involved in the simulation.

As far as the behavior of the personnel during the emergency is concerned, the following behavior have been implemented as part of the simulation:

- Individual behavior;
- Interaction with the environment;
- Group Behavior.

The individual behavior depends on the position of the person at the time of the event (e.g. at the time of the explosion causing the fire). The assessment of distance determines whether the person is dead, injured or safe according to a Fire Severity Zones circles approach (see Figure 7). The radius values for the death zone, injured zone and safe zone depend on the explosion and fire type; this variables have been set-up as parameters of the simulation model (to give the user the possibility to simulate multiple scenarios and carrying out what if analysis).

Once the distances are determined and evaluated for each person, the evacuation process begins. As far as the behavior of the personnel during the evacuation is concerned, two major approach have been used:

- a Queueing Behavior: people go to the safety zones in an orderly manner respecting distances, speeds and security actions
- a Competitive Behavior: people try to leave as soon as possible without respecting safety standards, this is due to the high stress, fear and instinct of survival present in each person. The effect produced by the Competitive Behavior is known “faster is slower effect”, the faster a person wants to go, the slower will be the entire process due to phenomena like clogging, impatience, etc. (Helbing et al., 2000).

It is worth mentioning that the situation evolves dynamically during the simulation; therefore during the evacuation, additional emergency situations are randomly generated (e.g. a new explosion and a new fire). This will affect again the people states (e.g. number of dead or injured people) and the evacuation process.

4. SIMULATION EXPERIMENTS AND RESULTS

The simulation model presented in the previous section has been used to carry out experimentations to determine how certain factors affect the evolution of the emergency situation in the plant and the evacuation.

4.1 Main Factors considered in the experimentations

The factors taken into account are described below:

- Task familiarity (A)
- Crowd Behavior (B)
- Human Error Mode (C)
- Fire Severity Zones circles Radius (D)
- Incident Gravity and secondary effects (E)

Task familiarity: it is the familiarity level with the task being performed by the operator. This task may assume two different levels:

- Totally unfamiliar (0), when the plant operator has no knowledge about the task that he executes; this condition arises when the operator does not receive the initial training therefore he does not have the necessary knowledge to perform the assigned task.
- Routine, highly-practiced (1), when the plant operator is knowledgeable about the task he executes, he is well-trained and he has all the necessary knowledge to perform the assigned task.

Crowd Behavior: it determines the behavior of people during the evacuation process. As mentioned in section 3.1, we have two different crowd behaviors: Queueing (0) and Competitive (1).

Human Error Mode: this factor expresses operator’s degree of concentration while performing the assigned task. It is assumed that the factor may have two levels: average-low (0) and average-high (1) degree of concentration.

Fire Severity Zones circles Radius: this parameter indicates the dimensions of the incidence radius according to the explosion type. Also in this case it is assumed that the factor may have two levels: low radius (0) between 20 and 30 meters; great radius (1) between 50 and 60 meters.

Incident Gravity and secondary effects: these parameters expresses the gravity of the incident and the probability to have additional side effects (e.g. secondary explosions/incidents). The factors has two different level: (0) it means low gravity incident with low probability of generating additional incidents and side effects; (1) it means high gravity incident with high probability of generating additional incidents and side effects.

4.2 Performance measures considered in the experimentations

A set of performance measures have been taken into account to understand how the previous described factors may affect the evolution of the emergency situation in the plant and the personnel evacuation. Namely, the following performance measures have been considered:

- Number of workers death.
- Number of workers injured.
- Number of non-injured workers.
- Number of errors that the operator commits due to its low concentration.
- The evacuation time, is considered the time in which an explosion occurs and the time that elapses until the last person arrives in the safety zone.
- The frequency of the evacuation time during the simulation time.

4.3 Simulation results and discussion

A sensitivity analysis has been carried out by considering the effects of the five factors on each performance measure. Figure 8 shows a Pareto Analysis for the number of workers death. The Pareto analysis reveals that the most relevant factors affecting the number of workers death are primarily the Incident

Gravity and the Fire Severity Zones circles Radius. While the dominant effects of factor E and D were expected, it is worth highlighting that there are second order as well as fourth order interactions that play an important role, namely:

- the interaction between the task familiarity and the crowd behavior;
- the interaction between the Human Error Mode and the Fire Severity Zones circles Radius
- the interaction between the Crowd Behavior and the Incident Gravity
- two fourth order interactions involving almost all the factors.

The second order interactions well explain how the number of workers death is jointly related to the operator experience but, during the evacuation, the number of workers death can be increased by a competitive behavior. Furthermore, the number of workers death decreases when the workers operate with higher degree of concentration but the decrease rate is smaller in case of larger Fire Severity Zone circles radius. Finally, the number of workers death increases when the Incident Gravity is larger, but the increase rate is even smaller when the evacuation happens according to a queueing behavior. As far as the fourth order interactions are concerned, these are quite difficult to explain and can be regarded as emergent situations due to the workers behavior, attitude and process and environmental factors.

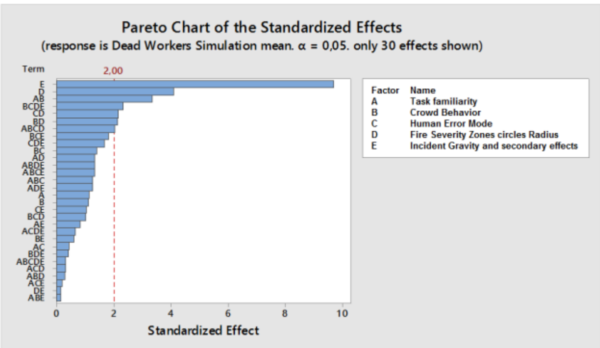


Figure 8. Pareto Chart Analysis for the number of workers death

Figure 9 shows the main effect plots for the Fire Severity Zones circle Radius and for the Incident Gravity factors. The percentage of workers death increase with the increase of the radius and decreases with the decrease of the incident gravity.

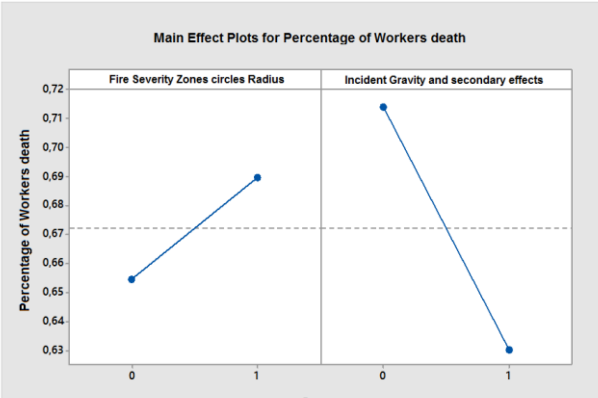


Figure 9. Main Effect Plots for the Percentage of workers death

Similar results have been obtained for the other performance measures. Figure 10 shows the Pareto Chart for the number of Injured people. Also in this case, several first order effects and higher order interactions can be observed. Figure 11 shows the Main Effect Plots for the percentage of injured workers as function of the three most significant factors.

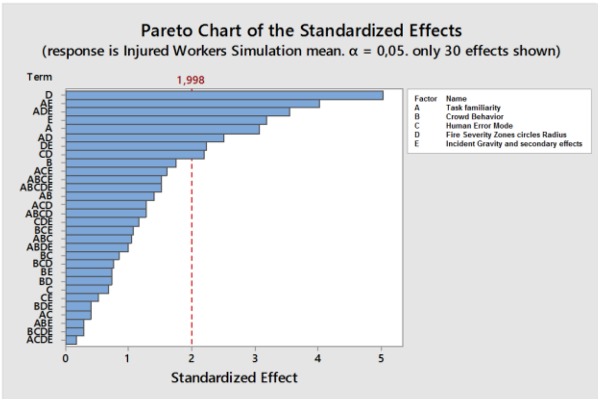


Figure 10. Pareto Chart Analysis for the number of workers injured

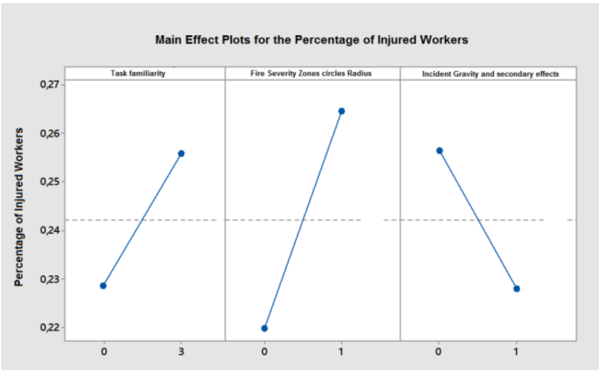


Figure 11. Main Effect Plots for the Percentage of injured workers

Finally, Figure 12 shows the Pareto Chart for the Evacuation time where it is possible to observe that there are multiple main effects and higher order interactions. While this result demonstrates that there are emergent situations that cannot be easily explained (e.g. having a significant fifth order interaction among all the factors considered) it also confirms that the simulation model is able to take into account correctly (from a logical point of view) how different factors may affect the evolution of a disaster in an industrial plant (e.g. the crowd behavior, the gravity and extension of the disaster, etc.).

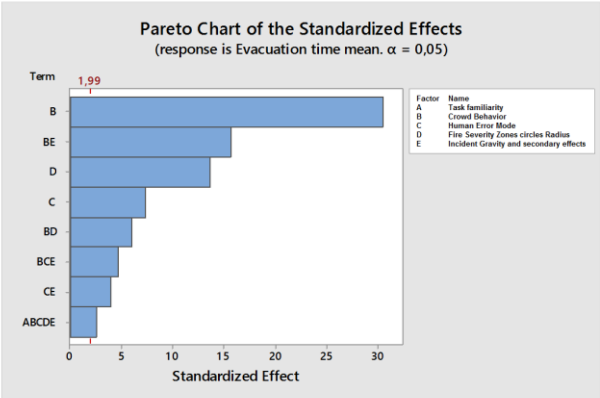


Figure 12. Pareto Chart Analysis for the Evacuation Time

5. CONCLUSIONS

The paper presents the results of a research work in which the main focus was the investigation of how the human behavior may affect the evolution of a disaster within an industrial plant. To this end the authors have developed an agent based simulation model including specific behavior models for agents (before and after the emergency, e.g. an explosion and a subsequent fire). The simulation model has been developed selecting as case study a real industrial plant (the Sansahuari Pumping Station) located in Ecuador. The simulation model has been used to carry out experimentations evaluating how different factors (related to workers capability and skills, workers behaviors, type of emergency/disaster) affect a set of performance measures (related to the disaster evolution such as number of workers injured and death, evacuation time, etc.). Simulation results shows both that the simulation model is able to recreate correctly the emergency evolution over the time (and the effects of the most relevant factors) and that there are some emergent situations (interactions among different factors) that cannot be easily explained (mostly due to the behavior of the different agents in the simulation and the interactions among them and with the environment).

REFERENCES

Beroggi G. E., Waisel L., & Wallace W. A., (1995). Employing virtual reality to support decision making in emergency management. Safety Science, Vol. 20(1), pp. 79-88.

Bessis N., Assimakopoulou E., Aydin M.E., Xhafa F., (2011). Utilizing next generation emerging technologies for enabling collective computational intelligence in disaster management. Studies in Computational Intelligence, Volume 352, pp. 503-526

Bruzzone A.G., (2013). Intelligent agent-based simulation for supporting operational planning in country reconstruction. International Journal of Simulation and Process Modelling, vol. 8, no. 2-3, pp. 145-159.

Bruzzone A.G., Massei M., (2010). Advantage of mobile training for complex systems. 9th International Conference on Modeling and Applied Simulation, MAS 2010, Held at the International Mediterranean and Latin American Modeling Multiconference, I3M 2010, pp. 57.

Cohen D., Sevdalis N., Taylor D., Kerr K., Heys M., Willett K., ... & Darzi A. (2013). Emergency preparedness in the 21st century: training and preparation modules in virtual environments. Resuscitation, Vol. 84(1), pp. 78-84.

Crichton M.T., (2009). Improving team effectiveness using tactical decision games. Safety Science. Vol. 47 (3), 330, pp. 0925-7535.

Davis M.T., Proctor M.D., Shageer B., (2017). Disaster Factor Screening using SoS Conceptual Modeling and an LVC simulation framework. Reliability Engineering & System Safety. Vol. 165, pp. 368-375.

Hashemipour M., Stuban S.M.F., Dever J.R., (2017). A community-based disaster coordination framework for effective disaster preparedness and response, Australian Journal of Emergency Management, vol. 32, no. 2, pp. 41-46.

Helbing D., Farkas I., Vicsek T., (2000). Simulating dynamical features of escape panic. Nature 407, 487-490

Mat R.C., Abubakar J.A., Aziz A.A., Yusoff M.F., Noorjasri N.A., (2017). 3D simulation for flood evacuation, Advanced Science Letters, vol. 23, no. 5, pp. 3883-3887.

Rosenthal, U., Boin, A., & Comfort, L. K. (2001). Managing crises: Threats, dilemmas, opportunities. Charles C Thomas Publisher.

Shi J., Ren A., Chen C., (2009). Agent-based evacuation model of large public buildings under fire conditions. Automation in Construction, vol. 18, no. 3, pp. 338-347.

Twaalfhoven S. F., Kortleven W. J., (2016). The corporate quest for zero accidents: A case study into the response to safety transgressions in the industrial sector. Safety science, Vol. 86, pp. 57-68.

Woods D.D., Dekker S., Cook R., Johannesen L., Sarter N., (2010). Behind Human Error, 2nd ed. Ashgate, Burlington, VT.

INNOVATIVE MODELLING OF SOCIAL NETWORKS FOR REPRODUCING COMPLEX SCENARIOS

Agostino G.Bruzzone [(a)], Riccardo di Matteo [(b)], Marina Massei [(c)], Ivan Vianello [(d)]

[(a) (c)] DIME University of Genoa, [(b)] Simulation Team, SIM4Future, [(d)] NATO Joint Warfare Center

[(a)] agostino@itim.unige.it, [(b)] dimatteo@simulationteam.com, [(c)] massei@itim.unige.it, [(d)] ivan.vianello@jwc.nato.int

ABSTRACT
The development of simulators able to reproduce the social network dynamics and their reactions to events and actions is a promising sector for extending the potential of Modeling in modern scenarios; in these contexts the simulations are usually devoted to support training and decision making. In particular it is interesting to reproduce the web social networks as interoperable simulators with main scenario models. The authors propose an approach for creating models of people and related web social networks to recreate their dynamics during simulation runs, experimentations and exercises. The paper presents an example of using this system to address complex scenarios where human factors are fundamental.

Keywords: Human Behavior Modeling, Social Network Simulation, Interoperable Simulation, CAX

1. INTRODUCTION

The major web social networks, currently ruling the life of a large part of the world population, have been launched at the beginning of third millennium (e.g. Linkedin 2003, Facebook 2004, Twitter 2006, Instagram 2009, Whatsapp 2010). Their diffusion turned viral and they evolved in harsh competition each other addressing specific needs and categories of people by changing theme as well as capabilities and characteristics (Kramer et al. 2017).

These social networks evolved very quickly thanks to the ideal set of boundary conditions and to the synergies with emerging technologies; indeed they turned to be a fundamental part of the social interactions for millions of people in few years (Larosiliere et al. 2017). Today is very evident the influence of social networks on public opinion and vice versa, so it could be expected that these contexts represent key elements for global decision makers in order to understand the situation and to control the effectiveness of their actions, considering that the networks are reacting dynamically to each event (Bond et al. 2012). Therefore, till today, in modern training processes, such as that one in CAX, it is much more common to adopt predefined scripts with main focus on traditional media respect dynamic web networks even due to the shortfall to model these entities (Cayirci & Marincic 2009). Indeed the main focus on traditional media (e.g. press releases, broadcasting, etc.) is dictated also by historical reasons and consolidated *forma mentis*.

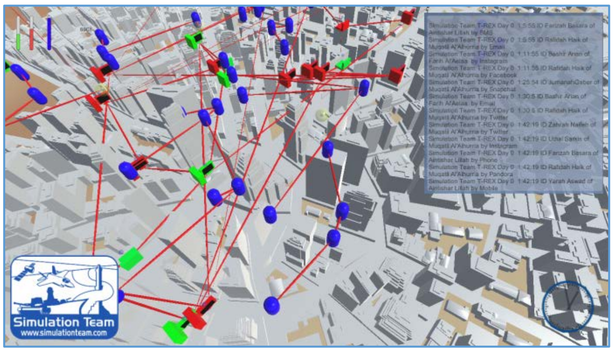


Figure 1 – T-REX Social Network Simulation

Therefore the experts identified the necessity to develop new models for properly address new frameworks, such as web social networks, as well as new challenges as Hybrid Warfare (Davis 2015; Bruzzone et al. 2016). It is evident that traditional media are still important and, obviously, they are interacting dynamical with social networks, but the principles ruling them are quite different (Keeton & McCann 2005). The new media channels are fast and diffused and their impact is very evident in politics and economics (Bond et al. 2012). In facts, the experience on specific cases such as recent USA elections confirmed that web social networks are often more effective respect traditional media, just for their capability to guarantee a continuous and quick connection with people and by targeting specific concepts and emotions (Enli 2017). The modern global scenarios are very strongly related to human factors: a very good example is the case of Hybrid Warfare (McCuen 2008; Baker 2015; Di Bella 2015; Lamb 2016). In this framework the use of models and technologies to develop effective capabilities, innovative doctrines and to create valuable training programs is fundamental (Gerasimov 2013 & 2016). Due to these considerations, it becomes critical to develop new simulators able to embed web social network into models to support decision making and training. This approach allows, with the modern complex scenarios, to extend the considerations carried out about Society Modeling since several decades (McLeod & McLeod 1984). In this paper, the authors present the analysis of the context and the conceptual model architecture as well as the preliminary results and scenario definitions for testing this approach.

2. WEB SOCIAL NETWORKS

In particular the smart phone evolution (e.g. computational power, memory, autonomy, camera capabilities, GPS, etc.) resulted to be the biggest driver for promoting, in practice, the potential of the IoT (Internet of Things) concepts (Silva et al. 2017); obviously the Internet Infrastructures acted also as a main stream, supporting this diffusion as well as being enhanced by the business opportunities generated by this approach itself (Wu et al. 2017). In facts, the web diffusion, availability, reliability and growth of mobile bandwidth allowed to develop almost real time interconnections much reliable than any other kind of previous communication/sensor networks (Rahman et al. 2017). This trend has further evolved in terms of Apps devoted to provide new services based on the Shared Economy approach (e.g. Amazon.com since 1994, Tripadvisor since 2000, Uber 2009) that further reinforced the interactions among people on cyber dimension (Turban 2015). The incoming future poses large emphasis on cloud solutions able to simplify and extend the capabilities of the different Hardware and Software systems, so it is expected to further promote this evolution in this direction (Hashem et al. 2015). The new I/O (Input/Output) devices proposing immersive Mixed Reality (MR), combining Virtual and Augmented Solutions, are promising a step towards new integration of the social networks with the human senses; this could produce breakthrough changes respect many different aspects that affect human behavior, especially for new generations (Cartwright et al. 1960; Parking 2016; Bruzzone et al. 2016; Boulus et al. 2017; Holyst 2017). Indeed, it is interesting to note that Psychologists and Sociologists are studying new phenomena and emerging behaviors in the Society and in the new generations that result in different attitudes, different trustiness concepts and different ethics (Luo 2009; Lehavot et al. 2010; Kahneman 2013; Harbeck Voshel et al. 2015).

In facts, it is also important to note that this picture corresponds to the creation of really immense Big Data soliciting many actors to use, support and sometime even abuse these opportunities (Chambers 2016).

In terms of Business several Applications have been very successful in this sense as they created connections with social networks to collect feedback, develop mutual trustiness, provide benefits to users and, obviously, promote services and products (Turban et al. 2015).

Vice versa, there are also several negative cases, such as that ones related to the use of camera and microphone enabled device “always on” capturing private media, often used among families or even minors (Gibbs 2015); these aspects raised, recently, several issues to the Federal Trade Commission in USA (Gray 2016). It is interesting to note that the different generations and geographical areas have approached web social networks in specific ways along the last decade with communalities and differences (Xenos et al. 2014; Fox & Moreland 2015; Dong 2017).

Today it is evident that social media and networks are fundamentals in influencing the politics in both domestic and overseas scenarios (Bond et al. 2012; Wolfsfeld et al. 2013; Kwon & Hemsley 2017); it is evident that this impact introduced also several issues about reliability of information collected on the social networks and capability to diffuse fake news and manipulate public opnion (Enli 2016; Paquet-Clouston et al. 2017). Due to these reasons it is currently pretty interesting to develop model of web social networks and the related diffusion of communications and examples (Bruzzone et al.2014b; Wang et al. 2017).

3 HYBRID WARFARE & POPULATION

Hybrid Warfare is a modern concept that is raising several issues, being sometime considered to be improperly named and referring to techniques and concepts consolidated since centuries (Hoffman 2009; Murray & Manson 2012). In facts, applied examples of Hybrid Warfare are dotting the human history, therefore today the reality, with its web social networks, IoTs and alternative media, is multiplying the effects of these warfare techniques by creating a really new framework pretty vulnerable to new technologies and socio-cultural attitudes (Gerasimov 2013 & 2016). Indeed the modern hybrid warfare actions are often addressing events with a very big media impact such as scandals and terrorist attacks (Keeton & McCann 2005; Bachmann & Gunneriusson. 2014; Krug 2017). Today, hybrid warfare addresses population and consensus in order to create critical conditions forcing the opponent even to capitulate on major goals of the attackers, without moving to traditional war (Galeotti 2016); a very good example is provided by the case of recent South China Sea disputes carried out as not armed conflicts (Blinka 2017). These considerations make it evident that the capability to simulate a complex modern scenario requires to consider the effect of these actions on population as well as the dynamics of web social networks as reactions to events and activities carried out on other layers, such as military operations, terrorist attacks, political, economic contexts, etc. (Bruzzone et al. 2016).

4 HUMAN BEHAVIOR SIMULATION & INTELLIGENT AGENTS

In order to properly simulate the web social networks is necessary to model the human factors and social interactions affecting the people perception and the mechanics of communication that are diffusing information among the population, over the different channels, from face to face to social media (Soller 2001; Abrahams et al. 2005; Santos et al. 2011; Bruzzone et al. 2014a). Considering these aspects, it results evident the opportunity to develop models adopting intelligent agents to reproduce the population and its dynamics (Wooldridge & Jennings 1995; Sarjoughian , Zeigler & Hall 2001; Bonabeau 2002; Sokolowski & Banks 2011; Massei et al. 2014).

Figure 2 – Communications and Diffusion on web

This opportunity could be further reinforced by adopting a distributed interoperable architectures that enable the people models to interact with other simulators during the dynamic evolution of the scenario (Hamilton et al. 1996; Smith 2002; Zacharewicz et al 2008; Bruzzone et al.2016). The authors used extensively the IA-CGF (Intelligent Agent Computer Generated Forces) developed by the Simulation Team to reproduce the population based on a multi layer structure (Ören & Longo 2008; Bruzzone et al.2014a).
This model includes Entities & Objects (e.g. as military units, PSYOPS centers, etc.), people (e.g. individuals and small groups composing the whole population) and interest groups (e.g. a leader, an economic group, a religion, etc.), all interacting mutually and with other models (Bocca et al. 2006). This structure is designed to be able to interoperate with other models by adopting HLA (High Level Architecture) and it has been successfully experienced in many different applications including CIMIC, Cyberwarfare, Urban Disorders and Joint Multicoalition operations (Bruzzone et al 2015; Massei & Tremori 2014b).

5 MODELING THE WEB SOCIAL NETWORK

The authors decided to develop models of the web social networks by making them interacting with the IA-CGF within a MS2G (Modeling, interoperable Simulation and Serious Game) application, named T-REX (Bruzzone et al. 2016). Indeed the combination of serious games and interoperable simulation has a great potential in education and training on new issues (Raybourn 2012; Bruzzone 2014b). In facts, T-REX simulation includes towns and villages as well as threat networks and population (see figure 1). The proposed scenario includes multiple towns with their specific Population and related Social Networks; the People Objects, reproducing the individuals, reacting to events, communicating and generating elements to populate the web social networks, while Trustiness and Belief evolve dynamically based on actions, decisions & events.
In facts, in the proposed model, the nodes of the threat networks as well as the structure of the population is mapped by using Intelligent Agents (IAs).

The IA are enabled to move over the terrain in consistency with their life cycle and characteristics and to interact each other and with other entities; each IA is characterized by a set of variables that includes among the others:

- Absolute Position (i.e. latitude, longitude & altitude)
- Logic Position (e.g. at home)
- Life Cycle Type (e.g. blue collar)
- Home Location (e.g. position)
- Work Location (e.g. position)
- Interpersonal Relationships (e.g. connections to family members)
- Gender (e.g. male)
- Age (e.g. 20 year old)
- Health Status (e.g. severely sick)
- Activity or Operational Status (e.g. working)
- Ethnic Group (e.g. Caucasian)
- Tribe/Clan/Urban Group (e.g. EAKL syndicate)
- Religion (e.g. Christian)
- Political Party (e.g. social democratic party)
- Social Status (e.g. worker)
- Educational Level (e.g. high school)
- Coolness (e.g. steady)
- Crisis Ready Level (e.g. unexperienced to face crisis)
- Human Behavior Modifiers (e.g. Aggressiveness, Stress, fatigue and fear)

Even the members of the threat network have such characteristics, indeed the threat network member have the following specific characteristics:

- Threat Type (e.g. terrorist, quisling, sympathizer, occasional contact, not involved)
- Threat Status (e.g. sleeper, active)
- Threat Awareness (e.g. unidentified, identified, tracked)

In this way the simulator reproduces the life cycle of the population and individual behaviors. At this point it could be defined the conditions that activate communications among the people with special attention to two different kinds:

- Communications among individuals in the Threat Networks: these communications are not restricted to that one among terrorists, several ones are going through people that is supposed to be part of the network due to occasional contacts, coincidences as well as intelligence failures; these additional communications introduce challenges in properly identify the real threats and tracking them.
- Communication among population reacting to scenario major events that are generated stochastically based on social models reproducing human behavior and diffused among the population.

In both cases it is created a specific kind of object that represent the information published in the web social network as well as the network itself with the relative characteristics of connectivity and member preferences. In this way, an information diffused over a social network diffuse inside it, as well as outside, considering the connections with the people objects.

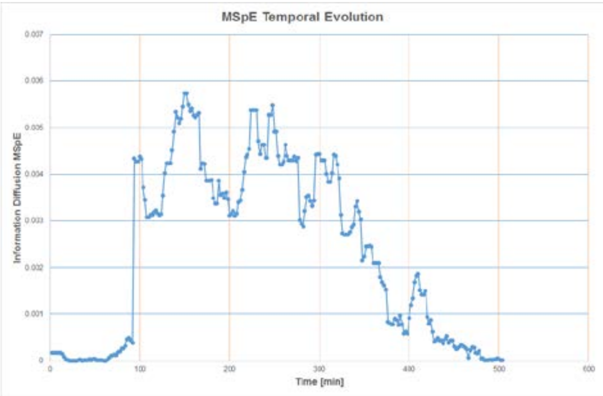


Figure 3 – Communications and Diffusion on web

The use of historical and statistical data support the fine tuning of these models even if obviously the validation is quite challenging and requires experimental tests (Balci 1994; Mastrosa et al. 2012; Liu et al.2016). In our model each web social network requires to be defined in terms of diffusion, average number of connections, user publishing intensity and frequency, trustability.
The authors experimented the first kind of communications on hybrid warfare scenarios as proposed in figure 2. In this case it has been used a scenario involving cyber attacks and media attacks over a region with a large city and four medium size town.
The attacks reproduce cyber actions on a critical infrastructure that involve a power station, a tank farm, a desalination facility and an oil port terminal (Bruzzone et al.2016).
In this context the scenario requires also to model the ICT infrastructure and cyber events and their impact on population (Theohary 2011; Damodaran & Couretas 2015). The dynamic validation has been based on temporal evolution of Mean Square Error related to the time required to diffuse the information of the attack over the population as proposed in figure 3.
Currently they are working on implementing the second type for reproducing specific web social networks in a complex scenario.

6 WEB SOCIAL NETWORK FOR CAX: SCENARIO FOR THE SIMULATOR

CAXs (Computer Assisted Exercises) are good examples where web social network simulation could provide a great benefits in different way:

- Capability to early recognize symptoms and effects of their to COE (Course of Actions) by the automatically generated response of the Social Network Simulator coupling population reaction
- Experiencing dynamically and interactively the dynamics of Web Social Networks as a main feedback to properly react to scenario evolution
- Capability to understand the dynamics of population in complex scenarios
- Develop skills in operations carried out in scenarios heavily affected by Human Factors and involving Populations and Social Networks

For these reasons currently the authors are defining a scenario and a set of Key Performance Indexes to test on the field the capabilities of the proposed Human Behavior Models (HBM) and Web Social Network Simulators.

CONCLUSIONS

The current paper outlines the opportunity to develop models of web social networks and their dynamics as well as the potential to be used in different applications. Currently the authors are experimenting, over different mission environments, the capability to reproduce these phenomena and to measure the related aspects in order to conduct investigative analysis on them; in addition the authors are fine tuning the proposed scenario as reference case study to be used for future researches in order to assess the potential of this approach within CAXs. In the next future, the authors are planning to further develop these models in order to extend the verification and validation of this approach and to diffuse the use of these simulators in real exercises to fill up the training gap related to web social networks.

REFERENCES

Abrahams A., Boisot M., Bharathy Gnana (2005) "Simulating the Knowledge Transfer Dilemma: Lessons for Security and Counter-Terrorism", Proceedings of SCSC, Philadelphia, NJ, July

Bachmann, S. D. O. V., & Gunneriusson, H. (2014) "Terrorism and Cyber Attacks as Hybrid Threats: Defining a Comprehensive Approach for Countering 21st Century Threats to Global Risk and Security", The Journal on Terrorism and Security Analysis

Baker B.D. (2015) "Hybrid Warfare with Chinese Characteristics", The Diplomat, September

Balci, O. (1994) "Validation, Verification, and Testing Techniques Throughout the Life Cycle of a Simulation Study," Ann. Oper. Res. 53, 121–173

Blinka J. (2017) "Hybrid Warfare in Asia: Potential not yet Unleashed", European Security &Defense, May

Bocca, E., Bruzzone, A. G., & Rocca, A. (2006) "Algorithms Devoted to Reproduce Human Modifiers in Polyfunctional Agents", Simulation Series, 38(4), 202

Bonabeau, E. (2002) "Agent-based modeling: Methods and Techniques for Simulating Human Systems", Proceedings of the National Academy of Sciences, 99, supplement 3, 7280-7287

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012) "A 61-million-person experiment in social influence and political mobilization", Nature, 489(7415), 295-298

Boulos, M. N. K., Lu, Z., Guerrero, P., Jennett, C., & Steed, A. (2017) "From urban planning and emergency training to Pokémon Go", Intternational Journal of Health Geographics, 16(1), 7

Bruzzone A.G., Massei M., Longo F., Cayirci E., di Bella P., Maglione G.L., Di Matteo R. (2016) "Simulation Models for Hybrid Warfare and Population Simulation", Proc. of NATO Symposium on Ready for the Predictable, Prepared for the Unexpected, M&S for Collective Defence in Hybrid Environments and Hybrid Conflicts, Bucharest, Romania, October 17-21

Bruzzone A.G., Massei M., Longo F., Nicoletti L., Di Matteo R., Maglione G., Agresta M. (2015) “Intelligent Agents & Interoperable Simulation for Strategic Decision Making on Multicoalition Joint Operations”, Proc.of DHSS2015, Bergeggi, Italy, September

Bruzzone A.G., Massei M., Poggi S., Bartolucci C., Ferrando A. (2014a) “Intelligent Agents for Human Behavior Modeling as Support to Operations”, Simulation & Modeling Methodologies, Technologies and Applications, Advances in Intelligent Systems and Computing, Springer, London, Vol. 319, pp 119-132

Bruzzone A.G., Massei M., Agresta M., Poggi S., Camponeschi F., Camponesch M. (2014b) “Addressing Strategic Challenges on Mega Cities through MS2G”, Proc. of MAS, Bordeaux, France, September 12-14

Chambers D. (2016) “Changing Media, Homes and Households: Cultures, Technologies and Meanings”, Routledge Taylor & Francis Group, NYC, USA

Cartwright D, Zander A. (1960) "Group Dynamics: Research & Theory", Tavistock, London, UK

Cayirci, E., & Marincic, D. (2009) “Computer Assisted Exercises and Training: A Reference Guide”, John Wiley & Sons, NYC

Damodaran, S. K., & Couretas, J. M. (2015) "Cyber Modeling & Simulation for Cyber-Range Events", Proceedings of the Conference on Summer Computer Simulation by Society for Computer Simulation International, San Diego, CA, July

Davis J.R. Jr (2015) “Continued Evolution of Hybrid Threats”, The Three Sword Magazine, N. 28, pp.19-25

Di Bella P. (2015) “Present and Futures Scenarios and Challenges for M&S terms of Human Behavior Modeling”, Invited Speech at I3M2015, Bergeggi, Italy, September

Dong, X., Chang, Y., Wang, Y., Yan, J. (2017) “Understanding usage of Internet of Things (IOT) systems in China: Cognitive experience and affect experience as moderator”, Information Technology & People, 30(1), 117-138

Fox, J., & Moreland, J. J. (2015) “The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances” Computers in Human Behavior, 45, 168-176

Enli, G. (2017) “Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election” European Journal of Communication, 32(1), 50-61

Galeotti, M. (2016) “Hybrid, ambiguous, and non-linear? How new is Russia’s ‘new way of war’?”, Small Wars & Insurgencies, 27(2), 282-301

Gerasimov V. (2016) "The Value of Science is in the Foresight: New Challenges Demand Rethinking the Forms and Methods of Carrying out Combat Operations ", Military Review, January-February

Gerasimov V. (2013) "The Value of Science is in the Foresight ", Voyenno-Promysh-lennyy Kurier online , February 26, (original version in Russian)

Gibbs S. (2015) “Samsung's Voice-recording Smart TVs breach privacy law, campaigners Claim”, The Guardian, February 27

Gray S. (2016) “Always On: Privacy Implications of Microphone-Enabled Devices”, Future of Privacy Forum, Washington DC, April

Hamilton John A. Jr., Nash David A., Pooch Udo W. (1996) "Distributed Simulation", CRC Press

Harbeck Voshel E., Wesala A. (2015) “Social Media & Social Work Ethics: Determining Best Practices in an Ambiguous Reality”, Journal of Social Work Values and Ethics, Volume 12, Number 1

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015) “The rise of Big Data on Cloud Computing: Review and open research issues” Information Systems, 47, 98-115

Hoffman, F. G. (2009) “Hybrid threats: Reconceptualizing the evolving character of modern conflict”, Institute for National Strategic Studies, National Defense University, Washington, DC, No.240, April

Holyst J.A. (2017) “Collective Emotions in Cyber Space”, Springer, NYC, USA

Kahnenam D. (2013) “Thinking, Fast and Slow”, Farrar, NYC, USA

Keeton, P., & McCann, M. (2005) "Information Operations, STRATCOM, and Public Affairs", Military Review, 85(6), 83

Krämer, N. C., Neubaum, G., & Eimler, S. C. (2017) “A brief history of (social) Cyberspace”, in Cyberemotions, Springer, NYC, pp. 11-35

Krug, G. J. (2017) “Fake information, fake warfare, and real destabilizations of new”, Russian Journal of Communication, 1-3

Kwon, K. H., & Hemsley, J. (2017) “Cross-National Proximity in Online Social Network and Protest Diffusion: An Event History Analysis of Arab Spring”, Proceedings of the 50th Hawaii International Conference on System Sciences, January

Lamb C., Stipanovich S. (2016) “Back to Basics on Hybrid Warfare in Europe A Lesson from the Balkans”, Joint Force Quarterly, 81, March 29

Larosiliere, G. D., Carter, L. D., & Meske, C. (2017) “How does the world connect? Exploring the global diffusion of social network sites” Journal of the Association for Information Science and Technology, 68(8), 1875-1885

Lehavot, K., Barnett, J., & Powers, D. (2010) “Psychotherapy, professional relationships and ethical considerations in the MySpace generation”, Journal of Professional Psychology: Research and Practice, 41(2), 160-166.

Liu, L., Li, Y. B., Wu, D., Zhang, L., & Wang, N. (2016) "Design and Developmental Research on the VV&A of Complex Simulation System", Proc. of MATEC Web of Conferences, Vol. 63

Luo, J. (2009) “The Facebook phenomenon: Boundaries and controversies”, Journal of Primary Psychiatry, 16(11), 19-21

Massei, M., Poggi, S., Agresta, M., & Ferrando, A. (2014a) "Development Planning Based on Interoperable Agent Driven Simulation", Journal of Computational Science, 5(3), 395-407

Massei M., & Tremori A. (2014b) “Simulation of an Urban Environment by Using Intelligent Agents within Asymmetric Scenarios for Assessing Alternative Command and Control Network-Centric Maturity Models”, the Journal of Defense Modeling and Simulation, 11(2), 137-153

McCuen, J. J. (2008) “Hybrid Wars”, Military Review, 88 (2), 107

McLeod, J., & McLeod, S. (1984) “Simulation of the Society”, Simulation, 43(6), 320-323

Murray, W., & Mansoor, P. R. (2012) “Hybrid warfare: fighting complex opponents from the ancient world to the present”, Cambridge University Press, UK

Ören T.I., Longo F., (2008) “Emergence, Anticipation and Multisimulation: Bases for Conflict Simulation”, Proceedings of the 20th European Modeling and Simulation Symposium, EMSS 2008, pp. 546-555

Paquet-Clouston, M., Bilodeau, O., & Décary-Hétu, D. (2017) “Can We Trust Social Media Data?: Social Network Manipulation by an IoT Botnet”, Proceedings of the 8th ACM International Conference on Social Media & Society, July

Parking S. (2016) “After the success of Pokémon Go!, what is the future for augmented reality?”, The Guardian, October 23

Rahman, M. A., & Hossain, M. S. (2017) “A Location-Based Mobile Crowdsensing Framework Supporting a Massive Ad Hoc Social Network Environment”, IEEE Communications Magazine, 55(3), 76-85

Raybourn, E.M. (2012) “Beyond Serious Games: Transmedia for more Effective Training & Education”, Proc. of DHSS2012, Rome, Italy

Santos, E. E., Santos, E., Wilkinson, J. T., Korah, J., Kim, K., Li, D., & Yu, F. (2011) "Modeling Complex Social Scenarios Using Culturally Infused Social Networks", Proc. of IEEE Int.Conference on Systems, Man, and Cybernetics (SMC), Anchorage, AK, October

Sarjoughian, H. S., Zeigler, B. P., & Hall, S. B. (2001b) "A Layered Modeling and Simulation Architecture for Agent-Based System Development", Proceedings of the IEEE, 89(2), 201-213

Silva, J. S., Zhang, P., Pering, T., Boavida, F., Hara, T., & Liebau, N. C. (2017) “People-Centric Internet of Things” IEEE Communications Magazine, 55(2), 18-19

Smith, R. (2002) "Counter Terrorism Simulation: A New Breed of Federation", Proceedings of the Spring 2002 Simulation Interoperability Workshop, Orlando, FL, March

Soller, A. (2001) "Supporting Social Interaction in an Intelligent Collaborative Learning System", Int. Journal of Artificial Intelligence in Education, 12 (1), 40-62

Sokolowski, J. A., & Banks, C. M. (2011) "Principles of Modeling and Simulation: a Multidisciplinary Approach", John Wiley & Sons, NYC

Theohary, C. A. (2011) "Terrorist use of the Internet: Information Operations in Cyberspace", DIANE Publishing, Darby, PA

Turban, E., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2015) “Electronic commerce: A managerial and social networks perspective”, Springer, NYC, USA

Wang, R., Rho, S., Chen, B. W. Cai, W. (2017) “Modeling of large-scale social network services based on mechanisms of information diffusion: Sina weibo as a case study”, Future Generation Computer Systems, 74, 291-301.

Wolfsfeld, G., Segev, E., & Sheafer, T. (2013). Social media and the Arab Spring: Politics comes first. The International Journal of Press/Politics, 18(2), 115-137

Wooldridge, M.; Jennings, N. R. (1995) "Intelligent Agents: Theory and Practice", Knowledge Engineering Review, 10 (2), 115–152

Wu, C. C., Wu, C. H., Huang, W. T., & Chen, C. H. (2017) “The Effects of The Use of Social Network Application on Interpersonal Relationship And Internet Addiction”, Proceedings of the 4th ACM Multidisciplinary International Social Networks Conference, July

Xenos, M., Vromen, A., & Loader, B. D. (2014) “The great equalizer? Patterns of social media use and youth political engagement in three advanced democracies”, Information, Communication & Society, 17(2), 151-167

Zacharewicz G., Frydman C., Giambiasi N. (2008) "G-DEVS/HLA Environment for Distributed Simulations of Workflows", Simulation, Vol.84, N.5, pp 197-213

AUTONOMOUS SYSTEMS & SAFETY ISSUES: THE ROADMAP TO ENABLE NEW ADVANCES IN INDUSTRIAL APPLICATIONS

Agostino G. Bruzzone [(a)], Marina Massei [(b)], Matteo Agresta [(c)], Riccardo di Matteo [(d)], Kirill Sinelshchikov [(e)], Francesco Longo [(f)], Letizia Nicoletti [(g)], Luciano di Donato [(h)], Laura Tomassini [(i)], Carla Console [(j)], Alessandra Ferraro [(k)], Marco Pirozzi [(l)], Daniele Puri [(m)], Leonardo Vita [(n)], Fabio Cassandra [(o)], Canio Mennuti [(p)], Giuseppe Augugliaro [(q)], Corrado Delle Site [(r)], Francesco Di Palo [(s)], Paolo Bragatto [(t)]

[(a) (b)] Simulation Team, DIME University of Genoa
[(c) (e)] Simulation Team, Liophant, (d) Simulation Team, SIM4Future,
[(f)] Simulation Team, MSC-LES, University of Calabria, [(g)]Simulation Team, CAL-TEK Srl
[(h) (i) (j) (k) (l) (m) (n) (o)] INAIL, Laboratorio macchine e attrezzature di lavoro
[(p) (q)] INAIL Laboratorio tecnologie diagnostiche per la sicurezza
[(r) (s) (t)] Laboratorio impianti a pressione e stabilimenti a rischio di incidente rilevante

[(a)] agostino@itim.unige.it, [(b)] massei@itim.unige.it, [(c)] agresta@simulationteam.com,
[(d)] dimatteo@simulationteam.com, [(e)] kirill.sinelshchikov@simulationteam.com, [(f)] f.longo@unical.it,
[(g)] l.nicoletti@cal-tek.eu, [(h)] l.didonato@inail.it, [(i)] l.tomassini@inail.it, [(j)] c.console@inail.it, [(k)] ale.ferraro@inail.it,
[(l)] m.pirozzi@inail.it, [(m)] d.puri@inail.it, [(n)] l.vita@inail.it, [(o)] f.cassandra@inail.it, [(p)] c.mennuti@inail.it,
[(q)] g.augugliaro@inail.it, [(r)] c.dellesite@inail.it, [(s)] f.dipalo@inail.it, [(t)] p.bragatto@inail.it

ABSTRACT

The paper addresses the safety issues related to the development of new solutions based on autonomous systems for industrial applications and the necessity to develop experimental environments for investigating these cases; a set of examples is proposed in order to provide cases and challenges as well as to suggest approaches to address these problems.

Keywords: Autonomous Systems, Safety, Industrial Plants, Security, Modeling and Simulation

1. INTRODUCTION

Industry 4.0 integrates all automation technologies and, among others, also the new category of the so-called UxVs (Unmanned Vehicles), i.e. all those remote ground, aquatic (surface or underwater) or aerial (fixed or rotary wing) vehicles operating with different levels of autonomy, from remotely operated to fully autonomous. Unmanned vehicles are characterized by many advantages, ranging from their agility and speed in reaching places that would be otherwise difficult to access, to their potential in replacing humans in the presence of dangers, to their expendable nature; due to these characteristics, UxVs are becoming popular and spreading exponentially among many different application fields, with special attention to the most light and less expensive models (Salvini 2017). Such versatility is reflected by their dissemination which is monitored and studied with great interest by the authorities and agencies involved in security and prevention in all areas, and therefore also by INAIL, in particular by DIT, the Department of Technological Innovations (Clarke et al. 2014; Di Donato 2017).

Figure 1 – IDRASS Simulator operating Autonomous Systems inside Industrial Plants

In Italy, DIT is also assessing its potential through research work aimed at promoting its diffusion for reducing the workers exposure to high risks and difficult tasks, where human presence was supposed to be indispensable up to now and to ensure safe inspections and monitoring to maintain the safety of hard-to-access structures and work equipment (Spanu et al. 2016).

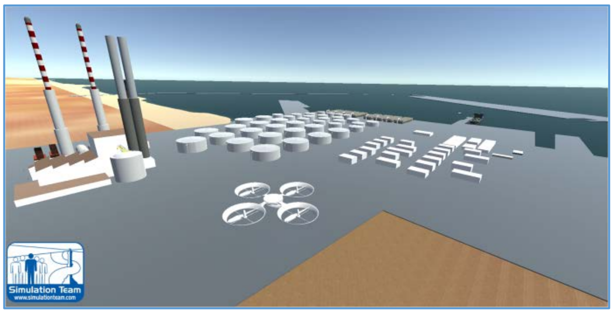


Figure 2 – UAV overview within Critical Infrastructure Protection Simulator, T-REX


Figure 3 – T-REX combined protection of Port Framework by USV and other Autonomous Systems

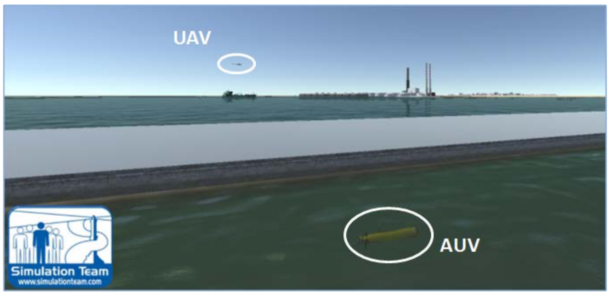


Figure 4 – AUV patrolling around the Critical Infrastructure in cooperation with UAV

Today efforts are made to promote the spread of Autonomous Systems, given their possible employment for prevention on safety, therefore it is not possible to ignore the fact that the introduction of such a new technology, even though one of its objectives is Risk reduction, can result itself a carrier of new risks and dangers that require to be investigated.
In this paper, it is proposed a set of case studies related to use of UAV (Unmanned Aerial Vehicles), UGV (Unmanned Ground Vehicle), USV (Autonomous Surface Vehicles) and AUV (Autonomous Underwater Vehicles) in industrial applications where safety and security are two main streams; these cases propose opportunities and challenges for the introduction of this technology in industrial sectors as well as the need to fix requirements for the equipment, the procedures and the training programs.

2 UxV & SAFETY

Safety is a major issue as well as one of the major driver, along with security, for the extensive use of UxV in civil and military applications; in facts their low operational cost and expendable nature make them ideal for being used in dangerous environments and almost all case studies proposed in the paper deal with these topics (Apvrille et al.2015; Merwaday & Guvenc 2015; Altawy et al. 2016). Due to the complexity of these context the use of M&S (Modeling and Simulation) is considered often the most promise methodology for investigating modern UxV problems (Bruzzone et al. 2016a).
In facts, Simulation Team is an International Network of Excellence working on up-to-date simulation technologies and has acquired a large experience in autonomous systems within a broad set of applications with special attention to collaborative multi domain cases (Bruzzone et al. 2014, 2016b). Along last years, the Simulation Team has activated several projects for Industry and major Agency in using UxV in industrial plants (Bruzzone et al. 2016a).
From another point of view, INAIL-DIT is a Department that, in line with its mission, is approaching these new UxV and remote pilot systems while also keeping in mind the safety aspect. The safety measures to be taken in the design and use of drones are just some of the topics that will need to be addressed in order to get products up to all applications where they can be effectively used; among these topics for instance it could be useful to develop studies on materials suitable for protecting the drone itself as well as the people around, or the development/adoption of sufficiently advanced safety equipment, beyond the ones required to fulfil the functional requirements of the drone (Valavanis et al. 2014; Sanchez-Lopez et al. 2016). Finally, it is desirable that regulatory and technical standards follow, or rather support, this innovative process within Industry 4.0 and its technologies, including remote pilot systems (Kehoe et al.2015). INAIL-DIT is committed to being ready to support the National Authorities in every step when collaboration would be necessary. For instance, the use of remote pilot systems involves remote control by a human and, sometimes, the presence of other personnel engaged in carrying out work in close proximity to the areas where autonomous vehicle operations take place; this introduces issues about training among the others. In addition this scenario implies also the need to apply the relevant legislation for the protection of the health and safety of workers, as well as the rules of transposition of applicable product and other relevant National or International regulations (Djellal & Gallouj 1999). This implies that the authorities and bodies involved are plural; in addition to the European Commission, which is responsible for the issuance of directives/product regulations, many National Departments and Ministries as well as Public Organizations are involved, such as, for instance in Italy, the Ministry of Labour and Social Policies, the Health and Safety of Workers, the Ministry of Labour Infrastructure and Transport, for Aviation Security or Navigation and the National Agency for Civil Aviation. In the following it is proposed an overview of different application fields for UxV where remotely operated vehicles and autonomous systems might positively affect the health and safety of workers. Given the succession of serious and fatal accidents occurred over the past few years during the conduct of activities in suspected or confined environments (Spillane et al. 2012; Nano et al.2013; Leão et al. 2015), which in many cases highlights an inadequate risk assessment of the possible presence of hazardous substances, one of the first uses of remote pilot systems is definitely about air quality control in confined environments such as silos, tanks, holds, and other environments (Valavanis et al. 2014).

In this case, it might be useful to equip drones with "smart" sensors for evaluation, for example, of the conditions that may allow operators to enter in dangerous areas (Floreano & Wood 2015). It is clear that, in order to operate these systems, it is necessary to carefully consider the characteristics that these systems to check consistency with presence of liquids, vapours or dust and, more generally, hazardous areas (such as areas subject to the ATEX Directive). These analysis and requirements are devoted to ensure safe and adequate use of UxVs under critical conditions that may occur in confined environments, or in general, within the areas where they should operate during emergencies; for instance in avoiding the accident escalation from fire scenarios in case of upper tier Seveso plants (Palazzi et al. 2017).

in order to ensure its safe and adequate use in difficult conditions that may occur in confined environments or in general in environments where drones should intervene for emergency management (e.g. in case of Accidents at high risk companies). Another useful application of UAVs for the protection and safety of workers is their use for inspections at relevant heights or at least in difficult-to-reach areas for structures and equipment in order to check their integrity and stability through a visual examination assisted by optical systems (cameras, thermal sensors etc.) or checks performed using other equipment (Jones 2006).In addition, UAVs can be utilized for environment surveys through high-resolution photographic capture, enabling visual 3D mapping and thus the knowledge of orographic features (e.g. slopes of the ground) as well as to detect, in real time, the presence of obstacles and particular conditions of danger determined, for example, by climatic factors which may change the orographic conditions already observed (Siebert & Teizer 2014). Another important, but less known application of UxV is monitoring of infrastructure using GPR (Ground Penetrating Radar). Unlike other mentioned types of sensors, the GPR allows to control conditions of infrastructures hidden under soil; for instance it is possible to detect flooding or voids, furthermore, being installed on a UAV, the GPR allows to perform this operation in short time (Kovacevic et 1l. 2016). Of course drones could operate not only individually, but also in a swarm which allows to install different kinds of sensors on the platforms and it could enhance drastically data acquisition capabilities of the whole system; this swarm collaborative use represents one of most promising directions of research in this field (Burkle et al. 2011). The data acquired in this way are available for being communicated instantly to the control unit of a bulldozer or any other moving machine that, through the help of a GPS system, could "alert" the driver if it is approaching to danger zones; it becomes possible also to develop a smart guiding support, in an assisted way, by providing automatic corrective actions, such as speed reduction until stopping or finding an alternative route, as a dynamic new risk point or area is approaching (Kim et al. 2015).

To do this, the remote pilot system should be equipped with appropriate instrumentation (sensors, radar, cameras, etc.), now largely present on the market, which can be utilized with the specific task.

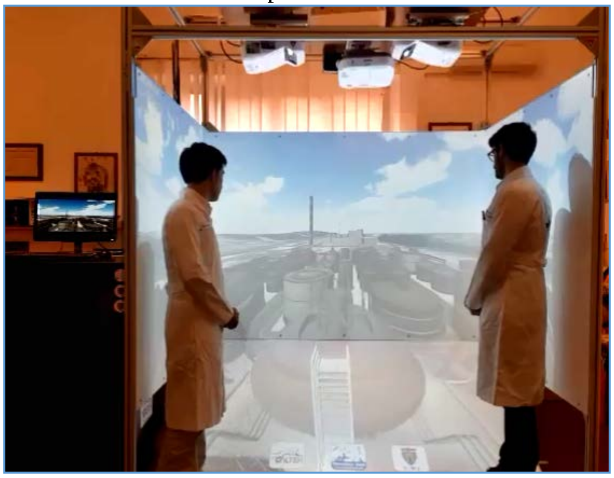


Figure 5 – Man on the Loop Supervising Operations within SPIDER special CAVE

As already mentioned, another useful area of employment is the acquisition of information through drones for the management of emergency relief activities (Doherty et al. 2007).

In facts, aerial reconnaissance with drones allows to have a direct view of the situation of places where it is not easy to access, at least for a first assessment process to drive first responders, thus facilitating assistance and recovery through, for example, identification of a possible access path to the affected area.

Other possible applications are those related to the evaluation of the various emissions of the machines and machineries (Gardi et al. 2016).

This includes also the use of microphone on drones that could prevent the placement of microphones on fixed positions difficult to be implemented due to the constraints of the machineries themselves (Ishiki et al. 2014). In addition, it has been studied the possibility to use drones to evaluate drift of fertilizers applied to irrigation machines in herbaceous and tree crops (Pulina et al. 2016).

This kind of testing could involve drift tracking using UAV equipped with high definition (HD) visual recording systems (Pizzarella 2014); the related images could be acquired by different profiles: from top, back and side, by using a water-based liquid mixed with a red powdered food for distribution analysis; once the drift motion is tracked the same is replicated graphically through the GIS support on an aerial photo.

More replicated tests in different climatic conditions could simulate different drift situations. In these cases further focused analysis could be carried out on the specific weight of the distributed product, as the weight of the treated molecules that have different behaviour even with same meteorological conditions.

In facts, it is now possible to extend the scope and use of new autonomous system technologies, including the remote pilot systems, to increase safety and health levels through specific studies and research that allow to evaluate and promote their effectiveness.

3 CASE STUDIES ON UxV SIMULATORS

The challenges presented in this paper represent often new application fields for UxV or specific implementations of new solutions; so it is evident that to complete their test and experiment and to evaluate related risks in terms of safety, it is necessary to recreate a realistic mission environments. Usually this requires to be addressed by M&S (Modeling and Simulation) in order to be effective (Bruzzone et al. 2014); in facts the adoption of MS2G paradigm (Modeling, interoperable Simulation and Serious Games) represents a very strategic advantage allowing to combining different models, simulators and also real equipment within a common synthetic environment. These simulation environments should be intuitive and interactive by using most advanced Mixed Reality solutions such as the SPIDER (Simulation Practical Immersive Dynamic Environment for Reengineering), developed by Simulation Team, in order to support the Subject Matter Experts (Bruzzone et al. 2016a). In the following case studies are proposed.

3.1 IDRASS

Indoor Operations in industrial Plants are critical especially in case the environment is contaminated, so they represent an ideal example to apply UxV; from this point of view, the support during disasters in industrial facilities is a very popular area for R&D on UxV due to the challenges represented by these environments (Bruzzone et al. 2016a); some of the authors developed IDRASS (Immersive Disaster Relief and Autonomous System Simulation) to address this context in case of CBRN (Chemical, Biological, Radiological and Nuclear) contamination due to accidents or man made disasters (see figure 1); IDRASS simulates both operations indoor and outdoor within different industrial facilities such as chemical and nuclear plants (Bruzzone et al.2016a). In these contexts it is usually necessary to introduce many actors to reproduce the whole crisis scenario and, obviously, the use of IA (Intelligent Agents) is a fundamental resource for being able to develop realistic mission environments. The safety issues in using drones within industrial facilities deals with the challenges due to these context (Mobley 2001): cables, cable trays, pipelines, tanks are physical obstacles that populate the area with high density. Therefore in several cases there also relief venting systems and safety valves that could create streams challenging for UxV in terms of blast as well as temperature; in the plants often the atmosphere could include dusts, corrosive agents as well as high temperature elements dangerous in terms of irradiations. It is evident that there are solid, thermal and gaseous barriers not easy to detect and creating complex environments; sometime the UxV could be required to operate also in confined environments where the air mix could turn to be dangerous for explosions respect the characteristics of some of the robotic system components. In facts, the whole industrial plant could include several systems that could create risks and domino effect in case of UxV collision or even just interaction; indeed the electronic interference between UxV controls and DCS (Digital Control System) could affect plant safety. In facts, it is also necessary to remind the very crucial aspect of electromagnetic compatibility. These aspects are common just to the presence of high voltage lines and equipment in the industrial plant and becomes even more intense in presence of ionizing radiation caused by nuclear spills and contamination, which could lead to loss of connection as well as drone malfunctions in case of its insufficient radiation hardening (McCurry 2017). In addition it is also necessary to consider the presence of “natural” communication barriers, caused by reflection and interference of electromagnetic waves due to the high density of metallic infrastructures; these are affecting UxV communications and operations.

From this point of view, it is expected to require an high level of autonomy to the UxV considering the risk to loose contact with central control.

In IDRASS also the issue related to battery autonomy are raised, considering that to move within an industrial plant and to carry out data collection and sensor measures it could challenging to have time to complete the whole mission, especially when moving indoor and/or in confined spaces. Finally the industrial plants include presence of humans and the UxV should be able to operate avoiding to injure them by collision or indirectly by generating other accidents.

3.2 T-REX

In references to the necessity to address On-Shore Critical Infrastructure Protection, Simulation Team developed an innovative simulation defined T-REX (Threat network simulation for REactive eXperience) that recreates, into a virtual immersive interactive environment, these complex scenarios (Bruzzone et al. 2016b). T-REX simulator has been used to address different problems and involves combined used of UAV, UGV, USV and AUV to protect an industrial complex that provide strategic services (i.e. power, water, oil) to a region involving different towns as proposed in figures 2, 3 and 4.

The multi domain coordination is very crucial in critical infrastructure protection that are located in coastal area requiring to take care of threats from air, ground, sea surface and underwater. In this extended maritime framework, it is fundamental to develop coordination capabilities (Tether 2009).

Very promising results have been achieved, along last decade, in UAV coordination arriving to complete successfully complex operations such as air refueling (Richards et al. 2002; Ross 2006).

Therefore today one major challenge is related to create a network of collaborating autonomous systems that operate over different domains (Vail & Veloso 2003; Tanner 2007a; Shkurti et al. 2009; Maravall et al. 2013). In facts in order to be successful, it becomes necessary to develop a capability to control an heterogeneous network of different resources (Feddema et al. 2002; Ferrandez et al. 2013; Bruzzone et al. 2013c).

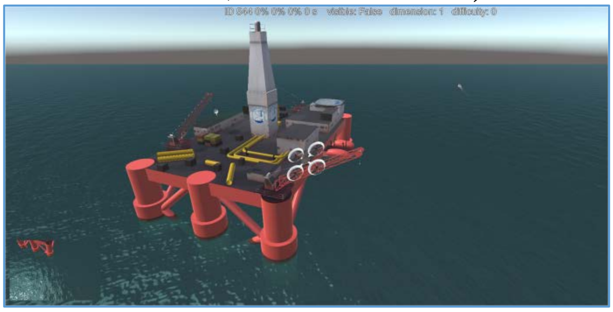


Figure 6 – SO2UCI with Drone supervision of the Off-Shore Platform

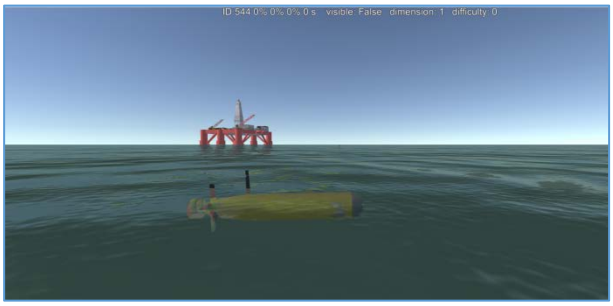


Figure 7 – AUV in SO2UCI patrolling around the Off-Shore Platform

From this point of view, an interesting idea is to assign the role of supervisor to humans instead than directly making them to pilot each single UxV as expressed by the man-on-the-loop concept supported by SPIDER solution as seen in figure 5 (Magrassi 2013).
This is a great opportunity, reducing drastically the number of people required to support complex operations, but also a challenge for safety and security and introduce new challenges in terms of technological solutions and training (Bruzzone et al. 2016a). In the case of critical infrastructure protection a major issue is related to false alarms; indeed, the large majority of suspect events are related to case dealing with harmless conditions due to general traffic, civilian unintentional infractions, birds or animal misclassification, etc (Bass 2000; Cardenas et al.2011). In addition to these considerations it is also possible to face sensor errors and failures that could create critical conditions in the protection systems (Merabti et al.2011).
In facts, this problem requires to cover 24/7 in presence even of challenging weather and boundary conditions (Kastek et al.2012).
Indeed the use of UxV could be useful to reinforce the protection of critical infrastructures, considering that could be a robust solution supported by multiplatform, multisensory data fusion and that could allow to conduct further investigations directly approaching to the alerts in order to discriminate real threats from false alarms.
Obviously these elements suggest specific requirements in terms of collaboration capabilities, redundancy and responsiveness of the multi UxV system.

3.3 SOU2CI

Operation on board of Off-Shore Platforms are very complex and the environment inside and outside the rig is very challenging due to the complexity of the platform and to the extreme boundary conditions where it operates (e.g. sea and weather); in this case the use of USV, UAV and AUV/UUV, in combined way, could be very useful (see figures 6 & 7); due to these reasons it is interesting to investigate the capabilities of heterogeneous networks of autonomous systems within this application field (Bruzzone et al. 2013c).
Indeed the coordination among UxV operating in different domains is a crucial element for addressing these kinds of complex missions (Stilwell et al. 2004; Shafer et al. 2008; Tanner et al. 2007b.
In facts in this case the use of AUV coordinated with air assets is fundamental (Grocholsky et al. 2006; Bruzzone et al.2016c); therefore the use of UUV (Underwater Unmanned Vehicles), such as gliders, introduces additional challenges due to the difficulties related to the underwater communications (Jans et al. 2006).
The collaboration capabilities are strongly related with the introduction of advanced AI solution and potentially by the human on the loop concept already mentioned (Sujit et al.2009; Bruzzone et al.2013c).

4 CONCLUSIONS

The proposed examples represent already a quite interesting set of cases where safety issues that need to be addressed in order to guarantee diffusion of the UxVs in these new roles and application fields.
Obviously, it is challenging to investigate all the different elements that concur in safety as well as the effectiveness of policies and technologies to prevent and mitigate risks.
Therefore M&S is for sure the most promising technique being able to run large and complex scenarios affected by stochastic factors involving new applications of UxVs and it is expected that these simulation environments will further evolve by applying intensively MS2G paradigm.
This approach is crucial to support the development of intensive collaboration among the different stakeholders in this context over a common and understandable virtual experimentation framework.

REFERENCES

Altawy, R., & Youssef, A. M. (2016) "Security, Privacy, and Safety Aspects of Civilian Drones: A

Survey" ACM Transactions on Cyber-Physical Systems, 1(2), 7
Apvrille, L., Roudier, Y., & Tanzi, T. J. (2015) "Autonomous drones for disasters management: Safety and security verifications", Proc. 1st IEEE URSI Atlantic, May
Bass, T. (2000) "Intrusion detection systems and multisensor data fusion" Communications of the ACM, 43(4), pp.99-105
Bruzzone A.G., Longo F., Massei M., Nicoletti L., Agresta M., Di Matteo R., Maglione G.L., Murino G., Antonio Padovano A. (2016a) “Disasters and Emergency Management in Chemical and Industrial Plants: Drones simulation for education & training”, Proc. of MESAS, Rome, June 15-16
Bruzzone A.G., Massei M., Longo F., Cayirci E., di Bella P., Maglione G.L., Di Matteo R. (2016b) “Simulation Models for Hybrid Warfare and Population Simulation”, Proc. of NATO Symposium on Ready for the Predictable, Prepared for the Unexpected, M&S for Collective Defence in Hybrid Environments and Hybrid Conflicts, Bucharest, Romania, October 17-21
Bruzzone A.G., Massei M., Maglione G.L., Di Matteo R., Franzinetti G. (2016c) “Simulation of Manned & Autonomous Systems for Critical Infrastructure Protection”, Proc. of I3M, Larnaca, Cyprus, September
Bruzzone A.G., Massei M., Agresta M., Poggi S., Camponeschi F., Camponesch M. (2014) “Addressing Strategic Challenges on Mega Cities through MS2G”, Proceedings of MAS, Bordeaux, France, September 12-14
Bruzzone, A.G., Berni, A., Fontaine, J.G., Cignoni, A., Massei, M., Tremori, A., Dallorto, M., Ferrando, A. (2013c) “Virtual Framework for Testing/Experiencing Potential of Collaborative Autonomous Systems”, Proc. of I/ITSEC, Orlando. FL USA
Bürkle, A., Segor, F., Kollmann, M. (2011). "Towards autonomous micro uav swarms". *Journal of intelligent & robotic systems*, *61*(1-4), pp. 339-353.
Cárdenas, A. A., Amin, S., Lin, Z. S., Huang, Y. L., Huang, C. Y., & Sastry, S. (2011) "Attacks against process control systems: risk assessment, detection, and response", Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, March, pp.355-366
Clarke, R., & Moses, L. B. (2014) “The Regulation of Civilian Drones' Impacts on Public Safety” Computer Law & Security Review, 30(3), 263-285
Di Donato (2017) “Intelligent Systems for Safety of Industrial Operators, the Role of Machines & Equipment Laboratories”, SISOM Workshop, Rome
Djellal, F., & Gallouj, F. (1999) “Services and the search for relevant innovation indicators: a review of national and international surveys”, Science and Public Policy, 26(4), pp. 218-232
Doherty, P., & Rudol, P. (2007) "A UAV search and rescue scenario with human body detection and geolocalization", Proceedings of the Australian Conference on Artificial Intelligence, Vol. 4830, December, pp. 1-13
Feddema, J.T.; Lewis, C.; Schoenwald, D.A., (2002) "Decentralized control of cooperative robotic vehicles: theory and application, "Robotics and Automation, IEEE Transactions on, vol.18, no.5, pp.852,864, Oct
Ferrandez J.M., De Lope H., De la Paz, F. (2013) “Social and Collaborative Robotics”, International Journal Robotics and Autonomous Systems, 61
Floreano, D., & Wood, R. J. (2015) “Science, technology and the future of small autonomous drones”, Nature, 521(7553), 460
Gardi, A., Sabatini, R., & Ramasamy, S. (2016) “Stand-off measurement of industrial air pollutant emissions from unmanned aircraft”, Proc. of IEEE International Conference on Unmanned Aircraft Systems, June , pp.1162-1171
Grocholsky, B., Keller, J., Kumar, V., Pappas, G., (2006) "Cooperative air and ground surveillance", Robotics & Automation Magazine, IEEE, vol.13, no.3, Septmber, pp.16-25
Ishiki, T., & Kumon, M. (2014) "A microphone array configuration for an auditory quadrotor helicopter system", Proc. IEEE International Symposium on Safety, Security, and Rescue Robotics, pp. 1-6
Jans, W., Nissen, I., Gerdes, F., Sangfelt, E., Solberg, C. E., & van Walree, P. (2006) “UUV covert acoustic communications– preliminary results of the first sea experiment”, in Techniques and technologies for unmanned autonomous underwater vehicles– a dual use view”, RTO Workshop SCI-182/RWS-016, Eckernförde, Germany
Jones, D. (2005) “Power line inspection-a UAV concept”, Proc. of the IEE Forum on Autonomous Systems, Ref. No. 11271, November
Kastek, M., Dulski, R., Zyczkowski, M., Szustakowski, M., Trzaskawka, P., Ciurapinski, W., Grelowska G., Gloza I., Milewski S, Listewnik, K. (2012) “Multisensor system for the protection of critical infrastructure of seaport” In Proc. of SPIE, Vol. 8288, May
Kehoe, B., Patil, S., Abbeel, P., & Goldberg, K. (2015) "A survey of research on cloud robotics and automation", IEEE Transactions on automation science and engineering, 12(2), pp.398-409
Kim, D. H., Kwon, S. W., Jung, S. W., Park, S., Park, J. W., & Seo, J. W. (2015) "A Study on Generation of 3D Model and Mesh Image of Excavation Work using UAV", Proceedings of the International Symposium on Automation and Robotics in Construction, Vol. 32, Vilnius, January
Kovacevic, M. S., Gavin, K., Oslakovic, I. S., & Bacic, M. (2016). "A new methodology for assessment of railway infrastructure condition". *Transportation research procedia 14*, pp. 1930-1939.
Leão, D. T., Santos, M. B. G., Mello, M. C. A., & Morais, S. F. A. (2015) "Consideration of

occupational risks in construction confined spaces in a brewery", Occupational Safety & Hygiene III, 343
Magrassi C. (2013) “Education and Training: Delivering Cost Effective Readiness for Tomorrow's Operations“,ITEC Keynote Speech, Rome, May
Maravall D., J. de Lopea,b, R. Domíngueza, (2013) “Coordination of communication in robot teams by reinforcement learning”, Robotics and Autonomous Systems 61, pp.661–666
McCurry Justin (2017) "Dying robots and failing hope: Fukushima clean-up falters six years after Tsunami", The Guardian, March 9
Merabti, M., Kennedy, M., & Hurst, W. (2011) “Critical infrastructure protection: A 21 st century challenge”, Proc. of IEEE Int.Conf. on Communications and Information Technology, ICCIT, March, pp. 1-6
Merwaday, A., & Guvenc, I. (2015) "UAV assisted heterogeneous networks for public safety communications", Proc. of IEEE Wireless Communications and Networking Conference Workshops, March, pp. 329-334
Mobley, R. K. (2001) “Plant engineer's handbook”, Butterworth-Heinemann, Oxford, UK
Nano, G., & Derudi, M. (2013) "A critical analysis of techniques for the reconstruction of workers accidents", Chemical Engineering, 31
Palazzi, E., Caviglione, C., Reverberi, A.P., Fabiano, B. (2017) “A short-cut analytical model of hydrocarbon pool fire of different geometries, with enhanced view factor evaluation”, Process Safety and Environmental Protection, August
Pizzella, L. A. E. (2014) “Contributions to the Configuration of Fleets of Robots for Precision Agriculture”, Thesis, Universidad Complutense, Madrid, Spain, May
Pulina, G., Canalis, C., Manni, C., Casula, A., Carta, L. A., & Camarda, I. (2016) “Using a GIS technology to plan an agroforestry sustainable system in Sardinia”, Journal of Agricultural Engineering, 47(s1), 23-23.
Richards A., J. Bellingham, M. Tillerson, and J. P. (2002) “How: Co-ordination and control of multiple UAVs”, Proc. of the AIAA Guidance, Navigation, and Control Conference, Monterey, CA, August
Ross, S., D. Jacques, M. Pachter, and J. Raquet, (2006) “A Close Formation Flight Test for Automated Air Refueling,” Proceedings of ION GNSS-2006, Fort Worth, TX, Sep
Salvini, P. (2017) “Urban robotics: Towards responsible innovations for our cities”, Robotics and Autonomous Systems, Elsevier
Sanchez-Lopez, J. L., Pestana, J., de la Puente, P., & Campoy, P. (2016) "A reliable open-source system architecture for the fast designing and prototyping of autonomous multi-uav systems: Simulation and experimentation", Journal of Intelligent & Robotic Systems, 84(1-4), pp.779-797
Shafer, A.J., Benjamin, M.R., Leonard, J.J., Curcio, J., (2008) "Autonomous cooperation of heterogeneous platforms for sea-based search tasks", Oceans, , September 15-18, pp. 1-10
Shkurti, F., Anqi Xu, Meghjani, M., Gamboa Higuera, J.C., Girdhar, Y., Giguere, P., Dey, B.B., Li, J., Kalmbach, A., Prahacs, C., Turgeon, K., Rekleitis, I., Dudek, G., (2012)"Multi-domain monitoring of marine environments using a heterogeneous robot team", Proc. of IEEE Intelligent Robots and Systems (IROS), vol., no., pp.1747,1753, October 7-12
Siebert, S., & Teizer, J. (2014) “Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system”, Automation in Construction, 41, pp.1-14
Spanu S., M. Bertolini, E. Bottani, G. Vignali, L. Di Donato, A. Ferraro, F. Longo (2016) "Feasibility study of an Augmented Reality application to enhance the operators’ safety in the usage of a fruit extractor", Proc. FoodOPS, Larnaca, Cyprus, September 26-28
Spillane, J. P., Oyedele, L. O., & Von Meding, J. (2012) "Confined site construction: An empirical analysis of factors impacting health and safety management", Journal of Engineering, Design and Technology, 10(3), pp.397-420
Stilwell D. J., A. S. Gadre, C. A. Sylvester and C. J. Cannell (2004) “Design elements of a small low-cost autonomous underwater vehicle for field experiments in multi-vehicle coordination”, Proc. of the IEEE/OES Autonomous Underwater Vehicles, June, pp. 1-6
Sujit, P. B., Sousa, J., Pereira, F.L., (2009) "UAV and AUVs coordination for ocean exploration", Oceans - EUROPE, vol., no., pp.1,7, 11-14 May
Tanner H. G. (2007a) “Switched UAV-UGV cooperation scheme for target detection”, IEEE International Conference on Robotics and Automation, Roma, Italy, April, pp. 3457-3462.
Tanner H.G., D.K. Christodoulakis, (2007b) “Decentralized cooperative control of heterogeneous vehicle groups”, Robotics and Autonomous Systems 55,pp 811–823
Tether, T. (2009) "Darpa Strategic Plan", Technical Report DARPA, May
Vail D. & M. Veloso, (2003) “Dynamic multi-robot coordination”, Multi-Robot Systems: From Swarms to Intelligent Automata, Vol II, pp. 87-100.
Valavanis, K. P., & Vachtsevanos, G. J. (2014) "Handbook of unmanned aerial vehicles", Springer Publishing Company, NYC

AUTHORS' INDEX

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.