THE 12TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION

SEPTEMBER 25-27 2013

ATHENS, GREECE

EDITED BY

MICHAEL AFFENZELLER

AGOSTINO G. BRUZZONE

FABIO DE FELICE

DAVID DEL RIO VILAS

CLAUDIA FRYDMAN

MARINA MASSEI

YURI MERKURYEV

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2013

ISBN 978-88-97999-17-1 (Paperback)

ISBN 978-88-97999-23-2 (PDF)

ISBN 978-88-97999-17-1 (Paperback)
ISBN 978-88-97999-23-2 (PDF)

THE 12TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION

SEPTEMBER 25-27 2013, ATHENS, GREECE

ORGANIZED BY

DIME – UNIVERSITY OF GENOA

LIOPHANT SIMULATION

SIMULATION TEAM

IMCS – INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF SIMULATION

DIMEG, UNIVERSITY OF CALABRIA

MSC-LES, MODELING & SIMULATION CENTER, LABORATORY OF ENTERPRISE SOLUTIONS

MODELING AND SIMULATION CENTER OF EXCELLENCE (MSCOE)

LATVIAN SIMULATION CENTER – RIGA TECHNICAL UNIVERSITY

LOGISIM

LSIS – LABORATOIRE DES SCIENCES DE L'INFORMATION ET DES SYSTEMES

MIMOS – MOVIMENTO ITALIANO MODELLAZIONE E SIMULAZIONE

MITIM PERUGIA CENTER – UNIVERSITY OF PERUGIA

BRASILIAN SIMULATION CENTER, LAMCE-COPPE-UFRJ

MITIM - MCLEOD INSTITUTE OF TECHNOLOGY AND INTEROPERABLE MODELING AND SIMULATION – GENOA CENTER

M&SNET - MCLEOD MODELING AND SIMULATION NETWORK

LATVIAN SIMULATION SOCIETY

ECOLE SUPERIEURE D'INGENIERIE EN SCIENCES APPLIQUEES

FACULTAD DE CIENCIAS EXACTAS. INGEGNERIA Y AGRIMENSURA

UNIVERSITY OF LA LAGUNA

CIFASIS: CONICET-UNR-UPCAM

INSTICC - INSTITUTE FOR SYSTEMS AND TECHNOLOGIES OF INFORMATION, CONTROL AND COMMUNICATION

NATIONAL RUSSIAN SIMULATION SOCIETY

CEA - IFAC

AFCEA, HELLENIC CHAPTER

TECHNICALLY CO-SPONSORED

IEEE – CENTRAL AND SOUTH ITALY SECTION CHAPTER

I3M 2013 INDUSTRIAL SPONSORS

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2013 MEDIA PARTNERS

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SIMULATION AND PROCESS MODELING

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF CRITICAL INFRASTRUCTURES

Inderscience Publishers – International Journal of Engineering Systems Modelling and Simulation

Inderscience Publishers – International Journal of Service and Computing Oriented Manufacturing

IGI GLOBAL – International Journal of Privacy and Health Information Management

Halldale Media Group: Military Simulation and Training Magazine

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

Euromerci

EDITORS

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

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

FABIO DE FELICE
UNIVERSITY OF CASSINO, ITALY
defelice@unicas.it

DAVID DEL RIO VILAS
UNIVERSITY OF A CORUNA, SPAIN
daviddelrio@udc.es

CLAUDIA FRYDMAN
LABORATOIRE DES SCIENCES DE L'INFORMATION ET DES SYSTEMES, FRANCE
Claudia.frydman@lsis.org

MARINA MASSEI
LIOPHANT SIMULATION, ITALY
massei@itim.unige.it

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

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2013

GENERAL CO-CHAIRS

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

PROGRAM CHAIR

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

THE 12TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION

GENERAL CO-CHAIRS

MICHAEL AFFENZELLER, *UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA, AUSTRIA*
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*

PROGRAM CO-CHAIRS

FABIO DE FELICE, *UNIVERSITY OF CASSINO, ITALY*
DAVID DEL RIO VILAS, *UNIVERSITY OF A CORUNA, SPAIN*

MAS 2013 International Program Committee

Michael Affenzeller, *Upper Austrian Univ. of AS, Austria*
Thècle Alix, *IMS Université Bordeaux 1, France*
Gabriel Aprigliano Fernandes, *LAMCE COPPE UFRJ, Brazil*
Naamane Aziz, *LSIS, France*
Eleonora Bottani, *University of Parma, Italy*
Agostino Bruzzone, *University of Genoa, Italy*
José M. Cecilia, *Universidad Católica San Antonio, Spain*
Diego Crespo Pereira, *University of A Coruna, Spain*
Gerson Cunha, *LAMCE COPPE UFRJ, Brazil*
Fabio De Felice, *University of Cassino, Italy*
David Del Rio Vilas, *University of La Coruna, Spain*
Claudia Frydman, *LSIS, France*
Andrea Grassi, *University of Modena and Reggio Emilia, Italy*
Yilin Huang, *Delft University of Technology, Netherlands*
Janos Sebestyen Janosy, *MTA KFKI Atomic Energy, Hungary*
Pasquale Legato, *University of Calabria, Italy*
Francesco Longo, *MSC-LES, University of Calabria, Italy*
Marina Massei, *Liophant Simulation, Italy*
Rina Mary Mazza, *University of Calabria, Italy*
Roberto Montanari, *University of Parma, Italy*
Letizia Nicoletti, University of Calabria, Italy
Federica Pascucci, *University of Roma 3, Italy*
Horacio Emilio Pérez Sánchez, *Universidad de Murcia, Spain*
Antonella Petrillo, *University of Cassino, Italy*
Gabriele Oliva, Campus Bio-Medico di Roma, Italy
NAdia Rego Monteil, *University of a Coruna, Spain*
Rosa Rios Prado, *University of a Coruna, Spain*
Adriano Solis, *York University, Canada*
Young Jun Son, University of Arizona, USA
Alberto Tremori, *University of Genoa, Italy*
Giuseppe Vignali, *University of Parma, Italy*
Gregory Zacharewicz, *IMS Université Bordeaux 1, France*

Tracks and Workshop Chairs

Inventory Management Simulation
Chairs: Adriano Solis, York University, Canada; Letizia Nicoletti, University of Calabria, Italy

Production Systems Design
Chair: David del Rio Vilas, University of La Coruna, Spain;

Decision Support Systems Applications
Chairs: Fabio De Felice, Univerisity of Cassino, Italy; Antonella Petrillo,Univerisity of Cassino, Italy

Simulation in Energy Generation and Distribution
Chair: Janos Sebestyen Janosy, MTA KFKI Atomic Energy Research Institute, Hungary

Promoting Enterprise Interoperability by Service Modeling & Simulation
Chairs: Thècle Alix, IMS Université Bordeaux 1, France; Gregory Zacharewicz, IMS Université Bordeaux 1, France

Workshop on Modeling and Simulation of Food Processing and Operations
Chairs: Andrea Grassi, University of Modena and Reggio Emilia, Italy; Giuseppe Vignali, University of Parma, Italy, Eleonora Bottani, University of Parma, Italy; Roberto Montanari, University of Parma, Italy

Simulation Based Design
Chairs: Yilin Huang , Delft University of Technology, Netherlands

Automation
Chair: Naamane Aziz, Laboratoire des Sciences de l'Information et des Systèmes, France

Workshop on virtual and augmented reality
Chairs: Gerson Cunha, LAMCE/COPPE/UFRJ, Brasil; Gabriel Aprigliano Fernandes, LAMCE/COPPE/UFRJ, Brasil

Simulation and Human Factors Engineering
Chair: Diego Crespo Pereira, University of La Coruna

Simulation based Optimization
Chairs: Pasquale Legato, University of Calabria, Italy; Rina Mary Mazza, University of Calabria, Italy

High Performance Simulation of biological systems
Chairs: Horacio Pérez-Sánchez, Universidad Católica San Antonio, (Murcia) Spain ; José M. Cecilia Universidad Católica San Antonio, (Murcia) Spain

General Co-Chairs' Message

Welcome To MAS 2013!

On behalf of the International Program Committee it is our pleasure to present to you the proceedings of the 12th International Conference on Modeling and Applied Simulation – MAS 2013.
MAS was established in 2002 as a joint event to HMS2002 Bergeggi, Italy in order to provide an opportunity for Project Meetings (i.e. IEPAL, LESNEX, WILD). Several editions of MAS events were held during the past years most of them within the framework provided by the International Mulltidisciplinary Modeling & Simulation Multiconference, I3M (since 2004).
The organizational structure of MAS 2013 is similar to other international scientific conferences. The backbone of the conference is the scientific program, which is complemented by workshops and open debates dedicated to special topics from the field of applied Modeling & simulation (M&S) and computer technologies; application fields include logistics, supply chain management, production control, business and industrial organization. Very recent trends in M&S focus on the integration of multiple components along the supply chain including predictive modeling techniques, simulation based optimization as well as meta- and metaheuristics for optimization. These recent trends are reflected in many of the MAS 2013 contributions.
We are very pleased to report that about 30 papers from more than 20 different countries were selected by the program committee for presentation and discussion during the conference and subsequent publication in the conference proceedings and international journals special issues (as part of the I3M Multiconference, best papers of MAS will have the opportunity to be selected and published in 7 different international journals special issues).
All submissions were reviewed by an international panel of at least 2 expert referees. We acknowledge the invaluable assistance of the program committee and the international referees, who opted to provide detailed and insightful comments to the authors.
Bridging the gap between theory and practice and incorporating theoretical results into useful products is still one of the main key issues for industrialized countries. Especially in the context of modeling and applied simulation it seems essential that researchers accept the challenge of solving real-world problems, making the science and technology based on mathematics, computer science and computational intelligence contribute to the progress of our developing communities.
MAS 2013 brings together fundamental and applied research contributions that reflect current trends and developments in modeling and applied simulation. We sincerely hope that the conference will encourage cross-fertilization between the various communities, bridging not only the gap between different fields but also, equally important, between different continents and cultures.
In closing, we would like to thank all authors for submitting their works, and all members of the Program Committee, listed on the following page, who contributed greatly to make the conference a success.
We wish you a fruitful and inspiring conference and a pleasant stay in Athens.

Michael Affenzeller
University of Applied Sciences, Upper Austria, Austria

Marina Massei
Liophant Simulation, Italy

Fabio De Felice
University of Cassino, Italy

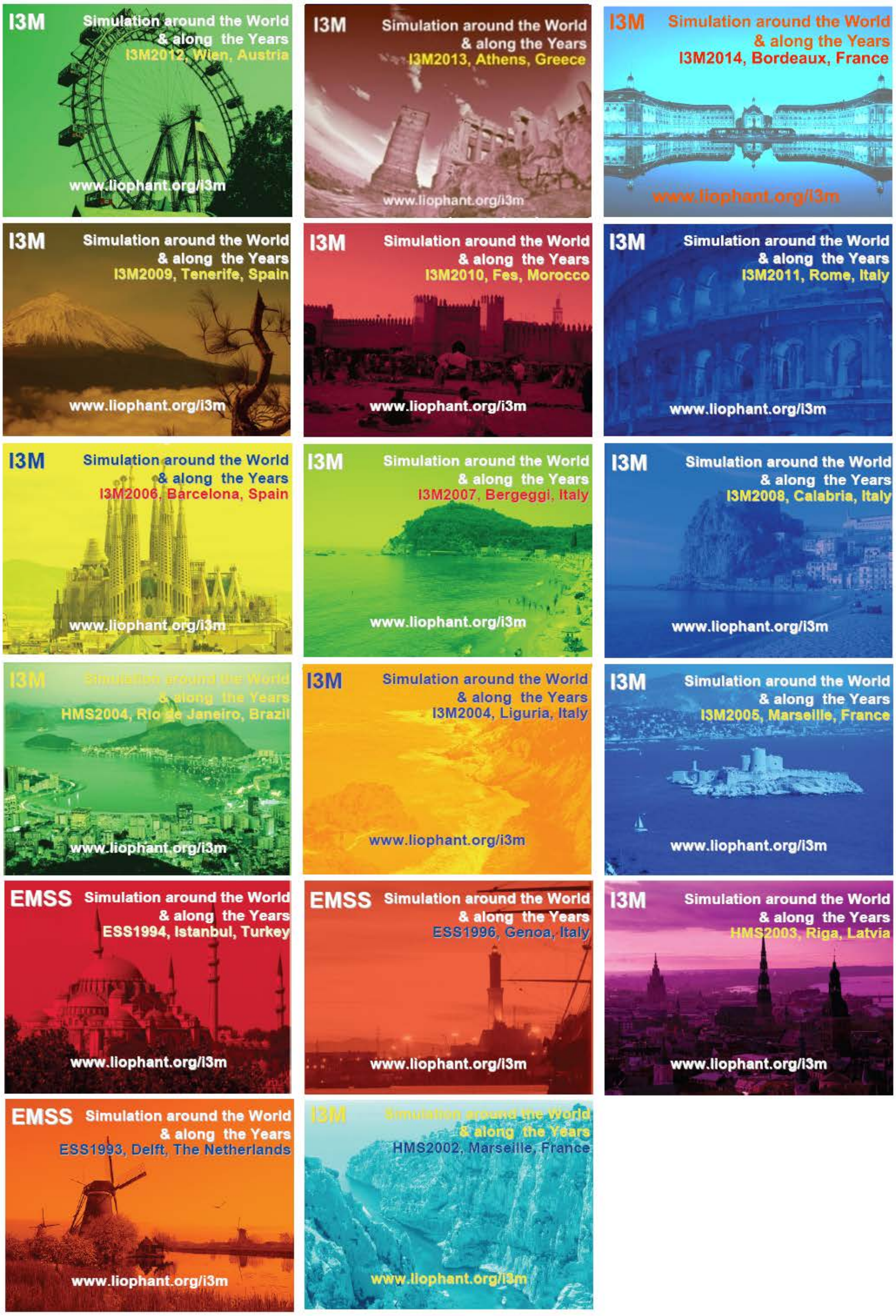
David Del Rio Vilas
University of a Coruna, Spain

Acknowledgements

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

Local Organization Committee

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



Index

PROPOSAL OF AGENT SIMULATION METHODOLOGY FOR THE PROSPECTIVE ANALYSIS OF MINERAL COMMODITIES MARKETS

Fenintsoa Andriamasinoro[(a)], Bruno Martel-Jantin[(b)]

[(a)(b)] BRGM - 3 Avenue Claude-Guillemin - BP 36009 - 45060 Orléans Cedex 2 - France

[(a)]f.andriamasinoro@brgm.fr, [(b)]b.martel-jantin@brgm.fr

ABSTRACT
The markets of mineral commodities for industrial use (MCI) risk supply shortages in the near future due to a possible restriction policy applied by producers. Therefore, the French government is not reassured because, over the coming decades, such situations may affect French industrial sectors using these products. By taking the world lithium market as an application example, this work aims to contribute to the elaboration of a multi-agent system (MAS) prospective tool, which allows decision makers to evaluate possible supply shortage periods in the world market and in France. The discussion is extended to the aggregates market, another kind of MCI market. The work also aims to evaluate to what extent MAS methodology is accepted in the literature regarding MCI system prospective analysis. This work concludes that a MAS approach could provide a new methodology for analysing MCI markets. However, convincing MCI sectors to use it remains a challenge.

Keywords: mineral commodities markets, multi-agent simulation, lithium market, aggregate resources market

1. INTRODUCTION

1.1. Thematic issues
The markets of mineral commodities for industrial use (MCI) risk supply shortages in the near future. This is true for aggregate resources (AR) markets (a subdivision of MCI markets), e.g. in France (Rodriguez-Chavez 2010) or in the UK (Brown, McEvoy and Ward 2011), and also for metal markets (another subdivision of MCI markets), regarding products such as lithium, indium or rare earths.

In AR markets (in which exchanges often occur at a regional/national scale), the main reason for such risks is objections to mining developments stemming from the perception of negative environmental and socioeconomic effects on surrounding communities and ecosystems (Graedel, et al. 2012). As for metal markets (in which exchanges occur at a national/global scale), risks would dramatically increase in the event of drastic changes to mining or commercial policies towards more restriction of the exportation quota by countries that currently dominate the world market, such as Chile for lithium (Daw and Labbé 2012) or China for rare earths (Roskill 2011, Giacalone 2012). Whereas in Chile this situation is still hypothetical but not impossible, in China such a restriction is already effective.

Given the above situations, the French government is not reassured because in coming decades, such situations may affect French industrial sectors using these products. These sectors are (for AR markets) building and public works or (for the metal markets) automobiles, glass, etc. Thus, in order to successfully implement a policy to deal with these situations, the government (in association with regional authorities in the case of AR markets) suggests the implementation of prospective tools that would help French industrialists regarding the choices to be made towards their future supplies. Such a tool is also expected by the authorities regarding future orientations they may undertake in the French industrial sector.

1.2. Objectives of the work
The objective of the work reported in this paper is twofold and concerns a thematic level and a methodological level.

By taking the world lithium market as an application example, the work at a thematic level aims to contribute to the elaboration of a prospective tool that allows the above actors to answer the following question: *given the uncertainty of supply, how long would a lithium supply shortage last (should the case arise) in the world market as well as in France?* The approach consists, via modelling and simulation by a multi-agent system (MAS) approach (Wooldridge 2009), in creating prospective scenarios of supply shortage in the lithium market due to restrictions decided by a producer, then (a) identifying the set of likely shortage periods that would correspond to these respective scenarios and (b) searching possible alternative supply scenarios to compensate the resulting shortages.

At a methodological level, the aim of the work is to evaluate and discuss the possible interests of applying the MAS approach to MCI applications and to what extent it is currently used in the literature regarding MCI prospective analysis via modelling/simulation. The aim is to show the possibility or not of a methodological transposition of the work on lithium, our application example, to other substances.

2. STATE OF THE ART

2.1. Presentation

In MCI markets and regarding economic prospective market analysis via modelling/simulation, various studies have already been carried out to deal with the supply shortage issue.

Regarding the lithium market, these studies were carried out either by the academic world (Yaksic and Tilton 2009, Gruber, et al. 2011), by lithium consultants (Roskill 2009), by banks ((McNulty and Khaykin 2009), on behalf of the Credit Suisse), or by producing companies ((De Solminihac 2010), for the Chilean company SQM, the current leading lithium producer in the world). In all these studies, production and consumption were respectively extrapolated in an independent manner and the results next compared arithmetically. Thus, there was no mutual driving between the evolution of the supply and the demand values. Furthermore, all the works adopted a global scale as the level of their studies. This approach is the same for metals other than lithium, such as rare earths (Roskill 2011) or copper and magnesium (Andriamasinoro and Angel 2012). In fact, initiating this kind of prospective analysis at a global level is necessary because mineral resources are unequally spread out over the Earth as a whole. However, this is not sufficient. Indeed, efforts to explore the criticality of metals should not consider only the global level, because organizational differences make a uniform analytical approach for all organizational (i.e., global, national and local) levels impractical (Graedel, et al. 2012). In the same way, the risks of distribution may be underappreciated when discussing resources at a global level (Kushnir and Sandén 2012). In particular, knowledge of the quantity available at the global level does not automatically imply that of the distribution per country. Likewise, if the period of likely shortage is known at a global level, nothing says that for a given consuming country it will be the same, since supply behaviour at a production side varies from one producing country to another, depending on its individual and collective interests (Andriamasinoro and Ahne 2013).

As for aggregate resource (AR) markets, the same situation can be observed at a national/regional scale. An example concerns a model called *Antag*, which analysed the supply shortage in AR on the French market: *Antag* was based on a dynamic systems approach (Rodriguez-Chavez 2010). The prospective analysis contented itself with observing the global market flow in France, whereas it has been known for a long time that the opening of production zones is decided at a regional (i.e. more detailed) subdivision level.

As a matter of fact, in MCI, global elements are important indications but need to be refined.

2.2. Proposal

Given this refinement objective, our proposal thus consists in making detailed scales of MCI markets more explicit, i.e. where it would be possible to better evaluate the impacts of the individual behaviour and constraints of producers on the supply shortage periods (if any) of consumers. This is important for the government of a consuming country, especially in a restriction policy context. In metal markets, it consists in passing from a purely world to a more national scale, where the interaction between countries is modelled. Likewise, in AR markets, it consists in passing from a national to a more regional level. Regarding this purpose, the MAS approach is suggested given its capability to represent the complexity of a system at any scale of a territory (Wooldridge 2009). This proposal follows the paradigm of Arthur, Durlauf and Lane (1997) stipulating that what happens in the market economy is actually determined by the interaction of dispersed heterogeneous agents, acting in parallel.

As one may see, the idea of using MAS to model a market is not new in itself (here, 1997). The approach has however never been considered by the MCI field. This absence of MAS in the field insofar can be explained by at least three reasons: (1) the method is not known by the studied field, (2) the method is known (because MAS has existed since the 1990s) but its use has always been considered as not indispensable or (3) its use has been considered as indispensable but its development is too difficult.

As announced in the introduction, the world lithium market has been chosen as an application example, given the importance of this metal in electric vehicle batteries (Gruber, et al. 2011). However, as the discussion throughout the paper will concern MCI markets in general, not only metal markets, the AR market situation, at a regional/national scale, will be also resumed in the discussion section (Section 5.2).

3. MODELLING OF THE EXAMPLE

3.1. Data sources

This lithium example uses international trade data from (GTIS 2012) as data sources. The GTIS data presents flows between producing countries and transit countries (i.e. countries connecting producers and consumers) as well as between transit countries and consuming countries. It should be noted that for various reasons (administrative, geographical, etc.), a given consumer can be supplied by the same producer via several transits.

The chosen data are on a quarterly timescale. This scale was preferred to an annual timescale because it increases the number of observations during the statistical tests.

Finally, the prospective period of the simulation here begins in 2013. The historic period is situated between 2005 and 2012, a period when the lithium data necessary for this work are available.

3.2. Hypotheses

As for the hypotheses of the model, the settings below have been adopted.

First, in order to better exploit the GTIS data, our market model will integrate not just producing and consuming countries obviously but also transit countries and will work "as if" consumers send their demands to transits even if, in reality, they directly address producers, which then send the product to the transits.

Second, for the moment, we only focus on the market of lithium carbonate (Li_2CO_3) and not on the other lithium compound markets such as lithium hydroxide (LiOH) or lithium chloride (LiCl).

Third, even if we integrate producing countries, we do not use either export or production figures. Indeed, in the GTIS data, the figures for exports and imports at a given time are not always identical for reasons of transport delay, administrative procedures, etc. Thus, we only say that a producing country *supplies* the quantity effectively received by a consuming country at that time (after possibly having applied preliminary restrictions, during the scenarios).

Fourth, in the model to be built, all chosen producing countries supply all chosen consuming countries, but with (obviously) different quantities, including 0.

3.3. Modelling

A country is modelled as an agent, which is either a producer (*pc*), a consumer (*cc*), or a transit (*tc*). The system contains η_{pc} producers, η_{cc} consumers and η_{tc} transits. A country may be in the following context: *Normal*, *Restriction*, *Compensation*, *Waiting* and *Maintenance*. At the beginning of the simulation, each country is in a *Normal* context.

The model also integrates agents called *ambassadors*. An ambassador $A(c1 \leftrightarrow c2)$ is a delegate agent that handles the flow exchanged between the countries *c1* and *c2*. The concept of ambassadors has been introduced because, given the complexity of the internal and external behaviours of a country (as will be detailed later) and the topology of the system in general, it seems difficult for us to describe the exchanges between *c1* and *c2*, in *c1* and *c2* at the same time; especially that, for a given *c1*, the mode of calculation of the exchanges changes from one *c2* to another. A decentralisation (delegation) of the description of the exchanges consequently seemed more appropriate to us.

Finally, communication between the agents is formally done via the exchange of events. They take the form *event*(*s*, *r*, <*q1*, …, *qn*>) where *s* is the sender, *r* the recipient and <*q1*, …, *qn*> a list of values to transfer from *s* to *r*.

3.3.1. Formalisation of a normal context

The normal context is the market context of a supply without restriction. Here, all agents are in a *Normal* context. In this context, the interaction between countries and ambassadors, at each time step of the market simulation, occurs by following the four stages below, in which the first two points concern the demand stage and the last two points concern the supply stage.

<u>Stage 1</u>: At the beginning of a time step, each consumer cc_k, $k \in \{1, \eta_{cc}\}$, asks its ambassadors $A(tc_j \leftrightarrow cc_k)$, $j \in \{1, \eta_{tc}\}$ to calculate the quantity $d(tc_j \leftarrow cc_k)$ to demand from all producers pc_i, $i \in \{1, \eta_{pc}\}$, the supply of which will next transit via the country tc_j. Once each $d(tc_j \leftarrow cc_k)$ is calculated, each $A(tc_j \leftrightarrow cc_k)$, $k \in \{1, \eta_{cc}\}$ sends that demand to tc_j. For now, a demand over time is calculated via two steps.

The first step consists in interpolating the time series $S_{d(tc_j \leftarrow cc_k)}$ of the GTIS data related to the demands from cc_k to tc_j between 2005 and 2012, in order to obtain a regression line, which would describe and prolong that demand evolution. Let us note this interpolated value $i(tc_j \leftarrow cc_k)$. Its evolution may take a linear, logarithmic, exponential or average shape.

The second step consists in removing, from the resulting interpolation, the current available stock $A(tc_j \leftrightarrow cc_k).s$ that $A(tc_j \leftrightarrow cc_k)$ already has. After this operation, the stock is naturally decreased. In case it becomes negative, it is set to 0 and the lacking quantity is included in the demand (given the intention to avoid supply shortage i.e. a negative stock).

The process of Stage 1 is summarised in Equation 1 where $cc_k.\sigma_d$ is the sum of the quantities to be demanded by a cc_k to all tc_j.

$$\begin{aligned} &\text{a) } d(tc_j \leftarrow cc_k) = max(0, i(tc_j \leftarrow cc_k) - A(tc_j \leftrightarrow cc_k).s) \\ &\text{b) } A(tc_j \leftrightarrow cc_k).s = max(0, A(tc_j \leftrightarrow cc_k).s - i(tc_j \leftarrow cc_k)) \\ &\text{c) } cc_k.\sigma_d = \sum_{j=1}^{\eta_{tc}} d(tc_j \leftarrow cc_k) \end{aligned} \quad (1)$$

<u>Stage 2</u>: When tc_j has received, from all the cc_k, the demands $\{d(tc_j \leftarrow cc_1), \ldots, d(tc_j \leftarrow cc_{\eta cc})\}$, the sum of which is noted $tc_j.\sigma_d$ (Equation 2.a), it transfers them to each ambassador $A(pc_i \leftrightarrow tc_j)$, $i \in \{1, \eta_{pc}\}$. The ambassador then calculates, from these demands, the part $d(pc_i \leftarrow tc_j)$ for which pc_i will have to respond. This part is here calculated as being a linear combination of all demands $d(tc_j \leftarrow cc_k)$, $k \in \{1, \eta_{cc}\}$. It is formulated in Equation 2.b in which K_{ij}^d and α_{ijk}^d are the parameters of the linear equation, obtained by a linear regression on the GTIS data corresponding to respective variables

$$\begin{aligned} &\text{a) } tc_j.\sigma_d = \sum_{k=1}^{\eta_{CC}} d(tc_j \leftarrow cc_k) \\ &\text{b) } \forall (i \in \{1, \eta_{pc}\}, j \in \{1, \eta_{tc}\}), \\ &\qquad d(pc_i \leftarrow tc_j) = K_{ij}^d + \sum_{k=1}^{\eta_{CC}} \alpha_{ijk}^d * d(tc_j \leftarrow cc_k) \\ &\text{c) } tc_j.\sigma_d' = \sum_{i=1}^{\eta_{pc}} d(pc_i \leftarrow tc_j) \\ &\text{d) } tc_j.\tau = tc_j.\sigma_d' / tc_j.\sigma_d \\ &\text{e) } \forall i \in \{1, \eta_{pc}\}, d(pc_i \leftarrow tc_j) = d(pc_i \leftarrow tc_j) * tc_j.\tau \\ &\text{f) } A(tc_j \leftrightarrow cc_k).h = d(tc_j \leftarrow cc_k) / tc_j.\sigma_d \end{aligned} \quad (2)$$

described by Equation 2.b. Next, as a linear regression generally generates a residual error, it often happens that the sum $tc_j.\sigma_d'$ of these (calculated) parts may be different from the initial sum $tc_j.\sigma_d$ of the demands from which these parts have been calculated An adjustment must then be made by tc_j regarding each $d(pc_i \leftarrow tc_j)$. Equations 2.d to 2.e show how to proceed, with $tc_j.\tau$ the ratio between $tc_j.\sigma_d'$ and $tc_j.\sigma_d$,. In parallel to all of these operations, $A(tc_j \leftarrow cc_1).h$ is calculated (Equation 2.f). It corresponds to the market share of the quantity supplied to cc_k via tc_j. The interest of this variable will be detailed in Stage 4.

Stage 3: When the demand arrives at each pc_i, the latter, in response, calculates the total supply $pc_i.\sigma_s$ it will provide all consumers. In the present normal context, $pc_i.\sigma_s$ is the same quantity as the sum $pc_i.\sigma_d$ of the demands $d(pc_i \leftarrow tc_j)$, $j \in \{1, \eta_{tc}\}$. This supply is then sent to all tc_j via their respective $A(pc_i \leftrightarrow tc_j)$. Equation 3 summarises the process. It should be noted that a pc_i has a maximal supply capacity $pc_i.\mu$. In this context, it corresponds to the maximum of supplies existing over time for pc_i. The interest of this variable will be detailed in Section 3.3.2).

$$
\begin{aligned}
&\text{a) } pc_i.\sigma_d = \sum_{i=1}^{\eta_{pc}} d(\,pc_i \leftarrow tc_j\,) \\
&\text{b) } pc_i.\sigma_s = pc_i.\sigma_d \\
&\text{c) } pc_i.\mu(t) = max(\,pc_i.\mu(t-1),\, pc_i.\sigma_s\,) \text{ with } pc_i.\mu(0) = 0
\end{aligned}
\tag{3}
$$

Stage 4: Finally, when tc_j has received the supplies from $pc_i.\sigma_s$, $i \in \{1, \eta_{pc}\}$, the sum of which is noted $tc_j.\sigma_s$, it calculates and transfers to each cc_k its part, via the ambassador $A(tc_j \leftrightarrow cc_k)$, $k \in \{1, \eta_{cc}\}$. The part to be transferred is determined by $A(tc_j \leftrightarrow cc_k).h$, obtained in Equation 2.f. $A(tc_j \leftrightarrow cc_k).h$ takes this opportunity to update its stock according to the supply it has obtained. Equation 4 summarises the process in which $cc_k.\sigma_s$ is the sum of the quantities supplying cc_k from the different tc_j, $j \in \{1, \eta_{tc}\}$ and $cc_k.s$ the total stock of cc_k.

$$
\begin{aligned}
&\text{a) } s(\,tc_j \rightarrow cc_k\,) = tc_j.\sigma_s \; * \; A(\,tc_j \leftarrow cc_k\,).h \\
&\text{b) } A(\,tc_j \leftarrow cc_k\,).s = A(\,tc_j \leftarrow cc_k\,).s + s(\,tc_j \rightarrow cc_k\,) - i(\,tc_j \leftarrow cc_k\,) \\
&\text{c) } cc_k.s = \sum_{j=1}^{\eta_{tC}} A(\,tc_j \leftarrow cc_k\,).s = cc_k.\sigma_s - cc_k.\sigma_d
\end{aligned}
\tag{4}
$$

3.3.2. Formalisation of a restriction context

Let us now assume that, as of an instant t_s, a producer pc_r decides to restrict its supply of $pc_r.\rho_s$ points (with $0 < pc_r.\rho_s \leq 1$). In this case, pc_r changes its context from *Normal* to *Restriction* and executes an action pc_r.*restrict*(). This action consists in sending an event *restrict*(pc_i, cc_k, < >) to each cc_k, $k \in \{1, \eta_{cc}\}$ to inform them about the restriction. It also consists in sending to each other producer pc_i, via their ambassador $A(pc_r \leftrightarrow pc_i)$ $i \neq r$, the not supplied partial quantity $pc_r.q_r$ resulting from this restriction. It is given in Equation 5.c. The quantity $pc_r.\sigma_s$ to be supplied to consumers during a restriction is the maximum supply of pc_r diminished by this rate $pc_r.\rho_s$ and potentially further diminished by the total demand $pc_r.\sigma_d$ arriving at pc_r (Equation 5.b).

Following this restriction, the stock $cc_k.s$ of each cc_k will naturally decrease (given Equation 5.b and, in this context, without compensation yet) and finally be in shortage. Let us note $cc_k.\lambda$ the cumulated shortage over time. When $cc_k.s$ is less than 0, it is added to $cc_k.\lambda$ (Equation 5.d). For a country cc_k, the shortage period is that during which its cumulated shortage $cc_k.\lambda$ is below 0 (cf. Equation 5.e, where $cc_k.t_{\bar{\lambda}}$ is the date of the end of the shortage). Finally, Equation 5.f indicates that the shortage to compensate cannot be, in a shortage period, greater than the demanded value. We call the *full shortage* the situation where the shortage is mathematically equal to the negative value of the demand.

$$
\begin{aligned}
&\text{a) } pc_r.\sigma_d = \sum_{i=1}^{\eta_{pc}} d(\,pc_r \leftarrow tc_j\,) \\
&\text{b) } pc_r.\sigma_s = min(\,pc_r.\sigma_d\,,\, pc_r.\mu * pc_r.\rho_s\,) \\
&\text{c) } pc_r.q_r = pc_r.\sigma_d - pc_r.\sigma_s \\
&\text{d) } cc_k.\lambda(t) = \begin{cases} cc_k.\lambda(t-1) + cc_k.s(t) & \text{if } cc_k.s(t) < 0 \\ cc_k.\lambda(t-1) & \text{otherwise} \end{cases} \\
&\text{e) } cc_k.t_{\bar{\lambda}} \text{ is } t > t_s \text{ such that } cc_k.\lambda(t_{\bar{\lambda}}-1) < 0 \text{ and } cc_k.\lambda(t_{\bar{\lambda}}) \geq 0 \\
&\text{f) } cc_k.\lambda(t) = max(cc_k.\lambda(t),\, -cc_k.\sigma_d(t))
\end{aligned}
\tag{5}
$$

3.3.3. Formalisation of a compensation context

On reception of the restriction imposed by pc_r, each cc_k, $k \in \{1, \eta_{cc}\}$ changes its context from *Normal* to *Compensation* and immediately executes the action cc_k.*makeup*(). It consists in sending, at each time step, and while $cc_k.\lambda < 0$, an event *demandMakingUp*(cc_k, pc_i, <$|cc_k.\lambda|$>) to all the pc_i, $i \in \{1, \eta_{pc'}\}$ with $\eta_{pc'} < \eta_{pc}$ and $i \neq r$. Each pc_i that receives the message, either immediately switches its context from *Normal* to *Compensation* and responds by executing pc_i.*makeup*() described below, or waits for a delay $pc_i.\delta$. In the latter case, it first switches its context from *Normal* to *Waiting* before switching from *Waiting* to *Compensation*, once this delay expires. This delay may be necessary for diverse reasons specific to pc_i: inability to immediately respond, speculation, etc.

During a pc_i.*makeup*(), the total quantity $pc_i.s$ to be supplied *a priori* is given by Equation 6.a. It is calculated as a function of the sum $pc_i.\sigma_d$ of the classical normal demand arriving at pc_i, the not supplied quantity $pc_r.q_r$, obtained from pc_r (cf. Section 3.3.2), the sum of the stocks $|cc_k.\lambda|$, $k=1\ldots\eta_{cc}$ to be compensated, obtained from all *demandMakingUp* events (cf. above), and the weight $pc_i.\omega$ (in %) of pc_i amongst all compensating producers.

The final value $pc_i.\sigma_s$ to be supplied (compensation included) is then evaluated in Equation 6.b, in which $pc.\rho_p$ is the compensation rate. Equation 6.b stipulates that if $pc_i.s$ is below the capacity $pc_i.\mu$,, it is the final quantity to be supplied. Otherwise, an augmentation of this supply capacity is first required. It corresponds to the minimum between an augmentation by the compensation rate and $pc_i.s$. Then, this minimum is taken for $pc_i.\sigma_s$. Finally, the maximal capacity $pc_i.\mu$ is updated accordingly (Equation 6.c).

When $t \geq t_s + pc_i.\delta$:

$$\text{a) } pc_i.s = pc_i.\sigma_d + (pc_r.q_r + \sum_{k=1}^{\eta_{CC}} |cc_k.\lambda|) * pc_i.\omega$$

$$\text{b) } pc_i.\sigma_s(t) = \begin{cases} pc_i.s & \text{if } pc_i.s < pc_i.\mu(t) \\ min(pc_i.\mu(t) * (1 + pc_i.\rho_p) * pc_i.s \end{cases} \quad (6)$$

$$\text{c) } pc_i.\mu(t) = \begin{cases} pc_i.\mu(t) & \text{if } pc_i.s < pc_i.\mu(t) \\ pc_i.\sigma_s(\text{t}) & \text{otherwise} \end{cases}$$

3.3.4. Formalisation of a maintenance context

On a consumer side, this context starts when a previous making up from compensating producers is finished, i.e. $cc_k.\lambda$ finally becomes ≥0 again while cc_k is still under the pc_r restriction. In a maintenance context, cc_k attempts to stay in the (new) equilibrium situation it has just obtained. Thus, cc_k changes its context from *Compensation* to *Maintenance* and then immediately executes the cc_k.*destock*() action to avoid a surplus stock (since $cc_k.\lambda$ is now >0) at each time step. This action consists in sending an event *demandDestorage*(cc_k, pc_i, <$cc_k.\lambda$>) to all the pc_i, $i \in \{1, \eta_{pc}\}$ and $i \neq r$, while $cc_k.\lambda > 0$. When $cc_k.\lambda$ becomes <0 again due to that action, cc_k.*makeup*() is executed again (but in a maintenance context, not a compensation context), etc. The two actions are alternatively executed over time, depending on the value of $cc_k.\lambda$, so that $cc_k.\lambda$ turns around 0 as closely as possible.

On a producer side, each pc_i that receives the message for the first time also switches its status from *Compensation* to *Maintenance* and alternatively executes pc_i.*destock*() and pc_i.*makeup*(), in reaction to the demands from cc_k. Note that pc_i.*destock*() is identical to pc_i.*makeup*() except that the sum of stocks $cc_k.\lambda$ is for it to be decreased (Equation 7.a) instead of to be increased (Equation 6.a).

$$\text{a) } ts = pc_i.\sigma_d + (pc_i.q_r - \sum_{k=1}^{\eta_{CC}} |cc_k.\lambda|) * pc_i.\omega \quad (7)$$

At this stage of the work, $pc_i.\delta$ and $pc_i.\omega$ are determined by the user, not by the agent. However, a $pc_i.\omega$ can be approximated via the weight provided by GTIS statistical data regarding all pc_i, plus a slight experimental adjustment so that each $cc_k.\lambda$ evolves around 0 during the maintenance context.

4. SIMULATION

4.1. Preamble: the simulation platform

While the statistical tests were performed with the proprietary tool SAS®, the simulation was implemented under the platform *Isatem* (Andriamasinoro 2012). *Isatem* is constituted of a set of components in interaction. Each component possesses a set of properties, a set of handlers, a set of output functions and a behaviour.

The set of handlers manages the input events. For a received event *eventX*, this handler is formally written *OnEventX* (*s*, <*pX1*, …, *pXn*>) where *s* is the component having sent the event and <*pX1*, …, *pXn* > the event parameters. Only a handler *OnTimeChange*() is generic; it allows a component to react to the simulation timer.

The output functions manage the output events. For an event *eventY* to be sent, this function is formally written *FireEventY*(*r*, <*pY1*, …, *pYm* >)) where *r* is the recipient component.

The behaviour possesses the same *OnEventX*() and *FireEventY*() as a component to which it is associated. Actually, a component does not handle an event directly. It sends it to the handler of its current behaviour, having the same name. This mechanism allows a component, at any time, to change its behaviour, which is the way it handles the events, without changing the communication mode between it and its behaviour. Each component possesses a default behaviour. It is then possible, by inheritance (as is defined by the object oriented concept), to *particularize* an *OnEventX*() or a *FireEventY*(). It is what allows two components to possibly have the same properties but totally different behaviours (e.g. a producing country and a consuming country). The body of functions in the default behaviour is either empty or groups the actions common to all the components of the same type (i.e. the *country* type and the *ambassador* type respectively).

4.2. Initialisation of the values for the simulation

The selected producing countries (pc_i) are Chile (*cl*), China (*cn*) and the United States (*us*). These countries are seen by GTIS as being those that regularly supply France with a high quantity (≥25 t/quarter) of Li_2CO_3. We also add a (virtual) country called the *rest of the world* (*rw*). The quantity supplied by *rw* is the world quantity, decreased by that provided by *cl*, *cn* and *us*. Thus, η_{pc}=4.

The selected consuming countries (cc_k) are France (*fr*), the subject of our study and, again, the rest of the world (*rw*). The quantity consumed by *rw* is the world quantity, decreased by that of France. Thus, η_{cc}=2.

Finally, the transit countries (tc_j) are Belgium (*be*), Germany (*de*), United Kingdom (*uk*), Italy (*it*) and the Netherlands (*nl*). The analysis of the GTIS data shows that it is via these transit countries that *cl*, *cn* and *us* transfer their quantities of Li_2CO_3 to France. We also add, again, the rest of the world (*rw*). The quantity transiting via *rw* is the world quantity, decreased by that of *be*, *de*, *uk*, *it*, *nl*. Thus, η_{tc}=6.

All the ambassadors are next naturally created to connect all these countries (*rw* included) in keeping with the formalisms previously described in this paper.

To comply with the data we have chosen in GTIS, the simulation time step is 3 months. Let us note, for example, 2/2019 quarter 2 of year 2019.

4.3. Prospective scenarios

The pattern of the proposed (and currently fictitious) prospective scenario is the following: one assumes that as of 2014 (=t_s), Chile restricts its supply rate by $cl.\rho_s$ points. Following this situation, China accepts to assure compensation at a rate of $cn.\rho_p$ points, and does so immediately, i.e. $cn.\delta$=0. The United States also accepts, with a rate of $us.\rho_p$ points, but only as of 2016, i.e. $us.\delta$=8 (quarters). The purpose of the simulation then consists in varying the values of these rates to find the shortage end date in France and in the rest of the world. A simulation will be formally written (example of France): $fr.t_{\bar{\lambda}} = fr(-cl.\rho_s, +cn.\rho_p, +us.\rho_p)$.

For each simulation performed, the result is read as follows: for each compensation $cn.\rho_p$ and $us.\rho_p$ and for each augmentation $fr.\rho_a$ and $rw.\rho_a$, the shortage end date for France and the rest of the world will be respectively $fr.t_{\bar{\lambda}}$ and $rw.t_{\bar{\lambda}}$. An inverse reading can also be performed: if one hopes that the shortage period does not extend beyond $fr.t_{\bar{\lambda}}$ for France and $rw.t_{\bar{\lambda}}$ for the rest of the world, the United States and China should increase their compensation rate by at least $cn.\rho_p$ and $us.\rho_p$ respectively.

Table 1 shows in detail the list of different scenario instances proposed in this paper. An instance is made of the scenario identifier (written in brackets), the value of

Table 1: List of all the Scenarios, an Instance Being Composed of a Restriction from Chile (cl) followed by a compensation from China (cn) and USA (us)

id	$-cl.\rho_s$	$+cn.\rho_p$	$+us.\rho_p$
(a)	-0.15	+0.3	+0
(b)	-0.4	+0.3	+0
(c)	-0.15	+0.1	+0
(d)	-0.4	+0.1	+0.5
(e)	-0.4	+0.1	+0.1
(f)	-0.1	+0.1	+0

the restriction from Chile and the value of compensation, respectively from China and USA. The value chosen in this table also allows a policy maker to analyse the sensitivity of the lithium market after a variation in important indicators (e.g. here, the diverse rates).

As complementary information regarding the rate variables, our data calibration with GTIS and our additional experiments allow us to allot the values of 35% and 65% respectively for $us.\omega$ and $cn.\omega$.

4.4. Results

Figure 1 first provides an example of how the stock of the rest of the world (i.e. $rw.\lambda$) may evolve throughout the different contexts. The examples taken are from

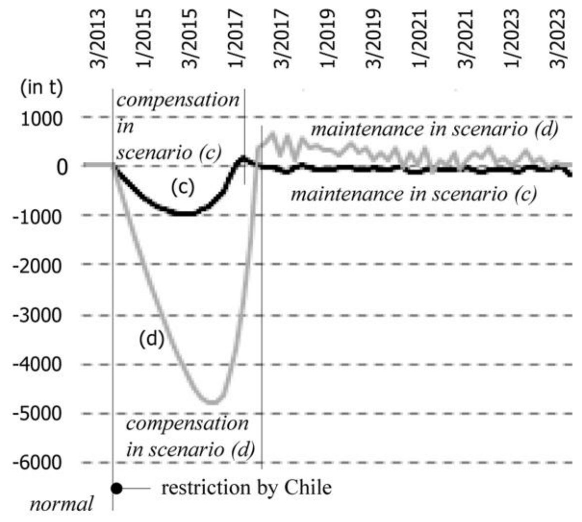


Figure 1: Prospective Evolution of the Lithium Stock in the Rest of the World in its Different Contexts during Scenarios (c) and (d).

scenarios (c) and (d). The figure approximately determines, for *rw*, the end of a normal context, the start and the end of a compensation context (e.g.: around 1/2016 for (c) and 3/2016 for (d)) and the evolution of a maintenance context, in both scenarios.

Figure 2 next shows the shortage end dates obtained for the rest of the world in all scenarios. In this figure, the value of 13,000 (in t/quarter), in absolute value, approximately represents the average demand of lithium of *rw* (according to the GTIS data). It means for example that in Scenario (f), at the peak time of a supply shortage period, there is still a minimal value of around (13,000-5,800) t/quarter of lithium (more than 50%) which are supplied to this consumer. It should be

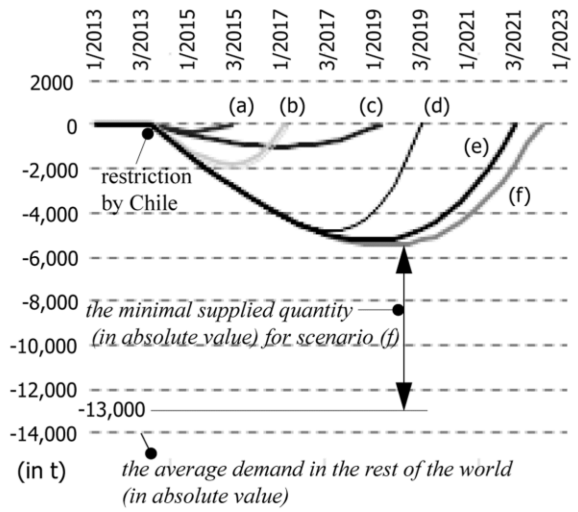


Figure 2: Prospective Evolution of the Lithium Stock in the Rest of the World, for all Scenarios

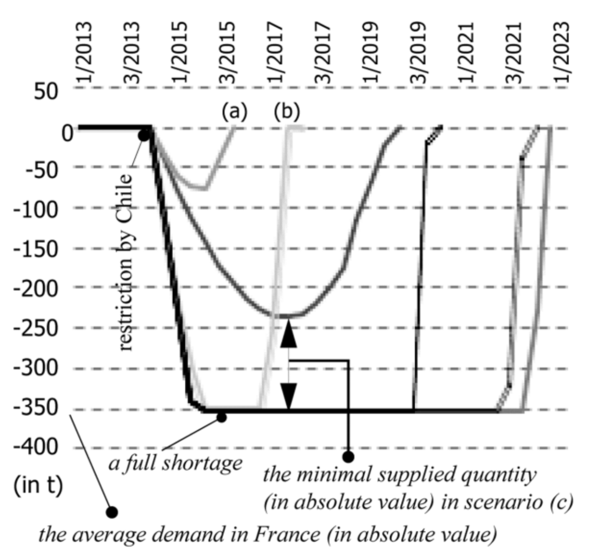


Figure 3: Prospective Evolution of the Lithium Stock in France, for all Scenarios

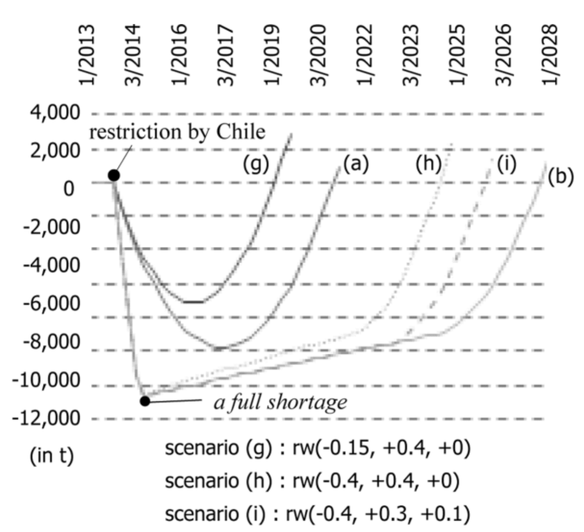


Figure 4: Prospective Evolution of the Lithium Stock in the Rest of the World according to the work (AH13)

noted that in Figure 2 the maintenance stage is no longer shown, for reasons of clarity, but actually, at that context, the shape of the curves of all other scenarios are approximately similar to that of the two scenarios presented in Figure 1.

Figure 3 is the "France equivalent" of Figure 2, with an average demand of around 350 t/quarter (according to the GTIS data). In this figure, France reaches a full shortage in all the scenarios where the Chile restriction is high (-0.4), i.e. (b), (d), (e) and (f), and with a different duration. The reason for this full shortage is that the linear regressions made on the GTIS data result in a behaviour where the model first handles the rest of the world (*rw*) and when the stock is close to 0 again for *rw*, there is afterwards an automatic consideration of France. It should however be observed, in Figure 3, that for France, the transition from a full shortage period to an equilibrium state is then very fast in all the scenarios where a full shortage occurs.

5. DISCUSSION

The results presented previously and in next sections are still being adjusted and validated by the domain experts. However, this current validation does not affect the fact that a progress has been made compared to the state of the art.

5.1. Thematic discussion

If we refer only to the results of these simulations, France does not need to worry about a possible prolonged shortage of its supply in lithium, even following a Chilean restriction. Indeed, even if Figure 3 shows many full shortage situations, at the same time, the rest of the world is still supplied at least around 50% of its needs and adding France to the list of supplied countries should not be a real issue since the French proportion is small compared to that of the rest of the world. This conclusion is also confirmed by the fact that a return from a full shortage to an equilibrium situation is fast for France. From this model to a real-world prospective analysis, it is possible to conclude that possibilities exist for France to avoid a supply shortage. (Daw and Labbé 2012) have already concluded that the risk of shortage is low for France, but they did not however analyse the effects of a scenario of an effective restriction in their study.

Regarding our modelling exercise stage, this work is a continuation of that previously carried out by (Andriamasinoro and Ahne 2013), noted AH13 for short. The simulation results of that work, regarding the rest of the world, are recalled in Figure 4.

The present work is a refined version of that work by (1) introducing a more dynamic interaction between a supply, a demand and a stock (Equations 1.b and 4.b) over the simulation while (2) introducing equations for adjustment (Equations 2.d and 2.e). The goal of the latter is to remove any residuals due to the linear regression processes and to be sure that the quantity demanded by consumers and passing via a transit country is exactly the same as the quantity demanded by this transit country to the producers.

These improvements allow us to implement the maintenance context, which did not exist in AH13, i.e. in AH13 the model could not consider what happened when having reached a return to an equilibrium state after a compensation stage. This possibility of remaining around an equilibrium state (as in the maintenance context) is possible only when considering the dynamic interaction between a supply, a demand

and a stock, which has been introduced only in this work.

These improvements also allow us to discover the possibility for obtaining a more optimistic market than that in AH13, regarding the supply shortage issue. Indeed, if we compare Figure 4 with Figure 2 (i.e. $rw.\lambda$ in this work), the following statements can be observed. First, the end of a supply shortage in the worst scenario of AH13 was 2028 against 2023 here. Second, the minimum supplied value in the best scenario of AH13 (Figure 4.g) practically corresponds to the minimum supplied value in the worst scenario of this work (Figure 2.f). Furthermore, it may be noted that the scenarios (g) (h) and (i) in Figure 4 are no longer represented in Figure 2. The reason is the corresponding supply shortage period of these scenarios is now very short. Thirdly, the worst scenario of AH13 (Figure 4.b) practically indicates an almost full shortage in the rest of the world when Chile restricts by “only” -0.4 (i.e. -40%). In the present work, a full shortage is far from reached, a statement more acceptable given that 60% of the supply is still realized even in a restriction.

Globally, we think as regards the rest of the world that the result in Figure 2 is more realistic than in AH13, at least if our hypotheses are verified. As for France, the conclusion in AH13 is practically the same as that presented at the beginning of this section.

All of these conclusions appear interesting and the present work seems to be an improvement. However, these conclusions should be interpreted with caution whether they concern the best or the worst scenarios. Indeed, as rightly recalled by (Feitosa, Bao Le and Vlek 2011), results obtained in any exercise of modelling complex systems do not represent either precise forecasts or deterministic answers. In addition, we think that in a crisis situation a model cannot always handle all circumstances that may happen and, as such, the situation in the previous work showing a more pessimistic result may occur anyway in a real world, even if it is more disputable at a conceptual level. All in all, the results obtained from this modelling exercise should mainly serve to feed the public debate concerning the subject and the scenarios provided here only aim to offer various potential situations, to make the debate as rich as possible.

5.2. Methodological discussion

Regarding the use of MAS in the MCI field, work carried out by Andriamasinoro, Orru and Pelon (2006) concluded at that time that MAS can be adopted to follow up *only* the possible dynamic evolution of MCI markets at a microeconomic scale (i.e. at a site scale), in which the issue is to follow a population activity and migration (Jonsson and Brycesson 2009). In the case of MCI markets, analysed at a more medium-macroeconomic scale, given that economic indicators are periodically displayed, decision makers already have a better idea of what to do in the future without inevitably using a method such as MAS. This statement has effectively been confirmed over recent years: market models used to handle the supply shortage issue in the MCI field generally rely on purely mathematical or statistical approaches. Such is the case of the works presented in the State of the Art of this paper (Section 2). Let us also acknowledge that in the field of macroeconomics, in which metal markets exist (i.e. at a world level), analysts have always traditionally preferred to rely on conventional economic equilibrium models such as DSGE (Fernández-Villaverde 2010). In no case has MAS been used. According to one policy adviser (Hamill 2010), a motivation towards better use of MAS requires proving to analysts that this method provides something better than they already have.

We think MAS is now in that case and should also be progressively adopted in both aggregate resource and metal markets, due to the reasons explained subsequently.

5.2.1. MAS and metal markets

With regards to metal markets, the economic crisis in 2008-2009, resulting from interactions between local players (individual banks, households, traders, etc.) and which then had repercussions on the world market, has made economists reconsider modelling at a macroeconomics level. Indeed, they observed that models such as DSGE were no longer sufficient to anticipate a crisis situation and that MAS could be a solution (Farmer and Foley 2009). The Economist (2010) review presents this insufficiency as follows: “if conventional models perform well enough in a business-as-usual economy, based only on the existence of an ideal state of equilibrium, there is no equilibrium during crashes. Agent-based models may be more suitable because they make no assumptions about the existence of efficient markets or general equilibriums. Instead, they are focused on the assignation of particular behavioural rules to each agent and large fluctuations and even crashes are inherent to the system”.

Supply shortage, the subject of the present work, may be a future market crisis occurrence. Indeed, it includes an important period of disturbances, resulting from individual decisions, but being able to trigger major consequences on the global market and on importing countries.

5.2.2. MAS and aggregate resources markets

With regards to aggregate resources (AR) markets, often analysed at a national/regional level, the main issue of the authorities concerning production in this market is to better identify the future geographical distribution of resources while regarding the environmental and societal constraints of the extraction activity. The idea of having the distribution is to favour the proximity of producers to consumers, thus reducing transport and environmental costs (Brown, McEvoy and Ward 2011). In order to quantify this proximity, a prospective model should then be spatially explicit; a map is in particular often used by geology analysts/authorities as a decision making support (Cassard, et al. 2008). However, a system dynamics approach, as used e.g. by Rodriguez-Chavez (2010) to model AR market flow in France, was not able to represent such a spatiality.

In these situations, we think MAS is more suitable. The idea is, at the start of a simulation, a modeller/user imports into the MAS environment the user thematic maps required by the application. Then, at any time during an on-going simulation where agents interact with the space, it should be possible for an analyst to export this dynamic spatial data as new temporal map data and to import this new data in a tool such as a GIS, for a more in-depth spatial analysis. This proposal is currently being implemented throughout an application in the Seine-Normandie region (France) (Andriamasinoro 2012). The application concerns the prospective analysis of AR flows in that region, by using thematic maps such as quarries (for production), cities (for consumption) and roads (for transport). That work does not currently consider other thematic maps such as geological resources or environmental constraints, necessary information to better identify the exploitable resources area. At a thematic level, this application needs to be improved. Nevertheless, at a methodological level, it already can demonstrate the interest of using MAS for the AR market analysis.

5.2.3. Suggestions for better acceptance of MAS

In addition to her first suggestion in Section 5.2, Hamill (2010) also suggests that it is necessary to demonstrate the value of MAS modelling by showing-by-doing and offering training projects. This is, we believe, the case of the present work: it is an additional MAS application (even if, we agree, not yet sufficient to convince). Yet, regarding MAS acceptance, Andriamasinoro and Angel (2012) suggest favouring the coupling between MAS models and mathematical or statistical models, instead of sticking to a solely MAS approach. This point of view is emphasized by Weyns, Helleboogh and Holvoet (2009), who stipulate that one of the reasons that MAS is not accepted in industry is research in MAS profiles itself as an isolated community and, as such, may create artificial thresholds in convincing people of its merits. A continuation in overcoming this situation should then be carried out over the next years.

6. CONCLUSIONS

The markets of mineral commodities for industrial use (MCI) risk supply shortages in the near future. This is true in aggregate resources markets, as well as in metal markets (lithium, indium, rare earth markets, etc.). Given this situation, the French government is not reassured because in the coming decades such situations may affect French industrial sectors using the products from these markets.

The objective of the work reported in this paper is twofold and concerns a thematic level and a methodological level. At a thematic level, it attempts to contribute to the elaboration of a public policy support tool to help French industrialists as well as the French government to ensure that in forthcoming decades there will always be a continuity of supply in the MCI markets. The application example taken is the world lithium market, but the discussion is also extended to aggregate resources markets. At a methodological level, the aim of the work is to evaluate and discuss the possible interests of applying the MAS approach to MCI issues and to what extent it is actually used in the literature regarding MCI prospective analysis via modelling/simulation.

As regards the thematic level, the model allows us to strengthen the idea that in the event of a supply restriction the risk of shortage in France remains very low or, at least, very limited in time. As for the rest of the world (i.e. the "sum" of the worldwide countries other than France), the shortage periods are more consequent. As an example, for a restriction starting in 2014, the return to an equilibrium state in the worst scenario (high restriction vs. very weak compensation) is 2023. However, this result is more optimistic than what has been found in AH13, a previous work (2030). All in all, both results would serve to feed the public debate concerning the subject.

As regards the methodological level, it has been argued that the MAS approach can implement MCI issues at world, national and regional levels thanks to its inherent hierarchical and spatially-explicit features. The interest in MAS has in particular increased since the economic crisis, generated by individual behaviour, and where conventional economic equilibrium models could not anticipate the crisis. In addition, its spatially-explicit features now enable modellers to introduce thematic maps to better localize the future potential resources where existing models have ignored these spatial aspects. However, despite this promising situation regarding MAS, convincing the actors of the MCI sectors to integrate models from MAS as a decision support tool remains a challenge. Explanations for this state of affairs and proposals for progress regarding MAS acceptance are provided.

7. FUTURE WORKS

In the first place will be the introduction of other variables to possibly explain supply and demand in France. We are thinking in particular of *mining reserves* and *prices*. Secondly, it will be necessary to make the system more complex by introducing the other producing and consuming countries (which will involve, in the model, their "withdrawal" from the virtual country "rest of the world"). Thirdly, a methodological transposition of all the work to other metals will be carried out. There are naturally strategic metals such as rare earths, but one should not completely forget major metals, with iron and steel to the forefront, but also aluminium or copper, which have an economic importance much greater than lithium. Likewise, works related to aggregate resources markets, precisely the Seine-Normandie (France) application, will be continued by adding other thematic maps such as geological resources and environmental constraints as factors to better localise future potential resources in that region.

REFERENCES

Andriamasinoro, F. 2012. Reinforcing the place of dynamic spatialised indicators in a generic socioeconomic model. In G. Borruso, S. Bertazzon, A. Favretto, B. Murgante, & C. M. Torre (Eds.), *Geographic Information Analysis for Sustainable Development and Economic Planning: New Technologies* (pp. 313-334). IGI Global.

Andriamasinoro, F., & Ahne, H. 2013. Prospective analysis of the world lithium market: contribution to the evaluation of supply shortage periods. *International Business & Economics Research Journal, 12*(3), pp. 359-372.

Andriamasinoro, F., & Angel, J.-M. 2012. Artisanal and small-scale gold mining in Burkina Faso: suggestion of multi-agent methodology as a complementary support in elaborating a policy. *Resources Policy, 37*(3), pp. 385-396.

Andriamasinoro, F., Orru, J.-F., & Pelon, R. 2006. Prospective modelling to monitor artisanal gold mining activity. *Proceedings of the 21st Colloquium of African Geology (CAG 21)*, (p. 340). Maputo, Mozambica.

Arthur, W. B., Durlauf, S., & Lane, D. 1997. Introduction: Process and Emergence in the Economy. In W. B. Arthur, S. Durlauf, & D. Lane (Ed.), *The Economy as an Evolving Complex System II* (pp. 1-11). Addison-Wesley.

Brown, T., McEvoy, F., & Ward, J. 2011. Aggregates in England — Economic contribution and environmental cost of indigenous supply. *Resources Policy, 34*(4), pp. 295–303.

Cassard, D., Billa, M., Lambert, A., Picot, J.-C., Husson, Y., Lasserre, J.-L., et al. 2008. Gold predictivity mapping in French Guiana using an expert-guided data-driven approach based on a regional-scale GIS. *Ore Geology Reviews, 34*(3), pp. 471-500.

Daw, G., & Labbé, J.-F. 2012. *Panorama 2011 du marché du lithium.* Rapport final, BRGM/RP-60460-FR. 157 p., 51 fig., 30 tab.

De Solminihac, P. 2010. SQM view of the lithium industry. *Lithium Supply & Markets 2010. 2nd.* Las Vegas, USA: Presentation, 21 p.

Farmer, J. D., & Foley, D. 2009. The economy needs agent-based modelling. *Nature, 460*(7256), pp. 685–686.

Feitosa, F. F., Bao Le, Q., & Vlek, P. 2011. Multi-agent simulator for urban segregation (MASUS): A tool to explore alternatives for promoting inclusive cities Computers. *Environment and Urban Systems, 35*(2), pp. 104–115.

Fernández-Villaverde, J. 2010. The econometrics of DSGE models. *SERIEs, 1*, pp. 3-49.

Giacalone, J. A. 2012. China, The World Trade Organization, And The Market For Rare Earth Minerals. *Proceedings of the 2012 International Academic Conference*, (pp. 445-452). October 15-17, Las Vegas, Nevada USA.

Graedel, T., Chandler, C., Barr, R., Chase, T., Choi, J., Christoffersen, L., et al. 2012. Methodology of Metal Criticality Determination. *Environmental Science & Technology, 46*(2), pp. 1063-1070.

Gruber, P. W., Medina, P. A., Keoleian, G. A., Kesler, S. E., Everson, M. P., & Wallington, T. J. 2011, October. Global Lithium Availability: A Constraint for Electric Vehicles? *Industrial Ecology, 15*(5), pp. 760-775.

GTIS. 2012. *Online market database*. Retrieved from https://www.gtis.com/gta/

Hamill, L. 2010. *Agent-based modelling: the next 15 years.* (JASSS, Ed.) Retrieved from Journal of Artificial Societies and Social Simulation: http://jasss.soc.surrey.ac.uk/13/4/7.html

Jonsson, J., & Brycesson, D. 2009. Rushing for Gold: Mobility and Small-Scale Mining in East Africa. *Development and Change, 40*(2), pp. 249-279.

Kushnir, D., & Sandén, B. A. 2012. The time dimension and lithium resource constraints for electric vehicles. *Resources Policy, 37*(1), pp. 93–103.

McNulty, J. P., & Khaykin, A. 2009. *Extracting the Details on the Lithium Market*. Retrieved from http://www.docstoc.com/docs/12415608/Lithium

Rodriguez-Chavez, M.-L. 2010. *Anticipation of the access to the aggregate resource by breaking present schemes in the long term.* Thesis (PhD), MINES ParisTech, Paris.

Roskill. 2009. *The Economics of Lithium.* (11th, Ed.)

Roskill. 2011. *Rare Earths & Yttrium: Market Outlook to 2015.* (14th, Ed.)

The Economist. 2010, July. *Agents of change: Conventional economic models failed to foresee the financial crisis. Could agent-based modelling do better?* Retrieved from http://www.economist.com/node/16636121

Weyns, D., Helleboogh, A., & Holvoet, T. 2009. How to get multi-agent systems accepted in industry? *International Journal of Agent-Oriented Software Engineering, 3*(4), pp. 383-390.

Wooldridge, M. 2009. *An Introduction To Multiagent Systems.* (2nd, Ed.) John Wiley & Sons.

Yaksic, A., & Tilton, J. E. 2009. Using the cumulative availability curve to assess the threat of mineral depletion: The case of lithium. *Resources Policy, 34*(4), pp. 185-194.

SIMULATION ANALYSIS FOR CASCADING FAILURE OF AN ELECTRICAL POWER GRID

Chulhan Kim[(a)], Kiwook Bae[(b)], Soojeong Bae[(c)], Tae-Eog Lee[(d)]

[(a)(b)(c)(d)]KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea

[(a)]chulhan.kim@kaist.ac.kr, [(b)]baekw92@kaist.ac.kr, [(c)]soojeongbae@kaist.ac.kr, [(d)]telee@kaist.ac.kr

ABSTRACT
In an electrical grid, destruction of generators or power transmission elements may lead to a serious power outage on an entire city or a country, not just on the region that the destroyed facilities exist. This is because of cascading failure of an electric grid. In order to prevent huge damage on a system, cascading failure needs to be analysed properly. In this study, we propose a systematic framework for examining cascading failure of an electrical grid with simulation. Procedure of cascading failure and mathematical models for simulation are introduced. In addition, demand shedding policies for reducing damage on a system are suggested. We also conduct simulation experiments as a case study which involves all the concepts that we present throughout the paper.

Keywords: electrical grid, cascading failure, simulation, demand shedding

1. INTRODUCTION

In modern countries, an electrical grid is one of the most important infrastructures to keep the societies alive. Electricity is involved in almost every part of life. It is one of basic resources for a society and used by houses, companies, schools, hospitals, and so on. In addition, other critical infrastructures including a traffic system, a finance system, and a gas/oil distribution system are highly dependent upon electricity. If there is a problem with supplying electricity to consumers, serious physical or financial damages may occur. Therefore, managing an electric grid is important for an entire society.

However, generators and transmission systems are vulnerable to many types of disasters by the nature and humans (e.g. tornado, flood, earthquake, explosion, and fire). Since the elements in a power distribution system such as power plants, transmission facilities, and consumers are tightly coupled as a form of a network, breakdown or destruction of a small part of the network can affect the whole network. The electrical failure tends to spread step by step causing blackouts and this process is called *cascading failure*.

We can find blackout cases caused by cascading failure of electrical power grids: In 2012, a typhoon called Bolaven hit South Korea and stopped distribution of electricity to approximately 2 million houses and several industrial facilities; an earthquake with magnitude 7.4 caused blackouts to 4 million houses in the year 2011 in Japan. In addition, power blackout in North America in 2003 inflicted 6 billion dollar worth of damage, and South Korea was suffered from national power outage on September in 2011. These two cases were not due to disasters, but they were enough to emphasise the strong influence of the cascading failure of a power distribution system.

Failure of a power distribution system has been studied widely. We first introduce some research on cascading failure of electrical power grids. Bienstock and Verma (2010) suggested a mixed integer programming model and a continuous nonlinear programming model to figure out whether a power grid can survive with k or fewer arcs where there are N arcs in a network. Even though this study did not consider cascading failure seriously, it gave an insight into modelling a power grid mathematically. Possible procedures of cascading failure in an electrical power grid have been presented in many articles (Carreras, Lynch, Dobson, and Newman 2002; Carreras, Lynch, Dobson, and Newman 2004; Chen, Thorp, and Dobson 2005; Dobson, Carreras, Lynch, and Newman 2007; Dobson, Carreras, and Newman 2005; Hardiman, Kumbale, and Makarov 2004; Nedic, Dobson, Kirschen, Carreras, and Lynch 2006; Pfitzner, Turitsyn, and Chertkov 2011). Most of the articles defined a series of power outage by assuming power grids to use direct current (DC) in order to simplify the problems. Since the studies deal with national-wide power distribution systems, DC approximation is enough to reflect the reality. They analysed several types of cascading failure, criticality, and so on. There are also studies of power grid failures by natural disasters. Han, Guikema, Quiring, Lee, Rosowsky, and Davidson (2009) proposed a model to predict power outages during hurricanes. On the other hand, decision making methodologies for recovery after blockouts to reduce damage have been studied as well as cascading failure itself (Guha, Moss, Naor, and Schieber 1999; Langevin, Perrier, Agard, Baptiste, Frayret, Pellerin, Riopel, and Trépanier 2009; Xu, Guikema, Davidson, Nozick, Çağnan, and Vaziri 2007). They mainly dealt with scheduling of recovery process after electric power distribution failed, especially after disasters including earthquakes and hurricanes. Lastly, Pinar, Meza, Donde, and Lesieutre (2010) suggested an optimal strategy to check the vulnerability of an electrical power grid.

In South Korea, there is a national research to design models for integrative disasters which may cause catastrophic damage to the critical social infrastructures including an electrical grid, a traffic system, a healthcare system, etc. As a part of the research, a cascading failure model needs to be developed taking the structure of the South Korean electrical system into consideration. As a first step, in this paper, we suggest a simulation framework for cascading failure which can be a basic reference for future studies. In addition, we introduce demand shedding policies for controlling balance of demand and supply, and minimising the loss of the total demand. Details of each subject are explained in the following sections.

The manuscript is organised as follows. In Section 2, we introduce a procedure of cascading failure of an electrical grid. In Section 3, internal mathematical models for finding stable flow of the electricity is presented. In Section 4, we suggest applying efficient demand shedding policies to reduce damage on the whole network.

2. A PROCEDURE OF CASCADING FAILURE

At first, we shortly introduce how the power outage spreads in an electrical grid. Figure 1 illustrates a flow chart for cascading failure of an electrical grid. As mentioned before, an electrical grid tries to be stable with balanced electrical flow. Stable flow is unique under assumption of DC approximation. We discuss how the stable flow can be obtained in Section 3. If an object in a grid is destroyed or harmed by a certain event, the structure of the network changes and the stability may also be broken. Then the grid tries to get balanced with new flow according to the physical property of the electricity. However, the new flow may not satisfy the pre-determined capacity of nodes or links because it ensures the safety of the network only in terms of the structure of the network, and does not consider detailed information of the elements in the network. In Figure 1, $f_{ij} < u_{ij}$ is the condition that the new flow does not violate the capacities of the elements, where f_{ij} is the balanced flow of the electricity from node i to node j, and u_{ij} is the capacity of the link from node i to node j. We do not consider the capacity of each node in a network in this paper. If the condition is violated for some elements, those elements are no longer functional in the network. In other words, the elements are considered to be destroyed and removed from the original network. This is why the cascading failure occurs. After the additional breakdown, the network repeats the same procedure until there is no breakdown.

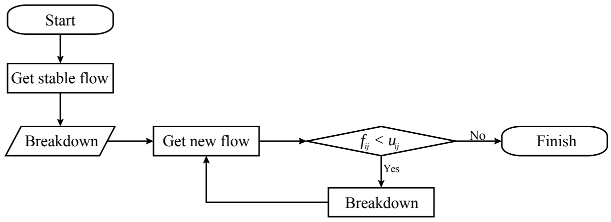


Figure 1: A flow chart for cascading failure

According to the procedure of cascading failure, one can expect huge failure of the entire network, even though an initial event destroys a single element. If the capacity of links are not enough to stand increasing burden, the extent of damage would be more serious.

3. A MATHEMATICAL MODEL FOR AN ELECTICAL GRID

As shown in Figure 1 and explained in the previous section, an electrical grid tends to remain stable due to its physical property. Therefore, modelling the balancing behaviour of the electricity in a network is one of the important jobs to be done. In this section, we explain the mathematical structure of DC-approximated networks and two alternative methodologies for obtaining stable flow of the networks. Some fundamental mathematical structure of the model has been borrowed in the previous study (Bienstock and Verma 1996).

3.1. A Mathematical Model

A network for an electrical power grid has three types of nodes: Generator, customer, and intermediate nodes. Generator nodes are responsible for generating the electricity and include various types of power plants. Though generator nodes may use the electricity themselves, we consider them as the nodes without consumption, and use the net value of the electricity as the amount of generation. Customer nodes are nodes that consume the electricity without generation. Intermediate nodes are nodes with net value of 0, which means the nodes are the medium connecting generators and customers. We model the supply/demand of the nodes in a vector $\begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix}^T$, where m is the number of nodes in a network, and b_i is the amount of supply/demand of the node i. If b_i is positive/negative/zero, node i is considered as a generator/customer/intermediate node.

The structure of a network is modelled in matrix $N = \left[N_{ij} \right]$ $(i = 1,\ldots,m, j = 1,\ldots,n)$. $N_{ij} = 1$ if node i is a source of arc j, $N_{ij} = -1$ if node i is a target of arc j, and $N_{ij} = 0$ otherwise.

In order to obtain flow of each link, say $f = \left[f_j \right]^T$ $(j = 1,\ldots,n)$, we need two additional matrices, X and θ. X is $n \times n$ diagonal matrix such that i th diagonal element indicates the reactance of node i . θ is $m \times 1$ column vector that represents the phase of each node.

In a DC-approximated electrical grid, we can eliminate a node assuming the phase θ of the node to be zero. This is because the phase of nodes in a network should be synchronised and a single node is able to be considered as the synchronisation node. Therefore, we can reduce the dimension of N , b , and θ into $\overline{N}$, $\overline{b}$, and $\overline{\theta}$, respectively, by removing the corresponding row of the matrices or vectors. Then, the phase and the flow of a network are uniquely determined with the following equations:

$$\overline{\theta} = (\overline{N}X^{-1}\overline{N}^T)^{-1}\overline{b} \text{, and} \tag{1}$$

$$f = X^{-1}\overline{N}^T\overline{\theta} \text{ .} \tag{2}$$

3.2. Obtaining Balanced Flow: Matrix Operation

The first method to get flow of a network is calculating the phase vector $\overline{\theta}$ and f using conventional matrix operation. This method is simple but involves a lot of matrix inversion and multiplication operations which require long computation time and large memory. In addition, those operations need to be repeated in every round of cascading, so there is no advantage on statistical analysis either. In small scale problems, however, it is still worth using the matrix operations since they can be easily implemented and solved in reasonable computation time.

3.3. Obtaining Balanced Flow: Linear Programming

Even though the balanced flow is unique, linear programming approach guarantees much faster computation time in many cases. This is because linear programming solvers such as CPLEX and Gurobi do not perform full matrix operations by reducing the original matrices, and tend to get a solution in the initialisation phase according to the properties of algorithms for linear programming (e.g. simplex method) (International Business Machines Corporation 2012, Gurobi Optimization 2013).

In addition, a linear programming model which has been built once is reused with the dual simplex method. As mentioned before, the flow vector is calculated from the first in every cascading round with pure matrix operations. However, minor changes of constraints and decision variables can be adopted without remodelling if the dual simplex method is used for the linear programming model. Hence, after a linear programming model is set, the single model can be used in all cascading round, and even in multiple replication of experiments. This approach should save huge amount of time. Since it is known that there is only one solution, setting the objective function is trivial. We set an arbitrary objective function that maximises the sum of flow. The linear programming model is written as follows:

$$\max \quad \sum_{i=1}^{n} f_i$$

subject to

$$\overline{N}f = \overline{b} \tag{3}$$

$$Xf = \overline{N}^T\overline{\theta} \tag{4}$$

Constraints (3) and (4) are equivalent to (1) and (2). Equation (3) describes the flow has to satisfy every supply/demand of the nodes, and (4) presents the relationships among the flow, the reactance, and the phase.

In experiments that we conducted and explained in next section, the linear programming based procedure showed much higher computational performance. For networks with 1,000 nodes, the linear programming approach finished simulation in average of 5 seconds, while matrix operation took more than 10 minutes in most cases. Therefore, it is wise to use linear programming model for national-wide networks which have numerous nodes and links.

4. SIMULATION CONSIDERING DEMAND SHEDDING

The cascading failure model that we presented in Section 2 with Figure 1 assumes that the total amount of supply and that of demand is the same. This assumption is quite intuitive in terms of the fact that the power grid cannot have surplus electricity and the demand is satisfied in a stable network. However, the balance of the supply and demand breaks when one or more elements of a network fail. In order to meet the balance of the supply and demand, generators reduce the amount of generation when the sum of the supply is greater than the sum of the demand, and demand nodes forcibly reduce the amount of demand by causing partial outage when the sum of the demand exceeds that of the supply. This procedure is called demand shedding.

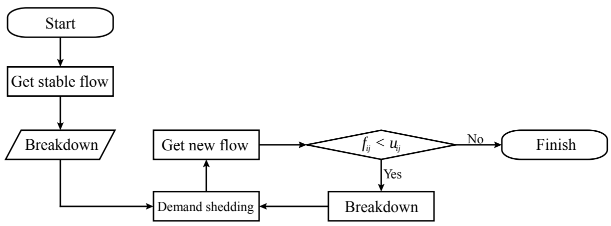


Figure 2: A flowchart for cascading failure considering demand shedding

The notable issue here is that the method to carry out demand shedding is not physically determined and there is room for humans to be involved as decision makers. That is, the damage caused by cascading failure can be lessened according to which demand shedding policy is used. Figure 2 shows the updated cascading failure flow chart applying a demand shedding policy. Most of the previous studies that we reviewed in Section 1 did not consider demand shedding seriously.

In this section, we review typical demand shedding policies that can be used in the real industry.

When power outage occurs, a decision maker needs to balance supply and demand manually. This shedding can be adjusting the amount of generation of a power plant or leading some demand nodes to be gone out. Since a power grid deals with the electricity and the tolerance of the network for standing unstable state may not be too long, the decision has to be made quickly to avoid severe damage on the network. If the decision making process is automated, the automated procedure should save a lot of time for shedding and eventually reduce damage on the network.

The demand shedding can be mathematically optimised to minimise the total damage on the network, it may need long computation time for a large network and not be able to meet a desired time limit for decision making. Therefore, we suggest to use pre-defined shedding rules based on empirical knowledge from the real system. We present four demand shedding policies on behalf of various possible policies.

4.1. Proportional Shedding Policy

This policy does not prioritise the nodes in a network. If some nodes fail, so the total amount of the demand exceeds the total supply, the policy decrease demand of all customer nodes based on the proportion of the demand of each node to the total amount of original demand. For example, let two customer nodes, say A and B, have demand of 100 and 200, respectively. If there is failure on some generator nodes and 60 should be reduced, A and B have 80 and 160 after the proportional shedding, respectively. This is because A has 1/3 of the total demand, B has 2/3 of the total demand, and the proportional shedding policy does not break the ratio of the demand of nodes to the demand of other nodes. The procedure is the same on the situation that the total supply is greater than total demand.

4.2. Largest-demand-first Shedding Policy

This policy gives priority to nodes according to the amount of their demand (supply). If imbalance occurs, the policy reduces the amount of demand (supply) in order of the priority. For instance, if there are two nodes A (100) and B (200), and 220 should be reduced, the nodes remain with demand of 80 and 0 after the shedding, respectively. Since the demand of node B is larger than the demand of A, B has higher priority. Therefore, the demand of B has been reduced to 0, and the demand of A has been reduced by the remaining 20.

4.3. Fewest-connection-first Shedding Policy

Fewest-connection-first policy is the same with the Largest-demand-first shedding policy in terms of prioritising the nodes in a network. The difference is that this policy gives higher priority to nodes with the fewer number of outgoing links. The electrical grid is modelled by a network with directed links (directed graph). This policy has come from the idea that avoiding demand shedding on hub nodes which are connected to many other nodes may preserve the original demand well.

4.4. Fewest-outgoing-first Shedding Policy

Fewest-connection-first policy introduced in Section 4.1.3 gives priorities to nodes considering the number of connected other nodes. Fewest-outgoing-first policy is similar with the Fewest-connection-first policy except that this policy only counts the number of outgoing links. Shedding a node with many outgoing links may affect a lot of other nodes, and this policy is intended to minimise damage by preserve those nodes as long as possible.

5. SIMULATION

We implemented the concepts explained in the previous sections into Java based software platform which can process existing network data or generate random network, and simulate cascading failure after an initial failure event (see Figure 3).

Figure 3: Cascading simulator

In the simulator, a user can customise the following conditions as input for cascading simulation.

- Input network: A user can put pre-defined input files describing a network or make the simulator generate an arbitrary network.
- Alpha (α): Exponential smoothing parameter for links (see Section 4.2.1).
- Tolerance (T): Tolerance of links in a network. Each link sets to have capacity of the initial stable flow multiplied by tolerance ($f_{ij}^{0} \times T$).
- Disaster scale (D): The scale of initial failure event. The simulator let D arbitrary elements in a network fail at first.
- Simulation speed.
- The number of replication.
- Demand shedding policy.

5.1. Simulation

In this subsection, we discuss details of the behaviour of the simulator.

5.1.1. Simulation Model

Figure 4 shows a simplified diagram representing the structure (objects) of a simulation model. A network object has node objects and link objects, and can be converted into a matrix introduced in Section 3.1. A Node object holds information of the location and the amount of power supply or demand, while a Link object has capacity, reactance, and its source and target objects (Nodes). A simulation model refers to a Setting object that contains input customization data mentioned at first in this section.

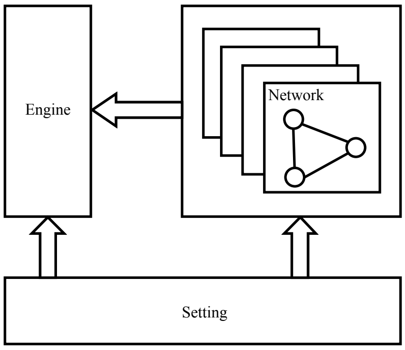


Figure 4: The structure of a simulation model and a simulation engine

As shown in Figure 4, we defined a simulation engine to be separated from a simulation model, so that the engine can manage simulation of general models (networks). The Engine object also uses a Setting to get configuration of simulation experiments.

5.1.2. Failure Model

In the real power grid, a link can stand during certain amount time even if the flow goes beyond the capacity. In other words, each link actually behaves based on the *effective capacity* or the *effective flow* which is the calibrated nominal capacity or flow. There can be many methods for determining whether a link fails or not such as moving average and exponential smoothing.

Moving average and exponential smoothing methods are to obtain the effective flow that can give links some extra sustainability on overflow by smoothing a certain degree of change. The moving average method gets the effective flow by calculating

$$\hat{f}_{ij}^{r} = \frac{1}{r+1}\sum_{k=0}^{r} f_{ij}^{k} \quad (5)$$

where f_{ij}^{k} is the flow from node i to node j at k th round. Exponential smoothing method also uses the effective flow obtained by

$$\hat{f}_{ij}^{r} = \alpha f_{ij}^{r} + (1-\alpha) f_{ij}^{r-1}. \quad (6)$$

On the other hand, we can give extra endurance to the links and use the effective capacity by multiplying extra endurance, say $\overline{T}$, to the capacity u_{ij} . That is,

$$\hat{u}_{ij} = \overline{T} \times u_{ij} = \overline{T} \times T \times f_{ij}^{0} \quad (7)$$

In the simulator, we applied exponential smoothing method to reflect all historical flow data into the effective flow of links.

5.1.3. Network Separation

During a simulation experiment, a network can be divided into several sub-networks after nodes or links are broken. The simulator that we designed considers such network separation as introducing new small networks working independently. That is, if a network is divided into two sub-networks, the simulator then deals with two individual networks. After splitting-up, each network set a new synchronisation node ($\theta_i = 0$) to sync the phase of nodes, and the simulator calculates the stable flow and demand shedding operations in parallel with the other networks.

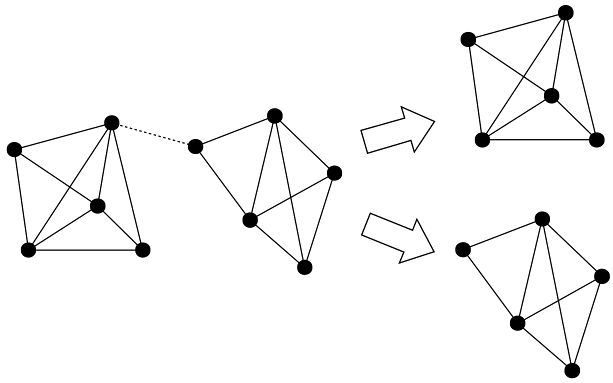


Figure 5: An example of network separation

Figure 5 illustrates an example of network separation into two networks. If a dotted link of the left network fails, the network is divided to two right side networks and each of them behaves independently upon the other network.

5.1.4. Node Failure

A node in a network fails due to the following reasons.

- Self-failure: During changing the amount of generation or demand, a node can fail with some internal errors. This kind of failure generally happens to generator nodes.
- Isolation: If all of connected links from/to the node fail, the node gets isolated and cannot do anything for the whole network, even though the node itself is still functional.
- Exhaustion: A node fails if the amount of demand or supply after converges to zero after shedding.

The simulator does not deal with self-failure of the node and assumes that all nodes can stand any types of demand/supply changes. In case of isolation, the simulator considers isolated nodes as failed nodes and removes from the network. Finally, generator and customer nodes that no longer have demand/supply (nodes that are exhausted) are treated as failed nodes.

The nodes may be able to function as intermediate nodes, but the simulator removes the nodes for consistency.

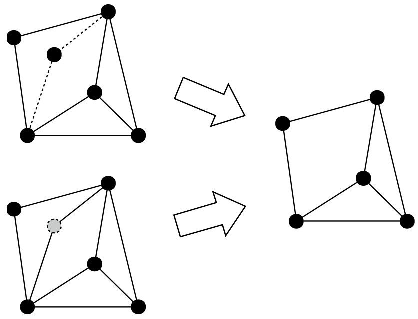


Figure 6: Examples of node failure

Figure 6 shows examples of node failure. The network on the top of left column describes isolation of a node. The node is removed after all of its connected links. The bottom network in the same column shows the case of self-failure and exhaustion. After the node is determined not to be functional, the node and its connected links are all removed.

5.2. Experiment: Comparing Demand Shedding Policies

We conducted simulation experiments with the simulator that is explained in the previous section to compare demand shedding policies introduced in Section 4. We simulated many types of networks as part of an effort to analyse the Korea electrical grid and present two virtual networks that are noteworthy.

The first sample network is tree-like and contains 102 nodes. The network has 4 generator nodes out of the 102 nodes and they are located in the root of tree. The second network is a much complex network that consists of 217 nodes and 285 links. In this network, each node is entangled by the other nodes more tightly.

We set the initial capacity of the nodes as 1.3 of the initial stable flow, which means that T is 1.3. In order to observe the dramatic effect of small failure, the scale of initial failure T was set to be 1. In addition, the value α was 0.5 to give the same proportion of historical data and the newly determined data. The simulation experiments were replicated 1,000 times for each network and demand shedding policy.

The simulation result is shown in Table 1. Proportional, LDF, FCF, and FOF in the first column of the table means the proportional shedding, the largest-demand-first policy, the fewest-connection-first, and the fewest-outgoing-first, respectively. The results for both the first and the second networks imply that the proportional shedding policy guarantees better performance compared to the other policies. In addition, LDF, FCF, and FOF policies do not show critical difference in terms of both the number of survived nodes and the percentage of preserved demand. Of course, these results do not cover all types of networks, but they are still valuable as the reference for the future studies. Besides, since the fundamental ideas (algorithms) and the simulator can deal with general networks provided in the pre-defined format, this study makes the future study much easier to achieve additional valuable results.

Table 1: Simulation Result

	Network 1	
	% of survived nodes	% of preserved demand
Proportional	87.3	87.0
LDF	77.3	78.1
FCF	76.8	77.4
FOF	76.3	76.6
	Network 2	
	# of survived nodes	% of preserved demand
Proportional	71.7	74.6
LDF	64.6	67.7
FCF	63.6	66.7
FOF	64.4	67.4

6. CONCLUSION

In this paper, we proposed a framework to conduct a simulation experiments for cascading failure of an electrical power grid. The cascading procedure, internal mathematical models and operations were introduced, and the simulator utilising the concepts had been implemented. In addition, we emphasised the importance of demand shedding policies and showed that cascading failure with different shedding policies end up differently in terms of the amount of preserved demand and the number of nodes.

Since this study deals with network models with capacity, the concepts that were introduced throughout the paper may be applicable to other reference systems such as the gas/oil system and the traffic systems like public transportation.

For further study, the simulation needs to consider self-failure of nodes which can cause serious problems and is decided with complex mechanism in real industry. Applying self-failure may include giving probability to fail to vulnerable nodes in the network and determine whether each node fails or not in every round of cascading. Besides, networks that have node capacity as well as link capacity should be considered for more realistic analysis.

ACKNOWLEDGMENTS

This research was supported by Integrative Modeling & Simulation System for Disaster Response System Design through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0006571).

REFERENCES

Bienstock, D. and Verma, A., 2010. The $N-k$ problem in power grids: New models, formulations, and numerical experiments. *SIAM Journal of Optimization*, 20 (5), 2352–2380.

Carreras, B., Lynch, V., Dobson, I., and Newman, D.E., 2002. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos*, 12 (4), 985–994.

Carreras, B., Lynch, V., Dobson, I., and Newman, D.E., 2004. Complex dynamics of blackouts in power transmission systems. *Chaos*, 14 (3), 643–652.

Chen, J., Thorp, J.S., and Dobson, I., 2005. Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model. *International Journal of Electrical Power & Energy Systems*, 27 (4), 318–326.

Dobson, I., Carreras, B., Lynch, V., and Newman, D.E., 2007. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos*, 17 (2), 026103.

Dobson, I., Carreras, B., Newman, D.E., 2005. A loading-dependent model of probabilistic cascading failure. *Probability in the Engineering and Informational Sciences*, 19 (1), 15–32.

Guha, S., Moss, A., Naor, J.S., and Schieber, B., 1999. Efficient recovery from power outage. *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pp. 574-582. May 1-4, New York (NY, USA).

Gurobi Optimization, 2013. *Gurobi optimizer reference manual*. Gurobi Optimization. Available from: http://www.gurobi.com/documentation/5.1/reference-manual [accessed 27 March 2013].

Han, S.-R., Guikema, S.D., Quiring, S.M., Lee, K.-H., Rosowsky, D., and Davidson, R.A., 2009. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. *Reliability Engineering and System Safety*, 94 (2), 199-210.

Hardiman, R.C., Kumbale, M., and Makarov, Y.V., 2004. An advanced tool for analyzing multiple cascading failures. *Proceedings of the 8th International Conference on Probabilistic Methods Applied to Power Systems*, pp. 629-634. September 12-16, Ames (IA, USA).

International Business Machines Corporation, 2012. *IBM ILOG CPLEX optimization studio: Getting started with CPLEX*. Version 12 Release 5. USA: International Business Machines Corporation.

Langevin, A., Perrier, N., Agard, B., Baptiste, P., Frayret, J.-M., Pellerin, R., Riopel, D., and Trépanier, M., 2009. Models and methods for emergency response logistics in electric distribution systems. *Proceedings of the 2nd International Workshop on Successful Strategies in Supply Chain Management*. January 8-9, Hong Kong (China).

Nedic, D.P., Dobson, I., Kirschen, D.S., Carreras, B.A., and Lynch, V.E., 2011. Criticality in a cascading failure blackout model. *International Journal of Electrical Power & Energy Systems*, 28 (9), 627-633.

Pfitzner, R., Turitsyn, K., and Chertkov, M., 2011. Statistical classification of cascading failures in power grids. *Proceedings of the 2011 IEEE Power & Energy Society General Meeting*, pp. 1-8. July 24-29, Detroit (MI, USA).

Pinar, A., Meza, J., Donde, V., and Lesieutre, B., 2010. Optimization strategies for the vulnerability analysis of the electric power grid. *SIAM Journal on Optimization*, 20 (4), 1786-1810.

Xu, N., Guikema, S.D., Davidson, R.A., Nozick, L.K., Çağnan, Z., and Vaziri, K., 2007. Optimizing scheduling of post-earthquake electric power restoration tasks. *Earthquake Engineering & Structural Dynamics*, 36 (2), 265-284.

AUTHORS BIOGRAPHIES

CHULHAN KIM received the B.S. degree in KAIST in 2010. He is in the Ph. D. program in the same university. His main research interests focus on feedback control design of timed event graph, optimization of timed Petri nets, and simulation of timed Petri nets. He is also involved in developing methodologies for minimising damages from disasters on networks, especially on electrical power grids.

KIWOOK BAE is a B.S. student in KAIST. He mainly studies on mathematical approaches to cascading failure of network systems.

SOOJEONG BAE is a B.S. student in KAIST. She focuses on research including demand shedding policies to reduce damages on electrical power grids.

TAE-EOG LEE is a Professor with the Department of Industrial and Systems Engineering, KAIST. His research interests include cyclic scheduling theory, scheduling and control theory of timed discrete-event dynamic systems, and their application to scheduling and control of automated manufacturing systems. He is also a head of Integrative Modeling & Simulation System for Disaster Response System Design project funded by the Ministry of Education, Science and Technology in South Korea.

FALLING STARS: SIMULATION OF THE EFFECT OF FIRM CULTURE ON EMPLOYEE PERFORMANCE

Roderick Duncan[(a)], Terry Bossomaier[(b)]

[(a)]Centre for Research in Complex Systems
Charles Sturt University
Bathurst NSW Australia
[(b)]Centre for Research in Complex Systems
Charles Sturt University
Bathurst NSW Australia

[(a)]rduncan@csu.edu.au, [(b)]tbossomaier@csu.edu.au

ABSTRACT
Many modern firms follow the strategy of recruiting talented employees from outside the firm in the hope of improving firm performance. This strategy is not confined to business alone, but can be found in such industries as sports, the arts and higher education. In order to gain a deeper insight into when this strategy would be successful, we designed an agent based model, featuring competition amongst firms requiring expertise along one or more dimensions. Firms can meet these expertise requirements through the formation of networks within the firm. These networks may make the recruitment of a star less successful than it might appear, as shown by Groysberg in a recent book on financial analysts.

Keywords: employee performance, firm performance, firm culture, simulation

1. INTRODUCTION
A common statement in human resource management research is that a firm's success depends on the talent of its employees, whether in production, sales or management. It is argued that a crucial decision for any firm is the choice to develop the skills of its employees or to acquire talented individuals from outside the firm (Lepak and Snell 1999, 2002). This decision has been likened to a "make or buy" decision (Miles and Snow 1984).

The perception is then that an advantageous corporate strategy may be to recruit external stars - highly productive employees outside the firm - in order to boost firm performance. There have been many occasions on which the market value of firms has soared upon the announcement of the hiring of a new senior manager, however bringing external stars into firms has been highly successful in some notable cases and highly disastrous in others (Groysberg, McLean and Nohria 2006). The importance of talented employees was highlighted by a group of McKinsey consultants (Michaels, Handfield-Jones and Axelrod 2001) who coined the term "the war for talent" arguing that firms are engaged in a war to recruit, develop and to retain their talented employees against competing firms.

Popular media has highlighted examples of industries which seem to be dominated by stars and in which star employee transfer is common. In contrast *The Economist* (Economist 2011) discussed the case of the Barcelona soccer club. In a sport that is dominated by clubs which recruit globally, the Barcelona club retained a developmental approach to acquiring talent, and at the time of the European Champions League final in 2011 - which Barcelona won - the majority of the players and the coach were from Catalan. What then determines whether an industry will be one in which talent is acquired or developed? Are there key features of industries which determine whether the industry will be dominated by stars?

2. THE STRATEGY OF HIRING STARS
One of the assumptions of this "hiring stars" strategy is that star employees of other firms can bring the key features of their performance over to the new firm (Groysberg, McLean and Nohria 2006). This "portability" of skills has been questioned in research on star security analysts on Wall Street (Groysberg, Lee and Nanda 2008, Groysberg and Lee 2009, Groysberg 2012). These analysts comment on and predict the performance of firms in various sectors of the US economy. Despite the seemingly generalized nature of the skills of these analysts, it was found that third party ranking of an analyst fell significantly for up to five years after transferring from one Wall Street firm to another. This research found that the star analyst's productivity was not immediately transferable across firms in the financial services sector.

In explaining what determined employee performance, Becker set out two forms of human capital – "general human capital" and "specific human capital" (Becker 1962, 1964). General human capital was those skills that are easily transferred between workplaces while specific human capital was not. Groysberg et al (2008) set out a typology of forms of human capital that differed according to their transferability between jobs.

The relative importance of these different forms of human capital in the new workplace will then determine how portable the new employee's skills are.

Firm culture and employee behavior are interdependent. In order to develop employee skills, firms have to transfer resources from more productive employees to less productive employees. Stars may refuse to do this and demand the entire value of their performance as salary. Becker pointed out that we would not expect firms to invest in employee skills that are highly portable, as the employee may simply leave the organization after training or demand an immediate increase in salary after becoming more productive. In industries dominated by stars, we would expect to see less emphasis on employee development, greater salary inequality across the firm and higher mobility of employees between firms. We call industries in which this thinking is dominant the "star" culture.

Firms may invest in employee skills that are firm-specific, which are of value to the firm but are not of value in the labour market outside the firm. In an industry that is not dominated by stars, we would expect to see more emphasis on employee development, less inequality in salary structure across the firm and lower mobility of employees between firms. Following the recent literature, we call this corporate culture "pro-social".

Yet there are industries in which we see a mix of firms with star structures and firms without star structures, such as the European soccer league. In those industries, firm culture and employee behavior support both a star equilibrium and a pro-social equilibrium. In this paper we produce a simulation of employee performance and firm culture to highlight the key features of industries in which we see labor markets with stars, without stars and with a mix of strategies.

3. THE SIMULATION MODEL

The model works on the basis that almost all firm based phenomena are emergent from the behavior of the individuals (agents), indexed by *i*, who make up the firm.

3.1. Agents

Agents have a single behavioral characteristic "pro-sociality" denoted by β, where $0< \beta <1$, which measures the extent to which agents are willing to share resources and help others. Agents with high prosociality are more willing to sacrifice their own salaries to develop other agents in the same firm. Agents with high prosociality are also less likely to change firms when offered a recruitment package by a competing firm.

Agents contribute to the income of the firm in which they are located directly through the agent's expertise, e_i, as well as indirectly through a "network effect" which depends on the strength of the ties between that agent and another agent in the firm and the expertise of both agents. The strength of these network effects is one of the parameters of interest in our model. Do network effects within firms in certain industries determine whether star systems can be successful, as has been suggested by Groysberg's research?

3.2. Industry

The industry in which the firms operate has two characteristics which are common across all firms. The first industry characteristic is the strength of network effects in firm income is represented by ρ, where $0<\rho<1$. The second industry characteristic is the fraction of firm profits taken by the owners of the firms. We assume that firms are owned by partners who take a fraction, $1-\gamma$, of firm profits, where γ is common across all firms and is determined by industry standards. The fraction γ is set at 0.33 for these simulations.

3.3. Firms and Firm Income

Firms are essentially networks of agents. Firms compete for jobs against other firms through networks of agents in each firm. The value of a job if won by agent *i*, J_i, is determined by the expertise of the agent *i* and the network relationships that agent i possesses within the firm, W_{ij}, moderated by the strength of network effects in the industry, ρ

$$J_i = \underline{e_i} W(\rho) \underline{e_i'} = e_i^2 + \frac{1}{2}\rho \sum_{\substack{j\in f \\ j\neq i}} e_i W_{ij} e_j . \qquad (1)$$

where *f* is the set of agents within the firm. These network relationships, W_{ij}, are either on or off depending on whether the agents have established a working relationship prior to the job being undertaken. Firm profits are the firm's share, γ, of the sum of the jobs won by agents in the firm less the sum of the salary costs of the agents in the firm.

As with agents there is a single behavioral characteristic for firms, prosociality. The value of prosociality for a firm, β, is derived from a random agent in the firm, who might be considered the founding agent of the firm. The level of prosociality for a firm determines the level of investment of the profits of the firm which is used for developmental and mentoring activities within the firm.

3.4. Agent Remuneration

The salary for each agent, s_i, is given by a base salary and a bonus salary component. Following practice in the finance industry (Groysberg 2012, p. 275), the base salary is comparatively low and uniform across agents, while the bonus is highly dependent on agent performance and varies greatly across agents.

For simplicity, the base salary is set to zero, while the agent's bonus depends on the job value which the worker generates for the firm, J_i, less the share γ taken as profits by the firm as well as the share of job value taken by the other agents in the firm networked to the winning agent.

Within firms, however, there are internal transfers to fund employee development and mentoring, so agents pay an internal "tax" to fund these activities based on the firm's prosociality. These internal transfers within the firm are reflected in the sharing rule which splits up the value of a winning job according to the expertise of the agents within the winning network and the prosociality of the firm in which they work. The share of the job value accruing to each agent in the winning network, *n*, net of the share taken by the firm is:

$$Share_i = \frac{e_i^{2(1-\beta)}}{\sum_{j \in n} e_j^{2(1-\beta)}}. \quad (2)$$

The effect of this sharing rule is to make the split of the job value between agents more even when firm prosociality, β, is higher. For β equal to one, each agent in the winning network receives the same share of the job value. For β equal to zero, agents receive a share proportionate to the square of their expertise over the sum of the square of the expertise of all other agents in the network.

3.5. Network Growth

The network weights between agents within a firm, W_{ij}, grow at a rate which is dependent on the developmental spending within the firm – the prosociality of the firm - and the expertise of the other agents to which the agent is linked within the firm.

3.6. Recruitment

Firms with low pro-sociality values may opt to hire a star agent from a different firm. If the prosociality parameter, β, of the hiring firm is greater than that of the star's current firm the star moves, thus a star knows that he/she will get a higher salary – net of internal transfers within the firm to fund employee development - by moving.

When a star moves into a new firm, the star's network weights to agents in the new firm are set to a low number. The star is assumed to commence in the new firm without strong relationships to other agents in the new firm. The cost of a star's movement between firms is the loss of the network relationships the star possessed in the old firm.

4. RESULTS OF THE SIMULATION MODEL

The simulations were conducted in Netlogo (Wilenky 1999). The simulations involved 2048 agents and 64 firms. As a comparison the Groysberg research on star security analysts involved 1500 agent in each of several sectors of the US economy.

In the Netlogo simulation, we implemented a visual representation of the model. Firms were represented by solid patches, and agents in firms by individual workers. Jobs to compete for are the patches between firms. Figure 1 shows a portion of the firms and agents in one of the simulations. The networks between agents are shown by the lines linking agents together.

Workers migrate to the outside of firm patches and compete for jobs against neighboring firms. In this simplified version of the model, agents only draw a salary based on the contribution of their expertise to the firm through jobs won, e_i^2, so bonuses are set to zero.

Investment in development and mentoring by firms is represented by faster development of network relationships between agents, which can only happen when agents encounter each other. Agents with higher levels of expertise are allowed to move each time step with greater probability and so can both reach the edge to compete for jobs faster than low expertise agents and also encounter other agents within the firm and form networks more quickly.

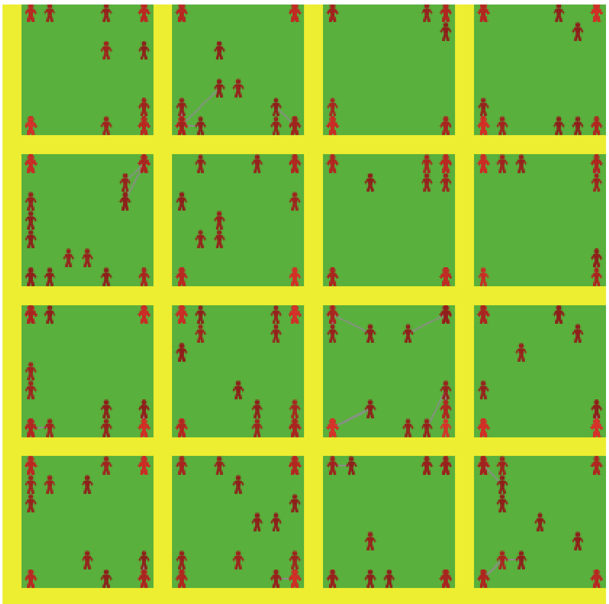


Figure 1: Portion of Netlogo Simulation of Firm Model.

Recruitment of stars is done locally. If an employee loses a competition for a job to the employee of another firm, the losing employee may opt to shift to the winning firm. Employees will only shift to winning firms if the firm offers higher salaries net of internal firm transfers and if the worker's own level of prosociality is low. The winning firm, however, will only recruit the losing employee if that employee has higher expertise than the average agent in the winning firm – only stars get recruited.

Simulations are run for 100 time steps to allow sufficient time for agents to generate network relationships within the firms, for agents to compete between firms for jobs and for agents to shift from less successful firms to more successful firms. The results for two typical simulations with different levels of network effects are presented.

Groysberg's research has suggested that network strength between agents and infrastructure within a firm are the key factors which explain why star systems may not boost firm performance. We would expect then that this would be reflected in the simulation by high prosociality firms performing better when the network effects within the firm, ρ, are strong. In that case, the presence of strong network effects means that internal staff development matters greatly. Conversely we would expect that when network effects are weak, low prosociality firms would out-perform high prosociality firms as the advantage of recruiting stars for low prosocial firms outweighs the small network effects.

Figure 2 shows the relationship between firm prosociality and firm profitability when network effects are strong. The developmental investments by highly prosocial firms mean that employees in those firms form strong relationship networks so that those employees remain successful in job competitions.

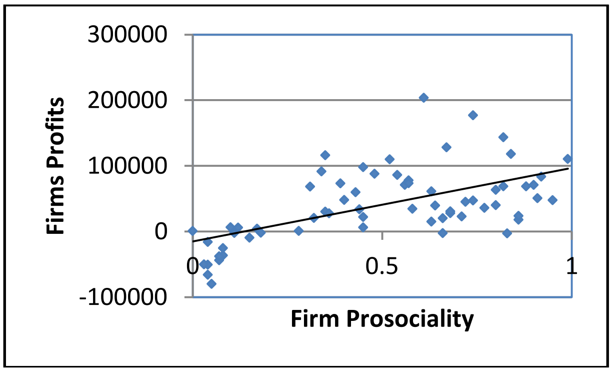


Figure 2: Relationship between Firm Prosociality and Firm Profitability under Strong Network Effects.

Figure 3 shows the relationship between firm prosociality and firm profitability when network effects are weak. When network effects are weak, the ability of the lower prosociality firms to steal stars from the high prosociality firms outweighs the faster development of relationship networks within the high prosocial firms. The number of agents in the high prosocial firms falls, and the profits of those firms falls as well.

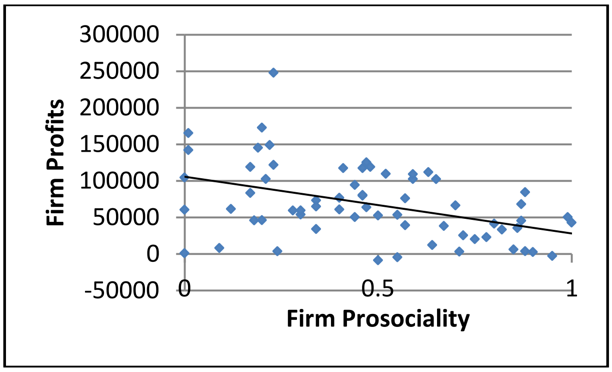


Figure 3: Relationship between Firm Prosociality and Firm Profitability under Weak Network Effects.

Groysberg (2012) indicated that the financial analyst industry was an industry with high network effects, given that stars who moved often took many years to regain the productivity they had in their original firms. In such a highly networked industry, Groysberg found that star analysts were less mobile than non-star analysts (Groysberg 2012, Table 10.1). This result matches the mobility of stars in Figures 4 and 5, where more agents remain in the higher prosocial firms in the simulation with strong network effects.

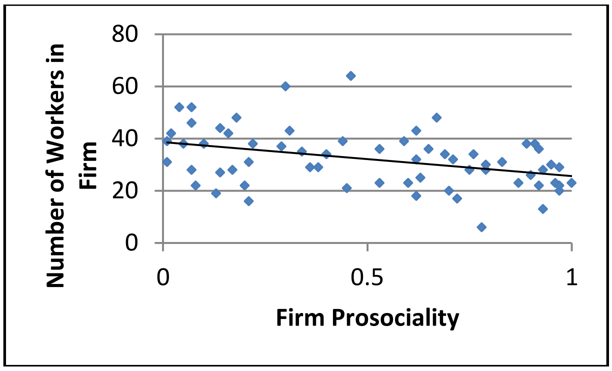


Figure 4: Relationship between Firm Prosociality and Number of Workers under Strong Network Effects.

As can be seen in Figure 4, with strong network effects the high prosocial firms can generate networks within their firm to prevent their star workers from being recruited away by low prosocial firms. However in Figure 5, with weak network effects, the high prosocial firms lose more of their star workers to recruiting efforts of the low prosocial firms.

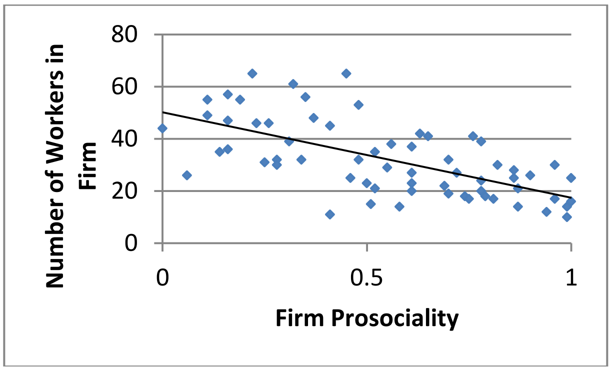


Figure 5: Relationship between Firm Prosociality and Number of Workers under Weak Network Effects.

These results suggest that a star system can work in an industry with low network strength. In that case, while recruiting stars does have the cost of losing the networks the star has in the former firm, the effect of the lost networks is low enough that hiring highly productive star workers pays off for the recruiting firm. However we also see that, in an industry with strong network effects, hiring stars might be counterproductive as the loss of the networks in the former firm can be costly to recover. A better strategy in the case of an industry with strong network effects may be to spend resources to develop their own workers.

5. CONCLUSIONS

It is common practice in many industries to seek performance improvements in organizations by recruiting talented outsiders rather than developing workers internally. Recent research by Groysberg and others has questioned whether this practice is effective in all industries, as Groysberg and co-authors have found evidence that hiring stars is not effective for Wall Street investment banks.

Groysberg (2012) has suggested that the star analysts are not as successful in the finance industry as would be expected, because the industry has strong network effects within firms. Stars transferring firms lose the network relationships those stars had in their former firm and developing new relationship in the new firm which hired them takes time – years, he suggests. The new firms who hire the stars will find that those stars who transfer with high salaries take a considerable time to get back to the levels of productivity they enjoyed at their former firm. Firms in industries with large network effects then may be better off developing their own employees, even though such development is costly.

We find that a simple agent based simulation, with just two parameters accounts for Groysberg's findings, when network effects are strong. This was hypothesized by Goldberg and our simulation provides a demonstration of the validity of this hypothesis. We compare the performance of firms which rely on hiring stars and spend less on employee development against firms which hire rarely and spend more on employee development. In an industry with strong network effects, firms which rely on hiring stars do less well than firms which concentrate on staff development. We find the opposite in industries with weak network effects.

REFERENCES

Becker, G.S., 1962. Investment in human capital: a theoretical analysis. *Journal of Political Economy*, 70(5), 9-49.

Becker, G.S., 1964. *Human Capital: a Theoretical and Empirical Analysis with Special Reference to Education*. 2nd ed. New York: National Bureau of Economic Research.

Groysberg, B., 2012. *Chasing Stars: the Myth of Talent and the Portability of Performance*. Princeton, NJ: Princeton University Press.

Groysberg, B., and Lee, L.E., 2009. Hiring stars and their colleagues: exploration and exploitation in professionsl service firms. *Organization Science*, 20(4), 740-758.

Groysberg, B., Lee, L.-E., Nanda, A., 2008. Can they take it with them? The portability of star knowledge workers' performance. *Management Science*, 54(7), 1213-1230.

Groysberg, B., McLean, A.N., Nohria, N., 2006. Are leaders portable? *Harvard Business Review*, May, 1-9.

Lepak, D.P., Snell, S.A., 1999. The human resource architecture: toward a theory of human cpaital allocation and development. *Academy of Management Review*, 24 (1), 31-48.

Lepak, D.P., Snell, S.A., 2002. Examining the human resouce architecture: the relationships among human capital, employment, and human resource configurations. *Journal of Management*, 28 (4), 517-543.

Michaels, E., Handfield-Jones, H., Axelrod, B., 2001. *The War for Talent*. Boston, MA: Harvard Business School Press, 2001.

Miles, R.E., Snow, C.C., 1984. Designing strategic human resource systems. *Organizational Dynamics*, 13, 36-52.

The Catalan Kings: the management secrets of the Barcelona football club. *Economist*, 19th May 2011.

Wilensky, U., 1999. *Netlogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. Available from: http://ccl.northwestern.edu/netlogo/.

USING CONVERSIVE HIDDEN NON-MARKOVIAN MODELS FOR MULTI-TOUCH GESTURE RECOGNITION

Tim Dittmar[(a)], Claudia Krull[(b)], Graham Horton[(c)]

[(a)(b)(c)] Otto-von-Guericke-University Magdeburg
P.O. Box 4120
39016 Magdeburg, Germany

[(a)]tim.dittmar@ovgu.de, [(b)]claudia.krull@ovgu.de, [(c)]graham.horton@ovgu.de

ABSTRACT
With the current boom of multi-touch devices the recognition of multi-touch gestures is becoming an important field of research. Performing such gestures can be seen as a stochastic process, as there can be many little differences between each execution. Therefore stochastic models like Hidden Markov Models have been already utilized for gesture recognition. Although the modelling possibilities of Hidden Markov Models are limited, they achieve an acceptable recognition quality. But they have never been tested with gestures that only differ in execution speed. Therefore we propose to use Conversive Hidden non-Markovian Models for multi-touch gesture recognition. This extension of Hidden Markov Models enhances the modelling possibilities and adds timing features. In this work two multi-touch gesture recognition systems were developed and implemented based on these two model types. Experiments with a set of similar gestures show that the proposed model is a good and competitive alternative and can even be better than Hidden Markov Models.

Keywords: HMM, CHnMM, gesture, recognition

1. INTRODUCTION

1.1. Background

Due to the big success of smartphones and tablets a ubiquitous presence of multi-touch devices is establishing itself around the world. While this multi-touch input method offers manifold possibilities to control these devices, almost all of them are usually controlled by using a fixed set of simple gestures like tap, drag and pinch. Such systems can be realized quite easily using heuristics but they are not very flexible.

To create a gesture recognition system using heuristics means that the system has to have all of the gestures previously implemented. Adding a new gesture afterwards could make code adaptions of the previously implemented gestures necessary, especially when the gestures are very similar. In order to create a more flexible gesture recognition system other methods need to be used that define a gesture by providing some performed examples. This way the user of a touch device could define the gestures that suit him the most for certain actions.

An existing flexible gesture recognition system for multi-touch devices is presented in the work of Damaraju and Kerne (2008). The system is based on Hidden Markov Models (HMM) and creates one HMM for each gesture according to sample inputs.

While there are other pattern recognition methods that are deployed for multi-touch gesture recognition, the current work focuses on HMM and on a quite new model class: Conversive Hidden non-Markovian Models (CHnMM). With this work we want to evaluate whether CHnMMs are applicable for multi-touch gesture recognition and how they perform in comparison to HMMs.

1.2. Motivation and Goals

The idea to use non-Markovian Models for gesture recognition was first brought into consideration by Bosse et al. (2011). The goal was to show that a system based on Hidden non-Markovian Models (HnMM) could distinguish gestures that are similar in shape but differ in execution speed. This differentiation has not been considered for HMM gesture recognition systems before. For that reason an HMM- and an HnMM-system were developed to recognize gestures performed with a Nintendo Wiimote and both systems were compared in their recognition quality.

Inspired by this, we want to apply a similar approach to multi-touch devices. Furthermore CHnMMs were used instead of HnMMs whose properties and differences are explained in Section 2.2 in more detail.

The general goal of this paper is to find out whether CHnMMs are applicable for multi-touch gesture recognition. Therefore the CHnMM-system needs to reach similar or better recognition rates than the HMM-system. Additionally the time needed for recognizing the gesture has to be competitive, so that the system could be used in real-time scenarios.

2. BACKGROUND INFORMATION

2.1. Hidden Models

In the context of this work the term Hidden Models stands for all model classes that are based on the concept that was introduced with Hidden Markov Models: inferring conclusions from a list of observations of a hidden system, a so called partially observable discrete stochastic (PODS) system (Buchholz 2012). The stochastic behaviour of such a system is known but its discrete states cannot be directly observed. Instead only arbitrary “messages”, so called symbols or signals, which are created by the system, can be recognized. Since the creation of these symbols is incorporated into the model, it is possible to infer conclusions about the state of the real system.

There are some common problems that can be solved with Hidden Models and two of them are relevant to this paper: Evaluation and Training. Evaluation is the task of calculating the probability that a trace O (a sequence of observations) was produced by the hidden model λ, formally $P(O \mid \lambda)$. This probability is often used to find out which model of a set of models was most likely responsible for a certain trace. The Training task is used to find a model λ, so that the result of the Evaluation task reaches a maximum according to a given trace O or a set of traces respectively:

$$\arg\max_{\lambda} = P(O \mid \lambda) \quad (1)$$

Since it is very difficult to create the best model by only knowing the trace, the training is often done by improving an already given model so that $P(O \mid \lambda_{new}) > P(O \mid \lambda)$.

2.2. Related Work

In HMMs the hidden model is a simple discrete time Markov chain with additional information about symbol creation in each of its states. So with every discrete time step a certain symbol will be created depending on the current state. Each possible symbol in that state has a fixed probability to be created. This model class has an obvious limitation: only Markovian systems can be modelled. Most real systems are not Markovian though and therefore a lot of extensions to HMMs are being developed that try to circumvent its limitations.

One of the latest developments in this research area has been achieved by Krull et al. (2009) with Hidden non-Markovian Models (HnMM), a powerful model class that is able to represent non-Markovian discrete systems like stochastic petri nets and combine them with the abilities of Hidden Models. This modelling power comes at the cost of complexity which can be reduced by focusing on a subclass of HnMMs. Buchholz (2012) has researched such a subclass in his work and named it Conversive Hidden non-Markovian Models (CHnMM). In this subclass the modelling potential is slightly restricted but the efficiency of computing solutions increases. For example transitions are only allowed if they produce a symbol. For this reason every state change can be recognized and the calculation of the solution does not have to cope with hidden state changes.

CHnMMs have already been employed in Bosse et al.’s (2011) work, even if they were not explicitly named as such. Furthermore they could not utilize the findings of Buchholz (2012) because they were simply not known at that time. Since this work is inspired by the work of Bosse et al. (2011), we have chosen to use CHnMMs as well. But in contrast to them we can also rely on the optimized algorithms and definitions that were developed by Buchholz.

2.3. Environment

For the experiments, a multi-touch tabletop, provided by the “User Interface & Software Engineering” (UISE) working group (at the computer science department of the Otto-von-Guericke-University Magdeburg), was used which was based upon the FTIR principle (Han 2005) and provided the touch data via the TUIO-protocol. Also a Java class by the UISE group that merges the TUIO messages to three simple events (stroke{Started, Updated, Finished}) was available. The software that processes camera images to TUIO messages has a frequency of approximately 15ms.

3. IMPLEMENTATION

In this section important aspects are covered that explain how both recognition systems are implemented and how they differ in certain details.

3.1. Symbol Generation

Symbols are very important for hidden models, because traces of them are used to deduce conclusions such as whether the trace was created by a certain model (evaluation) or which path of states the real system had probably taken (decoding). The symbol set used in this work consists of eight discrete directions: Up, UpLeft, Left, DownLeft, Down, DownRight, Right, UpRight. Each symbol represents the current direction of movement of a certain finger that touches the device, which is inspired by the symbol set used in the last experiment of Bosse et al. (2011).

Although both systems use the same symbol set, the way the symbols are created is different. For HMMs the symbol emission is connected to states, whereas with CHnMMs it is connected to state transitions. Therefore the emission for HMMs needs to be periodic in order to get clues about the current state. In contrast, a CHnMM works with symbol emissions that occur when the state of the system changes.

As a result the generation of symbols for the HMM system is done as follows: every time a strokeUpdated event is received, the vector from the last known position to the current position is calculated and the direction of this vector is discretised to one of the eight directions defined in the symbol set. This direction is directly emitted as a symbol after calculation. It is done this way for every active finger separately, i.e. two moving fingers would lead to the emission of two direction symbols per processing step. Timestamps of the emissions are not needed or cannot be used for HMMs although it seems to be an important source of information for gesture recognition with different execution speeds.

The calculation of symbols for the CHnMM-system is quite similar, but the symbols are only emitted if the current state or direction of the finger changes or if the finger touches the surface for the first time and therefore no previously moving state of the finger is known. This way a finger that moves in one direction should not emit any symbols until the direction changes. But in practice the data of a straightly moved finger can be noisy, so much so that symbols are emitted although the direction of the movement has not changed. To reduce this noise in the symbol stream, stroke updates are not considered for symbol emission if the calculated direction vector is shorter than an empirically defined threshold. For CHnMMs the timestamp is essential and it is taken from the last known point of a stroke that determines when this point was received as opposed to the time when the symbol was calculated.

Furthermore the two symbols GestureStart and GestureEnd were used for both systems to determine when the execution of a gesture starts and when it ends. For this work we determined that a gesture starts with the first touch of a finger and ends when the last finger is removed from the surface.

3.2. Modelling

A main aspect of both systems is the creation of models as this is what the recognition method is based on. There are quite a lot of possibilities to do the modelling for the recognition systems, e.g. the scope of what a model represents could be one of the following:

- A model represents the whole set of possible gestures
- A model only represents a single gesture
- Two or more models represent a single gesture

The last concept was used in the work of Bosse et al. (2011) where a single gesture was modelled with three models, one for each axis of the acceleration sensors. Our approach is to use one model per gesture and this is the case for both recognition systems.

The next step is to define what the states of the model will represent. Since symbols are used that give information about the direction of movement, it seems reasonable to separate a gesture into phases or states of movement directions. For example, a gesture consisting of a movement to the left and back to the right will be modelled with a single state for each of the movement phases left and right. While the HMM only consists of these movement states, the CHnMM is made of two additional states, namely the Start- and End-state. The Start-state is needed because CHnMMs are state change oriented, i.e. symbols are emitted when a state change occurs. If there were no state before the first movement phase, the first direction symbol could not be incorporated into the model, which would waste important evidence of the current state of the real system. The purpose of the End-state is mainly the ability of the model to refuse incomplete gestures. More details on that are given in Section 3.4.

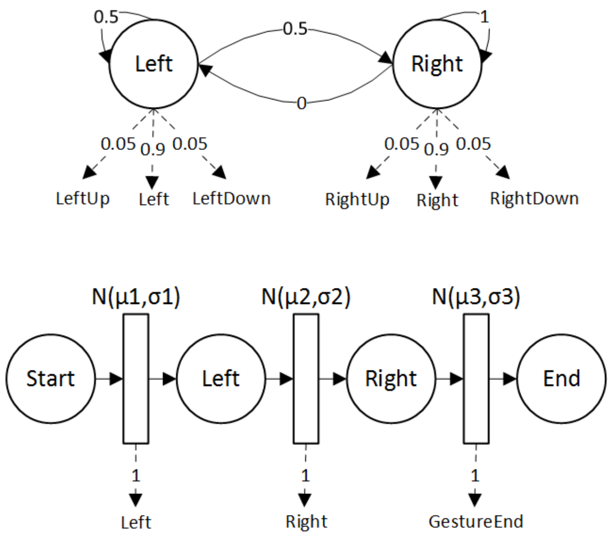


Figure 1: Example Models for a left-right Gesture (HMM: top, CHnMM: bottom)

Of course all of these aforementioned states need to be connected. For HMMs the connection is expressed with fixed probabilities that state how likely it is to stay in a state or to change to another one in every discrete time step, i.e. in this case with every periodic symbol emission. The CHnMMs however are connected with transitions and each of them is described by a certain probability distribution that determines when a state change is happening. We assume that the time for a certain state change between two movement phases is normal distributed, so each transition needs an expectation value and a standard deviation.

Figure 1 shows how an HMM and a CHnMM for a left-right gesture could look like. The dashed arrows mark possible symbol emissions and their probability. It can also be seen that the HMM only has fixed probabilities for state transitions whereas the CHnMM uses normal distributions in this case. Therefore an HMM is dependent on the discrete time step it was created or trained with and as a result will not work reliable with other devices that provide data with a different time step. Note that the HMM in Figure 1 tolerates noisy symbols by giving them a low probability to occur. For the CHnMM case the filtering is not done on model level but on symbol generation level as it was explained in Section 3.1.

It should be understandable now, how a certain gesture can be modelled with either an HMM or CHnMM. The last step to get a usable model is the training, which will be discussed in the next section.

3.3. Training

After a model for a gesture has been developed the parameters like state change and symbol emission probabilities need to be adjusted, so that they fit the finger movement of a user performing the gesture. This section describes how this was done for HMMs and CHnMMs.

The first step is to collect example data of the gesture, so it is performed twenty times with the HMM symbol generation and another twenty times with the CHnMM symbol generation method. Since a Java framework for HMMs is used, namely jahmm (François 2009), the training is simply done by creating a base model of the gesture (as it was described in the previous section) and executing the provided training algorithm with the example traces. The framework uses the Baum-Welch algorithm, a standard training algorithm for HMMs. With every iteration of this algorithm the model approaches a local optimum. That is why it is important to have a good base model to start with.

However the training for CHnMMs is more complex and cannot be done with the usual Baum-Welch algorithm. Buchholz (2012) developed an adapted version for CHnMMs that is nearly as powerful but sometimes creates even worse models. We therefore decided to train the CHnMMs manually similarly to how Bosse et al. (2011) proceeded in their works. This meant that sample traces were put into an excel sheet, movement phases or states were respectively identified and the expectation value and standard deviation for the normal distribution were calculated.

Eventually the process of creating ready-to-use HMMs and CHnMMs was complete.

3.4. Classifying the executed gesture

In this section the process of classifying an executed gesture are elaborated. This is the last part of the implementation before the experiments can be conducted.

The classification with HMMs is done by calculating the probability that the trace of the executed gesture O was created by the gesture model λ (evaluation) for each known gesture model. There is a function in the jahmm framework that does the evaluation based on the forward algorithm. The gesture, whose model created the highest probability, was returned as the classification result.

For the CHnMM system a similar approach was used but the fact that there is an End-state in each model is exploited. So instead of calculating P(O|λ) the forward probability of the End-state at the point of time of the last symbol emission T is taken as a measurement for the likeliness of a gesture. Formally this forward probability is defined as

$$\alpha_T(s_{end}) = P(O \cap q_T = s_{end} \mid \lambda) \tag{2}$$

with s_{end} being the End-state and q_T being the state of the model at time T.

With this approach only traces are considered that reached the End-state and therefore are likely to be complete gestures whereas the evaluation approach would also accept incomplete gestures. For example a trace of a movement to the left could be classified as a left-right gesture, because its model can create such a trace.

This technique could be easily used with HMMs too, but the jahmm framework is not suited for this technique. However, for the goal of this paper this circumstance is not a problem, because no incomplete gestures were executed in the experiments.

4. EXPERIMENTS AND RESULTS

Both implemented systems need to be evaluated in order to compare their recognition quality and performance. Since the recognition systems are classifiers, the commonly used Precision/Recall metric can and will be used, although the classifiers are not binary. The work of Sun and Lim (2001) shows how this metric is used for n-ary classifiers.

The precision of gesture G equals the number of gestures correctly classified as G divided by the number of gestures classified as G. Whenever the classifier returns G the precision value of this gesture states how trustworthy this result is. The number of gestures correctly classified as G divided by the number of performed G gestures is the recall of gesture G. It is a measurement of how complete the classification is.

For the experiments also a fixed set of gestures is needed. Figure 2 shows all six gestures that were chosen. They can be split into two groups of similar gestures: three DownUp and three circular gesture variants. These gestures are special, because they are similar in shape but different in execution speed.

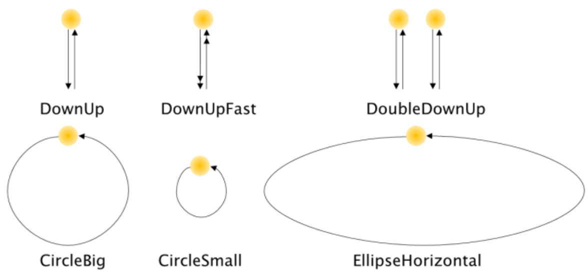


Figure 2: Gesture Set used for the Experiments

The expectation for the experiments is that the HMM-system will have problems to distinguish these similar gestures, whereas the CHnMM-system will have acceptable error rates. The needed computation time is expected to be quite similar. However the CHnMM-system might be even faster, because the number of computation steps is smaller due to the lower number of symbols created with the CHnMM-system.

4.1. First experiment

For the first experiment, each gesture was performed twenty times for each system and the symbol traces were recorded to produce training data for each gesture and system. With this data HMMs and HnMMs were created as described in Section 3 for each gesture.

The result is an HMM and a CHnMM recognition system for the defined gesture set. Again each gesture was performed twenty times for each system and the classified (most likely) gesture was recorded. In the case that no most likely gesture is present (all probabilities are zero) this attempt is recorded as not classified.

Table 1 presents the results obtained for the HMM-system that shows an overall precision of 0.737 and a recall of 0.725. These values are worse than usually measured with HMM gesture recognition systems. Especially the DownUp and DoubleDownUp gestures were problematic and could not be distinguished properly. Also the EllipseHorizontal gesture was classified as CircleBig five times. Other gestures, however, reached acceptable recognition rates. In only two cases no classification could be made, probably due to symbol traces that did not occur while training. This phenomenon is also known as overfitting.

Table 1: Results with HMM-System

Gesture	Correctly classified	Not classified	Pre-cision	Re-call
DU	10	-	0.526	0.500
DUF	17	-	1.000	0.850
DDU	9	2	0.409	0.450
CB	19	-	0.704	0.950
CS	17	-	0.944	0.850
EH	15	-	1.000	0.750
	87	2	**0.737**	**0.725**

The results of the CHnMM-system are presented in Table 2 and show a precision of 0.955 and a recall of 0.883, which are clearly better than the ones of the HMM-system. The DownUpFast gesture has the worst recall, because it was classified four times as a DownUp gesture which is why its precision is lower than the rest. The reason for this is probably that the gesture was performed faster for training than it was performed for classification.

Table 2: Results with CHnMM-System

Gesture	Correctly classified	Not classified	Pre-cision	Re-call
DU	20	-	0.833	1.000
DUF	15	1	1.000	0.750
DDU	19	1	1.000	0.950
CB	19	1	0.950	0.950
CS	16	4	1.000	0.800
EH	17	2	1.000	0.850
	106	9	**0.955**	**0.883**

Another observation is that the general precision is better than the recall. This basically means that you can have a good trust in the classification results but that not all gestures were classified. For real applications where a certain recognized gesture activates a certain command or behaviour the precision needs to be especially high. Otherwise a gesture is performed and the wrong command or behaviour is executed which will lead to a bad user experience.

4.2. Experiment with Training Data

As mentioned in the section before, a possible source of error is the discrepancy of gestures performed during training and classification. To avoid this, the experiment was reconducted, but instead of performing gestures the training data was used as input.

The results that can be seen in Table 3 and Table 4 show much better precision and recall values for both systems as expected. The CHnMM-system even reached a perfect precision; its recall though is slightly lower than the one of the HMM-system. This is because the gesture models of the CHnMM-system were modelled quite strictly, i.e. they did not take noisy data, which occurred a few times in the training, into account. This way the time consuming manual training was performed a little bit easier.

Table 3: Results with HMM-System and Training Data

Gesture	Correctly classified	Pre-cision	Re-call
DU	20	0.741	1.000
DUF	20	1.000	1.000
DDU	13	1.000	0.650
CB	20	0.952	1.000
CS	19	1.000	0.950
EH	20	1.000	1.000
	112	**0.933**	**0.933**

Table 4: Results with CHnMM and Training Data

Gesture	Correctly classified	Not classified	Pre-cision	Re-call
DU	20	-	1.000	1.000
DUF	20	-	1.000	1.000
DDU	18	2	1.000	0.900
CB	20	-	1.000	1.000
CS	16	4	1.000	0.800
EH	16	4	1.000	0.800
	110	10	**1.000**	**0.917**

Besides the good overall precision and recall values, the HMM-system had huge problems recognizing the DoubleDownUp-gesture. The reason seems to be that the HMMs of DownUp and DoubleDownUp are nearly the same although DoubleDownUp gesture emits quite exactly the doubled number of symbols. However, a HMM cannot express this movement because the probability to change from Down-state to Up-state is approximately the same in both gestures. This also explains the problems with this gesture in the first experiment and also shows a weakness of HMMs, even if it recognized the other gestures better than expected.

In both experiments the time from finishing the gesture to getting the classification result was measured. This time was always 0ms if the input data could not be classified. Otherwise the values were mostly 31 or 32ms and sometimes 47ms for both recognition systems. Since the reaction times appear fixed, we assume that this is because of the scheduling side-effects of the operating system (Windows) and that the pure calculation time of the recognition systems is even lower.

5. CONCLUSION

The goal of the paper was to show that CHnMMs can be used in the field of multi-touch gesture recognition and that they are an alternative to HMM recognition systems. The results of the experiments show that the CHnMM-system reached better, if not the same level of recognition quality as the HMM. Additionally both systems are on par regarding recognition speed, which makes CHnMM-systems also usable for real-time scenarios. Therefore all goals of this work have been achieved.

Further observations suggest that HMMs are not suited to distinguish certain gestures apart from others, as was seen with the DownUp and DoubleDownUp gestures, where both are described by nearly the same model. Nevertheless the HMM-system worked better than expected. It can be assumed that although HMMs are timeless, they get fed with timing information through the periodic emission of symbols (approx. every 15ms in our HMM experiments). This way they are able to even distinguish similar gestures with different execution speeds. Systems that do not emit symbols periodically will have more problems with that when using HMMs.

However a big advantage of HMMs is that they can be trained automatically with the Baum-Welch algorithm, making them easier to implement in comparison to manual training. For CHnMMs an efficient training algorithm was developed based on the Baum-Welch one by Buchholz (2012). The quality of this training algorithm in the field of multi-touch gesture recognition needs to be evaluated in future work though.

Further experiments and extensions to this work could be:

- a bigger gesture set,
- using other symbol generation methods,
- more complex multi-touch gestures,
- comparison to other gesture recognition methods (Artificial Neural Networks, Dynamic Time Warping, Support Vector Machines etc.).

As a consequence, CHnMMs could be used to develop flexible multi-touch recognition systems that are independent of the frequency of symbol emission and therefore are independent of the device. Furthermore there are less symbols generated resulting in lesser calculations than HMM or similar systems. This and the time step independency make a CHnMM recognition system suitable for mobile multi-touch devices.

REFERENCES

Bosse, S., Krull, C., Horton, G., 2011. MODELING OF GESTURES WITH DIFFERING EXECUTION SPEEDS: Are Hidden non-Markovian Models Applicable for Gesture Recognition. *Proceedings of the 10th International Conference on Modelling & Applied Simulation (MAS)*, 12th-14th September. Rome, Italy.

Buchholz, R., 2012. *Conversive Hidden non-Markovian Models*. Dissertation. Otto-von-Guericke-Universität Magdeburg.

Damaraju, S., Kerne, A., 2008. *Multitouch Gesture Learning and Recognition System*. Interface Ecology Lab at Texas A&M University.

François, J.-M., 2009. *jahmm*. Google Code. Available from: http://code.google.com/p/jahmm/ [29. September 2012]

Han, J. Y., 2005. Low-cost multi-touch sensing through frustrated total internal reflection. *Proceedings of the 18th annual ACM symposium on User interface software and technology (UIST '05)*, pp. 115-118, New York, USA.

Krull, C., Horton, G., 2009. HIDDEN NON-MARKOVIAN MODELS: Formalization and solution approaches, *Proceedings MATHMOD 09 Vienna*. Wien.

Lazarova-Molnar, S., 2005. *The Proxel-Based Method: Formalisation, Analysis and Applications*. Dissertation. Otto-von-Guericke-Universität Magdeburg.

Sun, A., Lim, E.-P., 2001. Hierarchical Text Classification and Evaluation, *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM 2001)*, pp. 521-528, California, USA.

A FLEXIBLE DECISION SUPPORT TOOL FOR MAINTENANCE FLOAT SYSTEMS - A SIMULATION APPROACH

Francisco Peito[a], Guilherme Pereira[b], Armando Leitão[c], Luís Dias[d], José A. Oliveira[e]

[a]Industrial Management Dpt., Polytechnic Institute of Bragança
[b]Research Centre ALGORITMI, University of Minho
[c] Industrial Engineering and Management Dpt., Faculty of Engineering, University of Porto
[d] Research Centre ALGORITMI, University of Minho
[e] Research Centre ALGORITMI, University of Minho

[a]pires@ipb.pt, [b]gui@dps.uminho.pt, [c]afleitao@ipb.pt, [d]lsd@dps.uminho.pt, [e] zan@dps.uminho.pt

ABSTRACT

This paper is concerned with the use of simulation as a decision support tool in maintenance systems, specifically in MFS (*Maintenance Float Systems*). For this purpose and due to its high complexity, in this paper the authors explore and present a way to develop a flexible MFS model, for any number of machines in the workstation, spare machines and maintenance crews, using *Arena* simulation language. Also in this paper, some of the most common performance measures are identified, calculated and analysed. Nevertheless this paper would concentrate on the two most important performance measures in maintenance systems: system availability and maintenance total cost. As far as these two indicators are concerned, it was then quite clear that they assumed different behaviour patterns, especially when using extreme values for periodic overhauls rates. In this respect, system availability proved to be a more sensitive parameter.

Keywords: Simulation, Discrete Event Simulation, Maintenance, Preventive Maintenance, Waiting Queue Theory, Float Systems.

1. INTRODUCTION

According to (Pegden et al., 1990), simulation can be understood as the process of construction of a real system representative model, as well as an experimental process aiming to a better understanding of their behavior and to assess the impact of alternative operations strategies. Thus, simulation may also be considered as a decision support tool that allows to predict and to analyze the performance of complex systems and processes as they are in many real systems. In addition, with the use of simulation we acquired a capacity to forecast and to achieve quickly the importance of taking some decisions about the system under analysis.

In some real systems like production areas, services such as transport companies, health service systems and factories, the main goal is to achieve high levels of competitiveness and operational availability. In this environment the need for equipment to work continuously is essential in order to maintain high levels of productivity. This is why MFS has an important role on equipment breakdown and production stoppage has a high and direct impact on production process efficiency and, as a consequence, on their operational results. Therefore, maintenance control and equipment use optimization become not only an important aspect for the mentioned reasons, but also for personnel security matters and to prevent negative environmental impact. This maintenance control and optimization of equipment utilization can be achieved implementing preventive maintenance actions that increase equipment control and avoid unexpected stoppage. However, to overestimate these actions makes the maintenance costs too high for the required availability.

The integration of the maintenance management with materials and human resources is an advantage in production systems that involve identical equipment such as float systems – involving the existence of spare equipment to replace those that fail or need review. Then, the direct and indirect costs due equipment stoppage are minimized and the level of production or service requirements fulfilled. Although the existence of spare equipment is important to maintain the production process working it is recommended to keep the number of spare equipment in an optimal level for economic reasons.

Mainly due to the non-existence of a specific simulator for the maintenance field, we had a great difficulty in choosing an appropriate simulation tool. However, (Dias et al., 2005) had a definite contribution as far as the simulation tool decision is concerned.

In fact, the choice of Arena® as a simulation language was based on the fact that its hierarchical structure offers different levels of flexibility, thus allowing the construction of extremely complex models, allied to a strong visual component (Kelton et al., 2004; Dias et al., 2011; Dias et al., 2006; Pidd, 1993 and Pidd, 1989).

Having referred the importance of studying MFS, the research background section of this paper will focus on the literature review on analytical models, but also on some type of simulation metamodels for this type of maintenance systems.

Next, the description of the MFS section, describes the MFS model used, which formed the basis for the development of our simulation model with the purpose of analysing system availability and total maintenance cost, as global efficiency measures.

The following section describes new developments on a previous simulation model towards flexibility. In fact, the model presented in (Peito et al., 2011) will gain the capacity to automatically generate a specific simulation program for each specific MFS desired. The program will then be adapted for specific situations with no need of further coding effort. In fact the new proposed tool is intended exclusively to give a response to a type-standard configuration of MFS. Nevertheless, within this type-standard configuration, the user could easily evaluate different strategies under different number of resources available (active machines, maintenance crews and spare machines). This way, the resulting MFS model aims to fill a gap in terms of computer solutions currently existing for this specific type of maintenance systems.

Then we present some results of both global efficiency measures under consideration, in order to evaluate its sensitivity, its precision and its robustness.

Conclusions and Future Developments are the closing sections for this paper.

2. RESEARCH BACKGROUND

A literature survey on the field of maintenance systems, regarding the use of discrete event simulation, shows a significant number of scientific publications. Recently, (Alabdulkarim et al., 2013) present a complete set of research works where maintenance costs, maintenance reliability, maintenance operations performance, are some of the most important issues discussed. (Chen and Tseng, 2003), however, are the only authors which main focus is MFS.

In this respect (MFS), (Lopes 2007) refers some studies where simulation has been used to produce results based on specified parameters. Due to the fact that these simulation models were only concerned with the input/output process, without dealing with what is happening during the simulation data process, some metamodels have emerged (Madu and Kuei, 1992b; Madu and Lyeu, 1994; Kuei and Madu, 1994; Madu, 1999; Alam et al., 2003). The metamodels express the input/output relationship through a regression equation. These metamodels can also be based on taguchi methods (Kuei and Madu, 1994) or on neuro networks (Chen and Tseng, 2003). These maintenance system models were also recently treated on an analytical basis by (Gupta and Rao, 1996; Gupta, 1997; Zeng and Zhang, 1997; Shankar and Sahani, 2003; Lopes, 2007). However, the model proposed by (Lopes, 2007) is the only one that deals, simultaneously, with three variables: number of maintenance teams, number of spare equipment, and time between overhauls, aiming the optimization of the system performance. Although this proposed model already involves a certain amount of complexity it may become even more complex by adding new variables and factors such as: a) time spent on spare equipment transportation, b) time spent on spare equipment installation; c) the introduction of more or different ways of estimating efficient measures; d) allowing the system to work discontinuously; e) speed or efficiency of the repair and revision actions; f) taking into account restrictions on workers timetable to perform the repair and revision actions; g) taking into account the workers scheduling to perform the repair and revision actions; h) taking into account the possibility of spare equipment failure; etc. Anyway these mentioned approaches would aim at ending up with MFS models very close to real system configurations. In fact, the literature review showed that most of the works published, involving either analytical or simulation models, concentrate on a single maintenance crew, or on a single machine on the workstation or even considering an unlimited maintenance capacity – thus overcoming the real system complexity and therefore not quite responding to the real problem as it exists.

As far as the model presented by (Lopes et al., 2005; Lopes et al., 2006; Lopes, 2007) is concerned it is assumed that systems works continuously, its availability is not calculated and the system optimization is only based on the total maintenance cost per time unit. Moreover, it considers that the total system maintenance cost is the same without taking into account the number of machines unavailable, which in many real situations it is not the best option. Finally the referred analytical model only allows that its failures occur under an homogeneous Poisson process (HPP).

Another important aspect on the companies' management strategic definition is to have their tasks correctly planned. To help this planning procedure it is important to know different indicators such as: machine availability, equipment performance and maintenance costs, among others. Therefore one should consider new factors that affect these float systems indicators, such as the possibility of some machine failure, efficiency, repair time, etc.

Moreover, when preventive maintenance policy is used, the time for individual replacement is smaller than time for group replacement. It means that the latter situation requires more machine on the process to be stopped, and also implies an increase for a certain time, on the maintenance crews.

In general companies' policy lies on using economic models to define their best strategies. Profits maximization or costs minimization are the most frequent goals used. However, strictly from the maintenance point of view availability is frequently used as an efficient measure of the system performance, and sometimes more important than the cost based process. In this work availability is calculated dividing the time the system is up (T_{up}) by the time the system is up plus the time the system is down (T_{down}) for maintenance reasons. Some authors, however, calculate availability through the ratio between MTBF and [MTBF+MTTR]. Being, MTBF the *Mean Time Between Failures* and MTTR the *Mean Time To Repair*.

3. DESCRIPTION OF THE MFS

Our model represents a typical *Maintenance Float System* (MFS) and it is composed of a workstation, a maintenance center with a set of maintenance crews to perform overhauls and repair actions and a set of spare machines (Figure 1). The workstation consists of a set of identical machines and the repair center of a limited number of maintenance crews and a limited number of spare machines. However, the model we have adopted, being a typical MFS, presents certain specificities both as far as the philosophy of the maintenance waiting queues are concerned, and related to the management of the maintenance crews.

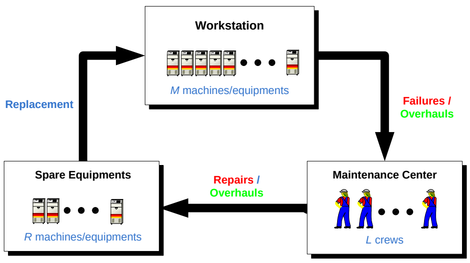


Figure 1: Typical Maintenance Float System

This model follows the one proposed and developed by (Lopes, 2005; Lopes et al., 2006; Lopes et al., 2007), considering *M* active machines, *R* independent and identical spare machines and *L* maintenance crews The active machines considered operate continuously. Machines that fail are taken from the workstation and sent to the maintenance park waiting queue, where they will be assisted according to arrival time. Machines that reach their optimal overhaul time are kept in service until the end of a period T without failures. However they will be also kept on a virtual queue to overhaul. If the number of failed machines plus the number of machines requiring overhaul is lower than the number of maintenance crews available, machines are replaced and repaired according to FIFO (*First In First Out*) rule. Otherwise if it exceeds the number of maintenance crews, the machines will either be replaced (while there are spare machines available) or will be sent to the maintenance queue. The machines that complete a duration period T or time between overhauls in operation without failures are maintained active in the workstation, where they wait to be assisted, and they are replaced when they are removed from the workstation, to be submitted to a preventive action. Its replacement is assured by the machine that leaves the maintenance center in the immediately previous instant. If an active machine happens to fail it waits for the accomplishment of an overhaul, then it will be immediately replaced, if a spare machine is available or as soon it is available.

In this version of our model it is assumed that the M active machines of the workstation have a constant failure rate while the model runs.

Time between failures are assumed as independent and identically distributed following an Exponential Distribution for all machines (failures occur under a *Homogeneous Poisson Process*). However, during a simulation run, this value could be adjusted based on time between overhauls. Obviously a smaller time between overhauls implies greater time between failures.

As far as time to overhaul and time to repair are concerned, we have assumed the *Erlang-2* distribution, even though considering overhaul time significantly lower than the repair time.

Now, for our MFS, the variables used are the following:

1. Number of active machines (M);
2. Number of maintenance crews (L);
3. Number of spare machines (R);
4. Machine- Overhauls rate (λ_{rev});
5. Machine-Initial Failures rate (λ_f);
6. Crews-Repair rate (μ_{rep});
7. Crews-Overhaul rate (μ_{rev});
8. Failure cost (C_f);
9. Repair cost (C_{rep});
10. Overhaul cost (C_{rev});
11. Replacement cost (C_s);
12. Cost due to loss production (C_{lp});
13. Holding cost per time unit (h);
14. Labour cost per time unit (k);
15. Time to convey and install spare machine ($T_{ConvInst}$).

The developed simulation model for our MFS allows us to estimate the following global efficiency measures:

a) Average system availability *(AvgSAv);*
b) Total maintenance cost per time unit *(AvgTCu);*

However, some other performance measures are also estimated, such as:

c) Average number of missing machines at the workstation *(AvgM$_{eq}$),*
d) Average number of machines in the maintenance waiting queue *(AvgLq)*;
e) Average waiting time in the maintenance waiting queue (*Avg Wt);*
f) Average operating cycle time (*AvgD***)**;
g) Probability of existing 1 or more idle Machines *(Prob$_{im}$);*
h) Probability of the system being fully active *(Prob$_s$)*;

and still, some individual efficiency measures per machine or maintenance crew, i.e.,

i) Utilization rate per machine;
j) Utilization rate per maintenance crew;
k) Number of overhauls and repair actions performed per maintenance crew;
l) Average availability per machine.

4. INCREASING FLEXIBILITY OF THE SIMULATION MODEL

The Arena® simulation language environment, used in the previous development (see details on Peito et al., 2011), has been now revisited, aiming to give flexibility to the previous model. The user, now, would be able to automatically generate a simulation program according to specific characteristics of the MFS, namely varying the number of active machines (*M*), the number of maintenance crews (*L*) and the number of spare machines (*R*). However, the steps towards the development of the previous simulation model were all kept and are presented in Figure 2, for a better understanding of the simulation model developed.

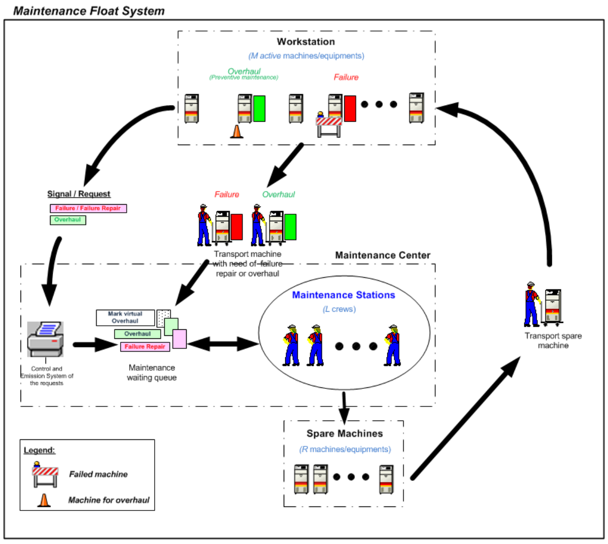


Figure 2: Steps for simulation model development

Figures 3 and 4 explicit the global logical simulation model before and after gaining flexibility, underlining its different developed components:

1. Active machines *(workstation)*;
2. Statistics 1 *(Recording Machines T_{up})*;
3. Maintenance queue;
4. Machines' transportation *(by the maintenance crews)*;
5. Spare machine request;
6. Maintenance center *(set of maintenance Stations)*;
7. Release machines to the set of spare machines;
8. Statistics 2 *(Recording Machines T_{up} and T_{down})*;
9. Spare machines *(in the start of the system).*

This logical model configuration choice was kept identical for the MFS (figures 3 and 4), providing again a clear global visualization of the undergoing operations and a great simplicity to make changes in the model. In fact the logical model, after increasing flexibility, will appear even more simplified – see Figure 4. The implementation of Arena resource sets, the inclusion of indexed variables and data arrays and also a set of control variables, replacing previous Arena internal variables, have definitely contributed to a simplified model.

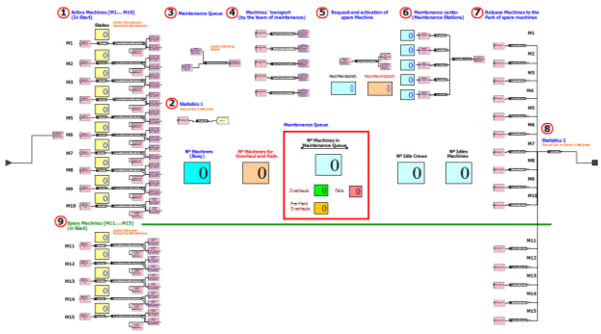


Figure 3: Arena® Logic Model before increasing flexibility

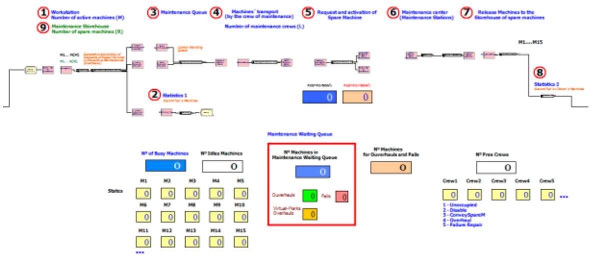


Figure 4: Arena® Logic Model after increasing flexibility

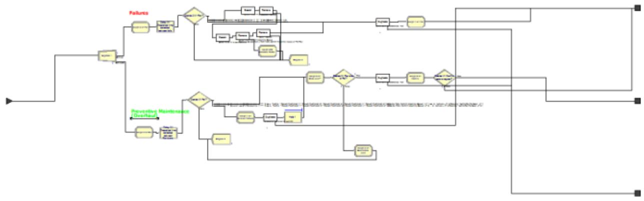


Figure 5: Generation and control system for repair and overhaul requests before increasing flexibility

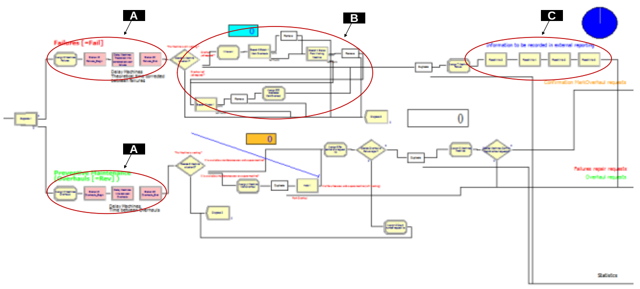


Figure 6: Generation and control system for repair and overhaul requests after increasing flexibility

The components 1 and 9 after increasing flexibility (Fig. 6) include now a generation and control system for all repair and overhaul requests of all machines, this was not the case in the previous model (Figure 5). For this control system to be effective, it would also be necessary to guarantee absolute independence of each type of request for every machine. For this purpose, a mechanism for attribute identification was developed. With this mechanism, it is now possible to identify the state of each machine and the occurrence of every type of machine request (failure or overhaul), at any instant – entity number and color (see Figure 6, zone A).

In Figure 6 (Zone B), a small change has occurred. In fact, some Arena Blocks have been replaced by Arena Modules. This way, planned changes to some parameters are now easy to implement once Arena shows data in a simple table format.

Finally, Figure 6, Zone C shows four *ReadWrite* Arena modules, allowing the registration, in an excel worksheet, of the failure instants and the number of failures for each machine.

The maintenance waiting queue is defined through a synchronization of events between the component 3 and 4. In the component 4 (figures 7 and 8) there is a "control mechanism", which only allows a request to proceed if there is a free maintenance crew. Component 4 will now include the use of an Arena *Resource Set* for the maintenance crews, selecting the available maintenance crew that has the least number of services allocated.

The rules for the maintenance queue management were all kept unchanged. In fact, FIFO (*First In First Out*) is the rule for the maintenance queue management, except for the case when the total number of maintenance requests (overhauls plus repair actions) exceed the number of maintenance crews available – in this case, machines requiring repair action have priority over machines requiring overhauls.

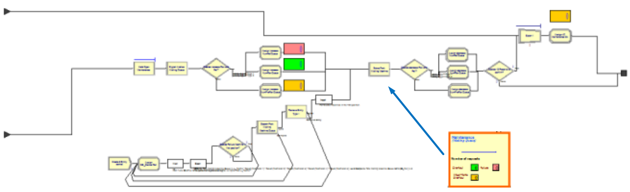


Figure 7: Maintenance waiting queue before increasing flexibility

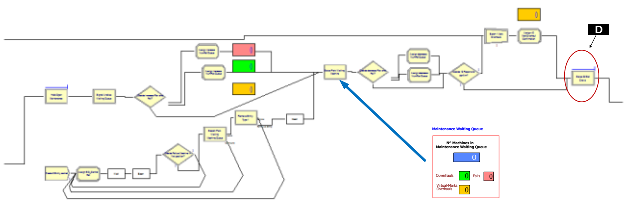


Figure 8: Maintenance waiting queue after increasing flexibility

Component 4 (figures 7 and 8) has also been changed and now includes an *Assign Module* in Zone D. Besides the identification of the maintenance crew and the machine transport state (for a spare machine or a failed machine or even a machine needing overhaul), this Module also updates the number of maintenance crews that are free.

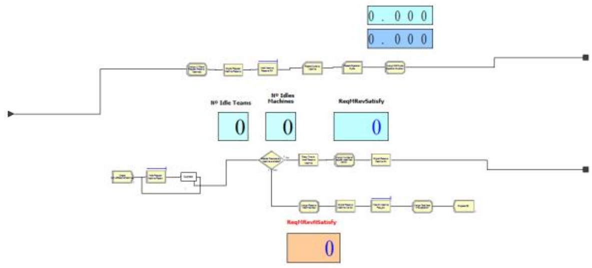


Figure 9: Request and activation of spare machines before increasing flexibility

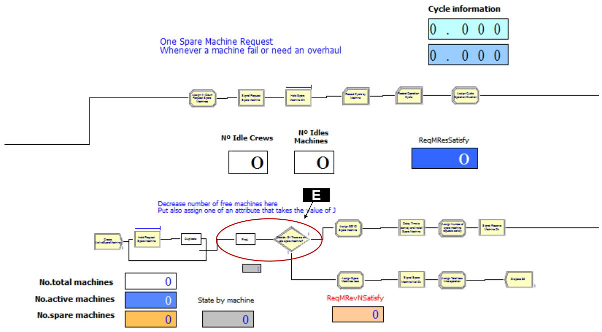


Figure 10: Request and activation of spare machines after increasing flexibility

In component 5 (figures 9 and 10) that performs the request of a spare machine, performed by a maintenance crew, there is only a small change in Zone E, that is related with the demand with one free available machine. Now the model includes a *Search Block* that searches for a free machine.

In component 6 (Figure 11), the change is in the structure of the component. In fact, the discrete variables are now indexed discrete variables – this way, it is possible to individually save a set of performance indicators for both types of maintenance operations.

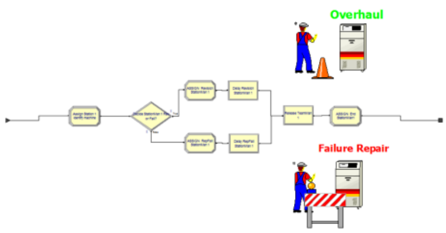


Figure 11: Identification and statistics of the states of the maintenance crew

In component 7, responsible for releasing machines under maintenance crew actions whenever they finish their work, either repairing or performing overhauls, all *Release Modules* have been replaced by a single *Release Block* – this was possible once now only a single indexed discrete variable is capable of saving all the information related to each machine.

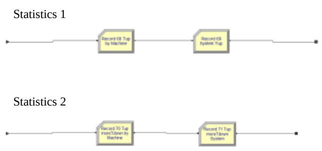


Figure 12: Record statistics

Components 2 and 8 (Figure 12) which are responsible to record fundamental statistical data to calculate adequate efficiency measures, do not suffered any change.

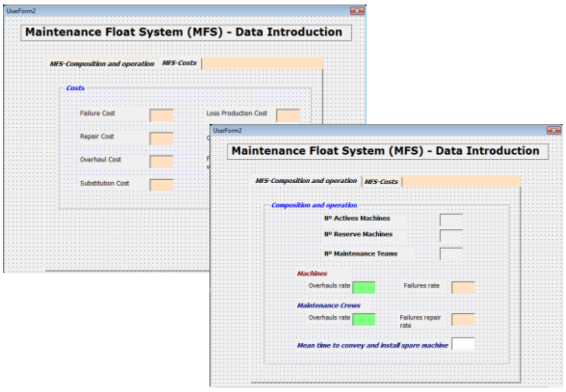


Figure 13: Screenshot of the data input area before increasing flexibility

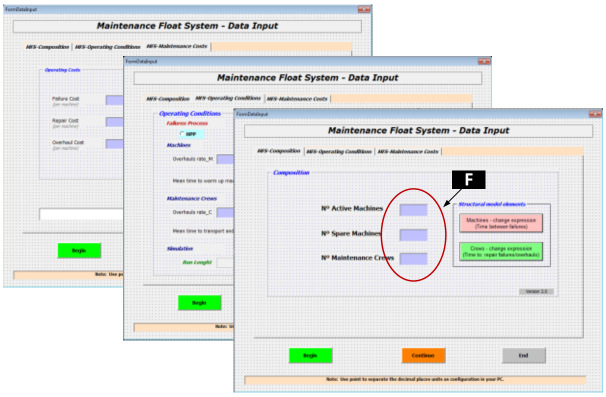


Figure 14: Screenshot of the data input area after increasing flexibility

This work, making previous simulation model gaining flexibility, allows the user to get a simulation model for any *Maintenance Float System* desired – regardless the number of active machines, the number of maintenance crews and the numbers of spare machines. After inputting these three values (Zone F, figures 14), the user will instantly get the appropriate simulation model automatically generated.

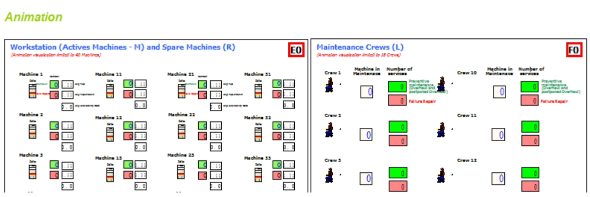


Figure 15: Screenshot of the detailed animation area of the *Workstation* (limited to M = 40, L = 18)

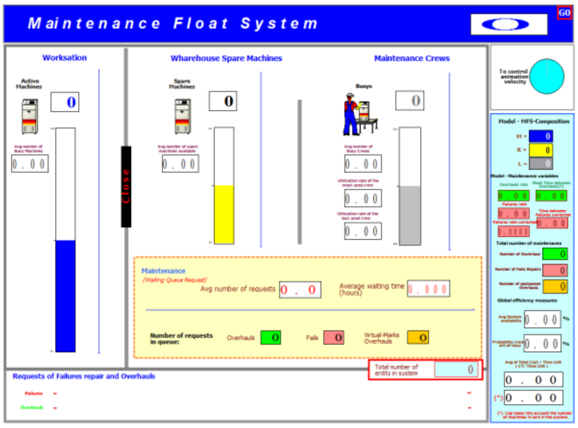


Figure 16: Screenshot of the global animation area of the *Workstation*

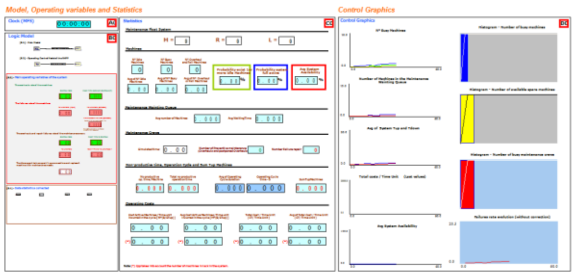


Figure 17: Screenshot of the Operating variables, Statistics and Graphics control area

The presentation of model animation (Figures 15 and 16) and output statistics (Figure 17) had changes relatively to the version presented in (Peito et al., 2011).

5. SIMULATION RESULTS AND DISCUSSION

This paper focus on the two general performance measures mentioned above – system availability and total maintenance cost per time unit, which were determined considering a Maintenance Float System with 10 active and identical machines (M), 5 spare machines (R) and 5 maintenance crews (L).

Simulation length was set to 9.000 hours (approximately one year) – warm-up period was set to 3.500 hours.

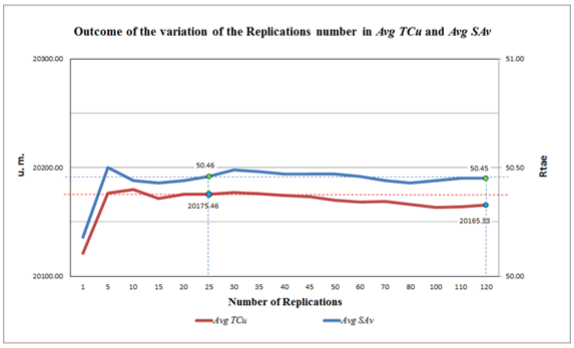


Figure 18: Outcome of the variation of the Replications number in *AvgTCu* and *AvgSAv* variables

For each set of input parameters and pattern for variables, the simulation output variables *AvgSAv*, *AvgTCu* and *AvgTCu(*)* were estimated based on 25 replications – for an adequate system stabilization and results robustness for both performance measures (Figures 18 and 19) and also due to computational time required to run the model.

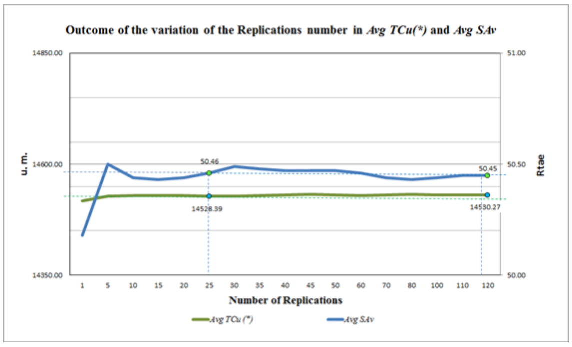


Figure 19: Outcome of the variation of the Replications number in *AvgTCu(*)* and *AvgSAv* variables

Bearing in mind the fifteen variables of MFS previously referred, simulation models were used to test and estimate the behavior of the two global efficiency measures mentioned on the previous section. Simulation models were carried out for (1-60) hypothetical scenarios with different overhauls rates (*λrev*). These different overhauls rates are associated with different times between overhauls (*T*) which are defined accordingly to the preventive maintenance policy aiming the best option.

Table 1: Global efficiency measures outcomes in the MFS model after 25 replications

(Values estimated by simulation after 25 replication)

Scenario	λ_{rev} (/hour)	*T* (hour)	*AvgSAv* (%)	*AvgTCu* (m.u./hour)	*AvgTCu(*)* (m.u./hour)
1	0,10	10,000	29,25	21064,60	17304,37
2	0,20	5,000	34,75	21127,96	16646,80
3	0,30	3,333	40,79	21065,24	15870,35
4	0,40	2,500	45,26	20915,39	15287,79
5	0,50	2,000	48,12	20751,46	14908,68
⋮	⋮	⋮	⋮	⋮	⋮
52	9,00	0,111	48,71	20195,00	14776,87
53	15,00	0,067	48,52	20212,24	14807,21
54	20,00	0,050	48,54	20228,63	14815,57
55	25,00	0,040	48,56	20223,63	14814,18
56	30,00	0,033	48,54	20230,24	14816,98
57	35,00	0,029	48,62	20226,82	14811,16
58	40,00	0,025	48,54	20229,16	14820,31
59	45,00	0,022	48,50	20228,80	14822,78
60	50,00	0,020	48,52	20226,42	14815,55

(*) *Considers that the cost of lost production changes in function of the number of active machines lacking in the system.*

A first global analysis of the values presented in tables 1 and 2 indicate that the precision obtained on the three efficient measures analysed is different. An individual analysis of each measure indicates that *AvgTCtu* shows the smaller variation (MPO lower). In Table 1 it can also be observed that when *T* takes very small values ($T \leq 0.111$ or $\lambda_{rev} \geq 9$) the three efficient measures *[AvgTCtu, AvgTCTu(*) and AvgSAv]* are kept practically unchangeable. This fact can be confirmed in Figure 20 or in Table 2 where the MPO for these values of *T* is extremely low, almost zero. On the other hand, when *T* assumes very high values ($T \geq 2,5$ or $\lambda_{rev} \leq 0,4$) the efficiency measures *AvgTCTu(*) and AvgSAv* present high MPO values in opposition to *AvgTCtu* that shows very small values. In Table 2 it can also be observed that *AvgTCtu* presents the lowest MPO average value of the three efficiency measures and that *AvgSAv* has the highest value.

Table 2: Observe percentage change in the global efficiency measures after 25 replications

MPO - Percentage change observed			
Scenario	*AvgSAv*	*AvgTCu*	*AvgTCu(*)*
1-2	15,83%	0,30%	-3,95%
2-3	14,80%	-0,30%	-4,89%
3-4	9,87%	-0,72%	-3,81%
4-5	5,95%	-0,79%	-2,54%
⋮	⋮	⋮	⋮
52-53	-0,14%	0,06%	0,10%
⋮	⋮	⋮	⋮
58-59	-0,07%	0,00%	0,02%
59-60	0,04%	-0,01%	-0,05%
Max.	**15,83%**	**-0,79%**	**-4,89%**
Mean	**0,80%**	**-0,07%**	**-0,27%**

In order to simplify the interpretation and analysis of these global efficiency measures, figures 20, 21 and 22 pinpoint the maximum and minimum values (table 2 and 3) as well as other points considered relevant for the analysis.

Table 3: Maximum values of the main efficiency measures

	Statistics (Maximum)	λ_{rev} (/hour)	*T* (hour)
AvgSAv (%)	50,70%	0,90	1,111
AvgTCu (m.u./hour)	21127,96	0,20	5,000
AvgTCu (*) (m.u./hour)	17304,37	0,10	10,000

Note: *Red points in the graphics*

Table 4: Minimum values of the main efficiency measures

	Statistics (Minimum)	λ_{rev} (/hour)	*T* (hour)
AvgSAv (%)	29,25%	0,10	10,000
AvgTCu (m.u./hour)	20096,90	1,80	0,556
AvgTCu (*) (m.u./hour)	14518,77	1,20	0,833

Note: *Yellow points in the graphics*

Tables 3 and 4 show that the T value corresponding to the minimum value of *AvgSAv* corresponds the maximum value, as expected, of *AvgTCu(*)*. When compared with the minimum of *AvgTCu*, there is a significant T gap (≈ 5 hours), although, its remains practically the same when the value of T changes from 5 to 10 hours (Figure 20).

When comparing the T value corresponding to the maximum value of *AvgSAv* with the T value corresponding to the minimum value of *AvgTCu(*)*, there is only a small gap, which is clearly higher in the case of the T value corresponding to the minimum of the *AvgTCu* (Figure 22).

Table 5: Correlation coefficients

	T	***AvgSav***	***AvgTcu***	***AvgTcu(*)***
T	1	-0,9021	0,8279	0,9017
AvgSav	-0,9021	1	-0,7980	-0,9986
AvgTcu	0,8279	-0,7980	1	0,8237
AvgTcu (*)	0,9017	-0,9986	0,8237	1

A careful analysis of the correlation coefficients of three efficiency measures, table 5, shows that T variations are better explained by *AvgSav* and *AvgTCu(*)* (≈90%). It is also verified that there is a high inverse correlation between *AvgSav* and *AvgTCu(*)* (≈99,8%). However when *AvgTCu* is compared with *AvgTCu(*)* or with *AvgSav*, the correlation coefficient decreases to 82,37% and to 79,80%, respectively. This partially explains why in tables 3 and 4 the T value corresponding to the maximum of *AvgSav* does not correspond exactly to the T value corresponding to the minimum of the *AvgTCu* and *AvgTCu(*)* and that difference being higher in the case of *AvgTCu(*)* (Figure 22).

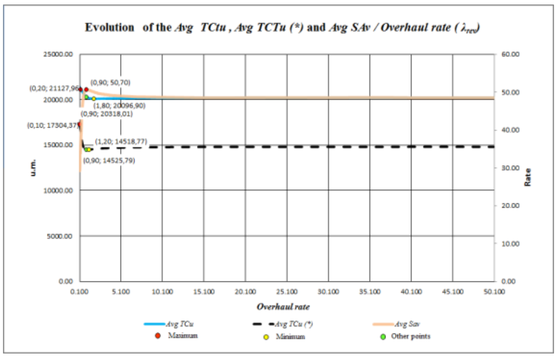


Figure 20: Evolution of the *AvgTCtu*, *AvgTCTu(*)* and *AvgSAv / Overhaul rate* (λ_{rev})

As it can be observed in Figure 20 and more clearly in figures 21 and 22, for the MFS analyzed, the three global measures of efficiency being studied only present small variations for values of λ_{rev} between 0,10 and 9,00 (or T between 10,000 and 0,111 hours). For values of λ_{rev} higher than 9.00 the three global measures of efficiency remain practically unchanged.

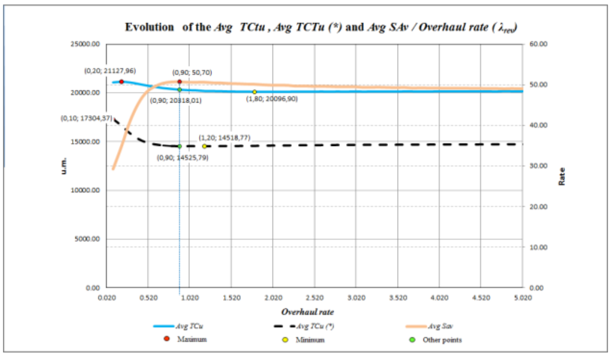


Figure 21: Evolution of the *AvgTCtu*, *AvgTCTu(*)* and *AvgSAv / Overhaul rate* (λ_{rev}) [Zoom Figure 20]

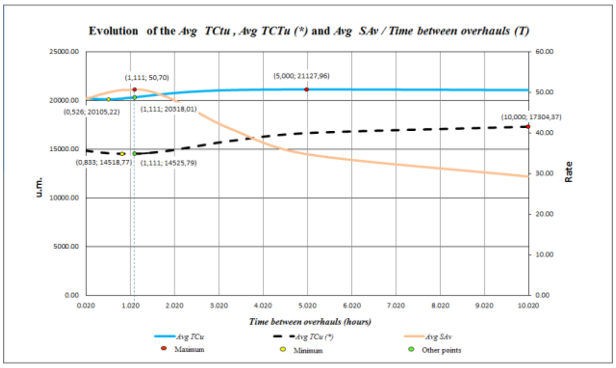


Figure 22: Evolution of the *AvgTCtu*, *AvgTCTu(*)* and *AvgSAv / Time between overhauls (T)*

Table 6: Comparison among the *AvgTCu* values estimates by the simulation model and analytic model (Lopes, 2007)

T	**Model**	***AvgTCu*** (m.u./hour)	***AvgTCu(*)*** (m.u./hour)	Δ_1 (%)	Δ_2 (%)	Δ_3 (%)
1,66	Simulation	20381,74	14554,5	-40%	-12%	-25%
	Analytic	18130,56	----	----		
3,33	Simulation	20126,65	14555,62	-38%	-19%	-17%
	Analytic	16968,39	----	----		
1,66	Simulation	20111,24	14628,02	-37%	-16%	-18%
	Analytic	17303,65	----	----		
3,33	Simulation	21065,24	15870,35	-33%	-16%	-14%
	Analytic	18167,34	----	----		
			Mean	**-37%**	**-16%**	**-18%**

Note: *M=10; R=5; L=5.*
(*) Considers that the cost of lost production changes in function of the number of active machines lacking in the system.
Δ_1 – Difference among Simulation AvgTCu and Simulation AvgTCu(*)
Δ_2 – Difference among Simulation AvgTCu and Analytic AvgTCu
Δ_3 – Difference among Analytic AvgTCu and Simulation AvgTCu(*)

In Table 6 there is a comparison between the values obtained from the simulation model developed by the authors in a former (Peito et al 2011) and the analytical model developed by (Lopes 2007). The sample size of the results presented and compared in this case was limited by the number of results presented by the author in her work (Lopes, 2007). In this table it can be verified that when the two global efficiency measures are both estimated from the simulation model the difference (Δ_1) is on average -37%, presenting

AvgTCu always higher values. However when *AvgTCu* is estimated through the analytical model that difference (Δ_3) is on average -18%. When the same efficiency global measure based on the analytical model is compared with the one calculated based on the simulation model, *AvgTCu*, this if calculated from the analytical model presents lower values, on average, of 16%. It is also observed that the analytical model always presents for its efficiency measure values that lie between the two efficiency measures estimated from the simulation model.

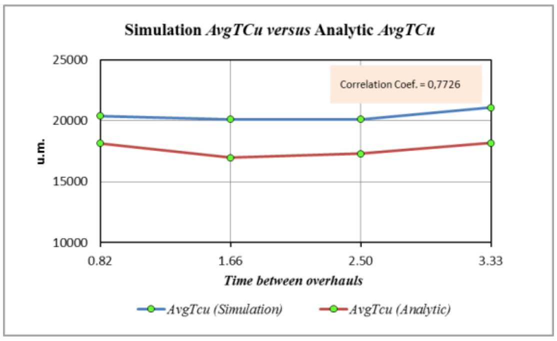


Figure 23: Comparison among the *AvgTCu* values estimates by the simulation model and by the analytic model (Lopes, 2007)

Finally, through Figure 23 it can be verified that the behavior of *AvgTCu* is identical in both models. However this results analysis lacks confirmation due to small sample size dimension.

6. CONCLUSIONS

Firstly, this paper shows how to develop an advanced simulation model, incorporating flexibility. This target would be reached by developing and incorporating new modules in our simulation tool, following past experiences found on literature (Dias et al. 2005, 2006 and Vilk et al. 2009, 2010) where the automatic generation of simulation programs enables desired model flexibility, i.e., making the model generating specific simulation programs for specific *Maintenance Float Systems*. This new development of our simulation model for our *Maintenance Float System* presents:

- More flexibility

  This was the main challenge for the work presented in this paper. The automatic generation of simulation models, depending on the three main maintenance system variables – *M*, number of active machines; *L*, number of maintenance crews; *R*, number of reserve machines. In fact, the user would just have to introduce *M*, *L* and *R* and, instantly, he will get the adequate simulation model to run and experiment.

- More interactivity

  Now the user has the possibility to interact with the simulation model during each simulation run. In fact the user can now modify all variables of the maintenance system under analysis and can, therefore, evaluate system behaviour under different maintenance strategies.

- Better information

  This model now offers much better maintenance information. Indeed, the strong visual aspect offered by the developed model clarifies the actual process inside the system. This allows a better understanding of the different interactions in the model and of the simulation results.

This paper also shows that the estimated values for the performance measures analysed (system availability and total maintenance cost per time unit) present similar values for the simulation model and analytical model, as far as a *Maintenance Float System* with *M*=10, *R*=5 and *L*=5 is concerned. Also, it is quite clear that variance is different for both global efficiency measures analysed, especially when using extreme values for periodic overhauls rates. In this respect, *AvgSav* is the most sensitive parameter. As expected, the least sensitive parameter is *AvgTCu*, as it does not take into consideration the number of available machines, i.e., the cost for production loss is constant, irrespective of the number of available machines in the system.

However, the greatest overall contribution of this paper is therefore related to the construction of a flexible simulation decision support tool for MFS, where several efficiency measures of MFS are involved. Thus, in any classic MFS (number of active machines, number of spare machines and maintenance crews) subject to preventative actions and accidental actions of maintenance, this tool deals with the evaluation of its efficiency in terms of costs and in terms of availability. Also, considering simultaneously preventive maintenance actions and accidental maintenance actions, represents another novelty, once the simulation models found on the literature would approach these issues individually. Moreover, this simulation tool enables to tackle large scale float systems, up to ±1000 actives machines, ±1000 spare machines and ±1000 maintenance crews. On the other hand, the diversity of efficiency measures calculated in the MFS simulation model really helps the decision maker to take the appropriate decisions. Finally this model presents the advantages usually associated with simulation models, namely a better understanding of the functioning of the system, the possibility of identifying the critical points of the system and the easy adaptation of the simulation model to reflect changes in the operating conditions of the system.

7. FUTURE DEVELOPMENTS

The simulation model here presented, incorporating analysis of usual performance measures, also drives its concern towards new efficiency measures, enabling new trends for the analysis and discussion of the best decisions as far as a specific *Maintenance Float System* is concerned. Nevertheless the authors are now aiming to the development of an advanced simulation model, incorporating still more flexibility. This target would be reached by developing and incorporating new modules in our simulation tool, in order to also incorporate maintenance systems where failure rates would also vary while the model runs, i.e., where a *Non Homogeneous Poisson Process* (NHPP) is present. These mentioned future developments also intend to potentiate the known capability of simulation to efficiently communicate with managers and decision makers, even if they are not simulation experts.

Acknowledgments. This work was funded by the "Programa Operacional Fatores de Competitividade - COMPETE" and by the FCT - Fundação para a Ciência e Tecnologia in the scope of the project: FCOMP-01-0124-FEDER-022674.

REFERENCES

Alabdulkarim A. A., Ball, P. D., Tiwari, A., 2013, *Applications of simulation in maintenance research,* World Journal of Modeling and Simulation, Vol. 9 (2013) No. 1, pp. 14-37.

Alam, Fasihul M.; McNaught, Ken R. and Ringrose, Trevor J., 2003, *Developing Simulation Metamodels for a Maintenance Float System*, AMORG, Engineering Systems Department, Cranfield University, RMCS Shrivenham, Swindon SN6 8LA, UK

Chen, M. C. and Tseng, H. Y., 2003, *An approach to design of maintenance float systems*, Integrated Manufacturing Systems, vol. 14, pp. 458-467.

Chen, M. and Tseng, H., 2003, *An approach to design of maintenance float systems*, Integrated Manufacturing Systems, vol. 14, no. 5, pp. 458-467.

Dias, Luis M. S; Rodrigues, A. J. M. Guimarães; Pereira, Guilherme A. B., 2005, *An Activity Oriented Visual Modelling Language with Automatic Translation to Different Paradigms*, Proceedings of the 19th European Conference On Modelling And Simulation (ECMS 2005), Riga, Letónia. Ed. Yury Mercuryev et al. Junho de 2005. pp. 452-461.

Dias, Luis S; Pereira, Guilherme A. B. and Rodrigues, A. J. M. Guimarães, 2006, *A Shortlist of the Most Popular Discrete Simulation Tools*, ASIM 2006 - 19th Symposium on Simulation Technique. SCS Publishing House. Ed. M. Becker and H. Szczerbicka. Hanover, Alemanha. pp. 159-163.

Dias, Luis S.; Pereira, Guilherme A. B. and Rodrigues, A. J. M. Guimarães, 2006, *Activity based modeling with automatic prototype generation of process based arena models*, EMSS 2006 - 2nd European Modeling and Simulation Symposium. Barcelona, Espanha. pp. 287-296.

Dias, L S, Pereira, G A B, Vik, P and Oliveira, J A V. *Discrete Simulation Tools Ranking – a Commercial Software Packages Comparison Based on Popularity*. Industrial Simulation Conference, Venice, Italy, June 2011: 5-10.

Gupta, V. and Rao, T., 1996, *On the M/G/1 machine interference model with spares*, European Journal of Operational Research, vol. 89, pp. 164-171.

Gupta, S. M., 1997, *Machine interference problem with warm spares, server vacations and exhaustive service*, Performance Evaluation, vol. 29, pp. 195-211.

Kelton, W. David; Sadowski, Randall P. and Strurrok, David T., 2004, *Simulation With Arena*, (3rd edition), McGraw-Hill, 1998-2004.

Kuei, C. H. and Madu, C. N., 1994, *Polynomial metamodelling and Taguchi designs in simulation with application to the maintenance float system*, European Journal of Operational, Reseach, vol. 72, pp. 364-375.

Lopes, Isabel S., 2007, *Técnicas Quantitativas no Apoio à Decisão em Sistemas de Manutenção*, Tese de Doutoramento, Universidade do Minho.

Lopes, Isabel S.; Leitão, Armando L. F. and Pereira, Guilherme A. B., 2005, *Modelo de Custos de Manutenção para um Sistema com M Unidades Idênticas*, in C. G. Soares, A. P. Teixeira e P. Antão (eds), Análise e Gestão de Riscos, Segurança e Fiabilidade 2: p 603-620, Lisboa: Edições Salamandra (ISBN 972-689-230-9).

Lopes, Isabel S.; Leitão, Armando L. F. and Pereira, Guilherme A. B., 2006, *A Maintenance Float System with Periodic Overhauls*, in Guedes Soares & Zio (eds), Safety and Reliability for Managing Risk 1: p 613-618, London: Taylor & Francis Group (ISBN 0-415-41620-5).

Madu, C. N. and Kuei, C. H., 1992b, *Simulation metamodels of system availability and optimum spare and repair units*, IIE Transactions, vol. 24, pp. 99-104.

Madu, C. N. and Lyeu, P., 1994, *On the use of simulation metamodeling in solving system availability problems*, Microelectronics Reliability, vol. 34, pp. 1147-1160.

Madu, I. E., 1999, *Robust regression metamodel for a maintenance float policy*, International Journal of Quality & Reliability Management, vol. 16, pp. 433-456.

Pegden, C.D.; Shannon, R. E.and Sadowski, R. P., 1990 *Introduction to Simulation Using SIMAN*, McGraw-Hill, New York, USA. v. 2.

Peito, Francisco; Pereira, Guilherme; Leitão, Armando; Dias, Luís, 2011, *Simulation as a Decision Support Tool in Maintenance Float Systems*, 17th European Concurrent Engineering Conference

(ECEC 2011). London, England. pp.68-74.

Pidd, Michael, 1989, *Computer Modelling for Discrete Simulation*, Wiley.

Pidd, Michael, 1993, *Computer Simulation in Management Science*, Third Edition, Wiley.

Zeng, A. Z. and Zhang, T., 1997, *A queuing model for designing an optimal three-dimensional maintenance float system*, Computers & Operations Research, vol. 24, pp. 85-95.

AUTHORS BIOGRAPHY

FRANCISCO PEITO was born in 1966 in Macedo de Cavaleiros, Portugal. He graduated in Mechanical Engineering-Management of Production in Polytechnic Institute of Porto. He holds an MSc degree Industrial Maintenance in University of Porto. His main research interests are Simulation and Maintenance.

GUILHERME PEREIRA was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management at 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.

ARMANDO LEITÃO was born in 1958 in Porto, Portugal. He graduated in Mechanical Engineering at University of Porto, Portugal. He holds an MSc degree in Production Engineering and a PhD degree in Production Engineering from the University of Birmingham, UK. His main research interests are Reliability and Quality Maintenance.

LUÍS DIAS was born in 1970 in Vila Nova de Foz Côa, Portugal. He graduated in Computer Science and Systems Engineering at 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, Operational Research and Systems Visual Modeling.

JOSÉ A. OLIVEIRA was born in 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 the University of Minho, Portugal. His main research interests are Optimization with Heuristic Methods in Systems Engineering.

FROM SAFETY REQUIREMENTS TO SIMULATION-DRIVEN DESIGN OF SAFE SYSTEMS

Alfredo Garro[(a)], Andrea Tundis[(a)], Lena Buffoni-Rogovchenko[(b)], Peter Fritzson[(b)]

[(a)]Department of Computer Engineering, Modeling, Electronics, and Systems Science (DIMES), University of Calabria, via P. Bucci 41C, 87036, Rende (CS), Italy
[(b)]Department of Computer and Information Science (IDA), Linköping University, SE-581 83 Linköping, Sweden

[(a)]{garro, atundis}@dimes.unical.it,[(b)]{olena.rogovchenko, peter.fritzson}@liu.se

ABSTRACT

System safety is an important aspect of System Dependability which should be taken in consideration during the whole system lifecycle. However, often systems are built by considering mainly their functional aspects and safety requirements are verified and validated in the latest stages of the development process. For this reason and due to the deep integration of modern systems in the daily life of people, regulatory standards have been defined and have to be applied during the development of critical systems to guarantee a minimum and acceptable level of safety. In this context, the paper proposes a model-driven process, inspired by ISO-26262, which provides a methodological support for the verification and validation of safety requirements. In particular, the proposed framework combines model-driven engineering tools and techniques with OpenModelica, an equation based simulation environment based on the Modelica language. The proposal is experimented through a case study concerning the safety analysis of an Airbag System.

Keywords: Model-Based Systems Engineering, Safety Analysis, Requirements Engineering, Verification and Validation, Modelica, Automotive.

1. INTRODUCTION

The modeling of system requirements deals with formally expressing constraints and requirements that have an impact on the behavior of the system so as to enable their verification through real or simulated experiments and/or analytical techniques. The need of models for representing system requirements as well as for methods and techniques, especially centered on model-based approaches, able to support the modeling, evaluation, and validation of requirements and constraints along with their traceability is today even more prominent (Krause, Hintze, Magnus, and Diedrich 2012; Peraldi-Frati and Albinet 2010; Tundis, Rogovchenko-Buffoni, Fritzson, and Garro 2013; Yu, Xu, and Du 2009). In particular, while the modeling and verification of functional requirements are well supported by several tools and techniques, there is still a lack of models and methods specifically conceived to deal with non-functional requirements (such as reliability, availability, maintainability, safety, security); as a consequence, their verification is often postponed to the late stages of the development process with the risk of having to revise already implemented design choices, and, consequently, to miss project deadlines and exceed the budget (Garro, Tundis, and Chirillo 2011; Garro and Tundis 2012c).

Among non-functional requirements, Safety, which represents an important requirement to be satisfied for a wide range of systems (Laprie 1992), becomes even more crucial in several industrial domains such as nuclear plants, medical appliances, avionics, automotive and satellite (Guillerm, Demmou, and Sadou 2010; Garro, Tundis, Groß, and Riestenpatt Gen. Richter 2013; Lahtinen, Johansson, Ranta, Harju, and Nevalainen 2010; Rierson 2013). In particular, in the automotive domain, although Safety has always played a key role, the importance that is attributed to it has become far greater in recent times (Herpel and German 2009; Garro and Tundis 2012a; Navinkumar and Archana 2011). In modern automobile design, Safety Requirements can be generally categorized in three main classes: (i) *Passive safety*, which aims to minimize the severity of an accident; examples of passive safety elements are seatbelts, crumple zones, airbags; (ii) *Active safety*, which aims to avoid accidents and to minimize their effects if they occur; examples of active safety elements are: predictive emergency braking, seatbelt pre-tensioning, anti-lock braking systems and traction control; (iii) *Functional safety*, which aims to ensure that both the electrical and electronic systems (such as power supplies, sensors, communication networks, actuators, etc.), also including all active safety related systems, function correctly. In other words, Functional safety aims to guarantee the absence of unacceptable risk due to hazards caused by malfunctioning behavior of electrical and electronic systems.

The increasing importance that Safety is gaining as one of the main selling points with which to differentiate between car manufactures has led these competitors to join together to foster the definition of safety standards for automotive such as ISO-26262 (ISO-26262 2011). Its basis resides in the more generic IEC-61508 (IEC-61508 2010) which has a broad field of application (industrial process, control and automation, oil/gas, nuclear, etc.). However, ISO-26262 is totally dedicated to the automotive sector and allows car manufacturers to indemnify themselves from liability in case a malfunction remains undetected when following the standard (Lahtinen, Johansson, Ranta, Harju, and Nevalainen 2010). At the process level, this standard allows to follow a clear guidance on the development and validation of electrical and electronic systems, avoiding errors in the design and implementation, which could otherwise induce more expensive production activities and delay during the development (Täubig, Frese, Hertzberg, Lüth, Mohr, Vorobev, and Walter 2012). Moreover, a well-defined and standardized development process, which goes from the Requirements Analysis phases up to the System Testing phases, allows supporting the traceability of Safety Requirements during all the intermediate development stages.

In this paper a comprehensive approach, inspired by the ISO-26262 standard, for the definition of *Functional Safety Requirements* of systems is proposed along with a mechanism to enable their traceability and support their verification through simulation. The approach is based on an iterative process which is an extension for the Safety Analysis of Physical Systems of that proposed in (Garro and Tundis 2012b; Garro, Tundis, Groß, and Riestenpatt Gen. Richter 2013) and is constituted by the following main phases (see Figure 1): *Requirements Analysis*, *System Modeling* and *Virtual Testing*. Both the *Requirements Analysis* phase and the *System Modeling* phase are based on UML/SysML (System Modeling language) and supported by related modeling tools (IBM Rational Rhapsody); whereas, the *Virtual Testing* phase is enabled by the *OpenModelica* environment (OpenModelica), an Open Source simulation environment based on the Modelica language which is an equation-based object-oriented language for representing physical systems with *acausal* features (Fritzson 2004).

The rest of the paper is structured as follow: Section 2 introduces the safety analysis discipline along with a brief survey on the most common related techniques; then, in Section 3, the proposed simulation-driven design process for the safety analysis of systems is presented; in Section 4, this process is exemplified through a case study in the automotive; finally, conclusions are drawn and future work delineated.

2. SYSTEM SAFETY ANALYSIS AND RELATED TECHNIQUES

Safety Analysis is a discipline of Safety Engineering whose aim is to ensure that engineered systems provide acceptable levels of safety through the identification of safety related risks, eliminating or controlling them by design and/or procedures, based on acceptable system safety precedencies (FAA 2000; NASA).

System safety uses systems theory and systems engineering approaches to prevent foreseeable accidents and minimize the effects of unforeseen ones. It considers losses in general, not just human death or injury. Such losses may include destruction of property, loss of mission and environmental harm. Safety of systems needs to be planned in an integrated and comprehensive engineering framework that requires experience in the application of safety engineering principles by exploiting well-known analysis techniques to perform safety analysis for the identification and the management of *hazards*. The general definition of Safety is based on the main concept of *risk* which is the combination of the probability of a failure event and the severity resulting from the failure.

Several techniques for performing quantitative and qualitative safety analyses are currently available. *Quantitative analysis* techniques are based on the identification and modeling of physical and logical connections among system components and on their analysis through statistical methods and techniques, but very often probabilistic information is not so relevant or desired, for example, when one wants to study the reachability of a state of the system, as a consequence *Qualitative analysis* techniques are often preferred (Rouvroye and Van den Bliek 2002).

The *Fault Hazard Analysis* (FHA) is a deductive method of analysis that can be used exclusively as a qualitative analysis or, if desired, expanded to a quantitative one (Pomeranz and Reddy 2009). The Fault Hazard Analysis requires a detailed investigation of the subsystems to determine component hazard modes, causes of these hazards, and resultant effects to the subsystem and its operation. This type of analysis belongs to a family of reliability analysis techniques which comprehends FMEA/FMECA (Failure Mode and Effects Analysis/Failure mode effects and criticality analysis). The main difference between the FMEA/FMECA and the Fault Hazard Analysis is a matter of depth. Wherein the FMEA or FMECA looks at all failures and their effects, the Fault Hazard Analysis deals only with those effects that are safety related.

Fault Tree Analysis (FTA) is a popular and productive hazard identification tool (Clifton 1999). A FTA is a deductive or backward logic representation which involves specifying a top event to analyze (a system failure), followed by identifying all of the associated elements in the system that could cause that top event to occur. It provides a standardized discipline to evaluate and control hazards. The FTA process is used to solve a wide variety of problems ranging from safety to management issues. This tool is used by the professional safety and reliability community to both prevent and resolve hazards and failures. Both qualitative and quantitative methods are used to identify areas in a system that are most critical to safe operation. The output is a graphical presentation providing a map of “failure or hazard” paths.

Event Tree Analysis (ETA) is an analysis technique for identifying and evaluating the sequence of events in a potential accident scenario following the occurrence of an initiating event (Kenarangui 1991). ETA is an inductive or forward logic representation, which starts from an initiating event and includes all possible paths, whose branch points represent successes and failures. The objective of ETA is to determine whether the initiating event will develop into a serious mishap or if the event is sufficiently controlled by the safety systems and procedures implemented in the system design. An ETA can result in many different possible outcomes from a single initiating event and it provides the capability to obtain a probability for each outcome.

Common Cause Failure Analysis (CCFA) is an extension of FTA to identify "coupling factors" that can cause component failures to be potentially interdependent (Liudong and Wendai 2008). Primary events of minimal cut sets from the FTA are examined through the development of matrices to determine if failures are linked to some common cause relating to the environment, location, secondary causes, human error, or quality control. A cut set is a set of basic events (e.g. a set of component failures) whose occurrence causes the system to fail. A minimum cut set is one that has been reduced to eliminate all redundant "fault paths". CCFA provides a better understanding of the interdependent relationship between FTA events and their causes. It analyzes safety systems for "real" redundancy.

Sneak Circuit Analysis (SCA) is a method for the evaluation of electrical circuits (Price and Hughes 2002). SCA employs recognition of topological patterns that are characteristic of all circuits and systems. The purpose of this analysis technique is to uncover latent (sneak) circuits and conditions that inhibit desired functions or cause undesired functions to occur, without a component having failed. The process converts schematic diagrams to topographical drawings and searches for sneak circuits.

The *Energy Trace* is a hazard analysis approach addresses all sources of uncontrolled and controlled energy that have the potential to cause an accident (Booya, Arghami, Asilian, and Mortazavi 2007). Examples include utility electrical power and aircraft fuel. Sources of energy causing accidents can be associated with the product or process. The purpose of energy trace analysis is to ensure that all hazards and their immediate causes are identified. Once the hazards and their causes are identified, they can be used as top events in a fault tree or used to verify the completeness of a fault hazard analysis. Consequently, the energy trace analysis method complements but does not replace other analyses, such as fault trees, sneak circuit analyses, event trees, and FMEAs.

Even though the above mentioned techniques are fairly popular for the safety static analysis of systems, nowadays, with the increase of complexity and heterogeneity of modern systems, more dynamic and flexible analysis techniques, based on simulation methods as well as compliant with international safety standards for specific domains, such as ISO-26262 in the automotive one (Aljazzar, Fischer, Grunske, Kuntz, Leitner-fischer, and Leue 2009; SAE 2003; Stapelberg 2008; Struble 2005), are even more required. As an example, the Process Deployment Advisory Service defined on ISO-26262 in order to help identifying gaps in the development processes, including requirements traceability and requirements based-testing, is fully supported by popular tools such as MatLab/Simulink (Mathworks).

3. A SIMULATION-DRIVEN PROCESS FOR THE DESIGN OF SAFE SYSTEMS

In this section a methodological process for the development of safe systems, based on the validation of the design through simulation, is presented.

As shown in Figure 1, such process which is inspired by the ISO-26262 standard, is defined in terms of three main iterative phases: *Requirements Analysis*, *System Modeling*, and *Virtual Testing*, which aim to provide a methodological support according to the ISO-26262 standard.

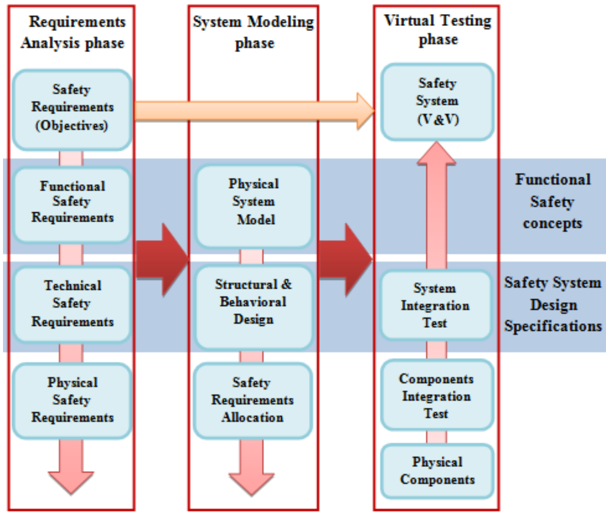


Figure 1: Main phases of the proposed simulation-driven process for the design of safe systems

In the *Requirements Analysis* phase the system safety objectives are analyzed and Safety requirements, in terms of Functional, Technical and Physical requirements, are identified (Rubio, Ponce, and Madrid 2011; Sommerville and Sawyer 2003). They may consist of properties and safety performances to be considered in order to eliminate the risk or to reduce it to an acceptable level. Specifically, a process for their elicitation, definition, formalization and validation is defined according to a meta-model proposed in (Tundis, Rogovchenko-Buffoni, Fritzson, and Garro 2013).

In particular, the first step consists in the requirements elicitation that, according to the proposed meta-model, is obtained through *RequirementAssertions*. An iterative process between the user and the analyst is typically executed in order to state all the requirements, as much as possible, by associating to each of them a *Name* for their identification along with a possible *Description* in a text format by using the natural language in order to provide an explanation of specific or salient aspects, characteristics, or features (e.g. functional, technical or physical) of the system in a detailed way. At the end of this step the so called *User Requirements (URs)* are generated according to the meta-model.

The second step consists in the refinement of the *URs* in order to generate *System Requirements (SRs)*. This step is very crucial to make URs machine readable and executable in order to enable their verifiability during the simulation, as a consequence, it is really important what to represent and how to do it as well as when to use such requirements. First of all a *RequirementAssertion* could be involved in several verification tasks grouped in different *RequirementModel*, so the membership of each requirement to at least one of those *RequirementModels* must be identified. Then, the output values, associate to the evaluation of requirements, for describing if a requirement has been not violated, violated, and so on, have to be fixed. At the end, a *Metric* needs to be specified for each *RequirementAssertion*. In particular, it specifies the purpose of a *RequirementAssertion* in terms of verification mechanism. In Figure 2 the relationships among the User Requirements, System Requirements and Safety Requirements are represented.

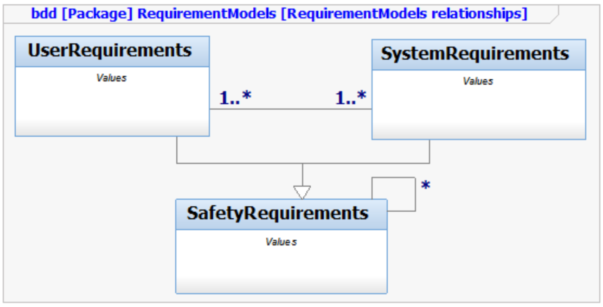


Figure 2: Relationships among User Requirements System Requirements and Safety Requirements

The representation of requirements is carried out by using Requirement diagrams available in SysML, a UML profile for modeling system, and exploiting tools such as IBM Rational Rhapsody (IBM) or Papyrus (Papyrus), in order to enable model-based system engineering.

It is worth to notice that not all the requirements can be formalized into something computable such as "a cable must be well connected", if the term "well connected" is not represented in a machine readable formalism.

In the *System Modeling* phase, a possible physical model of the real system in terms of its components is defined; in particular, the Structural and the Behavioral views are generated by breaking down the system in (sub)components.

Specifically the first step, according to the *Physical* side of the meta-model proposed in (Tundis, Rogovchenko-Buffoni, Fritzson, and Garro 2013), consists in building a possible *PhysicalSystemModel*, of the actual *PhysicalSystem* by specifying the *models* of its physical components (*PhysicalComponentModels*) and the related *Attributes* and, then, defining the *relationships* among them as well as their *behaviors*. In particular, the structural part of the system is described by using Block Definition Diagrams and Internal Block Diagrams in a top-down fashion. The behavior of the system, which is modeled by following a bottom-up approach, can be defined in terms of Activity, Sequence or Parametric diagrams in order to model the internal behavior of each system components as well as the flows of actions and interactions between components.

Then *SRs* belonging to the *RequirementModel* concerning Safety Requirements, can be further formalized in order to make them machine executable. In particular, a formal *Measure*, and its expected input and output values, can be associated to the defined *Metric*. Specifically, a *Measure* can be expressed by adopting an appropriate *ComputationalModel* which in turn could be represented through an *Algorithm*, a *Finite Automata*, a *Function*, a set of *Equation* or by their combination to enable the computational process.

Finally, the allocation between the *SafetyRequirements* and the *PhysicalSystemModel* is performed. Furthermore, inputs, required from the *Measure* of a *RequirementAssertion* for its evaluation, are explicitly included in the *PhysicalComponentModels*.

In the *Virtual Testing* phase, the Models of the system under consideration are transformed into executable models and represented in terms of the constructs offered by the OpenModelica platform (Open Source Modelica Consortium), an Open Source simulation environment based on the Modelica language, an equation-based object-oriented language for representing physical systems with *acausal* features, (Modelica and the Modelica Association). In particular, physical components are defined and integrated in order to build the physical system model and then the safety requirements to be verified are introduced into the overall model. Then, different simulation scenarios are set and simulations are executed; finally, simulation results can be analyzed on the basis of the system safety requirements identified in the first process phase. This analysis allows to evaluate the safety properties of the system, to compare different design choices for improving, possibly, the safety of the system under consideration.

As the process is iterative, if necessary, new partial or complete process iterations can be executed.

3.1. Relationships between the ISO-26262 standard and the proposed process

The above described process is inspired by the IEC-61508 standard and, in particular, by the ISO-26262 whose goal is to demonstrate the capability to develop certain products with acceptable risks.

ISO-26262 is organized in 10 parts as following:

- *Part 1 - Vocabulary*: which specifies the terms, definitions and abbreviated terms for application in all parts of ISO 26262;
- *Part 2 - Management of Functional Safety*: which specifies the requirements for functional safety management for automotive applications, including (i) project-independent requirements with regard to the organizations involved (overall safety management), and (ii) project-specific requirements with regard to the management activities in the safety lifecycle (i.e. management during the concept phase and product development, and after the release for production);
- *Part 3 - Concept phase*: which specifies the requirements for the concept phase for automotive applications (e.g. item definition, functional safety concept, etc.);
- *Part 4 - Product Development at system level*: which specifies the requirements for product development at the system level for automotive applications, such as the system design and system integration and testing;
- *Part 5 - Product Development at hardware level*: which specifies the requirements for product development at the hardware level for automotive applications (e.g. hardware design and hardware architectural metrics, hardware integration and validation);
- *Part 6 - Product Development at software level*: which specifies the requirements for product development at the software level for automotive applications such as software architectural design, software unit design and implementation, software integration and testing;
- *Part 7: Production and Operation*: which specifies the requirements for production, operation, service and decommissioning.
- *Part 8: Supporting Processes*: which specifies the requirements for supporting processes through qualified tools, system engineering approaches and best practices;
- *Part 9: Automotive Safety Integrity Level (ASIL)-oriented and safety-oriented analyses*: concerning the measures required to avoid unreasonable risks.
- Part 10: *Guidelines on ISO-26262*.

In the Table 1 the matching between ISO-26262 parts and the phases of the proposed process are shown, by indicating in which phase of the process a specific part of such standard should be considered.

In particular *Vocabulary* and *Management of Functional Safety Concept phase* can be considered in the *Requirements Analysis phase* for the definition, the organization and categorization of requirements; then *Product Development at system level*, *Product development at the hardware level* and *Product development at the software level* can be taken into account in the *System Modeling phase,* when the design of the system is under definition, whereas the *Supporting Process* part can be considered during the *Virtual Testing phase* of the proposed process.

Table 1: Matching between ISO-26262 and the proposed process.

Parts of the Standard ISO-26262	Simulation-Driven Process for the Design of Safe Systems
Vocabulary *Management of Functional Safety* *Concept phase*	*Requirements Analysis phase*
Product Development at system level *Product development at the hardware level* *Product development at the software level*	*System Modeling phase*
Supporting Process	*Virtual Testing phase*

4. FROM SAFETY REQUIREMENTS TO A SAFE DESIGN IN THE AUTOMOTIVE DOMAIN: A CASE STUDY

In this Section, a case study in the automotive domain concerning the modeling of an airbag system, and the validation and evaluation of its design according to the safety requirements through simulation, is analyzed following the proposed process. In particular, after a brief introductive description of the system under consideration, its safety analysis is performed.

4.1. Airbag description

Airbags are one of the most important components of a motor vehicle system for the occupant protection. It is used along with and as a supplement to the seatbelt restraint system to provide passenger protection in case of collision In addition to the standard airbags for the driver and front passenger, an increasing number of specialized airbag variants (such as curtain airbags, kneebags, etc.) are used.

Each airbag should be specifically designed and optimized for its intended purpose. In addition to the deployment technology, which can in principle be based on the uniform pressure approach or the more recent corpuscular method, this includes the selection of the inflow method (Wang-Nefske or hybrid approach, etc.) as well as the verification and validation of the associated inflow data. Moreover, the deployment behavior is also determined by the correct adjustment of contact, discharge opening and porosity parameters. As a consequence a sensible and comprehensive simulation of airbag behavior as part of a simulation of the entire restraint system is indispensable.

An airbag is typically made of synthetic material and equipped with holes in the rear; it is usually composed by different subsystems such as:

- a *sensor* that detects the abrupt deceleration of the vehicle caused by an impact and the pressure;
- an *Airbag Control Unit* (ACU) that monitors the readiness of the entire airbag system.
- a *detonator* that triggers the substance contained in the explosive capsule through an electric current or a bump of a ferrule;
- a possible second capsule (*GasSoure*) that contains pre-compressed inert gas which inflates the airbag;
- a *warning light* which is illuminated if a fault is detected.

Specifically the ACU receives the signal of the sensor, processes it and sends the command to switch on a detonator; which in turn blows up the capsule of the detonator by developing a large amount of gas, to inflate the container.

4.2. Requirements Analysis phase

In this phase of the proposed process all the possible user requirements need to be identified and elicited.

As an example, in the following some *URs* are reported: (*Req1)* when the car decelerates very quickly, as in a head-on crash, the electrical circuit has to be turned on for initiating the process of inflating the airbag; (*Req2*) the process, from the initial impact of the crash to full inflation of the airbags, takes less than 40 milliseconds; (*Req3*) when a sensor detects a collision an immediate trigger should be sent to enable the deployment of the airbag; (*Req4*) in order for the airbag to cushion the head and torso with air for maximum protection, the airbag must begin to deflate (i.e., decrease its internal pressure) by the time the body hits it, otherwise, the high internal pressure of the airbag would create a hard surface instead of a protective cushion; (*Req5*) the airbag is ignited within a well-define threshold.

Starting from the collected *URs* the next step consist into their rewriting in *SRs* for making them more formal and by identifying their belonging *RequirementModel.* For example:

- *AbruptDeceleration(Req1)*: when the deceleration *d* is greater than a *threshold*, a *signal* to switch on the electronic circuit has to be sent;
- *InflationTime(Req2)*: The time to inflate the airbag has to take less than *40ms*, *inflationTime<=40;*
- *CollitionDetection(Req3)*: when the collision is detected by the sensor, a *collitionSignal* has to be generated;
- *DeflationTime(Req4)*: the airbag has to be able to deflate in a *deflationTime* lesser than a deflation *threshold*.
- *Activation(Req5)*: after a crash the airbag is deployed in *delayTime*=45ms.

Specifically, the relationships among the above mentioned safety requirements are represented in Figure 3. In particular the status of the *DeflationTime* is not violated if at least the status of the requirement *InflationTime* is not violated. In turn the status of the *InflationTime* is not violated if at least the status of the *Activation* requirement is fulfilled at least by both the *AbruptDeceleration* requirement and the *CollitionDetection* requirement. That is to say, the status of both *AbruptDeceleration* and *CollitionDetection* must be not violated.

Moreover different scenarios can be analyzed, such as:

- the airbag is not ignited or is inflated too late even though a critical crash occurred;
- the airbag is deployed unintentionally, which means that it is ignited even though no crash at all or only a non-critical crash has occurred;

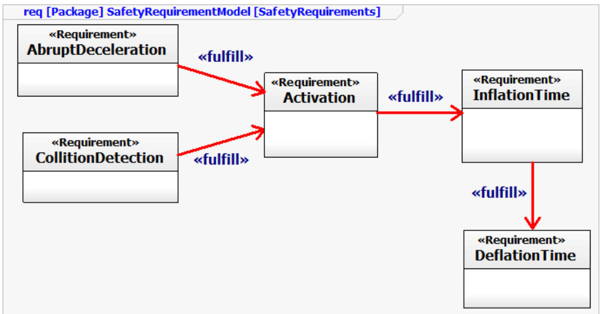


Figure 3: Safety System Requirement relationships

4.3. System Modeling phase

In this phase both the physical structure of the system is built by composing components and then the behavior of each single component is specified.

As it is shown in Figure 4, a Block Definition Diagram (BDD) of an Airbag System is depicted, in terms of its subsystems and ports. Then, the interactions among these components are better specified by using the Internal Block Diagram (IBD), as it is shown in Figure 5.

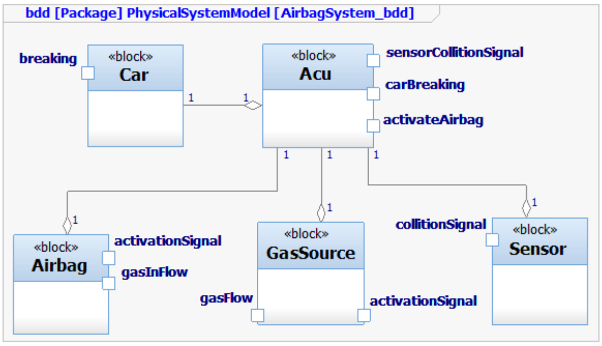


Figure 4: Physical System Model: Components of the Airbag System

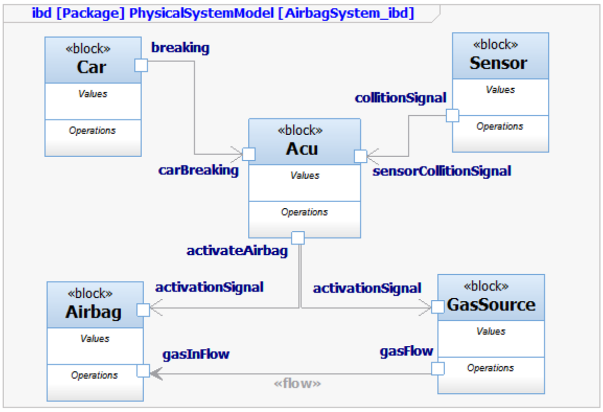


Figure 5: Physical System Model: Components interactions of the Airbag System

After the structure is built, Parametric diagrams are employed for representing the behavior of each subsystem as well as dynamic interactions among them, by exploiting a *Computational Model* based on *EquationsSet*. As an example, in Figure 6 the diagram concerning the behavior of the Airbag component is reported. In particular, in the first section of the diagram, the parameters taken in input from the model are defined, secondly a brief description about the use of such parameters is reported; then the behavior of the Airbag component, which exploits such input parameters, is represented in terms of equations.

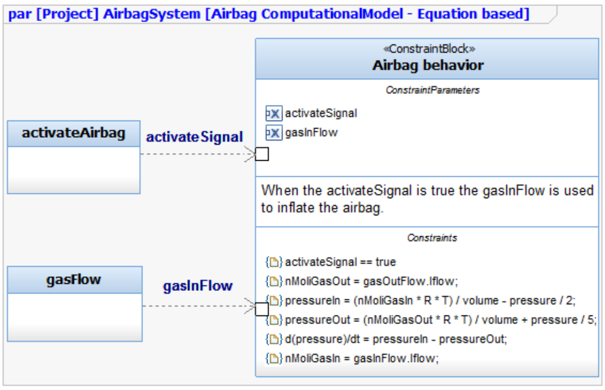


Figure 6: Computational Model of the Airbag component

Finally, requirements modeled in the previous phase, which need to be verified, are allocated to (i) a single physical component in order to check its behavior or (ii) a set of physical components in order to check if the interaction among them is or is not consistent as expected. In Figure 7 the allocation of some requirements to the airbag physical system model, is shown.

In particular, such a scenario wants to verify, the *InflationTime* of the airbag when a car-crash occurs. Specifically, the requirement is not violated when the status of the *Activation* requirement is not violated and both the *Acu* component and the *Airbag* component fulfill the internal rules specified by the *InflactionTime.*

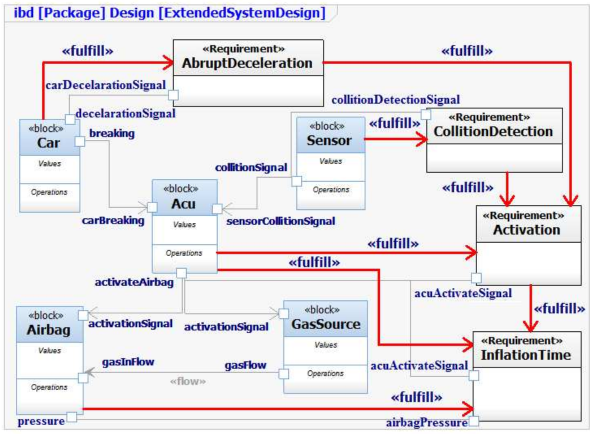


Figure 7: Allocation of Safety Requirements to the Airbag Physical System Model

4.4. Virtual Testing phase

In this phase the virtual testing is executed by exploiting the simulation in order to study the behavior of the system under consideration and analyze interesting aspects or, possibly, to discover some issues that are not immediately obvious when applying static analysis techniques. In order to enable the simulation, models generated in the previous phase need to be translated into the desired simulation platform in order to make them executable. In this case the OpenModelica environment has been chosen as simulation platform since: (i) it is equation based (by implementing the Modelica Language) and, as a consequence, compliant with the *Computational Model* which has been used to represent the behavior of the overall Airbag System; (ii) it is open source, thus allowing the possibility to extend both the language and the tool, to enable modeling of requirements and introduce allocation mechanisms.

In Figure 8, a fragment of source code in Modelica language, which represents the structure of the AirbagSystemDesign, is reported.

```
model AirbagSystemDesign
 import AirbagPhysicalComponentModel.*;
 Car car;
 Acu acu;
 Sensor sensor;
 Airbag airbag;
 GasSource gasSource(gasGain = 0.73);
 GasOut gasDestination(gasGain = 0.02);
equation
 connect(car.breaking,acu.carBreaking);
 connect(sensor.collitionSignal,acu.sensorCollitionSignal);
 connect(sensor.collitionSignal,acu.carBreaking);
 connect(acu.activateAirbag,airbag.activationSignal);
 connect(airbag.gasInFlow,gasSource.gasFlow);
 connect(airbag.gasOutFlow,gasDestination.gasFlow);
end AirbagSystemDesign;
```

Figure 8: Airbag System Design in Modelica

As we can see by looking at the source code, the transformation between SysML notation and Modelica constructs, is almost direct. In particular, each *SysML block* can be represented as a *Modelica Model*, whereas connections among *SysML blocks* can be enabled by the *connect* construct, which is already available in the Modelica language.

In Figure 9, a fragment of Modelica source code concerning the implementation of the behavior of the airbag component, is shown.

```
model Airbag
 Boolean activationSignal(start = false);
 GasFlow gasInFlow "Connector-flow";
 GasOut gasOutFlow "Connector-flow";
 Real nMoliGasIn(unit = "mol", start = 0);
 Real nMoliGasOut(unit = "mol", start = 0);
 constant Real volume(unit = "l") = 9.0;
 constant Real R(unit = "l*atm/mol*K") = 0.0821;
 constant Real T(unit = "K") = 295.0;
 Real pressureIn(start = 0.0, unit = "atm");
 Real pressureOut(start = 0.0, unit = "atm");
//Airbag Internal Pressure
 Real pressure(start = 0.0, unit = "atm") ;
equation
 if activationSignal == true then
   der(pressure) = pressureIn - pressureOut;
   nMoliGasIn = gasInFlow.lflow;
   nMoliGasOut = gasOutFlow.lflow;
   pressureIn = (nMoliGasIn * R * T) / volume - pressure / 2;
   pressureOut = (nMoliGasOut * R * T) / volume + pressure / 5;
 ...
  ...
 end if;
end Airbag;
```

Figure 9: Representation of the behavior of the Airbag component in Modelica

As we can see from the picture, the component behavior, which was described in the *System Modeling phase* through a SysML parametric diagram (see Figure 6), has been is translated in a set of equations by using the Modelica language.

Similarly, the requirements identified in the *Requirements Analysis phase* are formalized by exploiting some extensions of the Modelica language, proposed by the authors in (Rogovchenko-Buffoni, Fritzson, Garro, Tundis, and Nyberg 2013); specifically: (i) the *requirement* keyword is used for their representations, (ii) the *fulfill* relationship is used both for their allocation to the physical system and for their traceability, (iii) the *precondition equation* section is used to specify the conditions when the evaluation of the requirement has to be performed.

In particular, the source code of the formalized requirement *InflationTime* is reported in Figure 10, where the evaluation is based on the inflation time that the airbag takes to reach a specific safety level of pressure after the airbag is activated.

```
requirement InflationTime
  Real airbagPressure(unit="atm");
  Real safePressureLevel(unit="atm")=2.5;
  Boolean activateAirbag(start=false);
  constant Real maxInflationTime(unit="ms")=40;
  constant Real activationTime(unit="ms")=20;
precondition equation
 activateAirbag=true;
 (time < activationTime + maxInflationTime)=true;
equation
 (airbagPressure >= safePressureLevel)=true;
end InflationTime;
```

Figure 10: Formalization of the *InflationTime* requirement by using Modelica language extensions

The source code of the extend system design is reported in Figure 11, where the *fulfill* keyword is employed for creating the matching among the requirements as well as between requirements and physical components of the airbag system.

```
model ExtendedSystemDesign
      import SafetyRequirementModel.*;
 // System Design
 AirbagSystemDesign asd;
 // Safety Requirements
 AbruptDeceleration ad;
 CollitionDetection cd;
 Activation ac;
 InflationTime it;
equation
  // Fulfill equations
 {asd.car} fulfill ad;
 {asd.sensor} fulfill cd;
 {asd.acu, ad, cd} fulfill ac;
 {asd.airbag, asd.acu, ac} fulfill it;
  // Connect equations between Requirements
  // and PhysicalComponents
 connect(asd.airbag.pressure,it.airbagPressure);
 connect(asd.acu.activateAirbag,it.acuActivateSignal);
 connect(asd.car.decelarationSignal,ad.carDecelarationSignal);
 connect(asd.sensor.collitionSignal,cd.collitionDetectionSignal);
 connect(asd.acu.activateAirbag,ac.acuActivateSignal);
end ExtendedSystemDesign;
```

Figure 11: Formalization of the *ExtendedSystemDesign* by using Modelica language extensions

After obtaining the executable models, the tuning of the simulation parameters is performed in order to reach a safe working state of the system according to the specified requirements. Several simulations have been executed for testing virtually the System in different scenarios and evaluating its behavior. Moreover three possible values can be reached by a requirement.

In the considered experimentations (see Figure 12 and Figure 13) a pressure level (*safePressureLevel* in yellow color) of 2.5 atmospheres (atm) has been considered as minimum safe threshold, coupled with a maximum inflation time (*maxInflationTime*) of 40 milliseconds (ms) in order to reach such safe pressure level of the airbag, after that the activation airbag signal (*activateAirbag*) is arrived.

As it is shown in Figure 12, even though the *AbruptDeceleration* (in dark blue color) and *CollitionDetection* (in dark green color) requirements are satisfied as well as the *Activation* (in brown color) requirement, the *requirementStatus* of the *InflactionTime* (in red color) is negative, that is to say it is violated. Indeed, as we can see, the *airbagPressure* (in light blue color) is not able to reach the *safePressureLevel* (in yellow color) within the *maxInflationTime* (in 40 milliseconds) as expected.

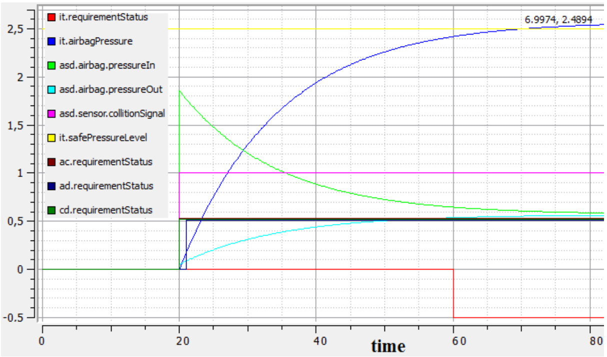


Figure 12: Violation of the *InflationTime* requirement

As it is shown in Figure 13, by setting opportunely some parameters of the airbag system, for example, by increasing the pressure to be provided in input to the airbag (*pressureIn* in light green color), it is possible to reach the necessary parameters tuning which fulfills all the requirements in the considered scenario.

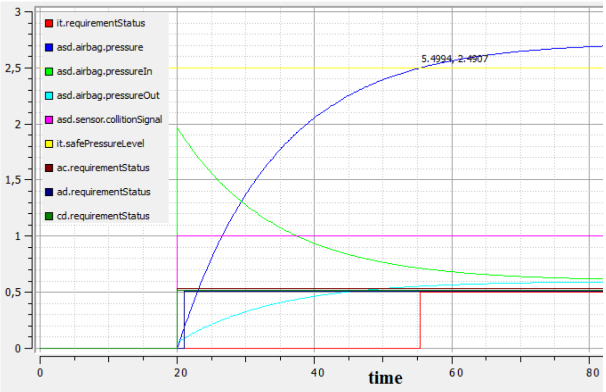


Figure 13: Fulfillment of the *InflationTime* requirement

5. CONCLUSIONS AND FUTURE WORKS

The paper has presented a model-driven process for supporting the safety analysis of systems which is inspired by ISO-26262 standard and exploits simulation techniques.

Two powerful languages for modeling systems have been combined in a comprehensive system engineering framework; specifically, SysML has been exploited for platform independent representation of the system; whereas, the Modelica language has been exploited for the executable representation of the systems according to an equation-based paradigm.

A prototype of the OpenModelica simulation platform, able to support both the modeling of requirements and their allocation, according to a well-defined reference meta-model, has been used for the simulation. Finally, a concrete experimentation has been conducted in the automotive domain which has allowed to point out both the flexibility and the effectiveness of the overall proposed process for safety analysis.

Future work includes both the improvements of the proposed model-driven process and its extension by introducing (i) some approaches and possible patterns for representing dysfunctional behavior and fault injection and, (ii) a probability model for enabling the representation of uncertainties in order to perform Fault Tree Analysis of a Modelica-based system model.

ACKNOWLEDGMENTS

Andrea Tundis has been supported by a grant funded in the framework of the “POR Calabria FSE 2007/2013”. This work is part of an ongoing research project, the MODRIO Project (ITEA 2), aiming at developing a model-based approach for system requirements verification and safety analysis through simulation.

REFERENCES

Aljazzar H., Fischer M., Grunske L., Kuntz M., Leitner-fischer F., Leue S., 2009. Safety Analysis of an Airbag System using Probabilistic FMEA and Probabilistic Counterexamples. *Proceedings of the 6th International Conference on the Quantitative Evaluation of Systems (QEST)*. September 17-18, Eger (Hungary).

Booya M., Arghami S., Asilian H., Mortazavi S., 2007. Safety analysis of a corn processing industry by energy trace and barrier analysis method: a case study. *Iran Occupational Health Journal*, 4 (3), 27-34.

FAA 2000. *System Safety Handbook*.

Clifton E., 1999. Fault tree analysis - a history. *Proceedings of the 17th Inernational Systems Safety Conference*, pp. 1-9. August 16-21, Orlando (Florida, USA).

Fritzson P., 2004. *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*, Wiley IEEE Press.

Garro A., Groß J., Riestenpatt Gen. Richter M., Tundis A., 2013. Reliability Analysis of an Attitude Determination and Control System (ADCS) through the RAMSAS method. *To appear in Journal of Computational Science*, Elsevier.

Garro A., Tundis A., 2012a. Enhancing the RAMSAS method for System Reliability Analysis: an exploitation in the automotive domain. *Proceedings of the 2nd International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH)*. July 28-31, Rome (Italy).

Garro A., Tundis A., 2012b. Modeling and Simulation for System Reliability Analysis: The RAMSAS Method. *Proceedings of the 7th IEEE International Conference on System of Systems Engineering (IEEE SoSE)*. July 16-19, Genova (Italy).

Garro A., Tundis A., 2012c. A model-based method for system reliability analysis. *Proceedings of the Symposium on Theory of Modeling and Simulation (TMS)*. March 26-29, Orlando (FL, USA).

Garro A., Tundis A., Chirillo N., 2011. System reliability analysis: a model-based approach and a case study in the avionics industry. *Proceedings of the 3rd Air and Space International Conference (CEAS)*. October 24-28, Venice (Italy).

Guillerm R., Demmou H., Sadou N., 2010. Engineering dependability requirements for complex systems - A new information model definition. *Proceedings of the 4th Annual IEEE Systems Conference*. April 5-8, San Diego (CA, USA).

Herpel T., German R., 2009. A simulation approach for the design of safety-relevant automotive multi-ECU systems. *Proceedings of the 4th IEEE International Conference on System of Systems Engineering (IEEE SoSE)*. May 30 - June 03, Albuquerque (New Mexico, USA).

IBM Rational Rhapsody web site - http://www.ibm.com/.

IEC-61508, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems, Parts 1-7.

ISO-26262, 2011. Software Compliance: Achieving Functional Safety in the Automotive Industry.

ITEA 2 Projects: MODRIO web site - http://www.itea2.org/.

Kenarangui R., 1991. Event-tree analysis by fuzzy probability. *IEEE Transaction on Reliability*, 40 (1), 120-124.

Krause J., Hintze E., Magnus S., Diedrich C., 2012. Model based specification, verification and test generation for a safety fieldbus profile. *Proceedings of the 31st International Conference on Computer Safety, Reliability and Security (SafeComp)*. September 25, Magdeburg (Germany).

Lahtinen J., Johansson M., Ranta J., Harju H., Nevalainen R., 2010. Comparison between IEC 60880 and IEC 61508 for certification purposes in the nuclear domain. *Proceedings of the 29th International Conference on Computer Safety, Reliability and Security (SafeComp)*. September 14-17, Vienna (Austria).

Laprie J.C., 1992. *Dependability: Basic Concepts and Terminology*, Springer-Verlag.

Liudong X., Wendai W., 2008. Probabilistic common-cause failures analysis. *Proceedings of Annual Reliability and Maintainability Symposium*, pp. 354-358. January 28-31, Las Vegas (NV, USA).

Mathworks MatLab/Simulink web site - http://www.mathworks.com/.

Modelica and the Modelica Association web site - https://www.modelica.org/.

NASA - http://askmagazine.nasa.gov/pdf/pdf31/NASA_APPEL_ASK_31i_introduction_to_system_safety.pdf

Navinkumar V.K., Archana R.K., 2011. Functional safety management aspects in testing of automotive safety concern systems (electronic braking system). *Proceedings of the 3rd International Conference on Electronics Computer Technology (ICECT)*. April 8-10, India.

OpenModelica - Open Source Modelica Consortium (OSMC) web site - https://www.openmodelica.org/.

Papyrus - Eclipse project web site - http://www.eclipse.org/papyrus/.

Peraldi-Frati M., Albinet A., 2010. Requirement traceability in safety critical systems. *Proceedings of the 1st Workshop on Critical Automotive applications: Robustness & Safety (CARS)*. April 27, Valencia (Spain).

Pomeranz I., Reddy S.M., 2009. Hazard-Based Detection Conditions for Improved Transition Fault Coverage of Functional Test Sequences. *Proceedings of the 24th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, pp. 358 - 366. October 7-9, Chicago (IL, USA).

Price C.J., Hughes N., 2002. Effective automated sneak circuit analysis. *Proceedings of Annual Reliability and Maintainability Symposium*, pp. 356 – 360. January 28-31, Seattle (WA, USA).

Rierson L., 2013. *Developing Safety-Critical Software: A Practical Guide for Aviation Software and DO-178C Compliance*, CRC Press.

Rogovchenko-Buffoni L., Fritzson P., Garro A., Tundis A., Nyberg M., 2013. Requirement Verification and Dependency Tracing During Simulation in Modelica. *To appear in proceedings of the 8th EUROSIM Congress on Modelling and Simulation*. September 10-13, Cardiff (Wales, UK).

Rouvroye J.L., Van den Bliek E.G., 2002. Comparing safety analysis techniques. *Journal of Reliability Engineering & System Safety*, Elseiver, 75 (3), 289–294.

Rubio D., Ponce S., Madrid F., 2011. ISO/IEC 17025 technical requirements in electrical safety laboratory for electromedical devices. *Proceedings of Pan American Health Care Exchanges (PAHCE 2011)*. March 28 – April 1, Rio de Janeiro (Brazil).

SAE International, 2003. *Airbags and Safety Test Methodology, Society of Automotive Engineers*.

Stapelberg R.F., 2008. *Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design*, 1st ed. Spinger-Verlag.

Sommerville I., Sawyer P., 2003. *Requirements Engineering: A good practice guide*. Wiley.

Struble D.E., 2005. *Advances in Side Airbag Systems, Society of Automotive Engineers Inc*.

Systems Modeling Language (SysML) web site - http://www.omgsysml.org/.

Tundis A., Rogovchenko-Buffoni L., Fritzson P., Garro A., 2013. Modeling System Requirements in Modelica: Definition and Comparison of Candidate Approaches. *Proceedings of the 5th International Workshop on Equation-Based Object-Oriented Modeling Languages and Tools (EOOLT)*. April 19, Nottingham (UK).

Täubig H., Frese U., Hertzberg C., Lüth C., Mohr S., Vorobev E., Walter D., 2012. Guaranteeing functional safety: design for provability and computer-aided verification. *Journal Autonomous Robots*, 32, Springer, 303-331.

Yu G., Xu Z., Du J., 2009. An Approach for Automated Safety Testing of Safety-Critical Software System Based on Safety Requirements. *Proceedings in the International Forum on Information Technology and Applications (IFITA)*. May 15-17, Chengdu (China).

LOOKING FOR THE EQUILIBRIUM OF THE SHIELDS PRODUCTION SYSTEM VIA SIMULATION MODEL

Pavla Ficova[(a)], Martina Kuncova[(b)]

[(a,b)]University of Economics Prague, Department of Econometrics, W.Churchill Square, 13067 Prague 3, Czech Republic
[(b)] College of Polytechnics Jihlava, Dpt. of Economic Studies, Tolsteho 16, 58601 Jihlava, Czech Republic

[(a)]Pavla.Ficova@gmail.com, [(b)]martina.kuncova@vspj.cz, kuncovam@vse.cz

ABSTRACT

The use of simulation increasingly becomes part of business planning and analysis. Using a computer program a model that faithfully mimics reality can be easily created and used to analyze the situation in the company. In this paper simulation is used to accurately determine utilization of shields manufacturing machines and also to analyze the entire production process, which begins with the arrival of the material and ends with finished product. The first task is to search for the optimal utilization of machines, number of shifts and employees. Based on the results of the analysis number of produced shields and production process could be modified. The aim is to determine the current state of utilization and to find the equilibrium or ceiling for possible future expansion of production. SIMUL8 has been used as a model background. The results obtained from the simulation model have given new information and invaluable knowledge of the process to the company.

Keywords: shields production, simulation model, discrete event simulation, SIMUL8

1. INTRODUCTION

Simulation methods belong to the suitable instruments that can be used in the real world situations to better understand the reality or to make a responsible decision. Simulation means a technique for imitation of some real situations, processes or activities that already exist in reality or that are in preparation – just to create a computer model (Banks 1998). Models can be created in variant software depending on the type of the model. Sometimes Monte Carlo simulation for iterative evaluation of a deterministic model is sufficient but real simulation is usually made via discrete event simulation model or continuous simulation.

When we talk about the production process model it is usual using discrete event simulation to model the process. It means to describe all activities that are necessary during the production process, their duration, resources needed and their sequence. Discrete event simulation is very useful when studying the behavior of the system or looking for the bottlenecks of the system. As production belongs to the areas where it is hard to find the problematic processes that limits the whole system, discrete-event simulation serves there as a good tool. O'Kane et al (2000) showed how discrete event simulation (and model in WITNESS software) can help with the decision how the output can be increased. The automotive industry is typical for simulation usage – for example Masood (2006) tried to find how to reduce the cycle times and increase the machine utilization in an automotive plant. Montevechi et al (2007) shows that simulation is also good for the creation and testing of different experiments or scenarios.

2. MATERIAL AND METHODS

Simulation can be used to show the eventual real effects of alternative conditions and courses of action. Key issues in simulation represent data acquisition of valid source of information, selection of key characteristics and behaviors, use of simplifying approximations and assumptions within the simulation, and also detection of fidelity and validity of the simulation outcomes (Law 2000). On the other hand simulation model can be created only thanks to the computer and simulation software. In this article we describe the simulation model of the shields production system created in the SIMUL8 software. Several case studies with similar approach to problem solving were published (O'Kane et al 2000, Aguirre, Mendéz 2005, Masood 2006, Montevechi et al 2007).

2.1. SIMUL8

SIMUL8 is a software package designed for Discrete Event Simulation (www.simul8.com).

Figure 1: SIMUL8 logo (www.simul8.com)

It allows user to create a visual model of the system under investigation by drawing objects directly on the screen of his computer. This software uses four main objects:

- Work Entry Points
- Storage Bin (or Queue)
- Work Center
- Work Exit Point

The characteristics of the objects are defined in terms of capacity, speed, etc. All these objects are link together by connectors that define the sequence of the activities and also the direction of movement of entities. Entities are other objects of the model. These dynamic objects (customers, products, documents) move through the processes and use various resources. Resources serve for modeling of limited capacities of the workers, material or means of production that are used during the activities.

Once the system is modeled a simulation can be undertaken. The flow of work items around the system is shown by animation on the screen, and for that reason the appropriateness of the model can be easily assessed. When the structure of the model has been confirmed, a number of trials can be run under different conditions and the performance of the system can be described statistically. Statistics of interest may be average waiting times or utilization of Work Centers or Resources (Shalliker and Rickets 2002).

SIMUL8 can be used for various kinds of simulation models (Concannon et al. 2007). The cases studies can be seen also on the web pages www.simul8.com or www.simul8.cz (in Czech Republic).

2.2. THE SHIELDS PRODUCTION PROCESS

Every production system consists of a lot of processes and activities that are sometimes hard to analyze by usual mathematical or analytical methods. Simulation is a suitable tool to do so. Our task is to create a model of the shields production so as to correspond with reality and afterwards analyze the effect of higher demand for these components. The main aim is to find the equilibrium for the company to meet the demand on the one hand and increase the number of machines minimally on the other hand.

The shields are produced in a Czech machine work company ZLKL as parts of electromotor alternators (Figure 2). Production of each shield is quite difficult process thanks to the requirements of accuracy.

Figure 2: Example of the shield for alternator usage (www.vltava2000.cz)

Before the raw material becomes a shield it must go through the whole production cycle. Raw material is transported towards the machine tools. After first treatment it follows to the procedures of turning, drilling, screwing and deburring. Then the shields must be degreased. Before packing they are randomly controlled if they are of the given size. The entire production process, as well as the expected form of the simulation model is shown in Figure 3

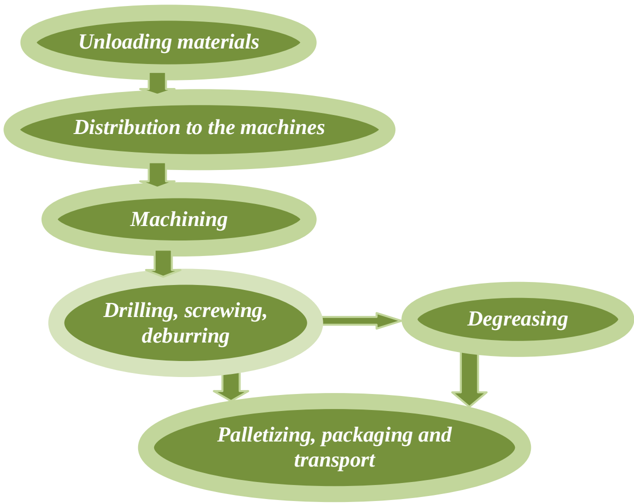


Figure 3: The conceptual model

Table 1: Real average monthly production and expected production after increase

Shield Type	Avg. Monthly production	plan for a year	plan for a month	fulfillment of a plan
21297015025D_B	1585	35 200	2 933	54,00%
21297002001D_C	1061	25 800	2 150	49,30%
21297064001D	1243	35 200	2 933	42,40%
21297009001D_C	1298	33 000	2 750	47,20%
51298132101001_B	1524	26 400	2 200	69,30%
51298160101001_A	415	13 200	1 100	37,70%
21297007040D_B	0	35 200	2 933	0,00%
21297084001D_A	0	33 000	2 750	0,00%
21297064001D	0	35 200	2 933	0,00%
32417311673200	0	30 800	2 567	0,00%

production in pieces

Before the creation of the model we have to know something about the production and the production times. Table 1 shows the average monthly production of shields and expected (planed) production after the increase of the demand. We can see that only for one type for the shield the company is able to satisfy more than 50% of the monthly average demand. Some shields are not produced now. The average yearly production is more than 85000 different pieces but the new demand asks for production at about 417000 of different shields so nearly 5 times more. For the simulation model we have to know the times for material distribution, machining, drilling, screwing, deburring, palletizing and packing. The machine times have been known before (Table 2) but the times where human power is necessary we have to measure and try to find the best distribution for material transportation (Figure 4), palletizing and packing.

Table 2: Times in minutes per different operations

Shield Type	Turning machine	Time for turning (min/piece)	Drilling machine	Time for drilling, screwing, deburring (min/piece)	Degreasing (min/1 piece in batch)
21297015025D_B	EMAG	2,1	V20/4	3,33	0,004
21297000001D_C	SPR 100	11	V20/4	1,5	0,23
21297064001D	EMAG	3	V20/4	3,33	0,004
21297009001D_C	SPR 100	8,2	V20/4	1,5	0,23
51298132101001 _B	SPR 100	11	OC Lilian	1,5	0,23
51298160101001 _A	SPR 100	10, 5	OC Lilian	4,5	0,64

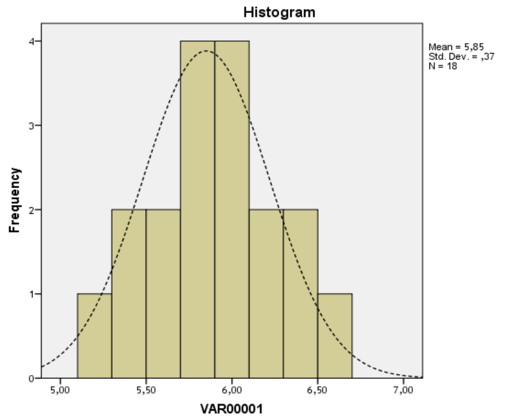


Figure 4: Time for material transport – distribution estimation

Simulation model uses those machine times known in advanced and corrected by the average scrap and delays caused by machine failures. The probability distribution has been set according to our own observations of the processes (normal distribution).

For the production it is necessary to comply with the number of shields of different types. The inputs of the basic model result from the updated information and material is split among the machines according to the percentage of the given type in the order. The model has been created so as each type of the shield would be shaped on different machine. The machine time for each shield is unique and this part is fundamental for the production process. Therefore it was necessary to create more machines within the simulation model than really exist and afterwards we had to summarize the capacity usage of these model machines. Before the summarization it was possible to see the workers utilization

3. RESULTS

According to the given conditions and know characteristics of the production process the first version of the model has been created (Figure 5).

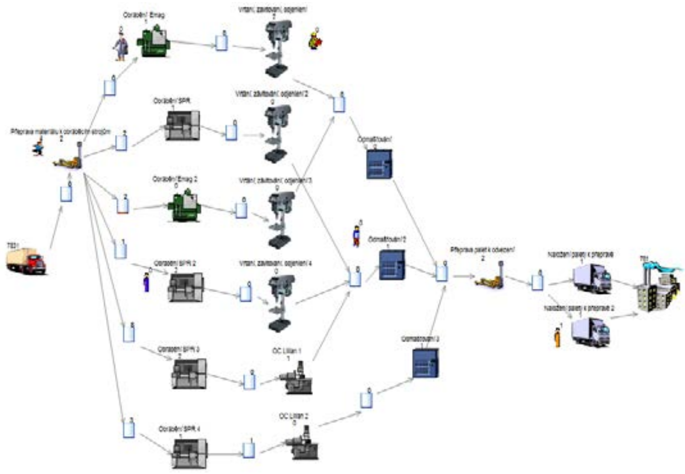


Figure 5: Simulation model of the current production

The results of the model agree with the real number of shields produces (Figure 6) and we can also see that the utilization of the machines is low except of one machine where the utilization is higher than 50% (but still it is not a problem). First one month and one shift simulation has been created and afterwards the results from the half-year simulation has been taken to validate the model (the warm-up time is included). The model imitates the real production and so it is good to use for the next analysis to find out whether the company is able to meet higher demand.

The next task is to check out the increase of the production. Today's shields production makes only 30% of the plan future expansion so production should be increased significantly. Whereas the first model supposed only one shift production in case of higher demand machines have to run for three shifts. The next model should show if it is necessary to use another machines and how and where to use it so as to be maximally efficient. The production expansion relates not only with the existing shields but also 4 new shield types should be produces, so instead of 6 types the company should produce 10.

The new model differs in number of EMAG machines – one machine should be used to produce three types of the shields. Also number of degreasers is four instead of three but both machines can be operated by one worker.

Except the added machines also other parameters have had to be changed such as the inter-arrival times of the material and percentage split among the shaped machines. Because of the different machine times of the drilling, screwing, deburring and degreasing some shields are processed on the same machines and so its routings in the simulation model merge and split up. Also material is split up according to the size of the shield types. The last change of the model has been the extension of the simulation time and warm-up period as instead of one shift production three shifts are used. The new model is on the Figure 7.

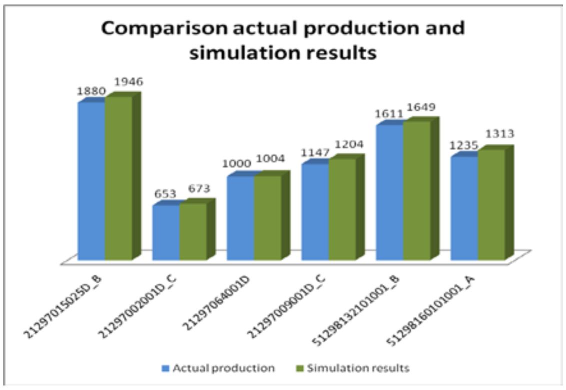


Figure 6: Comparison actual production and simulation results

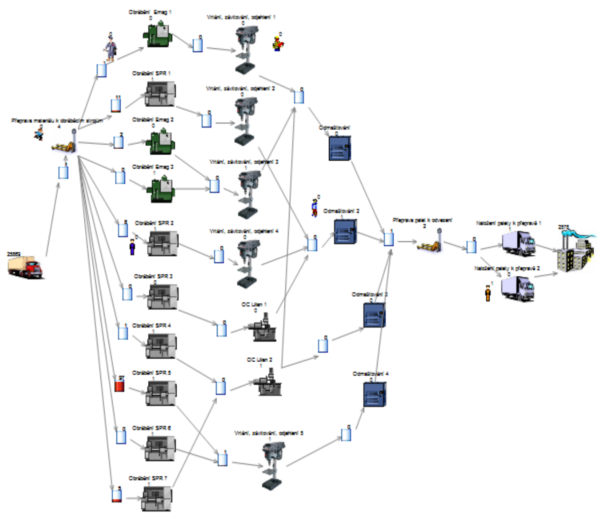


Figure 7: Extended simulation model

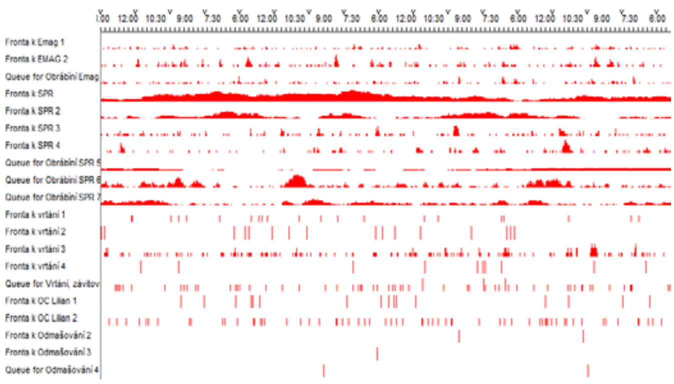


Figure 8: The queue size at a particular machine

Figure 8 shows the length of the queues in the production process after the production expansion. We can see that there are no serious problems except of the shaped machine SPR (the number of shields in a queue are described by connected red field). For the continuous production the maximum queue size is important as it is hard to store goods in process and it can stay in a way and complicate the movement in the hall.

Table 3 shows the length of the queues and the queuing time for each work center. We can see again that there are only two problematic places – shaped machines SPR 1 and SPR 5.

Table 3: Maximal and average queue size and time

Queue to Work Center	Average queue size (pcs)	Maximal queue size (pcs)	Average queue time (mins)	Maximal queue time (mins)
Transport Material	0,6	6,67	0,65	6,59
EMAG 1	0,32	6,17	3,03	130,61
EMAG 2	0,3	6,17	3,32	121,89
EMAG 3	0,36	6,83	3,39	41,26
SPR 1	11,65	33	116,65	8540,95
SPR 2	6,93	31,67	70,33	303,74
SPR 3	0,38	6,17	4,79	41,44
SPR 4	0,32	5,67	8,1	68,49
SPR 5	29,22	67,83	291,14	693,18
SPR 6	1,29	10	11,97	66,92
SPR 7	3,38	22	35,74	223,06
Drilling 1	0,03	1,83	0,28	2,85
Drilling 2	0,02	1	0,31	3,14
Drilling 3	0,29	4,17	1,36	13,83
Drilling 4	0,03	1	0,29	3,17
Drilling 5	0,08	3	0,39	4,65
OC Lilian 1	0,02	2	0,28	4,14
OC Lilian 2	0,11	2	0,78	7,58
Degreasing 1	0,01	2	0,02	0,81
Degreasing 2	0	2	0,01	0,78
Degreasing 3	0	1	0,01	0,27
Degreasing 4	0	1	0,01	0,64
Loading pallets to drop-off	0	2,33	0	2,42

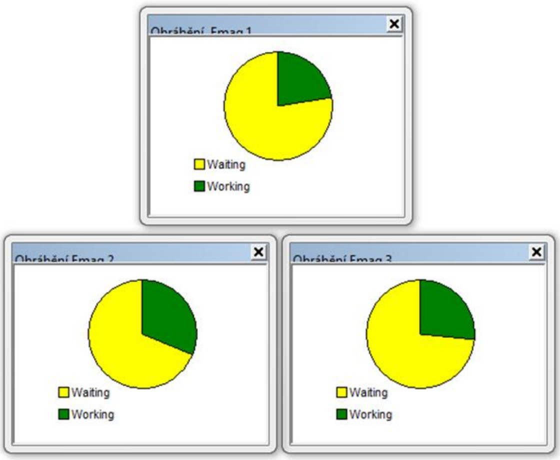


Figure 9: Utilization of shaping machine EMAG

It is important to check the utility of each work center and worker. Usually we can display the graph of results to see the utility of the machine but as it has been mentioned above model contains more work centers than number of real machines so we have to summarize utilities of more work centers. It can happen that that the total utility is higher than 100% and the simulation model should be changed. Figure 9 shows the time when the shaping center is working and when is awaiting for material. Although it may seems the machine works only a small part of the shift, the total utility is 80%.

We have checked also the utility of the next work centers, the results are in Table 4.

Table 4: Cumulative utilization of work centers

Work Center	Utilization
Working EMAG 1, 2 a 3	80,00%
OC Lilian	71,49%
Degreasing	13,72%
Drill	75,82%

Except for the degreasing machine the utility of machines is quite high. Similarly the utilization of workers is shown in Table 5.

Table 5: Utilization of employees

Work position	Worker utilization
Transportion worker	95,53%
Drilling operator	75,82%
SPR operator	96,91%
EMAG operator	80,00%
Degreasing operator	13,72%

The utility for the EMAG operator is the same as the utility of the EMAG machine itself – it is because both machines are operated by one worker on each shift.

We have made sure of the fact that the total utility is not higher than 100%. Now we have to verify that the condition of the number of shields made has been met.

As in the first model this simulation model perfectly imitates target values (Figure 9).

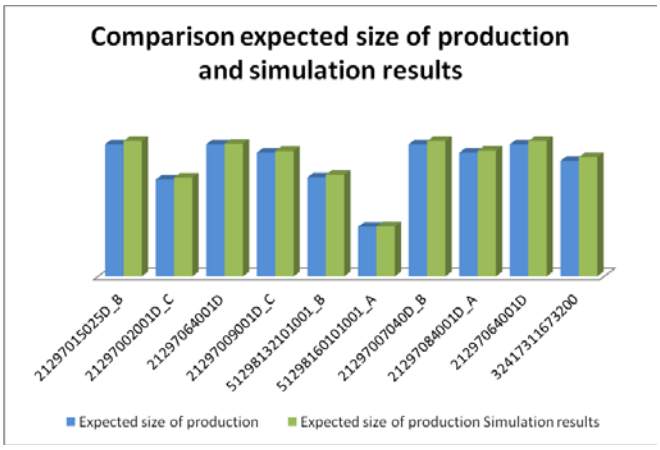


Figure 10: Comparison expected size of production and simulation results

Table 6 shows the expected production size and the results obtain from the simulation. It is clear that the all customers' requirements will be satisfied without problems also in case of higher waste.

Table 6: Comparing expected values and simulation results

Type of shield	Expected size of production	Simulation results
21297015025D_B	2 933	3 008
21297002001D_C	2 150	2 195
21297064001D	2 933	2 949
21297009001D_C	2 750	2 786
51298132101001_B	2 200	2 257
5129816010101_A	1 100	1 108
21297007040D_B	2 933	3 007
21297084001D_A	2 750	2 791
21297064001D	2 933	3 006
32417311673200	2 567	2 652

All the changes of the number of machines have been discussed with the management of the company and with the workers that use these machines so as not to cause any problems in future changes. We have tested also different number of machines and try to find the minimal number for the demand satisfaction but with acceptable machine utility. The best results (as we can see as an equilibrium) has been described here.

4. RECOMMENDATIONS

All simulation results show us that the company is able to satisfy the higher order (that has been derived from the new contract offered). In all production system there should not be longer queues and we expect that also the waiting times for the cutting, drilling, degreasing or other production activities will be small and acceptable. On the base of the simulation model the production process has been analyzed in detail and we have discovered some placed where the number of machines or workers can be decreased or somehow change the production to be more efficient.

The shaping and cutting center EMAG which is utilized on 80 % (Table 5) can produce three types of shifts numbering about 8964 pieces. This machine is very quick and precise compared to the other shaping machines SPR and its spoilage is nearly zero. If we compare the production of nine SPR machines for the same time we can see that they are able to made 16 795 pieces on average with similar utility. Average shaping time on EMAG is about 2,5 minute whereas on SPR machine is nearly 4 times higher. In case of buying new EMAG VL 7 center the time savings are very high. The next advantage can be the fact that only one worker can operate this new machine during each shift and so it can generate more money savings for the company. On the other hand the price of the machine is very high but could be decrease via selling the old and redundant SPR machines. This change could save the space in the hall and this is also very important as there is never enough space.

Thanks to the possibility of the graphical representation of the machines utility it is clear how much of the labor time the machines are working and how much time they are waiting for the next product. In the simulation model seven work centers for cutting, deburring and shaping are modeled. Two of them are the shaping centers Lilian where the utility is pretty good. The next five centers are drilling machines (in reality they are only three but according of the conditions of production it was necessary to model it by five centers). We can see on the Figure 11 their utility and the total sum is equal to 156%. It means that it is possible to use only two drilling machines instead of three and use the remaining machine for the different order.

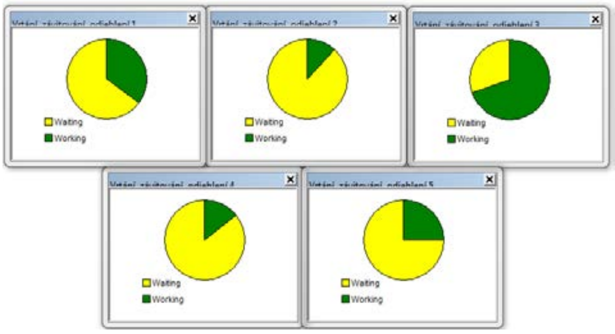


Figure 11: Utilization of drills

Degreasing is the last place where we can reduce production from three shifts to one. Because of the big batch degreasing at a time the utility of this workplace is very low (Table 4), only 14%. That is why the degreaser can work every third shift. Shields storing for a period of two shifts is acceptable for the company and the shields will not complicate the movement in the production hall. Some containers or palettes can be used nearby the degreaser to store the products. Moreover the degreaser can be used for another production and that is why it is possible to decrease the number of shifts (especially during the night because of extra charges that can lower the cost) of the following operations as pallet packaging and transport of the final products.

All parts and changes of the model has been discussed with the management of the company and we have also communicated with workers when measuring the working times. The changes meet the requirements of the new client and its new order (the higher production) but also it meets the requirements of the company (change the machines and number of machines so as to increase the utilization but acceptable for workers). From this point of view the equilibrium has been found.

5. CONCLUSIONS

The aim of this article is to find the equilibrium for the shields production system to produce ordered pieces with optimal number of machines. First we describe the creation of simulation model of the shield production process made in the software SIMUL8. The model corresponds with reality and so it has been possible to do some changes especially analyze the effects of higher demand. According to the simulation results some recommendations for the company relating workers and machines has been made.

REFERENCES

Aguirre, A., and Méndez, C. A., 2008: Applying a Simulation-based Tool to Productivity Management in an Automotive-parts Industry. In: *Proceedings of the 2008 Winter Simulation Conference* (Mason, S. J., Hill, R. R., Mönch, L., Rose, O., Jefferson, T., and Fowler, J. W., eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2005, p. 1838-1846.

Banks, J., 1998. *Handbook of Simulation*. USA, John Willey & Sons

Concannon, K. et al. 2007. *Simulation Modeling with SIMUL8*. Visual Thinking International, Canada 2007.

Law, A., 2000. *Simulation Modelling and Analysis. Boston (USA)*, MC-Graw Hill

Masood, S., 2006. Line balancing and simulation of an automated production transfer line, *Assembly Automation*, Vol. 26 Iss: 1, p. 69 – 74.

Montevechi, J. A. B., et al., 2007. Application of design of experiments on the simulation of a process in an automotive industry. *WSC'07 Proceedings of the 39th Conference on Winter Simulation*, IEEE Press Piscataway, NJ, USA 2007, p. 1601-1609.

O'Kane, J.F. et al., 2000. Simulation as an essential tool for advanced manufacturing technology problems. *Journal of Materials Processing Technology* 107/2000, p.412-424

Shalliker, J. and Ricketts C.: *An Introduction to SIMUL8, Release nine*. School of Mathematics and Statistics, University of Plymouth, 2002.

SIMUL8. 2012 SIMUL8 WEBMASTER. *Process Improvement with SIMUL8 Simulation Software*. Available from: http://www.simul8.com/about/ [accessed 10 April 2013]

SIMUL8-Clanky a pripadove studie 2013. Available from: http://www.simul8.cz/clanky-a-pripadove-studie [accessed 1 Jully 2013]

VLTAVA2000. 2012 In: Felicia, Available from: http://www.vltava2000.cz/interskoda/goods-047903015J-20-alternator-felicia-13-cz-047903015j-047903015j.html [accessed 13 April 2013]

ACKNOWLEDGEMENT

The research project was partly supported by Grant No. IG403033, Faculty of Informatics and Statistics, University of Economics, Prague.

AUTHORS BIOGRAPHY

Pavla Ficova: She is a final year student at the University of Economics in Prague. She obtained bachelor's degree in field Mathematical Methods in Economics in 2010. As a master's degree she has chosen branch of Econometrics and Operations Research, where she successfully passed the state exam in September 2012. In her study she is interested in the simulation models creation and her final thesis is aimed at this area. In 2010, she began working as a CRM / direct marketing specialist at Provident Financial. From July 2012 to the present she has worked at Česká spořitelna (important member of Erste Group) as a marketing research specialist.

Martina Kuncova: She has got her degree at the University of Economics Prague, at the branch of study Econometrics and Operational Research (1999). In 2009 she has finished her doctoral study at the University of West Bohemia in Pilsen (Economics and Management). Since the year 2000 she has been working at the Department of Econometrics, University of Economics Prague, since 2007 also at the Department of Economic Studies of the College of Polytechnics Jihlava (since 2012 as a head of the department). She is a member of the Czech Society of Operational Research, she participates in the solving of the grants of the Grant Agency of the Czech Republic, she is the coauthor of three books and the author of many scientific papers and contributions at conferences. She is interested in the usage of the operational research, simulation methods and methods of multi-criteria decision-making in reality.

A PARALLEL APPROACH OF INTEREST MANAGEMENT IN EXASCALE SIMULATION SYSTEMS

Elvis S. Liu

University College Dublin and IBM Research, Ireland

elvisliu@ie.ibm.com

ABSTRACT
Interest management in parallel/distributed simulation is a filtering technique which is designed to reduce bandwidth consumption of data communication and therefore enhances the scalability of the system. This technique usually involves a process called "interest matching", which determines what data should be sent to the participants as well as what data should be filtered. However, existing interest matching algorithms are mainly designed for serial processing which is supposed to be run on a single processor. As the problem size grows, these algorithms may not be scalable since the single processor may eventually become a bottleneck. In this research, a parallel approach of interest matching is developed, which is suitable to apply on exascale parallel/distributed simulation systems.

Keywords: Simulations, Gaming, Systems

1. INTRODUCTION
In the coming years we expect to reach a computational power equivalent to a thousandfold that of the current most powerful supercomputer (exascale systems). Computational advances have opened the way for a growing number of computer simulation applications across many fields. As the scale grows, providing scalable data distribution through interest management becomes one of the major design requirements of large-scale simulation systems. The basic idea of interest management is simple: all participants should only receive data that are of interest to them. This data filtering process, however, may introduce considerable computational overhead. If the cost of interest management is too high, it would degrade the overall performance of the simulation. Over the years, numerous interest management schemes have been proposed which sought to reduce the computational overhead and, at the same time, to maintain the high precision of data filtering. These schemes, however, are designed for serial processing which is supposed to be run on a single processor. As the problem size grows, these algorithms may not be scalable since the single processor may eventually become a bottleneck. Furthermore, large-scale simulations are executed on parallel/distributed high-performance computing systems, deploying the existing schemes on these systems would be unsuitable, and the performance cannot be guaranteed.

In this research, a parallel approach of interest matching algorithm is being developed, which is suitable for deploying on exascale parallel/distributed simulation systems. The new algorithm enables the multiple processors to work simultaneously and thus enhances the overall runtime efficiency of the matching process.

2. BACKGROUND AND RELATED WORK
Aura-based interest management (e.g., (Greenhalgh and Benford 1995)) use auras to represent the interests of each participant. When auras overlap, a connection between the owners of the auras is established and messages are exchanged through the connection. This approach provides a much more precise message filtering mechanism than the zone-based approaches (e.g., (Macedonia, Zyda, Pratt, Brutzman, and Barham 1995)); however, more computational effort is required for testing the overlap status for the auras. DIVE (Carlsson and Hagsand 1993) and MASSIVE (Greenhalgh and Benford 1995) are the early DVE systems that adopt this type of schemes. In the High-Level Architecture (HLA) (DMSO 1998), the Data Distribution Management (DDM) services allow the participants to specify "update regions" and "subscription regions" to represent their interests. These regions are similar to the auras except they must be rectangular and axis-aligned when using in a two- or three-dimensional space.

The interest matching algorithms are designed to solve the "trade-off" between runtime efficiency and filtering precision for aura-based interest management. They have usually been applied on high precision filtering schemes, such as HLA DDM, which ensures the participants receive the minimal set of data that are of interest to them. In addition, they provide a way to efficiently reduce the computational overhead of the matching process.

In an early paper (Van Hook, Rak, and Calvin 1994), Van Hook et al. pointed out that the matching process of the aura-based approach (referred to as "object-based approach" in the paper) could be computationally intensive. To solve this problem, Van Hook et al. proposed a crude grid-based filtering approach to cull out many irrelevant entities before a more compute-intensive procedure is carried out for finer discrimination. (Morgan, Storey, and Lu 2004) proposed a collision detection algorithm for aura-based interest matching. The algorithm uses aura overlap for determining spatial subdivision. The authors argued that it more accurately reflects the groupings of entities that may be interacting than existing collision detection algorithms and provided performance figures to demonstrate its scalability. Recently, more robust matching algorithms (Raczy, Tan, and Yu 2005; Liu, Yip, and Yu 2005; Pan, Turner, Cai, and Li 2007) based on dimension reduction were proposed. These algorithms are designed specifically for HLA-compliant systems, and thus adopt the use of rectangular auras (i.e., regions of the HLA). The basic idea of dimension reduction is to reduce the multidimensional overlap test to a one-dimensional problem, which is more computationally efficient than the original problem.

3. PARALLEL INTEREST MATCHING

This section describes a parallel interest matching algorithm which facilitates parallelism by distributing the workload of the matching process across shared-memory multiprocessors. The algorithm divides the matching process into two phases. In the first phase it employs a spatial data structure called uniform subdivision to efficiently decompose the virtual space into a number of subdivisions. We define as work unit (WU) the interest matching process within a space subdivision. In the second phase, WUs are distributed across different processors and can be processed concurrently.

For the sake of consistency, aura is hereafter referred to as "regions" as per the terminology of HLA DDM.

3.1. First Phase: Hashing

Uniform subdivision is a common spatial data structure which has long been used as a mean of rapid retrieval of geometric information. Over the years, it has been studied extensively in many fields such as computer graphics and robotics. The idea of using hashing for subdivision directory was first described in an early article written by Rabin (Rabin 1976) and was later discussed more generally in Bentley and Friedman's survey (Bentley and Friedman 1979).

During the simulation, regions are hashed into the hash table. The algorithm uses the coordinate of a region's vertex as a hash key. Given a key k, a hash value $H(k)$ is computed, where $H()$ is the hash function. The hash value is an n-dimensional index which can be matched with the index of a space subdivision, and therefore indicating that which subdivision the vertex lies in. Hence, the regions with hash key k are stored in slot $H(k)$. The hash function is given in **Definition 1**.

Definition 1. Let $[SMIN_d, SMAX_d)$ be the boundary of a space in d dimension, for $d = 1, 2, ..., n$. The boundary is uniformly divided into N_d sub-boundaries with unit length L_d. The hash function for transforming a key k_d into a hash value is defined as

$$H : \mathbb{R}^n \to \mathbb{Z}^n, H(k_d) = \lfloor \frac{k_d - SMIN_d}{L_d} \rfloor$$

There are two important properties of using a hash table for spatial decomposition. First, hash table collision means that regions in the same slot are potentially overlapped with each other; therefore, further investigation on their overlap status is required. This process will be left to the second phase of the algorithm. Second, if a region lies in multiple space subdivisions, it would be hashed into all of them. The algorithm assumes that the size of region is much smaller than a space subdivision. Therefore, a region would exist in at most four slots in the two-dimensional space (at most eight slots in the three-dimensional space). This assumption ensures that the computational complexity of the hashing process would be bounded by a constant.

Figure 1 illustrates the basic concept of the spatial hashing for two-dimensional space. In the figure, region A is hashed into slot (0,1); region B is hashed into slots (0,0), (0,1), (1,0) and (1,1); region C is hashed into slots (1,1) and (1,2); region D is hashed into (1,0), (1,1), (2,0) and (2,1). Note that if not all vertices of a region are hashed into the same slot, then the region exists in multiple subdivisions.

The hash table is constructed at the initialisation stage. During runtime, the position and size of regions may be frequently modified. Therefore, the algorithm needs to perform rehashing for the regions at every time-step. The complexity of this process is $O(n+m)$ where m is the number of subscription regions and n is the number of update regions.

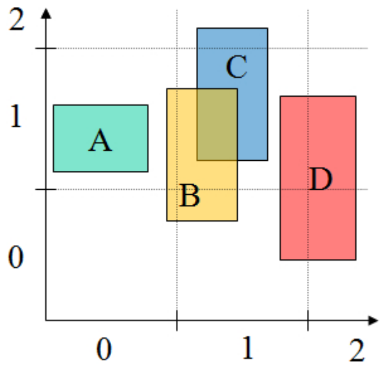


Figure 1: Hashing for Space Subdivisions

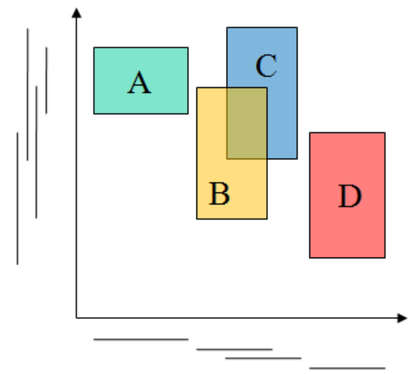


Figure 2: Dimension Reduction

3.2. Second Phase: Sorting

After the hashing stage, each slot of the hash table represents a WU which will be distributed across different processors. The algorithm then places the WUs on a task queue. Each processor fetches WUs from the queue and performs interest matching for the corresponding space subdivisions. Since only one processor has the authority to manage each space subdivision, there will be no ambiguous matching result.

The spatial decomposition approach essentially transforms the large-scale interest matching process into several individual sub-problems. When a WU is being processed, each processor carries out a matching process only for the regions within the WU. Since using a brute-force approach to determine the overlapping status of the regions would be time consuming, a sorting algorithm based on dimension reduction would help to increase the computational efficiency. The preliminary design of dimension reduction is presented in (Liu, Yip, and Yu 2005). It reduces the multidimensional overlap test to a one-dimensional problem, which is defined as follows:

Two regions overlap in n-dimensional space if and only if their orthogonal projections[1] *on the* 1^{st}, 2^{nd}, *..., and* n^{th} *dimension overlap.*

Figure 2 shows how the concept of dimension reduction works in two-dimensional space. In the figure, B-C overlap on x-axis; A-C, A-B, B-C, B-D, and C-D overlap on y-axis; hence, B-C overlap on two-dimensional space.

As discussed in (Dandamudi and Cheng 1995), the task queue approach is desired for task distribution and provides very good load sharing for shared-memory multiprocessor systems. When a processor finishes processing a WU, it would fetch another WU from the task queue immediately unless the queue is empty. Therefore, no processor would be idle until all WUs are fetched. The worst case happens only when all regions reside in a single space subdivision. In this situation, a single processor would be responsible for the matching of all of them.

[1]In the terminology of HLA, the orthogonal projection of a region is called "extent".

4. SUMMARY AND FUTURE WORK

Over the last few years, several matching algorithms were proposed to speed up the interest matching process. However, these efforts have focused on sequential algorithms. As the problem size grows, using these algorithms does not satisfy the scalability requirement of DVE since the single processor may eventually become a bottleneck. An approach to alleviate this problem is to exploit the inherent parallelism of the matching process and the high availability of parallel computation infrastructures. In this research, a parallel matching algorithm for exascale simulation systems is developed. The proposed algorithm can be run on a cluster of computers that enables them to work simultaneously and thus enhances the overall runtime efficiency of the matching process.

The future work will concentrate on performing experimental comparisons of the runtime efficiency of the proposed algorithm and the parallel algorithm presented in (Liu and Theodoropoulos 2009). We will also test and compare the performance the proposed algorithm under different entity behaviors, number of nodes, and occupation density.

ACKNOWLEDGEMENT

This work was supported, in part, by Lero - the Irish Software Engineering Research Centre (www.lero.ie).

REFERENCES

Bentley, J. L. and Friedman, J. H., 1979 . Data Structures for Range Searching. *ACM Computing Survey*, 11:397–409.

Carlsson, C. and Hagsand, O. (1993). DIVE - A platform for multi-user virtual environments. *Computers & Graphics*, 17:663–669.

Dandamudi, S. P. and Cheng, P. S. P. (1995). A hierarchical task queue organization for shared-memory multiprocessor systems. *IEEE Trans. Parallel Distrib. Syst.*, 6(1):1–16.

DMSO (1998). High Level Architecture Interface Specification Version 1.3.

Greenhalgh, C. and Benford, S. (1995). Massive: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer Human Interactions*, 2(3):239–261.

Liu, E. S. and Theodoropoulos, G. K. (2009). An Approach for Parallel Interest Matching in Distributed Virtual Environments. In *Proceedings of the 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications (DS-RT 2009)*, Singapore.

Liu, E. S., Yip, M. K., and Yu, G. (2005). Scalable Interest Management for Multidimensional Routing Space. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST) 2005*, pages 82–85.

Macedonia, M. R., Zyda, M. J., Pratt, D. R., Brutzman, D. P., and Barham, P. T. (1995). Exploiting Reality with Multicast Groups: A Network Architecture for Large-scale Virtual Environments. In *Proceedings of the IEEE Virtual Reality Annual International Symposium (VRAIS) '95*, pages 2–10.

Morgan, G., Storey, K., and Lu, F. (2004). Expanding Spheres: A Collision Detection Algorithm for Interest Management in Networked Games. In *Proceedings of the Entertainment Computing - ICEC 2004*.

Pan, K., Turner, S. J., Cai, W., and Li, Z. (2007). An Efficient Sort-Based DDM Matching Algorithm for HLA Applications with a Large Spatial Environment. In *Proceedings of the 21st International Workshop on Principles of Advanced and Distributed Simulation (PADS 2007)*, pages 70–82, Washington, DC, USA.

Rabin, M., 1976 . Probabilistic algorithms. In *Algorithms and Complexity: New Direction and Recent Results*, pages 21–39. Academic Press Inc, New York.

Raczy, C., Tan, G., and Yu, J. (2005). A sort-based DDM matching algorithm for HLA. *ACM Transactions on Modeling and Computer Simulation*, 15(1):14–38.

Van Hook, D. J., Rak, S. J., and Calvin, J. O. (1994). Approaches to relevance filtering. In *Proceedings of the Eleventh Workshop on Standards for the Interoperability of Distributed Simulations*, pages 26–30.

AUTHOR BIOGRAPHY

Dr. Elvis S. Liu is a Research Fellow of Irish Research Council and IBM Research, Ireland. His current research is focused on simulation technologies for Exascale high-performance computing systems. He received a B.Sc. from the University of Hong Kong, an M.Sc. from the University of Newcastle (UK), and a Ph.D. from the University of Birmingham (UK). His email address is `<elvisliu@ie.ibm.com>`.

SIMUSE: MODELING RECREATIONAL POLYDRUG USE THROUGH AN AGENT-BASED MODEL

François Lamy[(a)], Terry Bossomaier[(b)], Pascal Perez[(c)]

(a) CRiCS (CSU), Clersé (USTL), HEMA Consulting/DPMP
(b) CRiCS (CSU)
(c) SMART Infrastructure Facility (UoW), HEMA Consulting/DPMP

[(a)] flamy1978@gmail.com, [(b)] terry.bossomaier@csu.edu.au, [(c)] pascal.perez@uow.edu.au

ABSTRACT
This paper describes an attempt to capture individual and social characteristics of poly-drug use in a generic agent-based model called SimUse. We consider poly-drug use and its social context as a complex phenomenon and use a generative framework to create an iterative dialogue between qualitative fieldwork, theoretical constructs and computer simulations. The structure of SimUse includes five levels of influence. In a first time the context of recreational polydrug use and the rationales of this research is introduced; then, the second part of the paper describes the overall structure of the model before detailing key aspects of SimUse: (1) the neurological engine and its behavioural consequences; (2) the decision process and its different elements; and (3) the intra- and extra-individual reevaluation processes. We conclude with two examples of how the model reacts to illustrative parameter changes and external shocks and their consequences for decision-makers.

Keywords: drug use, agent-based model, social simulation, sociology of deviance

1. INTRODUCTION

Drug use is a major concern of western modern societies: the latest reports from the United Nations Office on Drugs and Crime (UNODC) estimate the number of problematic and injecting users around 27-59 million individuals and occasional consumers to 155-250 million (UNODC 2011). Furthermore, drug trafficking and consumption trends are subject to frequent and rapid evolutions (UNODC 2011; EMCDDA 2013). Indeed, for approximately twenty years, the drug market context has been characterized by the endemic presence of classic illicit drugs (cannabis, cocaine, ecstasy, heroin, amphetamine-type and numerous hallucinogens) (Faugeron & Kokoreff 2002), associated with the constant appearance of new psychoactive substances (known as "designer drugs" or "legal high"), and the augmentation of pharmaceutical substance misuse (EMCDDA 2013). This particular context, combined with the 'normalization' of drug use (Parker et al. 1998; Parker 2005) has favored polysubstance use (Fontaine et al. 2001), practice leading to the consumption of at least two psychoactive substances, sometimes concurrently. According to the EMCDDA (European Monitoring Centre for Drugs and Drug Addiction), polysubstance use is the actual "dominant pattern of drug use" and appears as a major social issue due to increased hazard risks and health-related harms (EMCCDA 2009).

According to these previous reports, drug use appears to be a complex social problem that needs to be seen through a multi-disciplinary prism. International institutions have called for developments of technology able to encapsulate such a dynamic complex phenomenon and hence, being able to evaluate the relevance and accuracy of public policies relative to this matter (EMCCDA 2009). To tackle this social issue and capture the complexity of polydrug use, we propose to create an artificial society via an agent-based model, SimUse, to run *in-silico* social simulations. This model purpose is twofold: on the one hand, it is used as a *mediator* framework, enabling the dialogue between several disciplines; and on the other hand, as a *predict*ive tool for public policy makers. Indeed, SimUse, attempts to encompass several levels of understanding in order to create an ontology of recreational polydrug use.

This paper is organized as follows: Section II legitimizes our approach by introducing key-concepts from the scientific literature concerning drug use and addiction. We then refer to social simulations built on the topic in order to reinforce our stance. The second part (III) of the paper presents three levels of modeling in SimUse: (1) the decision process concerning the choice of substances; (2) the neurological engine representing the behavioural and physiological responses to the consumption of the chosen substances; and (3), the intra- and extra-individual re-evalutation process following consumption. The last section (IV) presents two *what-if* scenarios illustrating the type of results that SimUse can produce.

2. DRUG USE AS A COMPLEX ADAPTIVE SYSTEM

The literature on substance use, misuse and addiction reveals that drug consumption results from a large set of risk/protective factors influencing individuals in their choices to consume drug(s). West (2006) indicates that every discipline, from genetics to economy has conceptual tools and operative theories to study and explain drug use and abuse. He also notes that these disciplines do not interact with each other, leading to the situation where there are "many theories but little progress" (West 2006). Based on this point-of-view, several researchers have called for a multidisciplinary approach to capture these risk/protective factors and understand their interactions (Unger et al. 2004). The review of the scientific literature establishes five main levels of analysis, starting from neurology and finishing with symbolic macrostructure. These levels are named here: *drug, intrapersonal, interpersonal, context,* and *symbolic.*

At the *drug* level, the neurosciences have highlighted the crucial role of *neurotransmitters* in mechanisms of pleasure, memory and mood changes (Koob and LeMoal 2006). Because each drug has a specific impact on the brain, we have to take in consideration these neuropharmacological differences to understand changes in user behaviours. Add to this preceding point, repeated and frequent intakes modify brain structures due to synaptic adaptation that can lead to alteration of both physiological and psychological state (Julien et al. 2008).
At the *intrapersonal* level, the beliefs and meanings attached by individuals to substances is conditioned by the set of *representations* constructed by these individuals through interactions and past experiences (Jodelet 2003). Therefore, user's beliefs and experiences about drugs influence and modify the decision process after each drug use' iteration. It also appears that users expect specific effects from psychoactive substances and infer these latter *functions* (Boys et al 1999). Moreover, each individual has a set of *capitals* (economic, symbolic and social) that affects his/her ability to find, afford and choose specific type of drugs (Boys and Mardsen 2003).

The *interpersonal* stratum should be understood as the level of interactions. Social learning theory indicates that individuals tend to mimic and incorporate behaviours they have witnessed. Concerning polysubstance use, this theory underscores the importance of *socialization* on drug social representations and opinions: parental drug consumption, *peer pressure,* and *peer influence* have been widely studied and appear to increase the risk of acute substance abuse. In the same way, belonging and identifying oneself with a drug user's group can induce a consumption reflecting group patterns (Sussman and Ames 2008). Furthermore, friends and acquaintances are generally the first sources of drug supply and are considered by neophytes as "safe keepers" insuring the safety of initiations.

On a *contextual* level, the intra- and interpersonal levels could be impacted by neighborhood conditions, economic deprivation, geographical relegation, economical and/or social inequalities (Rhodes et al. 2001). Obviously, *drug market structure* varies from one geographical area to another and facilitates or not the accessibility to certain drugs (Johnson et al. 1992). Consistent with these last two points, *geographical contexts* (specific suburbs, rural/urban areas, etc) give access to a more or less big panel of drugs. Some kinds of consumption can only take place in specific social context (Preble and Casey 1969).

Finally, the *societal* level condenses the legal and symbolic dimensions influencing the choices of drug users. *Legislation* and *global availability* define the ease of access to the licit and illicit drugs and, in turn, influence the price and penal risks of each drug (Sussman and Ames 2004). In a context of drug normalization (Parker 2005), *mass media, norms* and *social acceptance* play a major role on the beliefs of both users and non-users. Indeed, the repeated exposures to advertisements modifies preference and conduct (Theus 1994), movies or TV series could product a positive image of deviant behaviour (Villani 2005) and social goals such as cult of the performance or reconnaissance by wealth (Ehrenberg 1991; Simmel 1900) affect both consumer decisions and acts.

Polyuse accentuates the complexity of this social phenomenon. Indeed, most of the studies concerning polysubstance consumption are focused either on *simultaneous* polydrug use (SPU) as a social practice common to particular subpopulations (especially related to nightlife and rave groups) or on the adverse health effects of *concurrent* (life-based) polysubstance use (CPU) (Ives and Ghelani 2006). This research suggests that SPU and CPU are interdependent. This assertion is based on the fact that experiences arising from any drug(s) session could impact and transform representations attached by individuals to substances. These representations vary throughout the *career*, understood as the consolidated biographical experiences of polyusers, and orient the acts, which constitute the basis that orients further decisions related to substances use.

Therefore, the present research has studied these interconnected forms of polyuse by combining concepts coming from neuroscience (to capture the behavioural changes during SPU) to findings produces by a sociological investigation (to apprehend the changes that occurs throughout CPU). The former is informed by the literature on the subject (Solomon 1980; Koob and LeMoal 2006; Julien et al. 2008) and the latter by qualitative interviews conducted during fieldwork.

Considering the complexity and dynamics of drug use, several researchers have already proposed to study drug use as a *Complex Adaptive Systems* (Gorman et al. 2004; Perez et al. 2005) and to model this phenomenon through agent-based social simulations. Agar and Wilson created the first of these simulations. *SimTalk* was designed to capture the communication process existing between heroin users based on an ethnographic investigation amongst heroin injectors in Baltimore (Agar and Wilson 2004). Chattoe, Hickman & Vickerman have complexified *SimTalk* by inserting *non-users* agents inside the model: their model, *DrugChat,* introduces the role of network peers in agent's decisions (the communication in *SimTalk* was based on the spatial proximity). On the topic of heroin, Perez et al (2005) have built an agent-based model, *SimDrug,* to replicate the Melbourne heroin drought of 2000-2001. Agents in SimDrug are integrated into an informed spatial environment and other classes of agents (i.e. constables, dealer, outreach workers, wholesalers) were created to reproduce the social environment in which heroin users normally evolve into. Perez and colleagues (2012) have also developed an agent-based model concerning the consumption of amphetamine amongst young Australians, based on quantitative and ethnographic material. In *SimAmph,* the decisions of agents were shaped to acknowledge the key roles of individual perception, peers influence, and subcultural settings. Gorman and colleagues (2006) have produced a simulation on alcohol consumption in general population. Their model aimed to analyze the role of agent-environment interactions in the development and the continuation of alcohol use. Simulations have also been designed to study drug distribution markets (Agar and Wilson 2002; Romano, Lomax, and Richmond 2005; Hoffer, Bobashev, and Morris 2009).

However, most of these models focused on a single substance. Given the fact that our objective is to recreate the career of recreational polydrug users, by taking into account the five levels of analysis previously described, we needed to introduce several new features into our model. The next section details some of the novel components.

3. BUILDING SIMUSE: A MULTILAYER MODEL

SimUse has been developed in NetLogo 4.1.3 (Wilensky 1999). Social simulations enable us to carry out artificial social experiments to investigate the consequences of pre-defined conditions on a range of specific social and environmental conditions. SimUse investigates how agents take their decisions, consume recreational drugs, interact with other users and (a) evaluate their own actions and/or (b) judge the behaviours of other users. By doing so, SimUse aims to assess the impact of these drug practices on the social life of recreational consumers.

Agents representing recreational polydrug users are characterized by neurological, behavioural and social attributes. These attributes generate the agents' choices, actions and interactions, informed by qualitative interviews findings. These agents act in a preprogrammed routine inside a drastic simplification of an urban environment aggregating specific settings (Bar, Club, Bottle-Shop, etc). Other types of agents were included in the simulation (i.e., drug dealers, wholesalers and law enforcement) to recreate the context the in which the polyusers evolve.

From prior work and analysis of the interviews, carried out in Australia and France by the first author, it appears that the decision process regarding drug use is comparable to a "practical reasoning" (Bratman 1987). Indeed, the interviews reveal, consistently with other research (Boys et al. 1999; Boys and Mardsen 2003), that polydrug users have expectations regarding their consumption and infer "functions" and roles to the different substances they consume. These "functions" could be regrouped into four meaningful categories, namely "Sociable", "Relax", "Energy", and "Intoxicate". In brief, drugs providing the "Sociable" function are considered by interviewees as facilitating the communication with others and increase "fun" with peers. Substances with the "Relax" function attached to them are used for their analgesic or sedative properties and to establish a boundary between working and leisure time. "Energy" drugs are consumed for their stimulant effects allowing their users to stay awake longer and boost their physical capacities. Users that target the "Intoxicate" function generally cited drugs that produce intense rushes, hallucinations or analgesia.

Comparing findings from the qualitative interviews with the neuropharmacological consequences of each drug suggests that the substance choices of recreational polyusers are consistent with the functions they target. In other words, psychoactive substances, through their actions on the different neurotransmitter systems, are means employed by polyusers to obtain particular physiological and/or psychological effects and, in turn, achieve social-oriented functions. Considering the importance of the neurological properties of each substance on user decisions, SimUse needed to include a neurological component. This "NeuralBox" functions as an algorithm, where drugs are the inputs and behaviours are outputs. The translation between "drugs" to "behaviours" is taken care of by a set of modeled neurotransmitters. By drilling down to the neurotransmitter level, this model which we call the *neurological engine,* permits treating several substances at the same time and, therefore, to mimic polydrug use.

Given the neuropharmacology of the drugs most currently used, we decided to model eight neurotransmitters:

- **Dopamine** is involved in feelings of reward, self-confidence, talkativeness and happiness (Arias-Carrion and Pöppel 2007). It is also considered as the main cause of addiction (Wise 2002);
- **Cannabinoid** is a neuroregulator inhibiting the release of other neurotransmitters. It induces an analgesic effect, sensation of well-being, decreases in body temperature and potentiates opioid effects (Ashton 2001);
- **Opioid Peptides** generate analgesia and depression of the respiratory functions (Santiago and Edelman 1985);

- **Gamma-Amino-Butyric Acid (GABA)** is the principal inhibitory neurotransmitter in the brain reducing and regulating the activity of other neurons and neurotransmitters (Kuffler and Edwards 1957);
- **Glutamate** is the main excitatory neurotransmitter. It is involved in all aspect of brain function, including movement, language, learning, and memorization (Riedel 1996; Riedel, Platt, and Micheau 2002);
- **Norepinephrine** produces a host of changes including increasing arousal and attention, increasing body temperature, motor activity (Schwarze, Bingel, and Sommer 2012), respiration rate, blood pressure (Julien, Advokat, and Comaty 2008);
- **Serotonin (5-HT$_{1A}$ and 5-HT$_{2A}$)** Serotonin is involved in mood regulation and memory. Mild enhancement of 5-HT$_{1A}$ receptors brings euphoria and a sentiment of happiness. This neurotransmitter is implicated into prosocial behaviour (Crockett et al. 2010). A large dose of 5-HT$_{2A}$ in the brain leads to disorientation, confusion, and visual hallucinations (Manford and Andermann 1998).

Table 1 lists the substances considered, together with their related functions, and the neurotransmitters activated by each psychoactive substance.

Each drug carries a set of eight indicators (named, NeuralAction) characterizing the way they impact the corresponding eight neurotransmitters. By impacting specific neurotransmitter receptors, each drug induces a series of behavioural changes embedded in the "Behaviours" attribute of the user class. To trigger such reactions, the amount of neurotransmitters in the brain needs to exceed their "Tolerance-Threshold". These behavioural reactions vary accordingly to the level of neurotransmitters in the brain and important amounts of neurotransmitters lead to unexpected and unwanted behaviours and physiological/psychological damages. This functioning is illustrated with the following activity diagram (Figure 1):

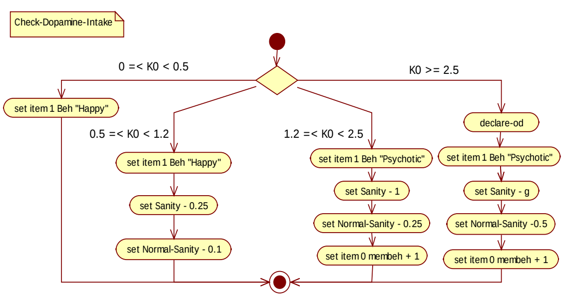


Figure 1: Impact of Dopamine dose Activity Diagram

The K value represents the difference between the actual amount of neurotransmitters and the "Tolerance-Threshold". The higher the value, the more likely the user will experience acute and severe reactions: for example, just reaching the Tolerance-Threshold turns the agent behaviour to "Happy", while a large difference between these two levels leads to "Psychotic" behaviour (here, "declare-od" means that the user is overdosing). SimUse "neurological engine" involves several other components (e.g., tolerance, comedown, and craving) that cannot be developed here.

Table 1: Relation between Substance, Targeted Functions, and Neurotransmitters

Substance	Function	Neurotransmitters
Alcohol	Sociable	Dopamine+ / 5-HT$_{1A}$+
	Relax	GABA+ / OpioidPeptide+ / Glutamate -
	Intoxicate	GABA+ / OpioidPeptide+ / Glutamate -
Cannabis	Sociable	Dopamine+ / 5-HT$_{1A}$+
	Relax	GABA+ / Cannabinoid+
	Intoxicate	GABA+ / 5-HT$_{2A}$+
Cocaine	Sociable	Dopamine+ / 5-HT$_{1A}$+
	Energy	Norepinephrine+ / Glutamate+
Crack	Intoxicate	Dopamine+
MDMA-type	Sociable	Dopamine+ / 5-HT$_{1A}$+
	Energy	Norepinephrine+ / Glutamate+
Opiate-type	Relax	OpioidPeptide+
	Intoxicate	OpioidPeptide+ / Dopamine +
Amphetamine-type	Energy	Norepinephrine+ / Glutamate+
Hallucinogens	Intoxicate	5-HT$_{2A}$+

Nevertheless, the decision process does not stop at the choice of substances based on their expected effects. The interviews showed that the representations users have of drugs condition their choices. Therefore, the interviews investigated these representations and their transformations taking into account that representations are socially constructed. These "Social representations" constitute the stock of information, beliefs and opinions that actors have produced about precise objects through their experiences and interactions (Jodelet 2003; Moscovici 2011). In SimUse, these social representations are modeled and formalized through numerical values representing the user attitudes towards each drug. The range of values goes from -5 to 5: a drug with a negative social representation will not be selected by the agent; neutral representation (0) could lead to consumption if the peers of the agent have a global positive representation of the drug; and a positive representation entails the selection of the related substance. This decisional process is modeled in SimUse as shown in figure 2:

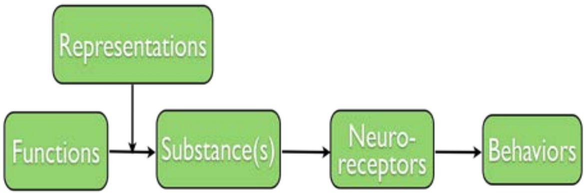


Figure 2: SimUse Drug Decisional Process

In the interviews, the respondents explained that their representations tend to be modified based on the behaviours they observe on themselves retrospectively and by judging the behaviours of other consumers. Indeed, the respondents indicate that they "measure" and balance the positive and negative effects substances have on them. Positive and expected effects appear to reinforce positively the social representation users have attached to the substances (e.g. becoming energetic and alert after the intentional consumption of amphetamine). Conversely, side effects and inappropriate behaviours entail a negative re-evaluation of the representation (e.g. displaying aggressive behaviours after the consumption of amphetamine), which in turn affects future drug-based decisions. Figure 3 provides a flowchart describing this process.

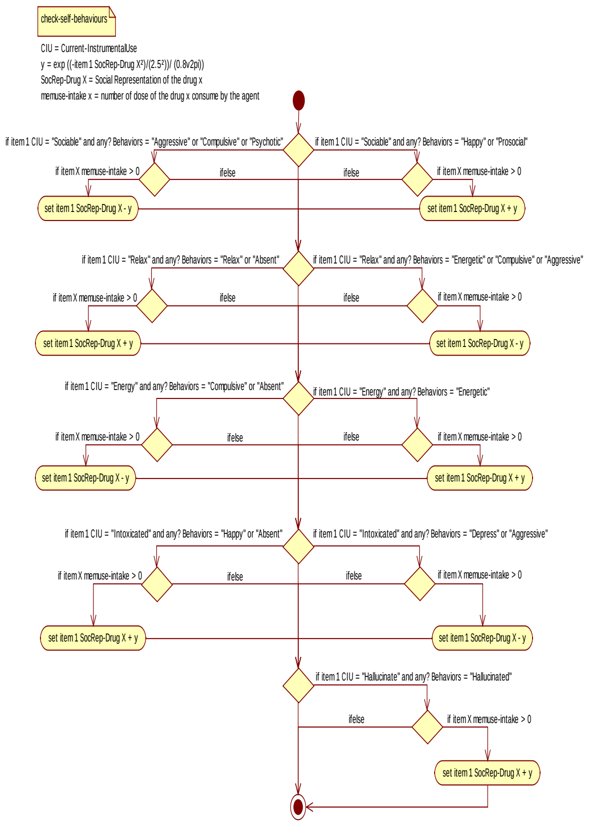


Figure 3: SimUse Activity Diagram for Social Representations Re-evaluation based on the Self Behaviours

The y value displays in this activity diagram is the result of a Normal distribution of mean 0, of variance 1.25 with x equal to the value of the social representation attached to the drug. In other words, agents with social representation values close to the extremes (either -5 or 5) see their representations feebly modified, while agents with a neutral representation (equal to 0) change substantially the way they perceive the drug incriminated.

However, self-reevaluation is not the only process that affects social representations. Based on the Symbolic Interactionist perspective, we consider that meanings, and so social representations, attached to objects are also modified throughout the interactions (Blumer 1998). Interview respondents explained that their opinions on particular drugs could change if they witness inappropriate behaviour from other users under the influence of these substances. Indeed, uncontrolled usage (i.e. compulsive use, being sick) and/or anti-social behaviour (e.g. aggressiveness) are negatively judged and stigmatized by recreational users. Conversely, witnessing expected effects and prosocial behaviours seems to modify positively user's representations. SimUse takes this second form of re-evaluation into account by modeling this process the following way (see Figure 4):

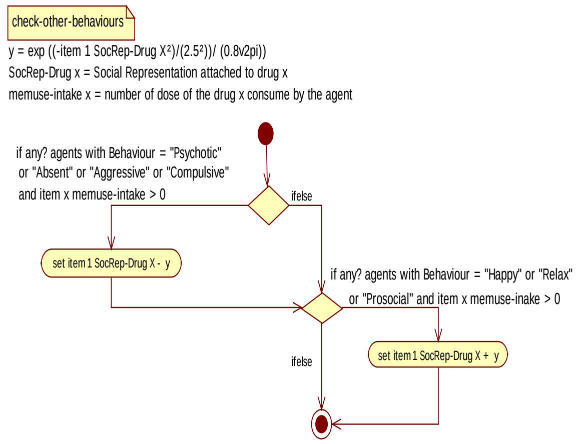


Figure 4: SimUse Activity Diagram for Social Representations
Re-evaluation based on Others Behaviours

It is important to note that repeated consumption of similar substance induce a neurologic *tolerance*. This tolerance reduces the response intensity from the neuroreceptors: in term of consumption, the higher the tolerance, the more users will need to consume to obtain the expected effects. Increased doses generate more intense side effects leading to inappropriate behaviours or unpleasant outcomes, modifying in turn the representation associated with the drug. Taking into account the re-evaluations processes and neurological tolerance, the decisional process shown in figure 1 is updated as follows (see Figure 5):

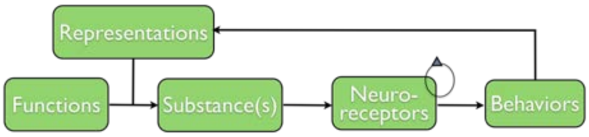


Figure 5: SimUse Iterated Decisional Process

The final model contains a larger quantity of operations and classes, but the complete description of SimUse would exceed the scope of this paper.

4. SAMPLE RESULTS

This section presents the results of two *what-if* scenarios aiming to test the reactions of the model to parameter variations and external shocks. These two scenarios share common initial parameters. Each simulation contains 500 agents and was run for 2400 ticks (which represents 200 virtual days).

The first scenario "EcstasyPrices" tests the impact of substance prices on the consumption rate of that substance. Five prices — 1, 10, 30, 60, and 100 — were tested to assess if the impact of price increases on consumption rates. To create equal conditions and reduce the impact of the irregular drug distribution, each agent knows a dealer of ecstasy at the initiation of the simulation. The first graph shows the impact of Ecstasy price on its consumption rates (cf. Figure 6):

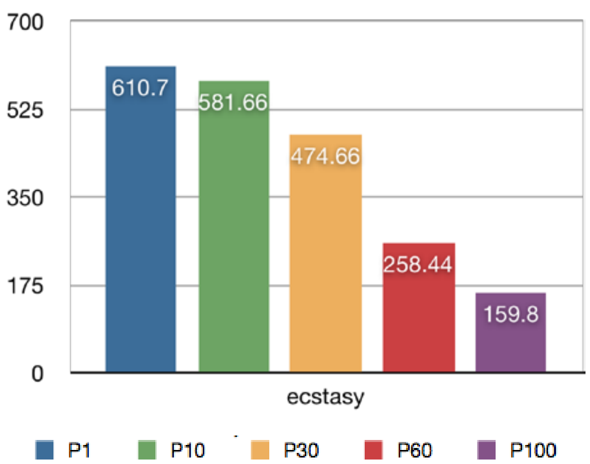


Figure 6: Impact of Ecstasy Price on its consumption rate

In Figure 6, the ordinate represents the number of Ecstasy dose consumed during the simulation. As it could be expected the ecstasy consumption decreases with increasing prices and, as indicated by (Figure 7), the rates of experimentations and regular uses diminish as well:

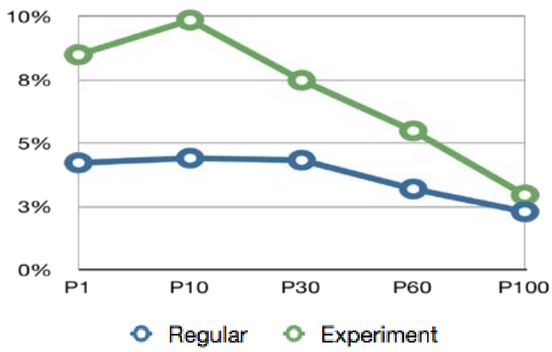


Figure 7: Impact of Ecstasy prices on experimentation and regular use frequencies

If increasing the price of Ecstasy induces a decrease in the rate of consumption, it also induces, in turn, a reduce number of positive experiences felt and reevaluated by potential consumers. Therefore, the ecstasy representation does not increase through the "check-self-behaviour" method, and, at the same time, the absence of visible users does not allow non-users to judge positively the substance through the "check-others-behaviour". In other words, the global social representation is less likely to increase if only a few agents can afford and test the drug. The results from the simulation tend to indicate that lowering the affordability of the drug reduces the number of regular users (certainly due to the financial aspect) but could also affect the number of future users.

The second scenario evaluates the plausibility of model reactions when a large augmentation in the potency of a particular substance is created. This scenario wants to mimic the arrival on the drug market of an "uncut" drug (with a high degree of purity) to assess the reactions of the *users* to that kind of shock. To mimic this sudden increase of potency, the purity of Cocaine is almost tripled after 1200 ticks. Purity is not modeled in SimUse, but could be reproduced by changing the values of the Cocaine's NeuralAction attribute. To judge the impact of this increased potency, we have run a "Standard" scenario (in blue) in which the NeuralAction of Cocaine remains unchanged. As indicated in the following graph, the quantity of cocaine consumed in the "CocainePurity" scenario is lower than in the Standard one (Figure 8):

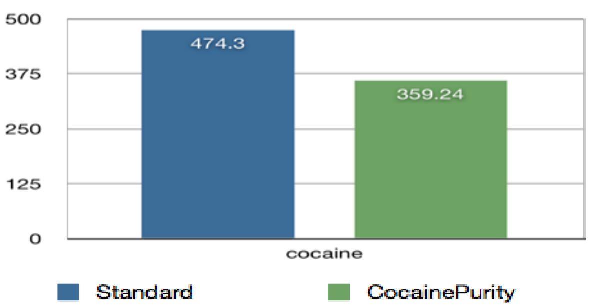


Figure 8: Impact of Increased Cocaine Purity on the Cocaine Consumption

This decrease could be explained by two facts: (a) *users* require lesser dosages to obtain targeted effects due to the increased potency of the substance, and; (b) this decrease could also be explained due to the accumulation of problematic situations or dramatic events following the purity augmentation (see Graph 4 below). Indeed, the number of "hazardous-acts" (i.e. irrational behaviours: putting oneself in danger, drive while intoxicated, attempt impossible actions), brawls ("assaults"), and users entering treatment ("treatnum") affect negatively the social representations of the agents experiencing these adverse reactions, as well as agents witnessing cocaine adverse consequences (Figure 9).

As shown by these preliminary results, changes in a single parameter at one level (price, purity) could entail modifications on several elements in the model directly or indirectly related to the parameters (for example, the diminution of the overall social representations or on the negative events experiences by the agent). Creating this type of ontology permits modeling this type of Complex Adaptive Systems. Indeed, by combining several levels of understanding, and, more importantly, by capturing their interactions and influences, this kind of agent-based model allows testing *what-if* scenarios.

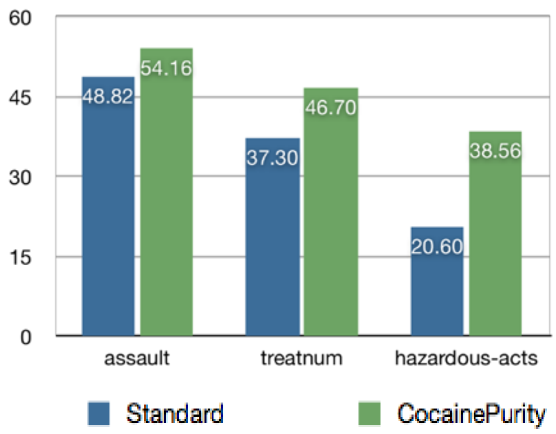


Figure 9: Impact of Increased Cocaine Purity on Negative Events

5. CONCLUSION

Again, as pointed by several institutions, drug policy is a complex topic due to the large number of protective and risk factors that can influence the choices and actions of drug users. This paper argues in favor of the utilization of social simulation to assist policy-makers in their choices, first because it helps to encompass and make interact different levels of analysis, and second because it could be used as a tool to test *what-if* scenarios.

This social simulation incorporates an original model of agent neurological reactions to psychoactive substances and allows encompassing polydrug use. In this paper, we have presented a decisional model of drug choice that consider both instrumental expectations and social representations attached to different substances, and a model of peers influences regarding drug users decisions.

SimUse is subject to numerous limitations (low number of agents, drastic simplification of neurological components and of the geographic area) and still needs further calibration regarding population statistics. However, given this complex agent response to polydrug use, we argue that the emergent social effects can be captured using an agent-based model. The experiments presented here offer an example of *what-if* scenarios and public policies testing achievable with this kind of agent-based social simulations.

REFERENCES

Agar, M. and Wilson, D., 2002. DrugMart: Heroin Epidemics as Complex Adaptive Systems, *Complexity*, 7(5).

Agar, M., 2005. Agents in Living Color: Towards Emic Agent-Based Models, *JASSS*, 8(1).

Arias-Carrion, O. and Poppel, E., 2007. Dopamine, learning, and reward-seeking behaviour. *Acta Neurobiologiae Experimentalis*, 67, 481-488.

Ashton, H.C., 2001. Pharmacology and effects of cannabis: a brief review. *The British Journal of Psychiatry*, 178, 101-106.

Blumer, H., 1998. *Symbolic Interactionism: Perspective and Method*, University of California Press, Berkeley.

Boys, A., Mardsen, J., Fountain, J., Griffiths, P., Stillwell, G. and Strang, J., 1999. What Influences Young People's Use of Drugs? A qualitative study of decision-making. *Drugs: education, prevention and policy*, 6(3), 373-387.

Boys, A. and Marsden, J., 2003. Perceived functions predict intensity of use and problems in young polysubstance users. *Addiction*, 98, 951-963.

Bratman, M. E., 1987. *Intentions, Plans and Practical Reason*. Harvard University Press, Cambridge.

Crockett, M.J., Clark, L., Hauser, M.D. and Robbins, T.W., 2010. Serotonin selectively influences moral judgement and behaviour through effects on harm aversion. *PNAS*, 17(40), 17433-17438.

Ehrenberg, A., 1991. Le culte de la performance. Pluriel. Hachette.

European Monitoring Centre for Drugs and Drug Addiction, 2009. Polydrug use: Patterns and responses. *Annual Report*. Lisbon. Portugal.

European Monitoring Centre for Drugs and Drug Addiction, 2013. *EU drug markets report: a strategic analysis*. Lisbon. Portugal.

Faugeron, C. and Kokoreff, M., 2002. *Société avec drogues: Enjeux et Limites*, Eres.

Fontaine, A., Fontana, C., Verchère, C. and Vischi, R., 2001. Nouvelles drogues, nouveaux usages. Evolution de la consommation de substances psychoactives en France et en Europe et particularités du milieu festif. *Revue Toxibase*, 4.

Gorman, D. M., Gruenewald, P.J., Hanlon, P. J., Mezic, I., Waller, L. A., Castillo-Chavez, C., Bradley, E and Mzic, J., 2004. Implications of Systems of

Dynamic Models and Control Theory for Environmental Approaches to the Prevention of Alcohol-and other Drug-use related Problems. *Substance Use & Misuse*, 39 (10-12), 1713-1750.

Hoffer, L.D., Bobashev, G. and Morris, J., 2009. Researching a Local Heroin Market as a Complex Adaptive System. *American Journal of Community Psychology*, 44 (3-4), 273-286.

Ives, R. and Ghelani, P., 2006. Polydrug use (the use of drugs in combination): A brief overview. *Drugs: education, prevention and policy*, 13(3), 225-232.

Jodelet, D., 2003. *Les représentations sociales*, PUF, Paris.

Johnson, B.D., Hamid, A. and Sanabria, H., 1992. Emerging Models of Crack Distribution, in Mieczkowski, T., *Drugs, Crimes and Social Policy*. Allyn and Bacon.

Julien, R. M., Advokat, C. D. and Comaty, J. E., 2008. *A primer in drug action: a comprehensive guide to the actions, uses, and side effects of psychoactive drugs*, Worth Publisher, 11th ed.

Koob, G. F. and LeMoal, M., 2006. *Neurobiology of Addiction*, Academic Press, Elsevier.

Kuffler, S.W. and Edwards, C., 1957. Mechanisms of Gamma AminoButyric Acid (GABA) action and its relation to synaptic inhibition. *Journal of Neurophysiology*. 21(6), 589-610.

Mandford, M. and Andermann, F., 1998. Complex visual hallucinations. Clinical and neurobiological insights. *Brain*, 121(10), 1819-1840.

Moscovici P., 2011. *Psychologie Sociale*, PUF, Paris.

Parker, H., Aldridge, J, and Measham, F., 1998. *Illegal Leisure: the Normalization of Recreational Drug Use*, London. Routledge.

Parker, H., 2005. Normalization as a barometer: Recreational drug use and the consumption of leisure by younger Britons. *Addiction Research and Theory*, 13(3), 205-215.

Perez, P., Dray, A., Ritter, A., Dietze, P., Moore, T. and Mazerolle, L., 2005. SimDrug: Exploring the Complexity of Heroin Use in Melbourne, DPMP. Monograph 11.

Perez, P., Dray, A., Moore, D., Dietze, P., Bammer, G., Jenkinson, R., Siokou, C., Green, R., Hudson, S.L. and Maher, L., 2012. SimAmph: An agent-based simulation model for exploring the use of psychostimulants and related harm amongst young Australians. *International Journal of Drug Policy*. 23, 62-71.

Preble, E. and Casey, J.J., 1969. Taking Care of Business: the Heroin User's Life on the Street. *The International Journal of the Addictions*, 4(1), 1-24.

Rhodes, T, Lilly, R., Fernàndez, C., Giorgino, E., Kemmesis, U.E., Ossebaard, H.C., Lalam, N., Faasen, I. and Spannow, K.E., 2003. Risk Factors Associated With Drug Use: the importance of 'risk environment'. *Drugs: education, prevention and policy*, 10(4), 303-329.

Riedel, G., 1996. Function of metabotropic glutamate receptors in learning and memory. *Trends Neurosciences*, 19(6), 219-224.

Riedel, G., Platt, B. and Micheau, J., 2002. Glutamate receptor function in learning and memory. *Behavioural Brain Research*, 140, 1-47.

Romano, D., Lomax, L. and Richmond, P., 2009, NARCSim: An Agent-Based Illegal Drug Market Simulation. *Games Innovation Conference IEEE*.

Santiago, T.V. and Edelman, N.H., 1985. Opioids and breathing. *Journal of Applied Physiology*. 59(6), 1675-1685.

Solomon, R.L., 1980. The opponent-process theory of acquired motivation: the costs of pleasure and the benefits of pain. *American Psychologist*, 35, 691-712.

Sussman, S., Earleywine, M., Wills, T., Cody, C., Biglan, T., Dent, C.W. and Newcomb, M.D., 2004. The Motivation, Skills, and Decision-Making Model of ''Drug Abuse'' Prevention. *Substance Use & Misuse.*, 39(10-12), 1971-2016.

Sussman S. and Ames S., 2008. *Drug Abuse: concepts, prevention, and cessation.* Cambridge University Press.

Schwarze U., Bingel U. and Sommer T., 2012. Event-Related Nociceptive Arousal Enhances Memory Consolidation for Neutral Scenes. *The Journal of Neurosciences*, 32(4), 1481-1487.

Simmel G., 1900. *La Philosophie de l'argent*, Paris. PUF. Quadrige.

Theus K.T., 1994. Subliminal advertising and the psychology of processing unconscious stimuli: A review of research. *Psychology and Marketing*, 11(3), 271-290.

Unger J.B., Baewconde-Garbanati L., Shakib S., Palmer P.H., Nezami E. and Mora J., 2004. What are the implications of structural/cultural theory for drug abuse prevention? *Substance Use & Misuse.* 39 (10-12), 1779-1820.

United Nations Office on Drugs and Crime. 2011. World Drug Report 2011. Vienna, Austria.

Villani S., 2001. Impact of Media on Children and Adolescents: A 10-Year Review of the Research. *American Academy of Child and Adolescent Psychiatry*. 40(4). 392-401.

West R., 2006. *Theory of Addiction.* Addiction Press. Blackwell Publishing.

Wilensky U., 1999. *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. Available from: http://ccl.northwestern.edu/netlogo/

Wise R.A., 2002. Brain reward circuitry: insights from unsensed incentives. *Neuron*, 36, 229-240.

EVALUATION AND MODELING THE AERODYNAMIC PROPERTIES OF MUNG BEAN SEEDS

Feizollah Shahbazi[(a)], Ali Dolatshah[(a)], Saman Valizadeh[(a)]

(a) Department of Agricultural Machinery, Faculty of Agriculture, Lorestan University, Iran

[(a)] *Corresponding author's e-mail: shahbazi.f@lu.ac.ir

ABSTRACT
Aerodynamic properties of solid materials have long been used to convey and separate seeds and grains during post harvest operations. The objective of this study was evaluation of the aerodynamic properties of mung bean seeds as a function of moisture content from 7.8 to 25% (w.b) and two grades of A and B referred to above and below a cut point of 4.8 mm in length. The results showed that as the moisture content increased from 7.8 to 25%, the terminal velocity of mung beans seeds increased following a polynomial relationship from 7.28 to 8.79 and 6.02 to 7.12 m/s, for grades A and B, respectively. Mung bean seeds at grade A had terminal velocities with a mean value of 8.05 m/s, while the seeds at grade B had a mean value of 6.46 m/s. The Reynold's number of both grades A and B increased linearly with the increase of seeds moisture content. While, drag coefficient decreased with the increase of moisture content. Mathematical relationships were developed to relate the change in seeds moisture content with the obtained values of aerodynamics properties. The analysis of variance showed that moisture content was significant at the 1% probability level on the all aerodynamics properties of mung bean seeds.

Keywords: aerodynamic properties, separation, post harvest operation, mung bean seed

1. INTRODUCTION

Mung bean (*Vigna radiata* (L.) Wilczek), also known as green bean, green gram, golden gram and mash (in Persian). It is primarily grown in Asia, Africa, South and North America, and Australia principally for its protein- rich edible seeds (Liu and Shen, 2007). Mung bean is similar in composition to other members of the legume family, with 24% protein, 1% fat, 63% carbohydrate and 16% dietary fiber (USDA, 2008). In addition to being an important source of human food and animal feed, mung bean also plays an important role in sustaining soil fertility by improving soil physical properties and fixing atmospheric nitrogen.

The behavior of particles in an air stream during pneumatic conveying and separation greatly depend on their aerodynamic properties. The aerodynamic forces which exist during relative motion between the air and the materials act differently on different particles. Separation of a mixture of particles in a vertical air stream is only possible when the aerodynamic characteristics of the particles are so different that the light particles are entrained in the air stream and the heavy particles fall through it. Knowledge of aerodynamic properties is therefore essential in the proper design of separating and cleaning equipments. When an air stream is used for separating a product such as mung bean seed from its associated foreign materials, such as straw and chaff, knowledge of aerodynamic characteristics of all the particles involved is necessary. This helps to define the range of air velocities for effective separation of the grain from foreign materials. For this reason, the terminal velocity has been used as an important aerodynamic characteristic of materials in such applications as pneumatic conveying and their separation from foreign materials (Mohsenin, 1978).

Several investigators determined the aerodynamic properties of various seeds such as rough rice by Arora (1991), African bread fruit seeds by Omobuwajo *et al.* (1999), amaranth seeds by Kram and Szot (1999), cheat seed by Hauhouot *et al.* (2000), chickpea by Konak *et al.* (2002), millet grain by Baryeh (2002), hemp seed by Sacilik *et al.* (2003), different varieties of rice, corn, wheat and barley by Matouk *et al.* (2005), pine nuts by Ozguven and Vursavus (2005), wheat kernel by Khoshtaghaza and Mehdizadeh (2006), makhana by Jha and Kachru (2007), pistachio nut by Razavi *et al.* (2007) and *Turgenia latifolia* seeds and wheat kernels by Nalbandi *et al.* (2010).

Information about the aerodynamic properties of mung bean seeds is limited. Hence, the objective of this study was to investigate the aerodynamic properties of mung bean seeds as a function of moisture content. Tests were conducted over a range of moisture contents from 7.8 to 25% w.b., which spans the moisture range of harvest to the post harvest operations.

2. MATERIALS AND METHODS

Samples of mung bean seeds at optimum maturity were harvested by hand in Lorestan province, Iran and cleaned in an air screen cleaner. The seeds were then classified into two grades based on their length, cut point being 4.8 mm. Grades A and B referred to above and below the cut point, respectively. The initial moisture content was 7.8% (wet basis), determined with ASAE S352.2 (*ASAE Standards*, 1988). Higher moisture content samples were prepared by adding calculated amounts of distilled water, then sealing in polyethylene bags, and storing at 5ºC for 15 days. Samples were warmed to room temperature before each test and moisture content was verified. Sample mass was recorded with a digital electronic balance having an accuracy of 0.001 g.

The major dimensions of the seeds (*L*, *W* and *T*) were measured using a digital caliper with an accuracy of ± 0.01 mm (Gupta *et al*., 2007). The true density of the seeds was measured using toluene displacement method (Chakraverty and Poul, 2001; Mohsenin, 1978).

To determine the terminal velocity value of mung bean seeds, a vertical wind tunnel was designed constructed and used. The apparatus is shown in Fig. 1. A centrifugal fan powered by one HP motor was used in the inlet of the wind tunnel to supply air flow. The air flow rate of the fan was controlled at inlet by adjusted by changing the velocity of the electric motor through an inverter set and a diaphragm. The final section of the wind tunnel consisted of a Plexiglas region where the terminal velocity of seed was measured. To determine the terminal velocity, each seed was placed in the centre of the cross section of the wind tunnel on the screen. The air flow was then increased until the seed flotation point. At this moment, when the rotational movement of the seed was lowest, the air velocity was measured using a hot-wire anemometer with an accuracy of 0.1 m/s. The terminal velocity of each seed was measured two times. For each condition the terminal velocity was calculated as the average of the velocity values obtained at the centre of the test section and at the four equidistantly distributed points on two orthogonal axes located at the test section. To determine the terminal velocity at each moisture content level, ten seeds were selected and used as ten replications in the statistical analysis. The values of air density and viscosity were taken as 1.2059 kg/m^{-3} and $1.816\times10^{-5} N.sec/m^2$, respectively, at room temperature of 20ºC.

In free fall, the object will attain constant terminal velocity (V_t) at which the net gravitational accelerating force (F_g) equals the resisting upward drag force (F_r) under the condition where terminal velocity has been achieved the air velocity which equal to the terminal velocity (V_t). Substituting for F_g and F_r, the expression for terminal velocity will be as follows (Mohsenin, 1984):

$$V_t = \sqrt{\frac{2mg(\rho_p - \rho_f)}{\rho_p\ \rho_f\ A_p C}} \quad (1)$$

And the drag coefficient can be derived as follows:

$$C_d = \frac{2mg(\rho_p - \rho_f)}{\rho_p\ \ \rho_f\ A_p V_t^2} \quad (2)$$

$$A_P = \frac{\pi}{4} LW \quad (3)$$

Where: A_p is projection area of the particle (m^2), C_d is drag coefficient (decimal), g is acceleration due to gravity (9.81 m/s^2), L is seeds length (m), m is mass of seeds (kg), V_t is terminal velocity (m/s), W is seeds width (m), ρ_f is density of air (1.2059 kg/m^3), ρ_p is density of seeds (kg/m^3).

In this study Reynold's number (*Re*) was calculated using the terminal velocity of each seed sample. Reynold's number (dimensionless) equations include a velocity term using the following relationship (Mohsenin, 1984):

$$\text{Re} = \frac{\rho_f V_t D_g}{\mu} \quad (4)$$

$$D_g = (LWT)^{1/3} \quad (5)$$

Where: D_g is geometric mean diameter of seeds (m), T is seeds thickness (m), μ is air viscosity at room temp (1.816×10^{-5} $N.sec/m^2$).

In this study, the effects of Mung bean seeds size (grades A and B) and moisture content (7.8, 12.5, 15, 17.5, 20 and 25%, wet basis) were studied on the terminal velocity, drag coefficient and Reynold's number of seeds. Tests were conducted over a range of moisture contents from 7.8 to 25% which spans the moisture range of harvest to the processing operations. The factorial experiment was conducted as a randomized design with three replicates. For each test, 10 seeds were selected randomly from each sample and tested by using the airflow device. Mean comparison of factors was carried out at 5% probability level. The terminal velocity, drag coefficient, Reynold's number and the moisture content data of different seed grades were fitted to linear, power, exponential and polynomial models. The models were evaluated according to the statistical criterion R^2 for verifying the adequacy of fit. The best model with the highest R^2 was selected to predict the terminal velocity, drag coefficient and Reynold's number of seeds as a function of the moisture content. Data were analyzed by SPSS 17 software.

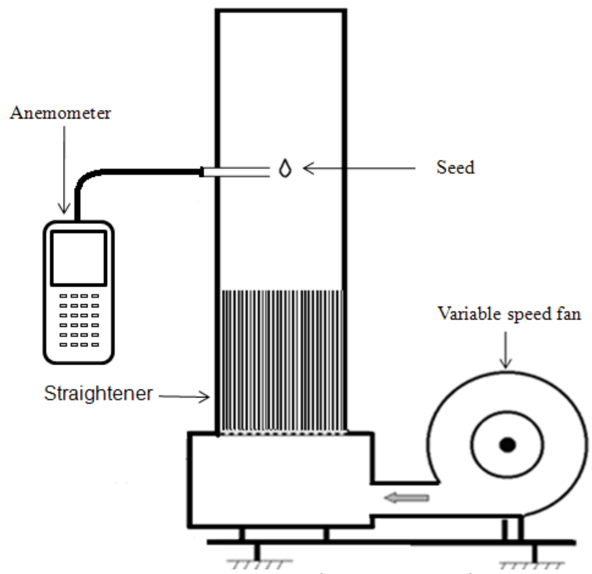


Fig. 1. Schematic diagram of wind tunnel for terminal velocity measurement.

3. RESULTS AND DISCUSSION

3.1. Terminal velocity

The analysis of variance showed that there was a significant difference between the terminal velocity of mung bean seeds at grades A and B. Also the effect of seed moisture content on this property was significant (Table 1). Terminal velocities for grade A was observed to be higher than those obtained for grade B. Mung bean seeds at grade A had terminal velocities with a mean value of 8.05 m/s, at different moisture contents, while the seeds at grade B had a mean value of 6.46 m/s. This result can be explained by the fact that the seeds of grade A were bigger than that of the grade B. Since the square of terminal velocity is directly related to particle size and shape, it follows that larger particles of similar shape need higher terminal velocities than smaller ones. Similar results were obtained by Kahrs (1994) on three fractions of wheat seeds. Wheat seeds > 2.8 mm had mean terminal velocity of 8.8 m/s while the fraction < 2 mm had mean terminal velocity of 6.4 m/s.

The terminal velocity of mung bean seeds increased with increasing moisture content (Table 2). The terminal velocity of mung bean seeds at grades A and B increased from 7.28 to 8.79 m/s and from 6.02 to 7.12 m/s, respectively, as the moisture content of seeds increased from 7.8 to 25% (Fig. 2). The maximum terminal velocity value (8.79 m/s) was obtained in grade A at a moisture content of 25% and the minimum amount (6.02 m/s), was obtained in grade B at a moisture content of 7.8%. These results are in agreement with published literatures for some seeds. Gupta *et al*. (2007) showed that in the moisture range of 6 to 14% d.b., the terminal velocity of NSFH-36, PSF-118 and Hybrid SH-3322 variety of sunflower seed increased from 2.93 to 3.28, 2.54 to 3.04, and 2.98 to 3.53 m/s, respectively. Zewdu (2007) measured the terminal velocity of Tef grains. He reported that it increased linearly from 3.08 to 3.96 m/s with increasing moisture content from 6.5 to 30.1% w.b. Hauhouot *et al*. (2000) showed that the mean value of terminal velocity of wheat seeds was 7.84 m/s. The terminal velocity of millet grain varied from 2.75 to 4.63 m/s for an increase in moisture content from 5 to 22.5% d.b. (Baryeh, 2002). Matouk *et al*. (2008) reported that the terminal velocity of sunflower, soybean and canola seeds increased from 5.34 to 5.91, from 10.16 to 10.38 and from 5.10 to 5.32 m/s with the increasing of seeds moisture contents from 7.35 to 23.7, 9.52 to 24.64% and 7.11 to 25.72% w.b., respectively. Similar results were reported for cotton seeds (Tabak and Wolf, 1998), coffee cherries and beans (Afonso *et al*., 2007), African yam bean (Irtwange and Ugbeka, 2003). The increase in terminal velocity with an increase in moisture content may be attributed to the increase in mass of an individual seed per unit frontal area presented to the air stream. The other reason is probably that the drag force is affected by the moisture content of particle.

Fig . 2 shows the variation of the terminal velocity with moisture content for grades A and B of mung bean seeds. The terminal velocity data for mung bean seeds in Fig. 2 were fitted as a function of moisture content to four mathematical models. These models were evaluated for verifying the adequacy of fit using the R^2 value. By comparing the average values of R^2, it was obvious that the polynomial model had the highest R^2 value. Accordingly, the polynomial model was selected as a suitable model to predict the terminal velocity of mung bean seeds as a function of moisture content. Razavi *et al*. (2007) developed a linear equation between the terminal velocity of pistachio nut and kernel as a function of moisture content. Zewda (2007) reported that the terminal velocity of Tef grain was linearly related to moisture content. However, Afonso *et al*. (2007) reported a nonlinear equation for the terminal velocity of coffee cherry and bean as a function of combination of moisture content and true density. Nalbandi *et al*. (2010) reported a polynomial relationship for the terminal velocity of wheat kernels a function of moisture content. The following equations were found for the relationship between the terminal velocity (V_t, m/s) and moisture content (M, %), for each mung bean seeds grade:

Table 1. Analysis of variance of the data of the aerodynamic properties of mung bean seeds.

Source of variation	df	Mean Squares		
		Terminal velocity	Drag coefficient	Reynold's number
Seed size (S)	1	104.533*	3.714*	5.592×10^8 *
Moisture content (M)	5	17.972*	2.302*	3.211×10^8 *
S×M	5	0.793*	0.357*	3.121×10^8 *
Error	108	0.001	.045	3922077.188

*- Significant difference at 1% probability level.

$$V_t = -0.001M^2 + 0.114M + 6.351$$

R^2=0.961 for: Grade A (6)

$$V_t = 0.002M^2 - 0.004M + 5.856$$

R^2=0.945 for: Grade B (7)

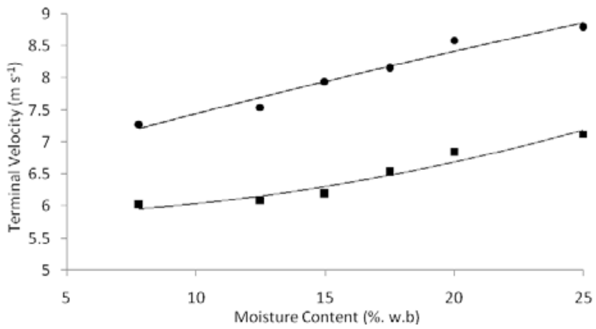


Fig. 2. Terminal velocity variation versus seed moisture content: ● grade A, ■ grade B.

3.2. Drag coefficient

The values of the drag coefficient and the projected area of mung bean seeds were calculated using equations (2) and (3) by measuring the terminal velocity, true density and the two principal dimensions (length and width) of seeds (Table 2). The analysis of variance showed that there was a significant difference between the drag coefficients of mung bean seeds at different moisture content. But the drag coefficients were not affected significantly by mung bean seeds grade. The results showed that the drag coefficient of mung bean seeds decreased as moisture content increased. Afonso *et al*. (2007), Gupta *et al*. (2007) and Irtwange and Ugbeka (2003) reported similar results for coffee cherries, sunflower seed and African yam bean (cv. TSS 138), respectively. However, some odd results have been reported for some products. Irtwange and Ugbeka (2003) reported that the drag coefficient of African yam bean (cv. TSS 137) increased as moisture content increased from 4 to 16% w.b. Afonso *et al.* (2007) showed that the drag coefficient of coffee beans (cv Catual), coffee cherries and beans (cv. Conilon) increased as moisture content increased.

The drag coefficient values of mung bean seeds in grade A were found to be 0.845, 0.827, 0.773, 0.700, 0.619 and 0.565 (with a mean value of 0.722 and standard deviation of 0.113) for the moisture contents of 7.8, 12.5, 15, 17.5, 20 and 25%, respectively. In grade B, the drag coefficients of seeds were found to be 0.846, 0.826, 0.776, 0.703, 0.644 and 0.575 (with a mean value of 0.729 and standard deviation of 0.106) over this same moisture contents (Fig. 3). Hauhouot *et al*. (2000) reported that the drag coefficient of wheat seeds is 0.74. Matouk *et al*. (2008) reported that the drag coefficient of sunflower, soybean, and canola seeds, decreased from 0.75061 to 0.6178, from 0.6841 to 0.6829 and from 0.6301 to 0.5687, with the increasing of seeds moisture contents from 7.35 to 23.7, 9.52 to 24.64 and 7.11 to 25.722%, respectively.

Fig. 3 shows the variation of the drag coefficient with moisture content for two grades of mung bean seeds. The values of this interaction varied from 0.565 to 0.846 that occurred in the grade A at the highest moisture content and in the grade B at the lowest moisture content, respectively. The models fitted to the data using the regression technique showed that the drag coefficient decreased linearly with increases in the moisture content for two grades of mung bean seeds. Similar results were also reported by Matouk *et al.* (2005) for rice, corn, wheat and barley. They stated that, the relationship between terminal velocity and moisture content may be described by an exponential model while, drag coefficient and Reynold's number has linearly relationships. So the following equations were found for the relationship between drag coefficient (C_d) and moisture content (M, %), for each grade of mung bean seeds:

$$C_d = -0.018M + 1.021 \qquad R^2=0.969$$
for: Grade A (8)

$$C_d = -0.017M + 1.011 \qquad R^2=0.945$$
for: Grade B (9)

3.3. Reynold's number

The values of the Reynold's number and the geometric mean diameter of mung bean seeds were calculated using equations (4) and (5) by measuring the terminal velocity and the three principal dimensions (length, width and thickness) of seeds (Table 2). The analysis of variance showed that there was a significant difference between the Reynold's number of mung bean seeds at grades A and B. Also the effect of seed moisture content on this property was significant. The results showed that the Reynold's number of mung bean seeds increased with moisture content. Similar results were reported by Arora, (1991) for three varieties of rough rice and Matouk *et al*. (2005) for rice, corn, wheat and barley. The Reynold's number values of mung bean seeds in grade A were found to be 2281.402, 2423.524, 2579.805, 2714.513, 2937.160 and 3129.219 (with a mean value of 2677.604 and standard deviation of 317.422) for the moisture contents of 7.8, 12.5, 15, 17.5, 20 and 25%, respectively. In grade B, the Reynold's number of seeds were found to be 1494.057, 1524.342, 1579.693, 1673.520, 1793.911 and 1934.339 (with a mean value of 1666.644 and standard deviation of 170.588) over this same moisture contents (Fig. 4). Matouk *et al.* (2008) reported that the Reynold's number of sunflower and soybean seeds in the ranges of 2226.476 to 2571.506 and 4379.706 to 4652.204, with the increase of seeds moisture contents from 7.35 to 23.7% and from 9.52 to 24.644%, respectively.

Fig. 4 shows the variation of the Reynold's number with moisture content for two grades of mung bean seeds. The models fitted to the data using the regression technique showed that the Reynold's number increased linearly with increases in the moisture content. Similar results were also reported by Matouk *et al.* (2005) for rice, corn, wheat and barley. So the following equations were found for the relationship between the Reynold's number (R_n) and moisture content (M, %), for each grade of mung bean seeds:

$$R_n = 52.42M + 1823, \ R^2=0.988, \text{ for: Grade A} \quad (10)$$

$$R_n = 27.50M + 1218, \ R^2=0.973, \text{ for: Grade B} \quad (11)$$

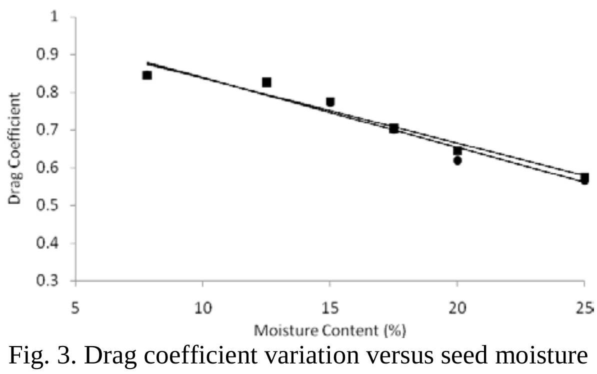


Fig. 3. Drag coefficient variation versus seed moisture content: ● grade A, ■ grade B.

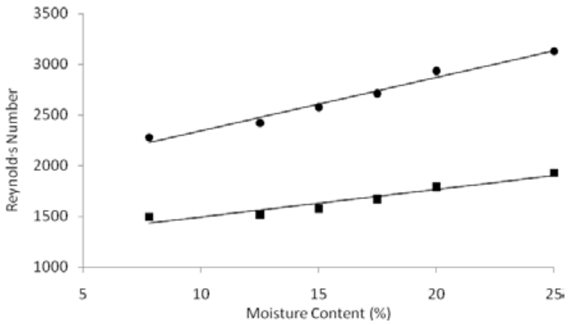


Fig. 4. Reynold's number variation versus seed moisture content: ● grade A, ■ grade B.

4. CONCLUSIONS

From the results of this study, the following conclusions can be drawn:

1. The analysis of variance showed that there was a significant difference between the terminal velocity, and Reynold's number of mung bean seeds at both grades, and at different moisture contents.
2. Terminal velocity of mung beans seeds increased following a polynomial relationship from 7.28 to 8.79 and 6.02 to 7.12 m/s, for grades A and B, respectively, as the moisture content increased from 7.8 to 25%. Mung bean seeds at grade A had terminal velocities with a mean value of 8.05 m/s, at different moisture contents, while the seeds at grade B had a mean value of 6.46 m/s.
3. There was significant difference between the drag coefficients of mung bean seeds at different moisture content. But the drag coefficients were not affected significantly by the mung bean seeds grade.
4. The Reynold's number of mung bean seeds increased linearly with the increase of seeds moisture content. While, drag coefficient decreased with the increase of moisture content.
5. Mathematical relationships were developed to predict the terminal velocity, drag coefficient and Reynold's number of seeds as a function of the moisture content.

REFERENCES

Afonso Jr. P.C., Correa P.C., Pinto F.A.C., and Queiroz D.M. (2007). Aerodynamic properties of coffee cherries and beans. Biosystems Eng., 98, 39-46.

Arora S. (1991). Physical and aerodynamic properties of rough rice (Oryza sative). Indian Journal of Agric. Eng. 1:1, 17-22.

ASAE (1998). ASAE Standards. St. Joseph, MI, USA.

Baryeh E.A. (2002). Physical properties of millet. J. Food Eng, 51: 39-46.

Chakraverty A., and Paul S.R. (2001). Post Harvest Technology: Cereals, Pulses and Vegetables. Sci. Publ., India.

Gupta R.K., Arora G., and Sharma R. (2007). Aerodynamic properties of sunflower seed (Helianthus annuus L.). J. Food Eng, 79: 899-904.

Hauhouot O.M., Criner B.R., Brusewitz G.H., and Solie J.B. (2000). Selected physical characteristics and aerodynamic properties of cheat seed for the separation from wheat. Agric. Eng. Int, 2: 1-14.

Irtwange S.V. and Ugbeka J.C. (2003). Effect of accession and moisture content on aerodynamic properties of African yam bean (Sphenostylis Stenocarpa). Appl. Eng. Agric, 19(3): 321-328.

Jha S.N. and Kachru R.P. (2007). Physical and aerodynamic properties of makhana. J. Food Proces. Eng, 21: 301-316.

Kahrs J. (1994). Aerodynamics properties of weeds seeds. Int. Agrophysics, 8: 259-262.

Khoshtaghaza M.H. and Mehdizadeh R. (2006). Aerodynamic properties of wheat kernel and straw materials. Agric. Eng. Int, CIGR Ejournal, 8: 1-10.

Konak M., Carman K., and Aydin C. (2002). Physical Properties of Chick Pea Seeds. Biosys. Eng, 82: 73–78.

Kram B. and Szot B. (1999). Aerodynamic and geometric properties of amaranth seeds. Int. Agrophysics, 13: 227-232.

Liu W.J. and Shen Q. (2007). Studies on the physicochemical properties of mung bean starch from sour liquid processing and centrifugation. J. Food Eng, 79: 358-363.

Matouk A.M., Abd El-latif S.M., and Tharwat A. (2008). Aerodynamic and mechanical properties of some oil crops. J. Agric. Sci. Mansoura Univ, 33: 4195 – 4211.

Matouk A.M., El-Kholy M. M., Hamam A. S. and Ewis T. R. (2005). Aerodynamic characteristics for different varieties of some cereal crops. Misr J. Agric. Eng, 22 (3): 1086-1102.

Mohsenin N.N. (1978). Physical Properties of Plant and Animal Materials. Gordon Breach Sci. Press, New York, USA.

Nalbandi H., Seiiedlou S., and Ghassemzadeh H.R. (2010). Aerodynamic properties of Turgenia latifolia seeds and wheat kernels. Int. Agrophysics, 24: 57-61.

Omobuwajo T.O., Akande E.A., and Sanni L.A. (1999). Selected physical, mechanical and aerodynamic properties of African breadfruit (Treculia africana) seeds. J. Food Eng, 40: 241-244.

Ozguven F., and Vursavus K. (2005). Some physical, mechanical and aerodynamic properties of pine (Pinus pinea) nuts. J. Food Eng, 68: 191-196.

Razavi S.M.A., Rafe A., and Akbari R. (2007). Terminal velocity of pistachio nut and its kernel as affected by moisture content and variety. African J. Agric. Res, 2: 663-666.

Sacilik K., Ozturk R., and Keskin R. (2003). Some Physical Properties of Hemp Seed. Biosystems Eng, 86: 191–198.

Tabak S. and Wolf D. (1998). Aerodynamic properties of cottonseeds. J. Agric. Eng. Res, 70: 257-265.

USDA. (2008). National Nutrient Database for Standard Reference, Release 18.

Zewdu A.D. (2007). Aerodynamic properties of tef grain and straw material. Biosystems Eng, 98: 304-309.

Table 2. Average values of dimensions, projected area, geometric mean diameter, true density and terminal velocity of mung bean seeds at different moisture contents.

Moisture Content (%)	Length (mm)	Width (mm)	Thickness (mm)	Projected area (mm^2)	Geometric mean diameter (mm)	True density (kg m^3)	Terminal velocity (m s^{-1})
				Grade A			
7.8	4.339 (0.24)*	3.701 (0.19)	3.251 (0.14)	12.606 (1.03)	3.737 (0.11)	1252.97 (13.63)	6.021 (0.21)
12.5	4.423 (0.32)	3.712 (0.21)	3.262 (0.23)	12.888 (1.06)	3.769 (0.08)	1303.13 (12.24)	6.092 (0.35)
15	4.512 (0.31)	3.779 (0.20)	3.329 (0.35)	13.384 (0.98)	3.843 (0.14)	1346.24 (17.98)	6.198 (0.12)
17.5	4.523 (0.21)	3.790 (0.33)	3.346 (0.29)	13.486 (1.12)	3.859 (0.19)	1376.16 (11.85)	6.531 (0.46)
20	4.632 (0.42)	3.871 (0.24)	3.421 (0.58)	14.075 (1.23)	3.943 (0.15)	1404.23 (10.12)	6.851(0.22)
25	4.793 (0.37)	4.012 (0.34)	3.561 (0.52)	15.092 (1.31)	4.091 (0.23)	1425.13 (11.02)	7.120 (0.52)
				Grade B			
7.8	5.541 (0.65)	4.613 (0.44)	4.112 (0.63)	20.065 (2.12)	4.719 (0.36)	1186.33 (13.65)	7.280 (0.41)
12.5	5.821 (0.56)	4.617 (0.54)	4.219 (0.52)	21.100 (1.05)	4.840 (0.27)	1193.52 (10.87)	7.542 (0.52)
15	5.905 (0.43)	4.671 (0.34)	4.235 (0.63)	22.652 (2.04)	4.886 (0.35)	1202.42 (11.25)	7.951 (0.31)
17.5	6.102 (0.75)	4.807 (0.39)	4.302 (0.32)	23.025 (1.25)	5.015 (0.52)	1296.36 (18.98)	8.150 (1.02)
20	6.152 (0.81)	5.022 (0.46)	4.419 (0.25)	24.252 (0.98)	5.149 (0.41)	1325.63 (14.52)	8.591 (0.58)
25	6.715 (0.88)	5.098 (0.54)	4.501 (0.22)	26.827 (1.54)	5.361 (0.47)	1369.87 (14.25)	8.792 (0.62)

*Standard deviation.

USING ANALYTIC HIERARCHY PROCESS TO EVALUATE HUMAN PERFORMANCE

Fabio De Felice[(a)], Antonella Petrillo[(b)], Michele Tricarico[(c)]

[(a)] Department of Civil and Mechanical Engineering - University of Cassino, Italy
[(b)] Department of Civil and Mechanical Engineering - University of Cassino, Italy
[(c)] Horseracing Italian Agency, Italy

[(a)]defelice@unicas.it, [(b)]a.petrillo@unicas.it, [(c)]m.tricarico@mpaaf.gov.it

ABSTRACT
This article proposes a reference model to assess performance of human resource for the supervision and management of horse racing in the appraisal context. Performance appraisal is defined as the formal process of evaluating organizational members. The present paper uses Analytic Hierarchy Process (AHP) to evaluate human performances and provides a way to rank the alternatives of the problem by deriving priorities.

Keywords: AHP, human resources, decision support system, performance appraisal

1. INTRODUCTION

In the global economy, the modern commercial and industrial organization needs to develop better methods of assessing the performance of the human resource than simply using performance measures such as efficiency or effectiveness (Albayrak and Erensal, 2004; Mani, 2002).

Thus, performance appraisal is a human resource management tool that has received much attention for more than seven decades (Erdogan, 2002). Fairness of performance appraisals has been identified as an important criterion in judging their effectiveness and usefulness for organizations. This problem is complex and, like most real world problems depend upon a number of tangible and intangible factors which are unique to each problem. The complexity stems from a multitude of quantitative and qualitative factors influencing location choices as well as the intrinsic difficulty of making numerous trade-offs among those factors (Suwignjo et al., 2000).

One analytical approach often suggested for solving such a complex problem is the Analytic Hierarchy Process (AHP) introduced by Saaty (Saaty, 1980). The AHP enables the decision maker to structure a complex problem in the form of a simple hierarchy and to evaluate a large number of quantitative and qualitative factors in a systematic manner under conflicting multiple criteria (De Felice, 2012). It is developed and designed to solve complex problems involving multiple criteria. It is a highly flexible decision methodology that can be applied in a wide variety of situations (De Felice *et al.*, 2012)

There are two types of measurement involved in the AHP, absolute and relative. The first requires a standard with which to compare elements, but mostly alternatives at the bottom of the hierarchy. The process leads to absolute preservation in the rank of the alternatives no matter how many are introduced. The second is based on paired comparisons among the elements of a set with respect to a common attribute. This process is essential for comparing intangible attributes for which there are no agreed upon measures.

At the level of alternatives new elements (i.e. alternatives) do introduce new information generated by the changing number in the set and by their measurement which essentially rescales the criteria and hence can lead to reversals of previous rank orders.

Absolute measurement is used on standardized problems whereas relative measurement is used in new learning situations (Saaty, 2005). Absolute method is typically used in a decision situation, which involve selecting one (or more) decision alternatives from several candidate decision alternatives on the basis of multiple decision criteria of a competing or conflicting nature (McCarthy, 2000; Roberts, 2003).

In this paper, we have developed a case study to evaluate human performance using AHP absolute model. Though AHP has been applied in numerous real settings, but there is few evidence that AHP has been applied in human performance evaluation (De Felice and Petrillo, 2013; Sun *et al.*, 2008). This paper attempts to fill up the gap. This article proposes a multicriteria decision model of antecedents and consequences of justice perceptions in the appraisal context based on Erdogan's model (Endorgan *et al.*, 2001) in order to develop a flexible decision model useful for evaluating the performance of human resources.

A real case study applied for evaluate the performance of human resource appraisal for the horse racing supervision and management is proposed.

The paper is structured in section 2 in which problem statement is analyzed; section 3 in which methodological approach is presented; section 4 in which the case study is presented. Finally conclusions are reported.

2. PROBLEM STATEMENT

Performance appraisals are essential for the effective management and evaluation of staff. There is increasingly a need for performance appraisals of staff and especially managers, directors and CEO's. The performance appraisal process is an interactive process between the supervisor and the employee meant to assess and summarize the work performance of the employee as well as set new goals and identify new career development plans and training (UCSD, 2005; Liden *et al.*, 2004).

The performance appraisals process is a very difficult process because involves different aspects and problems. We based our study on Erdogan's model reported in Figure 1 (see appendix).

According to this model it is possible to differentiate between four types of justice perceptions in performance appraisals (Wayne *et al.*, 1997; Masterson *et al.*, 2000; Bauer et al., 2012). The model introduces several antecedents of justice perceptions:

- *Proposition 1a:* Components of adequate notice will be differentially related to system and rater procedural justice such that, communication of appraisal criteria and involvement in development of appraisal criteria will be positively related to system procedural justice, whereas frequent feedback during appraisal period will be positively related to rater procedural justice.
- *Proposition 1b:* Components of fair hearing will be differentially related to system and rater procedural justice such that, having a rater familiar with ratee's work will be positively related to system procedural justice, whereas allowing ratees input in decision-making will be positively related to rater procedural justice.
- *Proposition 1c:* Components of judgment based on evidence will be differentially related to system and rater procedural justice such that, existence of effective appeal mechanisms will be positively related to system procedural justice, whereas consistent application of standards and explaining the decision to the ratee will be positively related to rater procedural justice.
- *Proposition 2a:* POS - Perceived Organizational Support- before the appraisal will be positively related to system procedural justice perceptions during the appraisal.
- *Proposition 2b:* Organizational culture will be related to perceptions of rater procedural justice such that, in constructive cultures compared to passive–defensive or aggressive–defensive cultures, the highest levels of rater procedural justice will be observed.
- *Proposition 2c:* LMX - Leader Member Exchange Quality - quality before the performance appraisal will be positively related to perceptions of rater procedural justice.
- *Proposition 3:* The rater's use of job focused impression management tactics will be negatively related to interactional justice perceptions, whereas the use of supervisor and subordinate focused tactics will be positively related.
- *Proposition 4:* Pre-appraisal LMX quality will be positively related to perceptions of interactional justice during performance appraisal.
- *Proposition 5:* The relationship between ratings and distributive justice perceptions will be moderated by LMX quality such that, for high LMX employees, there will be a stronger positive relationship between ratings and distributive justice perceptions.
- *Proposition 6a:* When ratees do not know the performance ratings of their coworkers, they will believe that those with higher LMXs are more likely to receive higher performance ratings.
- *Proposition 6b:* The perception that the leader forms LMXs based on work-related factors will be positively related to distributive justice *perceptions.*
- *Proposition 6c:* The perceived type of information used in appraisals will be related to distributive justice perceptions, such that the use of consistency and distinctiveness of information will be positively related to distributive justice perceptions, whereas the use of consensus information will be negatively related.
- *Proposition 7a:* System procedural justice will be positively related to post-appraisal POS.
- Proposition 7b: Post-appraisal POS will mediate the relationship between system procedural justice and organizational outcomes.
- *Proposition 8a:* Rater procedural justice and distributive justice perceptions would be positively related to post-appraisal LMX.
- *Proposition 8b:* The relationship between rater procedural justice, interactional justice, distributive justice, and leader-related outcomes will be mediated by post-appraisal LMX.
- *Proposition 9a:* Distributive justice perceptions in performance appraisals will be positively related to perceived accountability.
- *Proposition 9b:* The relationship between distributive justice perceptions and performance related outcomes will be mediated by perceived accountability.

The above model presents some weaknesses this is the reason because we propose an "integration" with AHP.

3. METHODOLOGICAL APPROACH

The aim of our paper is to explain the uses of multi-criteria prioritization, and in particular the use of absolute measurement in the optimal evaluation of human performance.

Below are the steps of absolute measurement process adopted (De Felice and Petrillo, 2011):

- Step 1: Definition of the experts team.
- Step 2: Identification the criteria, subcriteria for evaluation and put them into the AHP hierarchy.
- Step 3: Identification of the alternatives.
- Step 4: Calculate the weights of the decision criteria by the relative measurement of AHP.
- Step 5: Evaluation of consistency analysis.
- Step 6: Division of each subcriterion into several intensities or grades.
- Step 7: Measurement of performance intensity under each subcriterion.

In figure 2 (see appendix) is shown the methodological approach.

4. CASE STUDY

In this paragraph we analyze the AHP model adopted in order to evaluate human performance. In particular, following, the different steps are detailed.

Step 1: Definition of the experts team.

First of all experts team were defined. The experts team consisted of 4 Gallop experts and 3Trot experts. The experts team developed the AHP Model. In figure 3 is shown AHP Model (see appendix).

As is shown in Figure 3 the criteria C1.3, C2.1 and C2.2 are outlined because as we will underline in step 6 for these criteria the experts team defined a different scale of intensity.

Step 2: Identify the criteria, subcriteria.

The experts team defined criteria and subcriteria in order to assess the human performance. Here below criteria and subcriteria are detailed.

- *C1 - Core competencies*: skills without which it is not possible to perform the functions of the components of the direction racing:
    - *C1.1 - Knowledge of Regulation*: knowledge of all regulations and their updates;
    - *C1.2 - Technical knowledge*: Excellent knowledge of the peculiarities of horse racing and the technical elements;
    - *C1.3 - Qualification*: Evaluation of the process of training for the qualification competition judge;
    - *C1.4 - Problem solving:* Proactive attitude and management capabilities.
- *C2 - Complementary skills:* skills that enhance the actions by making them more effective:
    - *C2.1 – Experience and CV* : evaluation of the Curriculum Vitae;
    - *C2.2 - Education*: evaluation of training and qualification;
    - *C2.3 - Professional ethics*: formal and informal attitudes appropriate for the respect and the fulfillment of the institutional role;
    - *C2.4 - Availability*: attitude to hold the post received as a priority over the needs / preferences or otherwise professional.
- *C3 - Relational skills:* ability to interact optimally with regard to the context in which a person works:
    - *C3.1 - Authority and Charisma*: ability to exert a strong influence on other people;
    - *C3.2 - Teamwork*: ability to interact with the different positions, dealing with different opinions and find a constructive synthesis;
    - *C3.3 - Interpersonal relationships*: ability to manage external relations.
- *C4 - Skills for implementing Regulation:* ability to enforce formal rules and regulations in a uniform manner with respect to the context.
    - *C4.1 - Written and verbal presentation skills*: ability to represent verbally and / or in writing in a clear and concise;
    - *C4.2 - Correct formulation of regulation*: ability to accurately report the infringements;
    - *C4.3 - Personal integrity*: the ability to apply the regulation adequately;
    - *C4.4 - Perseverance and determination*: ability to enforce the regulation always with moderation and with the same commitment and willingness.

Step 3: Identify the alternative or "guide profiles".

In the present step the experts team defined the different alternatives characterizing the human performance that they called "guide profiles". In detailed the experts team defined:

- 4 guide profiles for Trot: President; Starter, Junta member and Commissioner.
- 3 guide profiles for Gallop: Commissioner; Official and Starter.

Step 4: Calculate the weights of the decision criteria and subcriteria.

The experts team developed pairwise comparison matrices to determine the criteria weights. We note that in the AHP paired comparisons are made with judgments using numerical values taken from the AHP absolute fundamental scale of 1-9. In particular the constructed the pairwise comparison matrix for all the criteria and compute the normalized principal right eigenvector of the matrix. This vector gives the weights of the criteria and subcriteria. Then in similar way weights for subcriteria were calculated. Finally these weights were multiplied by the weights of the parent criteria.

In Figure 3 is shown an example of pairwise comparison.

C1	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	**C1**
C1	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	**C1**
C1	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	**C1**
C2	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	**C2**
C2	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	**C2**
C3	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	**C3**
CI = 0.07475																		

Figure 3: Pairwise comparison – Gallop - Commissioner

The judgments of all the experts were aggregated using the geometric mean. Here below are the weights calculated for Gallop – profile Commissioners.

Table 1: Weights- Gallop - Criteria

Gallop - Commissioners					
Criteria	Exp.1	Exp.2	Exp.3	Exp.4	Weights
C1	0,565	0,366	0,356	0,577	0,491
C2	0,041	0,047	0,044	0,103	0,059
C3	0,205	0,155	0,474	0,104	0,215
C4	0,187	0,430	0,125	0,213	0,233

Table 2: Weights- Gallop – Sub Criteria

Gallop – Commissioners					
SubCriteria	Exp.1	Exp.2	Exp.3	Exp.4	Weights
C1.1	0,454	0,200	0,401	0,215	0,316
C1.2	0,252	0,327	0,102	0,349	0,247
C1.3	0,040	0,092	0,092	0,112	0,070
C1.4	0,252	0,379	0,450	0,322	0,364
C2.1	0,048	0,155	0,247	0,335	0,173
C2.2	0,048	0,049	0,062	0,177	0,078
C2.3	0,653	0,575	0,375	0,400	0,533
C2.4	0,249	0,219	0,314	0,086	0,214
C3.1	0,500	0,559	0,708	0,549	0,591
C3.2	0,250	0,352	0,178	0,209	0,246
C3.3	0,250	0,088	0,112	0,242	0,161
C4.1	0,166	0,170	0,135	0,097	0,140
C4.2	0,166	0,185	0,180	0,164	0,175
C4.3	0,333	0,363	0,331	0,376	0,352
C4.4	0,333	0,279	0,352	0,360	0,331

In a similar way the weights for:

- Gallop – profile Officials and Starters
- Trot – profile Presidents, Starters, Junta members and Commissioners

were obtained.

Step 5: Consistency analysis.
After all pairwise comparison the consistency index (CI) of the derived weights was calculated by Equation (1):

$$CI \frac{\lambda_{max} - n}{n-1} \quad (1)$$

In general, if CI is less than 0.10, satisfaction of judgments may be derived.

Step 6: Divide each subcriterion into several intensities or grades.
To implement the absolute measurement model in AHP, each criterion is divided into several intensity ranges to differentiate the qualifications of the candidates with respect to that criterion (Saaty, et al. 2007).
Experts team defined 4 different intensity scales:

- *Scale 1*: For the evaluation of the criteria C1.1, C1.2, C1.4, C2.3, C2.4, C3.1, C3.2, C3.3, C4.1, C4.2, C4.3 and C44 the experts team defined the following intensities (see Table 3).
- *Scale 2*: For C1.3 the experts team defined intensities reported in Table 4.
- *Scale 3*: For C2.1 the experts team defined intensities reported in Table 5.
- *Scale 4*: For C2.2 the experts team defined intensities reported in Table 6.

Also in this case the judgments of all the experts were aggregated using the geometric mean.

Table 1: Scale 3 – Intensities - Gallop

Score	Intensity	Weight
8	Exceptional	0,420
6	Exceed Expectations	0,309
5	Good	0,162
3	Satisfactory	0,071
1	Poor	0,035

Table 4: Scale 2 – Intensities - Gallop

Score	Intensity	Weight
8	Absolutely agree	0,455
6	Partly agree	0,300
5	Neutral	0,119
3	Slightly disagree	0,890
1	Strongly disagree	0,442

Table 5: Scale 3 – Intensities - Gallop

Score	Intensity	Weight
8	Absolutely agree	0,402
6	Partly agree	0,273
5	Neutral	0,165
3	Slightly disagree	0,098
1	Strongly disagree	0,059

Table 6: Scale 4 – Intensities - Gallop

Score	Intensity	Weight
3	Level 1	0,067
5	Level 2	0,124
6	Level 3	0,270
8	Level 4	0,537

In a similar way the scales for trot were obtained (see Table 7, 8, 9 and 10).

Table 7: Scale 1 – Intensities - Trot

Score	Intensity	Weight
8	Exceptional	0,472
6	Exceed Expectations	0,195
5	Good	0,183
3	Satisfactory	0,102
1	Poor	0,045

Table 8: Scale 2 – Intensities - Trot

Score	Intensity	Weight
8	Absolutely agree	0,501
6	Partly agree	0,256
5	Neutral	0,105
3	Slightly disagree	0,789
1	Strongly disagree	0,057

Table 9: Scale 3 – Intensities - Trot

Score	Intensity	Weight
8	Absolutely agree	0,511
6	Partly agree	0,227
5	Neutral	0,120
3	Slightly disagree	0,083
1	Strongly disagree	0,057

Table 10: Scale 4 – Intensities - Trot

Score	Intensity	Weight
3	Level 1	0,068
5	Level 2	0,129
6	Level 3	0,253
8	Level 4	0,548

Step 7: Measure performance intensity under each subcriterion.

In this step performance intensity under each subcriterion was calculated. The process was repeated for all types of profile (see Table 11 and Table 12).

Table 11: Performance intensity - Gallop

Gallop			
SubCriteria	Commis.	Officials	Starters
C1.1	15,57%	7,82%	4,39%
C1.2	12,19%	6,96%	8,19%
C1.3	3,46%	2,22%	2,33%
C1.4	17,93%	9,01%	10,46%
C2.1	1,03%	1,78%	3,48%
C2.2	0,47%	1,03%	1,35%
C2.3	3,17%	5,24%	6,75%
C2.4	1,28%	2,56%	3,57%
C3.1	12,78%	10,49%	29,56%
C3.2	4,65%	17,90%	7,25%
C3.3	4,17%	7,53%	12,59%
C4.1	3,27%	4,12%	1,25%
C4.2	4,09%	4,05%	1,24%
C4.3	8,23%	10,80%	4,23%
C4.4	7,74%	8,50%	3,38%

Table 12: Performance intensity - Trot

Trot				
SubCriteria	Presid.	Commis.	J. members	Starters
C1.1	15,26%	6,53%	12,68%	9,63%
C1.2	14,75%	11,04%	12,16%	8,86%
C1.3	2,30%	3,67%	5,10%	3,13%
C1.4	5,14%	14,29%	5,17%	13,38%
C2.1	6,25%	11,55%	10,89%	4,35%
C2.2	1,81%	3,05%	2,44%	1,36%
C2.3	4,94%	11,77%	5,32%	3,54%
C2.4	2,72%	5,88%	4,42%	2,46%
C3.1	5,19%	6,02%	6,63%	8,48%
C3.2	1,64%	8,78%	8,63%	1,96%
C3.3	1,97%	8,20%	5,61%	6,04%
C4.1	5,83%	1,40%	2,81%	6,07%
C4.2	5,81%	0,77%	4,06%	5,78%
C4.3	9,39%	2,63%	7,46%	14,44%
C4.4	16,99%	4,42%	6,62%	10,53%

In Figures 4 and 5 are shown final weights for subcriteria for each profiles both Gallop and Trot. We can note that regarding:

- GALLOP: the parameter most representative is C1.4 (17,93%) for Commissioners; C3.2 (17,90%) for Officials and C3.1 (29,56%) for Starters.
- TROT: the parameter most representative is C4.4 (16,99%) Presidents; C4.3 (14,44%) for Starters; C1.1 (12,68%) for Junta members and C1.4 for Commissioners (14,29%).

5. CONCLUSIONS

Performance appraisal is a performance management mechanism that has broad implications for attitudes and behaviors in organizations. We propose a simple and effective appraisal system that emphasizes continuous professional development enhances a firm's overall performance.

The model proposed based on AHP, definitely, has a positive impact on the proposition 1a, 2a, 2b, 6a and 6c. Furthermore the model presented wants to identify areas in which there is a need for more research.

In fact, we analyzed different aspects but in our opinion it is necessary to differentiate between different forms of justice perceptions. Further research will cover this gap.

REFERENCES

Albayrak, E., Erensal, Y.C., 2004. Using analytic hierarchy process (AHP) to improve human performance: An application of multiple criteria decision making problem. *Journal of Intelligent Manufacturing*, August 2004, Volume 15, Issue 4, pp 491-503.

Bauer, T. N., & Erdogan, B., 2012. *Organizational socialization outcomes: Now and into the future*. In C. Wanberg (Ed.), Organizational socialization. (pp. 97-112). Oxford, UK: Oxford University Press.

De Felice, F., Petrillo, A., 2013. Absolute measurement with analytic hierarchy process: A case study for Italian racecourse. *International Journal of Applied Decision Sciences*. Volume 6, Issue 3, 2013, Pages 209-227.

De Felice, F., Petrillo, A., 2011. Methodological Approach for Performing Human Reliability and Error Analysis in Railway Transportation System. *International Journal of Engineering and Technology*,Vol.3 (5), 2011, 341-353.

De Felice, F., 2012. Editorial Research and applications of AHP/ANP and MCDA for decision making in manufacturing. *International Journal of Production Research*, Vol. 50, No. 17, 1 September 2012, 4735–4737.

De Felice, F., Petrillo, A., Tricarico, M., 2012. Optimal allocation of economic resources using the AHP absolute model. *Proceedings of the International Conference on Modeling and Applied Simulation* 2012, 19-21 September 2012 Vienna. ISBN 978-88-97999-02-7 (Paperback), pp 202-211.

Erdogan, B., Kraimer, M. L., & Liden, R. C., 2001. Procedural justice as a two-dimensional construct: an examination in the performance appraisal context. *Journal of Applied Behavioral Science*, 37, 205–222.

Erdogan, B., 2002. Antecedents and consequences of justice perceptions in performance appraisals. *Human Resource Management Review*, 12:555–578.

Liden, R. C., Bauer, T. N., & Erdogan, B., 2004. T*he role of leader-member exchange in the forming and dynamic relationship between employee and employer*. In J. Coyle-Shapiro, L. Shore, S. Taylor, & L. Tetrick (Eds.). The Employment Relationship: Examining Psychological and Contextual Perspectives. (pp. 226-250). Oxford University Press.

Mani, B.G., 2002. Performance Appraisal Systems, Productivity, and Motivation: A Case Study, *Public Personnel Management*, 31(2), 141-159.

Masterson, S. S., Lewis, K., Goldman, B. M., & Taylor, M. S., 2000. Integrating justice and social exchange: the differing effects of fair procedures and treatment on work relationships. *Academy of Management Journal*, 43, 738–748.

McCarthy, J., 2000. How to Conduct Productive Performance Appraisals. *Journal of Property Management*, 22-25.

Roberts, G.E., 2003. Employee Performance Appraisal System Participation: A Technique that Works, *Public Personnel Management*, 32(1), 89-98

Saaty, T.L., 1980. *The Analytic Hierarchy Process*. third ed. McGraw-Hill, New York.

Saaty, T.L., 2005. *Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks*. RWS Publications, 4922 Ellsworth Ave., Pittsburgh, PA, 2005, p. 15213.

Saaty, T.L., Peniwati, K., Shang, Jen S., 2007. The analytic hierarchy process and human resource allocation: Half the story. *Mathematical and Computer Modelling* 46, 1041–1053.

Sun, Y.-H., Ma, J., Fan, Z.-P., Wang, J., 2008. A group decision support approach to evaluate experts for R&D project selection. *IEEE Transactions on Engineering Management*, 55 (1), 158–170.

Suwignjo, P., Bititci, U.S. and Carrie, A.S., 2000. Qualitative Models for Performance Management System, *International Journal of Production Economics*, 64, 231-241.

University of California, San Diego, 2005. *Guide to performance management*. UCSD Human Resources Department. California (USA).

Wayne, S. J., Shore, L. M., & Liden, R. C., 1997. Perceived organizational support and leader–member exchange: a social exchange perspective. *Academy of Management Journal*, 40, 82–111.

APPENDIX

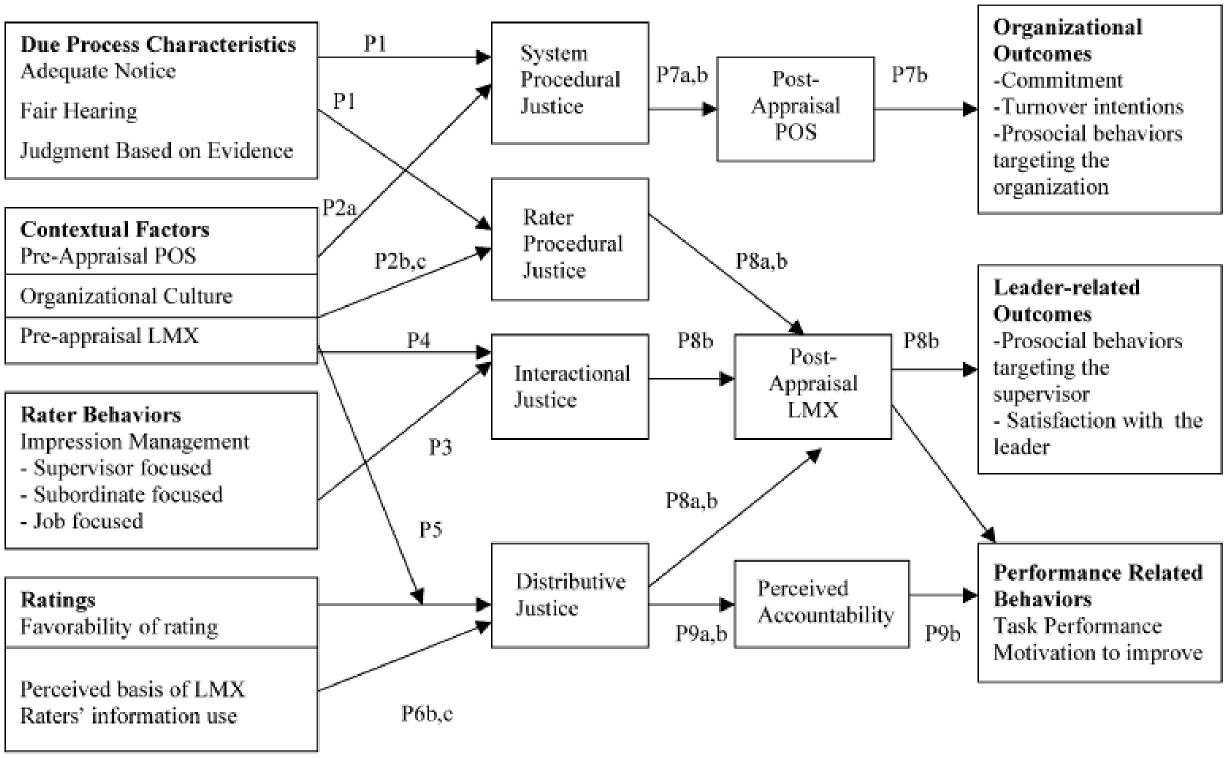


Figure 1: Antecedents and consequences of justice perceptions in performance appraisals (source B. Erdogan, 2002)
POS - Perceived organizational support, organizational culture
LMX – Leader Member Exchange Quality

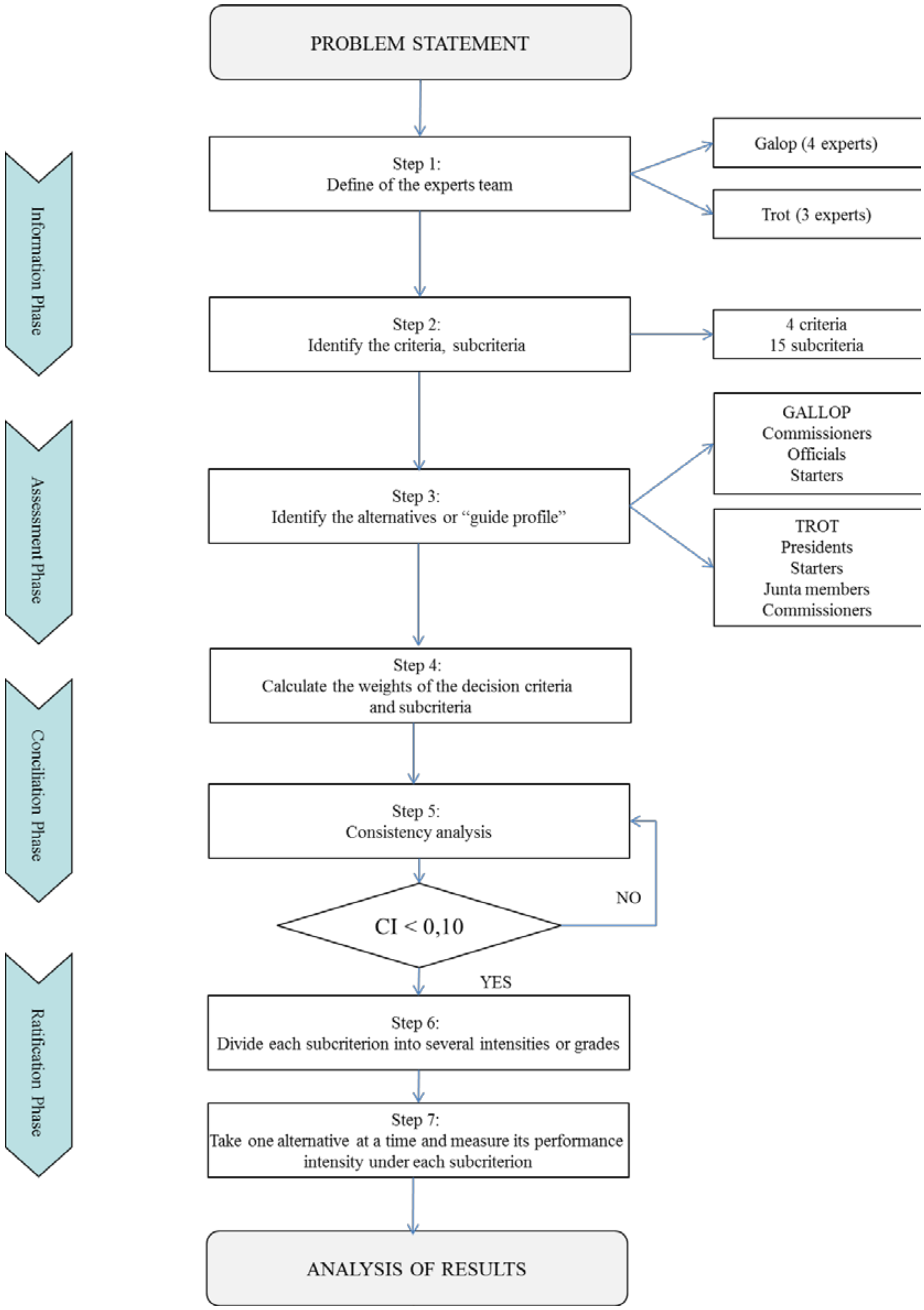


Figure 2: Methodological approach

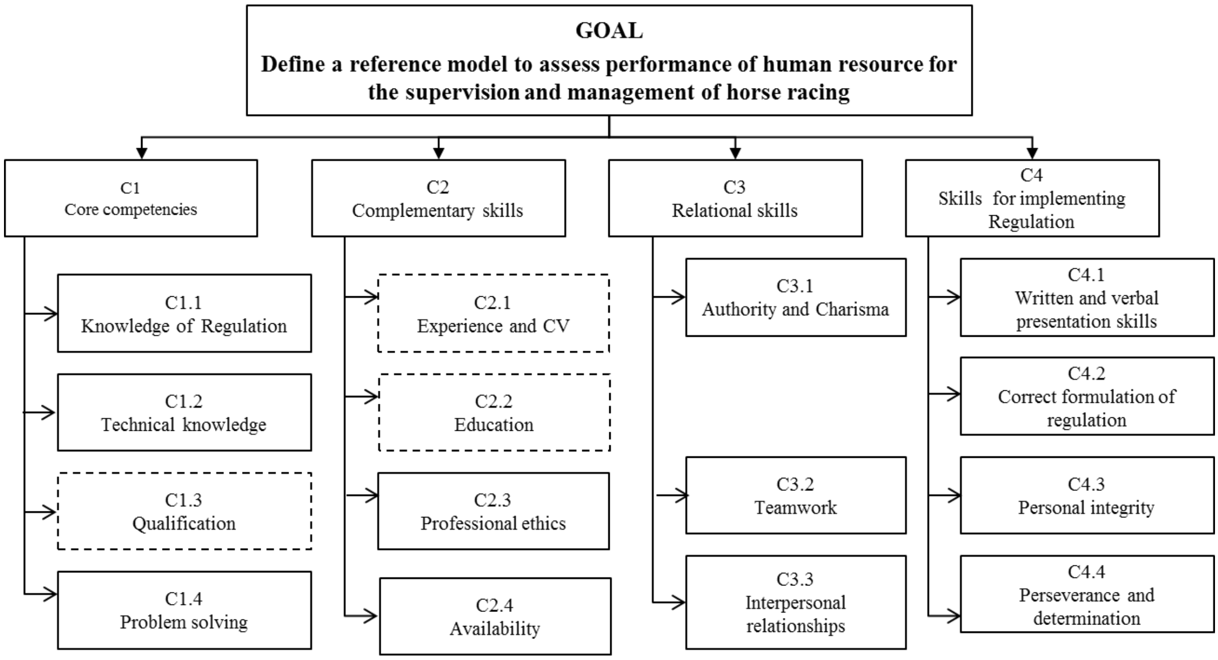


Figure 3: AHP Model

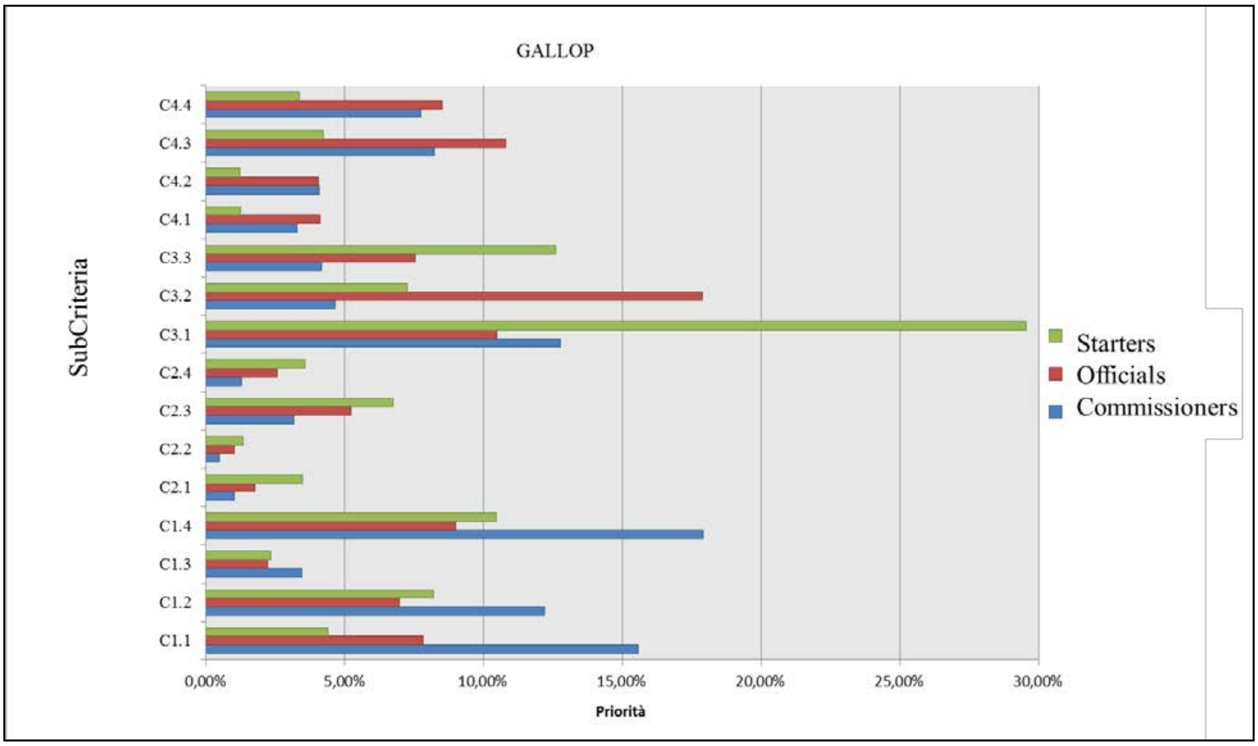


Figure 4: Final weights for subcriteria for each profiles- Gallop

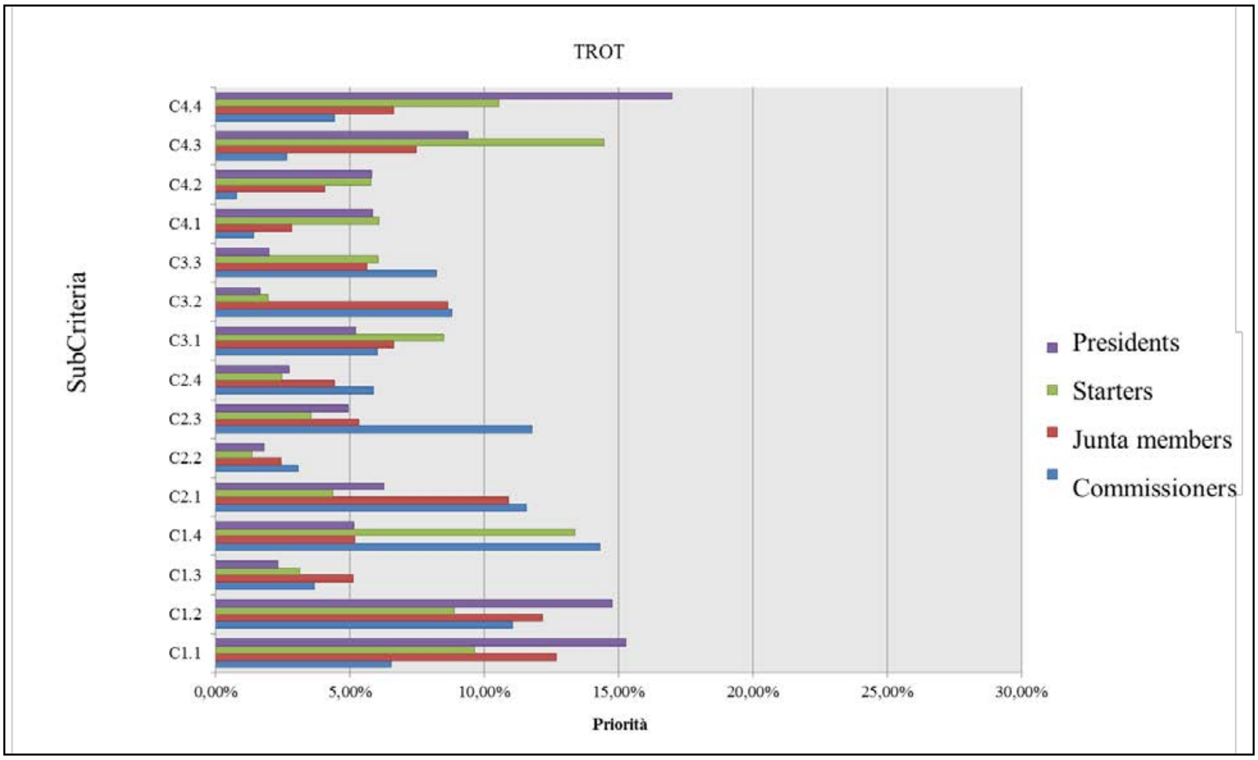


Figure 5: Final weights for subcriteria for each profiles- Trot

PROBLEMS OF INFORMATION SYSTEMS INTEGRATION IN LARGE TRANSPORT COMPANIES

Eugene Kopytov[(a)], Vasilijs Demidovs[(b)], Natalia Petukhova[(c)]

[(a), (b), (c)] Transport and Telecommunication Institute,
1 Lomonosova Str., Riga, Latvia
[(b), (c)] State Join-Stock Company "Latvian Railway"
21Turgeneva Str, Riga, Latvia

[(a)] kopitov@tsi.lv , [(b)] dem@ldz.lv , [(c)] natalia@ldz.lv

ABSTRACT
The article considers the issues of arranging the information systems (IS) interaction and their integration in large transportation company (exemplified by Latvian Railway Company). There presented the comparative analysis of integration technologies on three levels: on the level of data, on the level of messages and on the level of services. The Analytic Hierarchy Process (AHP) is employed as a tool of analysis. The evaluation of the efficiency of application of three IS integration technologies is presented for implementation in Latvian Railway Company.

Keywords: railway, information system, integration technology, Service-oriented architecture, Analytic Hierarchy Process

1. INTRODUCTION

The transport sector plays a leading role in the economy of any developed country; it consumes considerable labour, material and financial resources. In European Union the transport industry directly employs more than 10 million people, accounting for 4.5% of total employment, and represents 4.6% of Gross Domestic Product (GDP) (European Commission 2013). The search for optimal solutions in transport enterprises is greatly hindered by many factors influencing the processes of their activities. The following factors can be mentioned within them: high dynamics and speed of the processes of passenger- and freight transportation; high requirements towards reliability and safety of transportation and their regularity; substantial involvement of financial, labour and material resources; random demand for transportation; significant dependence on the variety of random factors, and first of all on meteorological conditions, large distances between related objects; a complex, far-flung routes network, and others. The above listed factors put forward their requirement towards the employment of computer engineering and information technologies for planning and managing the transportation enterprise activities. As a result, the transportation industry has a leading position among other economy sectors in implementing information technologies in its operations for many years. The recent decades are rich in constructing various IS in big transportation companies, and they are still operating along with the simultaneous implementation of new information systems (Ambrosino et al. 2010)..

Nevertheless, employment of IS based on different hardware platforms and implementing information processing technologies significantly complicates the process of their interaction with business planning and managing processes. It predetermines the importance of integrating the functioning IS for many transport companies. The procedure of integrating the IS typically gives rise to numerous problems, having their peculiarities for every company; the solution of these problems depends on the specific characteristics of employed information systems. It can be exemplified with the problems arisen in the process of integration of IS of transportation company "Latvian Railway".

Nowadays there used various IS integration technologies (Manouvrier and Ménard 2007). The choice of the optimal technology is rather complicated task and should be done with considering the peculiarities of both utilized IS and the area of their employment. It is worth mentioning that there are numerous researches oriented on substantiation of used integration technologies, see, for example (Bussler 2003; Krafzig et al. 2004; Sauser 2008; Sauser 2010). According to the authors' opinion, the choice of the best technology should be considered as a task of multi-criteria choice, taking into account the variety of indicators describing the efficiency of integration processes and entire IS.

The presented article considers the issues of arranging the IS interaction and their integration in big transportation company (exemplified by Latvian Railway Company). There presented the comparative analysis of integration technologies on three levels: on the level of data, on the level of messages and on the level of services. AHP method is employed as a tool of analysis. The assessment of integration technologies is considered for implementation of IS in Latvian Railway Company.

2. INFORMATION SYSTEMS OF THE LATVIAN RAILWAY

In spite of relatively insignificant operating length of the main railway tracks, which is under 2000 km., the freight turnover of Latvian Railway is more than 21 billion tonnes-km in 2011, and the prevailing freight turnover is East-West transit, taking more than 90% of total carried freight; and considering import and export, it is about 98% (Basic Performance Indicators 2012). The good coordination of activities is necessary for provision of this freight turnover; it should take place not only at the level of railway and its enterprises, involved in transportation process directly, but at the level of other railway administrations, as well as sea ports, providing transhipment of freights from marine transport to railway transport and vice versa. The eastern border of Latvian Republic is simultaneously the border of the European Union, and, consequently, Latvian Railway should cooperate with the united European custom system and governmental institutions. One of the Latvian Railway priorities is provision of the high-quality service for the customers of the railway. All abovementioned factors require the availability of various IS and interaction between them.

The first Computer Based Information System appeared at Latvian Railway in 1962. The computers belonging to Latvian State University and to Riga Civil Aviation Institute were used in its operations. The first ECM "Minsk 32" belonging to railway appeared only in 1969, and in 1972 the Information Technology Centre (ITC) of Latvian Railway has been created. Since that time the rapid growth of IS usage has started and it has resulted in availability of more than 100 big, medium sized and small IS at railway today (Kopytov, Demidovs, and Petukhova 2012). ITC provides 80 IT services for supporting the operation of railway. All employed IS were developed by and bought from different companies and in different periods of time; the different technologies, software and databases were used for their creation.

Nowadays ITC supports IS operating on various hardware (Mainframe, IBM pSeries, Oracle Sparc, Intel) and software platforms (Operation systems: z/OS , IBM AIX, Oracle Solaris, DOS, Windows and Linux; Databases: IBM DB2 (5 different versions), MS SQL (3 different versions), Oracle, Ingres, PostgreSQL, MySQL, DBF). There used various technologies of clusterisation (HCMP, MS Clustering services, MS Network Load Balancing Services, IBM WebSphere Network Deployment, Oracle Clusterware) and virtualisation (VMWare, Oracle, IBM). Figure 1 demonstrates the principal IS employed in the Latvian Railway Company, software platforms and spheres of their implementation. The official IS abbreviation is shown inside blocks. Presented IS are divided into three groups according to the number of users (marked by colour): minor (under 50 users), middle (from 50 to 200 users) and big (more than 200 users).

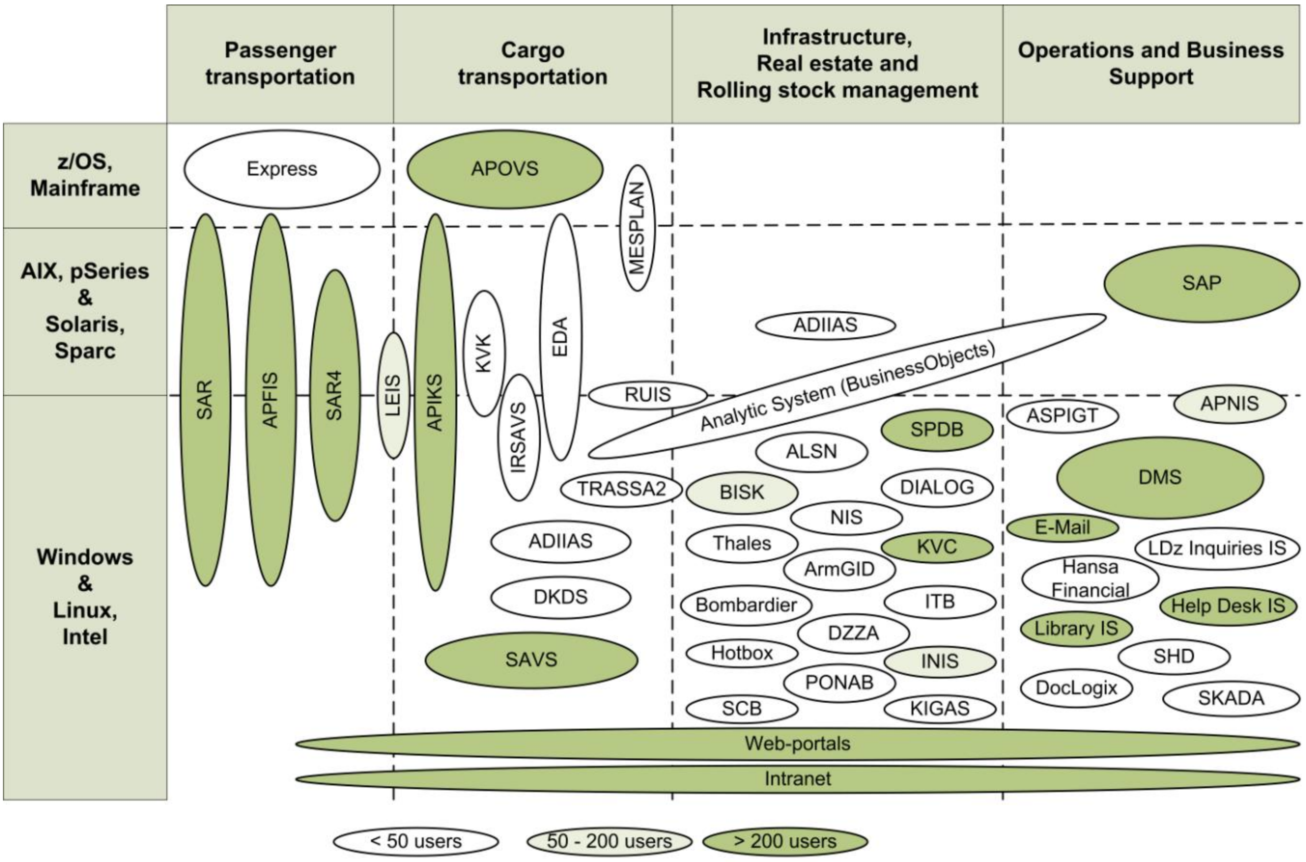


Figure 1: The principal IS in the Latvian Railway

This variety has impact on not only complexity of supporting these IS, but also on the problem of integrating these systems into the uniform information resource of an enterprise. The problem of integrating the information technologies resources at Latvian Railway exists for a rather long time. The attempts of integration of these systems employing various technologies and approaches were taken at all times. The approaches towards integration can be grouped in two directions: data-level integration and integration on the basis of business processes and messages. There is detailed consideration of these directions in IS integration at Latvian Railway.

3. INTEGRATION OF LATVIAN RAILWAY INFORMATION SYSTEMS

The principal integration procedures employed nowadays for providing the information exchange between information systems is asynchronous exchange with messages, for instance, on the basis of product IBM WebSphere MQ and files by scheme “point-to-point”. The direct addressing to the system database in a reading mode is used for the systems constructed with employment of DBMS with SQL support.

The first fundamental attempt of systems integration was taken in 1997; the extended analysis of available IS was provided and the concept of development of Latvian Railway IS for the period 1998-2002 was developed. The IS were grouped by the principal directions of railway activities: cargo transportation, passenger transportation, infrastructure, real estate, rolling stock, financial activities (see Figure 2). The concept presupposed centralisation of the IS on the basis of creating the high-speed network of data transmission, and integration of operational information systems (OLTP) on the basis of common database (see Figure 3), as a uniform resource of an enterprise with common system of railway classifiers, and simultaneous reduction of number of IS by integration of different IS with similar functionality into uniform IS (for instance, at that moment there were about 20 accounting departments in the company with different financial systems).

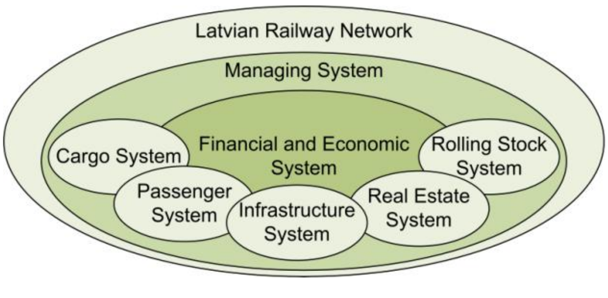


Figure 2: Main groups of IS in the Latvian Railway

Another important direction in integration of information technologies resources was development of the Data warehouse (Kopytov, Demidovs, and Petukhova 2003). This approach, oriented on creation of Decision Support System (DSS), was successfully implemented in IS for analysis and forecasting of the statistical and financial indicators of company activities in the sphere of passenger transportation. Nevertheless, the integration of OLTP systems requires another approach, namely interaction of IS at the level of business processes, implemented by these systems.

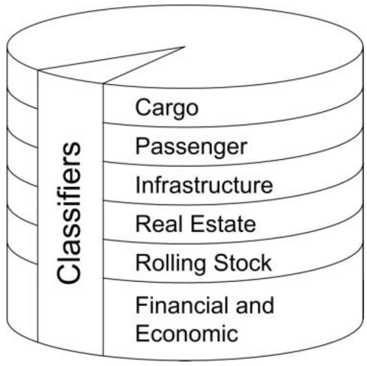


Figure 3: Common database for IS of Latvian Railway

Despite the performed work on integrating OLTP information systems during all these years, the number of problems with IS integration has increased. The total integration of IS on the basis of uniform integrated database was not completed. There can be noted several reasons of failure:

- The integration process was prolonged for more than ten years, and there was a need to implement the migration of newly developed systems to the latest versions of the Database Management System (DBMS). Unfortunately, problems with migration of some IS turned into appearance of even bigger number of databases with different versions;
- Entrance of Latvia into the European Union required prompt actions oriented on restructuring the railway, and consequently resulted in buying new IS and integration of IS not only within Latvian Railway but also together with European IS;
- Peculiarities of business of Latvian Railway, oriented on Russian and Belorussian directions, require information integration and interaction with Russian and Belorussian IS, which are promptly developing and changing;
- These peculiarities require buying or implementing of the finished IS, developed according to different technologies and implemented on various DBMS;
- Rapid dynamics of changing business environment, limited resources for development, and simultaneous necessity to improve the previous IS and to upgrade the new-developing IS for exchanging the old ones resulted in procrastination of the moment of transition to the new IS. There are examples of parallel implementation of two IS, the old and the new one, during the recent ten years.

Despite the above presented facts, IS in Latvian Railway have the minimum level of integration necessary for supporting the transportation process; nevertheless, this integration is very complicated since it employs different technologies, has different structural complexity and extremely difficult for support. Figure 4 demonstrates the instance of integration of several principal IS in the sphere of freight transportation in Latvian Railway, having the intended purposes as follows.

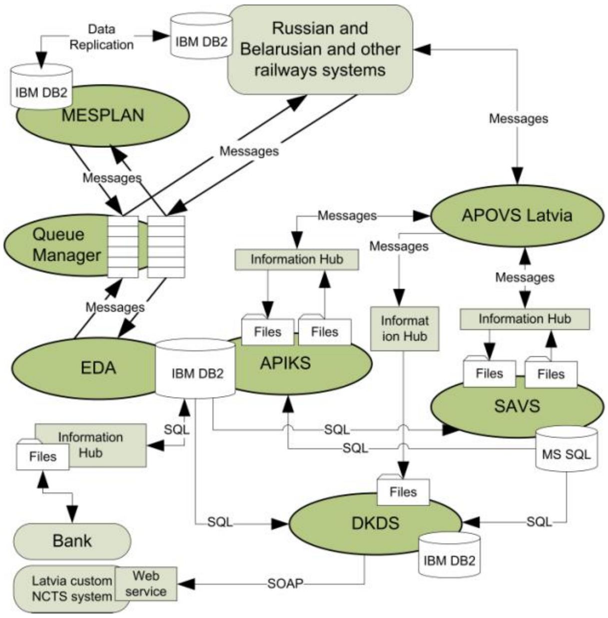


Figure 4: Integration of CARGO information systems

- APOVS is an Industrial Control System of freight transportation operational management. It is the basic system in the area of managing the transportation process. The system integrates the entire information about the procedure of transportation process as an aggregate of interconnected dynamic modules of all objects participating in the transportation process. APOVS comprises about six thousand programmes.
- APIKS is an Industrial Control System of controlling the freight transportation revenues; the system embraces the following modules: loading planning; handling the consignment notes; services sales records; arranging the bank documents; analytical information system; bookkeeping the materials – registration of accountable forms.
- EDA is the system of electronic data exchange (electronic consignment notes) between Latvian Railway and Russian Railways, and between Latvian Railway and Byelorussian Railways (according to MQSeries protocol).
- SAVS is an Industrial Control System of managing the stations.
- DKDS is an Industrial Control System of electronic reporting the railway freights in transit; the system complies with the peculiarities of Latvian Railway; it provides the information exchange within the frameworks of New Computerised Transit System (NCTS) between Latvian Railway and State Revenue Service following the agreement "On order of registering the custom procedure: transit of railway freight transportation" (NCTS, pan-European computer-based system of managing the transit freights, worked out on the basis of UN/EDIFACT standard for electronic documents circulation).
- MESPLAN is a system of arranging the month plan of freight transportation.

The considered example of IS integration implements four levels of integration: level of data, level of messages, level of services and mixed integration, which are presented below.

Integration on the level of data.

- Using the general database for several IS. Every IS implements its set of functions, and the data of these systems are saved in unified database with logical division. The data exchange between the IS is implemented via the common tables. The instance of such type of integration for EDA – APIKS is shown in Figure 2;
- Employment of SQL queries or stored procedures for data extraction. Every system has its own database, but there are special tables, views and stored procedures intended for access from other systems for data exchange within the database of these systems. Examples: APIKS – SAVS; and APIKS – DKDS;
- Data replication on the database level. Examples: MESPLAN of Latvia – MESPLAN of Russia;
- Exchange with files of .txt, .xml and so on types. Example: APOVS – APIKS; SAVS, DKDS, and APIKS – IS of the Bank.

Integration on the level of messages.

- System APOVS has its own API and a protocol of massages, used for massages exchange between APOVS and other IS (APIKS, SAVS, and DKDS);
- Employment of separate system of messages manager, for instance, IBM WebSphere MQ Series. Examples: EDA of Latvia – EDA of Russia; EDA of Latvia – EDA of Belorussia; and MESPLAN of Latvia – other IS of Russia;

Integration on the level of services.

- Employment of activations of web-services IS NCTS from DKDS (see Figure 4), providing the exchange with information within the frameworks of NCTS between Latvian Railway and State Revenue Service.
- Service-oriented architecture (SOA) (Cummins 2009; Krafzig et al. 2004). In year 2012 Latvian Railway initiated the project with employment of SOA technology for integrating information systems EDA, APIKS, SAVS, APOV, and MESPLAN. Another objective of the project is creating the portal via which the users can utilize the functions of these IS (see Figure 5).

Mixed integration.

- Integration on the level of messages and data (by implementing the files exchange). The exchange between the IS takes place with employment of the Information Hubs. Examples: APOVS – APIKS; SAVS, DKDS, and APIKS – IS of the Bank (see Figure 2.). Information Hub transforms the received massages from information system APOVS into files and places them on file-server in the directories of corresponding IS. Files, placed in directories for transmitting to APOVS, are transformed into the messages, which are sent to APOVS;

- Integration on the level of messages and data (employing SQL queries to the databases). Example: instead of transformation into files, Information Hub exchanges with SQL queries with IS APIKS.

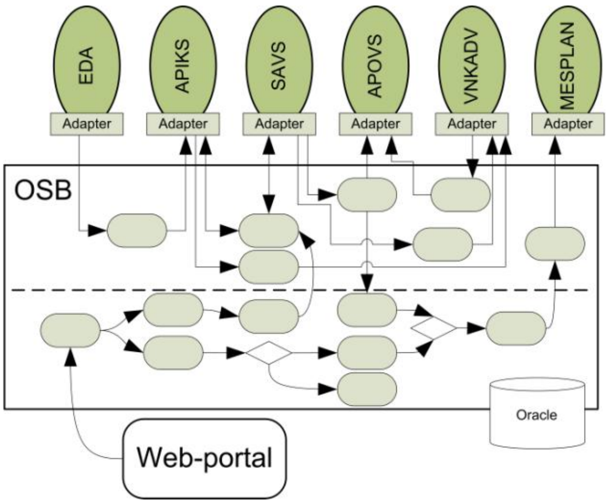


Figure 5: Integration of Latvian Railway IS with employment of Service-oriented architecture

Employment of various technologies in the process of IS integration in Latvian Railway is explained by the set of reasons, including objective and subjective ones; it also depends on the developers of integration projects. The authors suggest implementing the multi-criteria approach for comparative assessment of considered technologies; the approach is based on AHP method and described in the Section 5.

4. PROBLEMS ARISEN IN THE PROCESS OF INTEGRATING IS OF DIFFERENT CLASSES

Integration of IS of Latvian Railway, based on the exchange with messages and files, is in process for at least twenty years. But this integration is implemented between the independent systems with limited functionality and directed mainly on obtaining the necessary information from the IS source and subsequent processing and saving this information at special-purpose IS. Quite often some IS simultaneously serve as both: an information source and target information system. This interaction requires creating various interfaces for every IS involved into the integrating process. Typically these interfaces are limited with implementation of necessary functionality and they cannot be used for integration with other IS.

In 2012 Latvian Railway began the integration of IS with employment of SOA technology. Within the framework of the project it is planned to design information portal and integrator allowing not only obtaining the information important for the customers from available IS but also performing various activities within these systems via integrator equally to the users of these systems (see Figure 5). Integration of business processes on the basis of servers is very popular and well-developed technology, but the developers faced the certain problems at the very first steps towards integration.

The integration of IS was developed according to different technologies and have different users' interfaces without considering the possibilities of integration at the level of services and there is no Application Programming Interface (API) which can assist interaction with these systems at the software level. It seems to be rather strange, but the most suitable for integration are the IS which have evolved from the ones developed in the beginning of the 80es according to the centrally-distributed architecture. These IS have their own protocol for inter-computer interaction (inter-nodes) and for communication with users' terminals. This protocol was also used before for interaction with these systems according to the scheme "Point-to-point".

For integrating the IS without own API to the system kernel the authors have researched the approaches as follows:

- Development of API of the available information system;
- Organisation of interaction at the data-level;
- Development of functionality of this information system in the integrator environment.

The first approach is time- and money-consuming in its implementation, if there is a developer of the system. In case there is no developer any more, for example, the company ceased from the market or a developer does not work at the company any more, the problem of availability and actuality of source code can arise. In this case upgrade of the system can become impossible. Implementing this approach and the problems appearing in this process are considered by authors on the example of integration of information services provided for the passenger; they include: seats reservation, tickets buying, hotel reservation, car rent, and others. The peculiarities of integrating these IS lie in the fact that they were developed by different developers, they belong to different companies using different payment systems.

The second approach can result in the conflict with the developers of this IS who are tracking it. The developer can refuse to support the IS if the access to the software is direct but not via their software. There also can be the problems with blocking data at the database, appearance of phantom data, etc. This approach implementation at Latvian Railway can be exemplified by considering the system of support of freight transportation process.

The third approach implementation is also connected with additional time and money expenditures. It can also involve the problem with integrating the rewritten system with other IS previously integrated with this system. Implementation of this approach and associated problems are exemplified by development of new version of data processing system by customers of freight transportation, since actual information system operates for many years and has become obsolescent. However, new version of this system requires transformation of numerous other IS using the information accumulated at original information system in on-line regime. The compromise variant suggests itself; it supposes that new information system by customers of freight transportation, developed in the integrator environment, goes into service while the part of the previous IS continues performing simultaneously with it, supplying the numerous available systems with necessary data. The additional issue on synchronization of data between the new and the old systems appears. Evidently it requires the substantial financial expenditures.

5. COMPARATIVE ASSESSMENT OF INTEGRATION TECHNOLOGIES EFFICIENCY

In practice the search for an optimal integration technology for a particular set of IS should be performed taking into account the different criteria determining the efficiency of the integration technology on the whole. In present research the authors have been focused on choosing the better integration technology for Latvian Railway IS in the sphere of freight transportation considered in Section 3. This choice has been made taking into consideration the requirements of various categories of enterprise employers, including top managers, developers, supporting specialists, database administrators and IS users.

In the process of the criteria system formation the authors have been focused on five groups of criteria: Costs, Time, Structural Complexity, Information Security, and Technology Universality. 20 criteria have been selected as a result of investigation. The hierarchy of criteria used in the given research is presented in Figure 6.

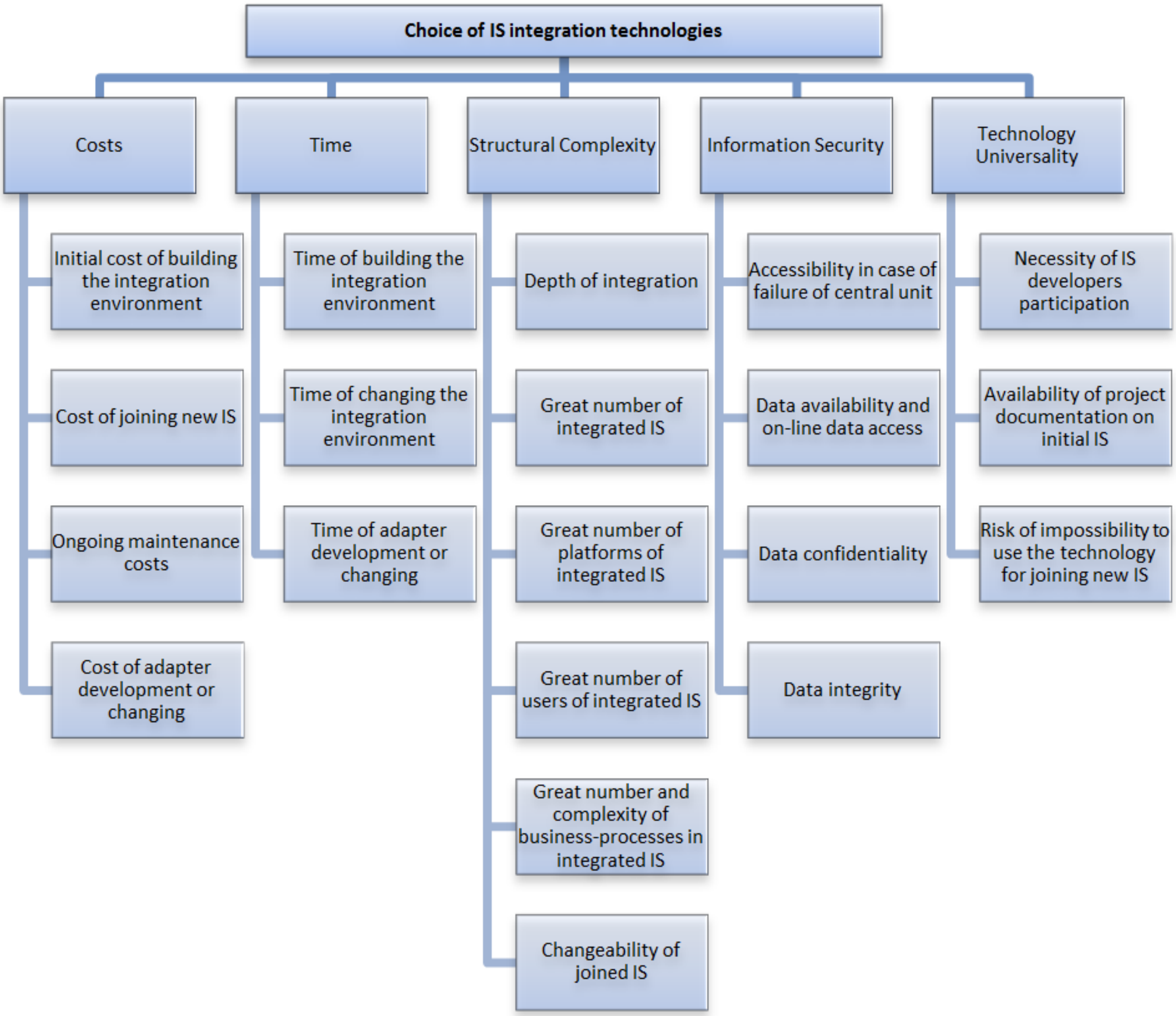


Figure 6: Hierarchy of the criteria for evaluating the considered IS integration technologies

- Group “Costs” considers expenses which the customer of the integration project has on the stages of creation, implementation and further operating the integration system;
- Group “Time” comprises temporal criteria influencing the terms of development and implementation of integration system;
- Group “Structural complexity” includes indicators describing the complexity of joined IS and the entire integration system;
- Group “Data Security” embraces the criteria, connected with the process of data access, data confidentiality, and data integrity;
- Group “Technology Universality” includes the indicators describing the dependence of technology on the initial projects of integrated IS.

To perform the calculations of criteria, the authors have used standard algorithms of the AHP method with the commonly used pairwise comparison scale from 1 to 9 (Saaty 2001). This scale has the following values: 1 – if two alternatives *A1* and *A2* are equal in importance; 3 – if *A1* is weakly more important than *A2*; 5 – if *A1* is strongly more important than *A2*; 7 – if *A1* is very strongly more important than *A2*; 9 – if *A1* is absolutely more important than *A2*; and 2, 4, 6, and 8 are intermediate values between the two adjacent judgments.

The summary data of the pairwise comparisons for the criteria of the first hierarchy level are presented in Table 1. It is quite visible, that experts from Latvian Railway gave high value to the criteria of groups “Data Security” and “Structural Complexity” in the procedure of choosing integration technology. Simultaneously the significance of groups “Costs” and “Time” criteria have substantially lower values. The experts’ evaluation is also reflected in calculated values of priorities vector; the priority of group “Data Accessibility” criteria is 0.3874; this value is by an order of magnitude higher than the priority 0.0362, calculated for group “Costs”. This fact supports the idea of urgency of integration problem solution for the company; the customers are ready to bear substantial expenses for obtaining the integration system of higher quality.

Table 1: Paired comparison matrix for criteria (first hierarchy level)

Group of criteria	Costs	Time	Structural Complexity	Data Security	Technology Universality	Priority vector
Costs	1	1/2	1/8	1/9	1/6	**0.0362**
Time	2	1	1/5	1/6	1/3	**0.0654**
Structural Complexity	8	5	1	1	2	**0.3364**
Data Security	9	6	1	1	3	**0.3874**
Technology Universality	6	3	1/2	1/3	1	**0.1745**

The criteria significance inside the criteria groups was evaluated by experts of different specialisation: top managers, systems developers, and IS supporting specialists. Table 2 presents an example of calculating the priorities of the second level criteria of group “Costs”. Obviously the experts give the highest significance to the indicator “Initial cost of generating the integration environment” with the highest priority 0.5450. Similar calculations were made for all other the second level criteria “Time”, “Structural Complexity”, “Data Security”, and “Technology Universality”.

Table 2: Paired comparison matrix for criteria “Costs” (second hierarchy level)

Criteria	Initial cost of building the integration environment	Cost of joining new IS	Ongoing maintenance costs	Cost of adapter development or changing	**Priority vector**
Initial cost of building the integration environment	1	4	5	3	**0.5450**
Cost of join-ing new IS	1/4	1	2	1/2	**0.1385**
Ongoing maintenance costs	1/5	1/2	1	1/3	**0.0837**
Cost of adapter development or changing	1/3	2	3	1	**0.2329**

Next step of assessment is calculating the matrices of evaluations of the priority vector for the suggested integration technologies based on the evaluation of the criteria priority vector of two levels of the hierarchy. Table 3 gives an example of the results of calculating the priorities of considered integration technologies for the second level criteria “Cost”. Similar calculations were made for all other the second level criteria.

To perform the verification of the correctness of judgments in the criteria evaluation, the consistency ratio (Saaty, 2001) has been calculated; its values are from 0.79% till 8.04 % for different groups of criteria. The values of consistency ratio under 10% indicate that the experts’ judgments are sufficiently consistent.

The final results of the evaluations of the global priority vector for the suggested integration technologies are shown below in Table 4 and Figures 7-8. The value of the global criteria priority for technology “Integration on the level of services” is **0.4208**, and it is significantly higher than the final evaluation of technologies “Integration on the level of data” and “Integration on the level of messages”, which criteria priorities are equal **0.3102** and **0.2374** respectively. So, technology “Integration on the level of services” can be recommended for usage.

Table 3: Matrix of evaluations of the vector of the criteria priorities of the "Costs" group for Integration technologies

Alternatives	Criteria				Priorities in group "Costs"
	Initial cost of building the integration environment	Cost of joining new IS	Ongoing maintenance costs	Cost of adapter development or changing	
	Numerical value of priority vector				
	0.5450	*0.1385*	*0.0837*	*0.2329*	
Integration on the level of data	0.7418	0.0852	0.7703	0.2583	**0.5407**
Integration on the level of messages	0.1830	0.6442	0.1618	0.6370	**0.3508**
Integration on the level of services	0.0752	0.2706	0.0679	0.1047	**0.1085**

Table 4: Evaluation results of integration technologies (for implementation in Latvian Railway IS)

	Criteria					Global priority vector
	Costs	Time	Structural Complexity	Data Security	Technology Universality	
	Numerical value of priority vector					
	0.0362	*0.0654*	*0.3364*	*0.3874*	*0.1745*	
Integration on the level of data	0.5407	0.3377	0.0851	0.5080	0.2472	**0.3102**
Integration on the level of messages	0.3508	0.3009	0.2500	0.1702	0.3149	**0.2374**
Integration on the level of services	0.1085	0.3614	0.6649	0.3219	0.4379	**0.4524**

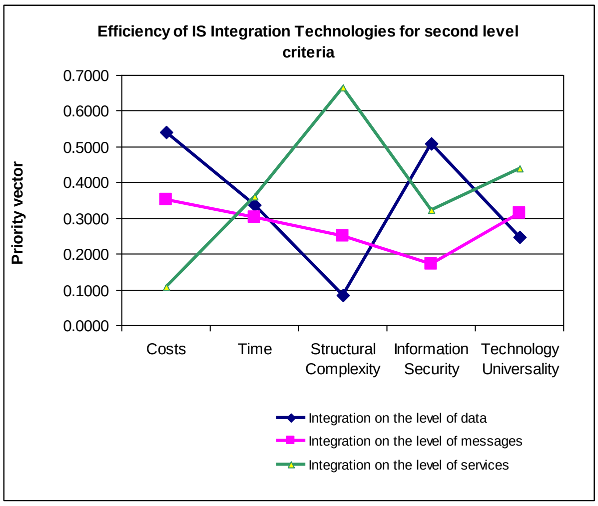


Figure 7: Evaluation results of integration technologies for groups of criteria (second level)

For groups "Structural Complexty", and "Technology University" the criteria values (priorities) of "Integration on the level of services" are greater than these criteria values of other technologies. The special note for "Integration on the level of services" should be given to the criteria "Structural Complexty" with value 0.6649. "Integration on the level of data" evaluation exceeds other technologies evaluation for groups "Cost" and "Information Security". In group "Time" all three technologies have practically the same values: 0.3371, 0.3009 and 0.3614. Consequently, the evaluation results show that the "Integration on the level of services" has the highest value of global priority and is recommended as the best integration technology for Latvian Railway IS in the sphere of freight transportation.

Nevertheless, the "Integration on the level of data" is recommended for an employment in case, when expenses (group of criteria "Costs") are very important.

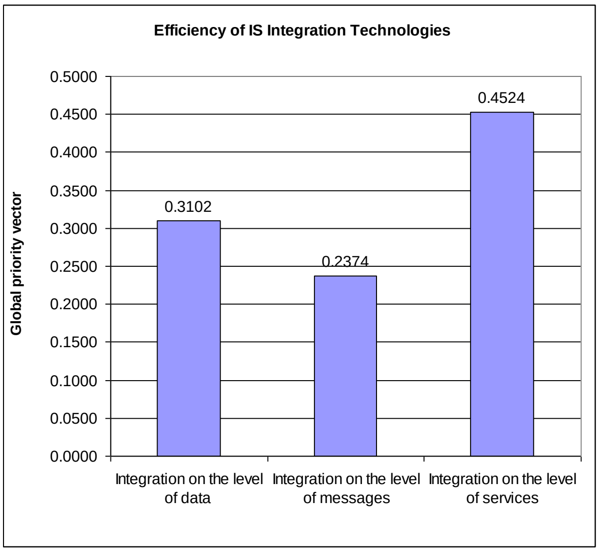


Figure 8: Evaluation results of integration technologies

CONCLUSIONS

The paper considers the evolution of development of Latvian Railway Company IS; there also have been determined the problems appearing in the process of integration. Despite the substantial investments in IS integration, allocated by company during long time, the integration of Company IS has not been finished yet. Separate integration tasks are solved with employment of various integration technologies, choice of which is often subjective and sometimes is not quite reasonable.

The paper suggests the multi-criteria approach for assessing and choosing the integral technologies. The implementation of this approach has been done with employment of Analytic Hierarchy Process. With the use of the AHP method this research fulfils the evaluation of the efficiency of application of three Integration technologies for the IS of Latvian Railway. To determine the optimal technology, a two-level hierarchy system of criteria has been developed with ranging expert evaluations.

It is worth mentioning that actually every big transportation company faces the majority of considered problems; that is why the results of the offered research can be implemented extensively. First of all it relates to the system of criteria for evaluating the comparative efficiency of integration technologies. Obviously, the expert evaluations of criteria significance can be considerably different depending on the investigated company.

REFERENCES

Ambrosino, G., Boero, M., Nelson, J.D., Romanazzo M., 2010. *Infomobility Systems and Sustainable Transport services*. Italian National Agency for New Technologies, Energy and Sustainable Economic Development: ENEA.

Basic Performance Indicators 2011. 2012 LDz: Economics Unit Marketing Communication Unit,. http://www.ldz.lv/texts_files/LDZ_darba_raditaji_2011_en_print.pdf [accessed 20 May 2013]

Bussler, C., 2003. *B2B integration: Concepts and architecture*, Springer.

Cummins, F.A., 2009. *Building the agile enterprise with SOA*, BPM & MBM, Elsevier.

European Commission: Mobility and Transport, 2013. http://ec.europa.eu/transport/strategies/facts-and-figures/transport-matters/index_en.htm [accessed 15 May 2013]

Kopytov, E., Demidovs, V., Petukhova, N., 2012. Improving of Railway Information Analytical Systems Using Newest Databases Technologies. *Proceedings of the 2nd International Conference on Stochastic Modeling Techniques and Data Analysis International Conference (SMTDA 2012*), pp.401-409. June 5 – 8, Chania, Crete Greece.

Kopytov, E., Demidovs, V., Petoukhova, N. Principles of Creating Data Warehouses in Decision Support Systems of Railway Transport, 2003. *Computing Anticipatory Systems: CASYS 2003 - Sixth International Conference*. *The American Institute of Physics, Conference Proceedings, 718*, 497-507.

Krafzig, D., Banke, K., Slama, D., 2004. *Enterprise SOA: Service-Oriented Architecture Best Practices*, Prentice Hall.

Manouvrier, B., Ménard, L., 2007. *Application integration: EAI, B2B, BPM and SOA*, Chippenham, Wiltshire: John Wiley & Sons Inc.

Saaty, T.L., 2001. *Decision making for leaders: The Analytic Hierarchy Process for decisions in a complex world.* Pittsburgh, PA: RWS Publications.

Sauser, B.J., Ramirez-Marquez, J.E., Magnaye, R. and Tan W., 2006. A systems approach to expanding the technology readiness level within defense acquisition," *International Journal of Defense Acquisition Management, 1*, 39-58.

Sauser B.J., Gove R., Forbes E. and Ramirez-Marquez J.E., 2010. Integration maturity metrics: Development of an integration readiness level. *Information Knowledge Systems Management, 9*.17-46.

AUTHORS BIOGRAPHY

EUGENE A. KOPYTOV was born in Lignica, Poland and went to the Riga Civil Aviation Engineering Institute, where he studied Computer Maintenance and obtained his engineer diploma in 1971. Candidate of Technical science degree (1984), Kiev Civil Aviation Engineering Institute. Dr.sc.ing. (1992) and Dr.habil.sc.ing. (1997), Riga Aviation University. Professor (1998). Present position: Head of Department of Computer Science of Transport and Telecommunication Institute, professor. Member of International Telecommunication Academy. Fields of research: statistical recognition and classification, modelling and simulation, modern database

technologies. Publication: 290 scientific papers and teaching books, 1 certificate of inventions.
His e-mail address is: kopitov@tsi.lv

VASILIJS V. DEMIDOVS was born in Murmansk, Russia and went to the Riga Polytechnic Institute (Riga Technical University), where he obtained his engineer diploma in 1987. He studied in Transport and Telecommunication Institute under the Doctor's program from 2001 to 2004 and obtained the scientific degree of Doctor of Science in Engineering (Dr.sc.ing.) in 2006 by scientific research direction “Transport and Logistics”. He has got a docent since 2007. Present position: a Head of Database Management System Unit of Information Technology Centre of State Joint-Stock Company “Latvian Railway”; docent of Computer Science Department of Transport and Telecommunication Institute. Fields of research: modern database technologies, Business Intelligence, Data warehouse technologies. Publication: 33 scientific works in Belgium, Estonia, Italy, Latvia, Lithuania, Russia, USA, including 20 scientific papers.

His e-mail address is: dem@ldz.lv

NATALIA Y. PETUKHOVA was born in Riga, Latvia. She studied Computer Sciences in Transport and Telecommunication Institute and obtained her MS in 2001, obtained the scientific degree of Doctor of Science in Engineering in 2011 by scientific research direction “Transport and Logistics”. Present position: leader specialist and DB administrator of Database Management Systems department of Computer Centre of Latvian Railway; docent in Transport and Telecommunication Institute. Publications: 22 scientific papers. Scientific activities: databases, data warehouses, business intelligences.

Her email address is: natalia@ldz.lv

SIMULATING CONTINUOUS TIME PRODUCTION FLOWS IN FOOD INDUSTRY BY MEANS OF DISCRETE EVENT SIMULATION

Fabio Bursi[(a)], Andrea Ferrara[(b)], Andrea Grassi[(c)], Chiara Ronzoni[(d)]

[(a,b,c)] Dipartimento di Scienze e Metodi dell'Ingegneria
Università di Modena e Reggio Emilia
Via Amendola 2, 42122 Reggio Emilia, Italy

[(a,b,c,d)] Logistics & Automation Consulting srl
Via G. V. Catullo 22, 42124 Reggio Emilia, Italy

[(a)]fabio.bursi@unimore.it ,
[(b)]andrea.ferrara@unimore.it,
[(c)]andrea.grassi@unimore.it,
[(b)]chiara.ronzoni@lac-consulting.eu

ABSTRACT

The paper presents a new framework for carrying out simulations of continuous-time stochastic processes by exploiting a discrete event approach. The application scope of this work mainly refers to industrial production processes executed on a continuous flow of material (e.g. food and beverage industry) as well as production processes working on discrete units but characterized by a high speed flow (e.g. automated packaging lines).
The proposed model, developed adopting the DEVS formalism, defines a single generalized base unit able to represent, by means of an event scheme generated by state changes, the base behaviors needed for the modeling of a generic manufacturing unit, that is, (i) breakdowns and repairs, (ii) speed and accumulation, and (iii) throughput time. Moreover, the possibility to keep trace of additional measures of parameters related to the process and the flowing material (i.e. temperature, concentration of pollutant, and so on) is also considered. Since these parameters can change over time in a continuous manner with respect to some laws that depend on contingent conditions, the possibility to transmit those laws as functions is introduced in the model.

Keywords: continuous flow simulation, DES, DEVS.

1. INTRODUCTION

It is widely known that simulation is effectively applied in industry to address design and management issues in complex production systems that cannot be easily represented mathematically.

The use of simulation delivers added value to customers both in the deployment of a new production plant as well as in analyzing existing ones. Typical applications regards the definition of work centers capacity and buffers dimensioning and location, of process control rules, of layout configuration, accomplishing with specific design target in terms of system performances also considering different scenarios. Specially, Discrete Event Simulation (DES) has been widely adopted in the industrial context thanks to the advantages deriving from the discretization of time, that is, the possibility to speed up computation time and to ease model building activity.

As a consequence, in last decades the birth and the development of numerous discrete event simulators has been seen. However, there are industrial contexts of great relevance for which the discrete event simulation is not the best approach to represent the system since approximations are typically requested to keep the simulation time in line with industrial requirements. These sectors are, for instance, fluid processing (e.g. food and beverages industry) or high-speed automated lines (e.g. packaging lines) that, for the high processing speed, the system behaves as the same as it were a fluid process.

While a large number of works addressing the discrete event simulation of manufacturing systems have been produced in years by scientists (see for a comprehensive review Jahangirian et al. 2010), only recent papers have focused on the problem of defining simulation models able to consider the production flow as a if it were a fluid, that is, continuous (Praehofer 1991, Tamani et al. 2009). Hence, further studies to develop new approaches for the simulation analysis of production system adopting a continuous flow approach are of interest for both academics and practitioners.

The aim of this paper is to present a new modeling framework for the simulation of flow manufacturing processing following an approach that aims to reproduce the behavior of a continuous-time stochastic process. The innovation of the proposed paper resides in the definition of a generalized model able to represent a continuous-time process by using a discrete event approach, in which the events are signals related to state changes of the simulation units. The structure of the basic model is so that all of the minimum requirements needed for modeling an industrial process are met and, moreover, the possibility to consider continuous functions for process parameters is also introduced. This results in a very efficient and scalable modeling framework.

The Discrete Event system Specification (DEVS) formalism, firstly introduced by Zeigler (Zeigler 1976, Zeigler 1984, Zeigler et al. 2000), is used in this paper to define the base unit model. DEVS allows the development of robust model representation based on the concept of atomic models and on the concept of higher-level models coupling. Several applications of DEVS for the definition of models for simulating manufacturing systems have been presented in literature. Among them, interesting contributions are Giambiasi and Carmona (2006) and Pujo et al. (2006).

The remaining of the paper is organized as follows. Section 2 defines the problem statement, while Section 3 develops the model for the base unit. Finally, Section 4 provides concluding remarks.

2. PROBLEM STATEMENT

The aim of this paper is to present a new modeling framework for the simulation of flow manufacturing processing following an approach that aims to reproduce the behavior of a continuous-time stochastic process.

This aspect is of particular interest for modeling manufacturing processes acting on a continuous flow of material (i.e. food and beverage industry), as well as processes working on discrete units but flowing at a high rate (e.g. packaging lines). For the latter, discrete event simulation is typically adopted to carry out performance analysis and to address design tasks, involving a huge overhead in terms of computation time. In fact, simulating a high capacity production line by means of a discrete event approach involves the need to manage a large number of events just to represent the flowing of the units.

Conversely, a continuous-time stochastic approach determines states and transitions probabilities, considering the manufacturing process as it were working on a continuous flow. Typically, mathematical modeling is used to model the system and to obtain the closed form solutions. The base unit model is the so-called two-machines one-buffer building block, whose first models were proposed by Zimmern (1956), Gershwin and Schick (1980), Yeralan and Tan (1997), and that obtained several improvements in years to enhance its capability to represent the behavior of real systems (Tan and Gershwin 2009, Tolio 2011, Tan and Gershwin 2011, Gebennini et al. 2011, Gebennini and Gershwin 2013). This building block is able to represent a simple series of two machines (a simple line) in which the decoupling effect of a buffer is also considered, and then it is the minimal requirement need to model a flow based manufacturing system. To model more complex systems, i.e. lines with more than one buffer, decomposition techniques have been introduced (Tan and Yeralan 1997, Gershwin and Burman 2000, Levantesi et al. 2003).

The approach proposed in this paper uses a discrete event mechanism to reproduce the behavior of a continuous-time stochastic process. To reach this goal, the base unit model (see Figure 1) has been conceived so as to manage signals, coming from the other connected units, that are delivered following a discrete event scheme. A signal transmits information about a state change in the upstream or in the downstream, then producing changes in the internal states and parameters of the unit itself. In this way, the need to model the very production flow is avoided, and then computational time is saved, while the accurate behavior of the system is granted by the transmission of the only signals needed to determine state changes.

To be able to represent production processes in industry, the basic modeling unit has been engineered so as to be able to represent three basic behaviors with which the most general real-world working unit can be modeled. In particular, those behaviors are:

1. failures and repairs;
2. working speed and accumulation;
3. throughput time.

Failures and repairs represent the operational state of the unit and are related to the Time-To-Failure (TTF) and the Time-To-Repair (TTR) profiles, that typically are random variables. Those random variables determine the occurrence of breaking and repair events, then putting the unit in down and up states, respectively. Working speed and accumulation make it possible to model changes in working speed as a consequence of state changes in the upstream and the downstream, on one side, and in the internal accumulation level on the other side. Moreover, internal accumulation level can involve changes in working speed, and this is the reason why those two aspects have to be jointly considered. It has to be pointed out that accumulation is here related to decoupling capability, so the capability of the unit to vary its content of material from zero to a maximum value. The throughput time represents a delay that has to be applied to a signal exiting from the downstream and generated as a consequence of an signal entering from the upstream.

Referring to Figure 1 the three aforementioned basic behaviors have to be considered in the reported sequence (1, 2, and 3), given the implicit interdependence among them. By means of those three basic behaviors, we are able to produce a general base unit model that can represent the main categories of working units found in real applications, such as:

- work centers continuously operating on the flow, characterized by a specific maximum production speed, zero accumulation and zero throughput time;

- buffers, characterized by a maximum speed, an accumulation greater than zero, and a throughput time depending on the buffering strategy (FIFO, LIFO, mixing, etc.);
- conveying units (i.e. conveyors, belts, pipes, etc.), characterized by a fixed speed, zero accumulation, and a throughput time depending on the length and the speed.

Summarizing, the basic behavior allows for the correct representation of the basic operation of a generic line, making also possible the computation of the classical performance measures such as throughput, efficiency, stay time in different status, and so on.

Moreover, the modeling approach here proposed provides the capability to include additional parameters whose values need to be tracked during the simulation and along the production flow. Classical examples, referring to the food industry, are the temperature of the product, or the concentration of pollutant substances that can be generated by some unwanted situations in a process unit and then propagate in some way along the production flow. The need to track those kind of parameter is clear since the use of a flow simulator in food industry is mainly related to the definition of control policies in processes and of product traceability strategies, as well as product waste and net efficiency estimation.

Hence, the modeling approach proposed in this paper allow for the addition of every parameter to monitor the user need to trace, thanks also to the a general scheme to define interactions among the monitored parameter and the basic behaviors. Moreover, the model has been conceived so as to allow the exchanging of functions between units, rather than simple values, by means of which parameter values can be calculated as a function of time without the need to generate additional events for updating parameter values.

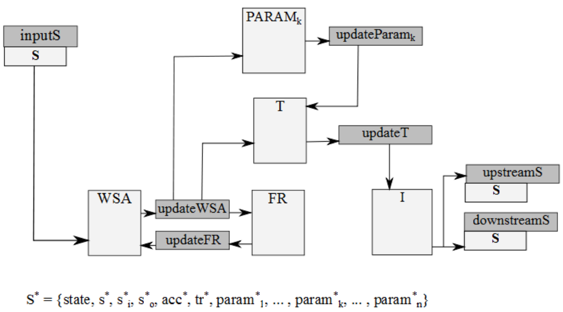


Figure 1: The Base Unit Model.

In other words, we have also adopted the continuous-time approach to model parameters value variations.

To understand the importance of this last aspect we can refer to a typical case that can be found in food industry, that is, an accumulation unit working with a pure mixing strategy (i.e. a tank) having an entering product flow coming from an upstream process and an exiting flow sent to a downstream process. If, at a certain time, the upstream process start to fail producing a product characterized by a constant concentration of pollutant, the concentration of pollutant inside the accumulation unit begins to change over time following a law that depends on the amount of product present in the unit and the flow of pollutant entering the unit itself. As a consequence, the product sent to the downstream by the accumulation unit will be characterized by a concentration of pollutant following the same law, thus continuously varying over time. From here the need to allow the possibility to transmit functions among modeled units. If we can transmit functions, we have only to generate events related to a state change and then transmitting the new functions to trace parameter values, thus avoiding the need to generate polling events just to update the values of those parameters that are varying over time following continuous laws.

3. THE BASE UNIT MODEL

3.1. Base unit object

3.1.1. Informal description

The base unit model is composed by three atomic models representing the behaviors (i) failures and repairs, (ii) working speed and accumulation, (iii) throughput time, plus one interface and n additional parameter models. This purpose of this section is to illustrate the interaction between them.

The Figure 1 shows the interactions between external signals, coming from the upstream and the downstream flows, and internal signals. The base unit is composed by standard objects and a interface whose task is to generate output signals. Considering a simple flow system, the base unit is provided with one input and two output ports for external signals, in order to send system variations both upstream and downstream the flow.

Forward signals processed among internal objects follows a static logical scheme represented by a matrix of dependencies (see Figure 2) composed by:

1. the minimal matrix managing signals among standard objects;
2. the external signals;
3. the set of rows and columns managing additional parameters signals.

	FR	WSA	T	Interface	Output	$Param_1$	…	$Param_k$	…	$Param_n$
Input	0	1	0	0	0	0	0	0	0	0
FR	0	1	0	0	0	0	0	0	0	0
WSA	1	0	1	0	0	1	1	1	1	1
T	0	0	0	1	0	0	0	0	0	0
Interface	0	0	0	0	1	0	0	0	0	0
$Param_1$	0	0	1	0	0	-	-	-	-	-
…	0	0	1	0	0	-	-	-	-	-
$Param_k$	0	0	1	0	0	-	-	-	-	-
…	0	0	1	0	0	-	-	-	-	-
$Param_n$	0	0	1	0	0	-	-	-	-	-

Figure 2: The Dependency Matrix.

The vector S^* represents the internal set of data that are used for local storing purposes.

The signal generated by the failures and repairs object (FR) is related to a state change due to the occurrence of a TTF or of a TTR, that is, a change of the operative conditions of the unit. A signal (updateFR) is then sent to the working speed and accumulation object (WSA) for updating purposes. The opposite signaling direction (updateWSA) is also introduced, since a change in the speed of the unit can imply a variation in the way the TTF is consumed (i.e. if the failures are operation dependent or not). Updates in WSA conditions are also sent to the additional parameter objects (ADD_k) to allow the update of their values and functions.

The throughput time object (T) acts as a sort of final gateway by which every signal has to pass before being sent to the external units. The reason is that the object T is devoted to the computation of the delay with which external units located in the downstream and in the upstream realize that something has changed in the considered unit. Finally, the interface object (I) takes care of the broadcasting of the signal to the downstream and the upstream.

3.1.2. Formal description

$$\text{'base unit'} = < X_{bu}, Y_{bu}, T_{bu}, \delta_{ext}, \delta_{int}, \lambda, t > \qquad (1)$$

Input event variable Xbu = (inputS) where:

- ‘inputS’ = S where S = { f , s, $param_1$, ... , $param_k$, ... , $param_n$}
  - *f* ={*d*,*u*} is the flow parameter, where ? *f* = ‘*d*’ represents a message coming from a downstream unit while ? *f* = ‘*u*’ represents a message from an upstream unit;
  - *s* is the working speed;
  - *paramk, for k = 1, ... ,n, is the function describing the variation of the additional parameter k over time.*

Output event variables Y_{bu} = (*downstreamS*,*upstreamS*) where:

- ‘downstreamS’ = {‘u’, s*,param1, ... ,paramk, ... , paramn }
  - ‘*u*’ indicates the unit in the downstream that the signal is coming from an unit in its upstream;
  - s^* is the value of the actual working speed;
  - $param_k$, for *k* = 1, ... ,*n*, is the function describing the actual law of variation of the additional parameter *k* over time.
- ‘upstreamS’ = {‘*d*’, s^*, $param_1$, ... , $param_k$, ... , $param_n$}
  - ‘*d*’ indicates the unit in the upstream that the signal is coming from an unit in its downstream;
  - s^* is the value of the actual working speed;
  - $param_k$, for *k* = 1, ... ,*n*, is the function describing the actual law of variation of the additional parameter *k* over time.

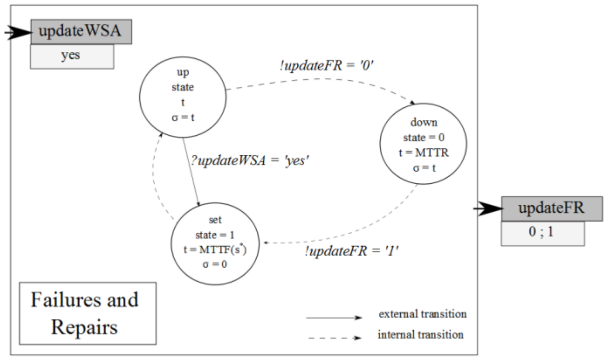


Figure 3: The Availability Object.

3.2. Failures and repairs object model

3.2.1. Informal description

The failures and repairs object (FR) models the base unit availability, switching the system from ‘up’ state to ‘down’ state and vice-versa (see Figure 3).

Let us describe the behavior in the case of the unit entering the ‘up’ state. Considering the ‘set’ phase, the base unit is set in a working state, meaning state variable set to ‘1’, t variable set to the result of TTF profile function and s to t. After that initialization, the system switches to the ‘up’ state. Since then, two cases can happen:

1. a working speed chage signal arrives from the WSA object;
2. an internal state change happens.

In the first case, the incoming signal represents a notification of a speed change and it makes the system jump from the state ‘up’ to the state ‘set’ in order to update the TTF function. Since this is a transitional state, the system returns immediately to the previous state, until the amount of time s (the TTF) has passed. After this time, an internal transition from the ‘up’ to the ‘down’ state happens.

In the second case, the state is set to ‘0’ and σ is set to the TTR value. An internal update message is created in order to notice this operative change. After σ has passed, the system makes an internal transition from the ‘down’ state to the ‘set’ state where a new TTF is computed and the state variable is set to ‘1’. After that, the system switches to the ‘up’ state for a time equal to σ. By changing the state from ‘down’ to ‘set’ an internal output signal is created. All internal signals generated are sent to the WSA object.

3.2.2. Formal description

$$\text{'}\boldsymbol{failures\ and\ repairs\ object}\text{'} = < X_{fr}, Y_{fr}, T_{fr}, \delta_{ext}, \delta_{int}, \lambda, t > \qquad (2)$$

Input event variable: X_{fr} = (*updateWSA*) where:

- 'updateWSA' = {yes} indicates a work speed and accumulation variation.

State variables: T_{fr} = (phase, state, t, σ) where:

- 'phase' = {up,down, set} is a name representing the situation in the real world;
- 'state' = {0,1} represents the availability of the base unit, where ?state = '0' means the base unit is down and needs to be repaired, and ?state = '1' means the base unit is up;
- ' f ' ={TTR,TTF} represents the time to repair and the time to failure functions, that are probability distribution functions also considering the speed s^* for TTF consuming computations;
- $\sigma \in \mathbb{R}^+_0 \cup \infty$ is the life time of the current state.

Output event variable: Y_{fr} = (updateFR) where:

- 'updateFR' = {0,1} indicates a variation in the system operativity.

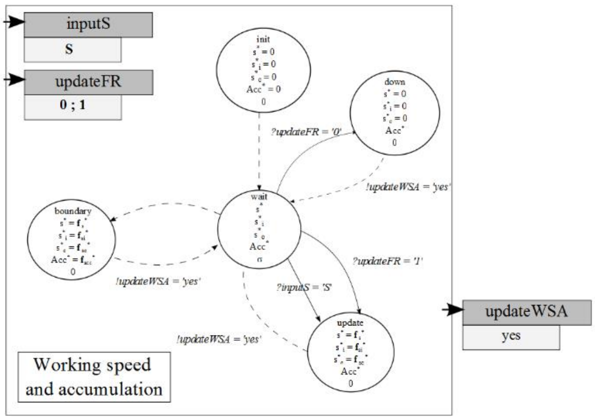


Figure 4: The Working Speed And Accumulation Object.

3.3. Working speed and accumulation object model

3.3.1. Informal description

The purpose of this object (see Figure 4) is to model the operative conditions of the base unit system. Considering the 'init' state as initialization of the object variables, all parameters are set to '0' and, by means of an internal transition, the system jumps in the 'wait' state. Thus, the working speed and accumulation object receives an !updateFR='1' and the system switches to the 'update' state: s^*_i , s^*_o , s^*,and Acc^* variables are updated to the new calculated values obtained by reading previously stored values of s*i , s*o , s*,and Acc*. In this case, the base unit restarts calculating parameters using the stored ones.

After the update have been done, the system switch to the 'wait' state creating an output signal !updateWSA = 'yes' in order to notice the changes made. The same path is followed also when there is an external signal ?inputS = 'S' entering the unit, where information about speed changes, both upstream and downstream, are stored and then used to update the unit operative condition.

As a transitional state, the system switch to the 'wait' state and remain in this state for a time that is the minimum value between infinite and t_b, the time when the base unit will be at one accumulation boundary. Boundaries are possible only when the base unit is configured as a buffer, so that boundaries are the empty level of the buffer and a level of the buffer equal to its capacity.

At the boundary state, the actual level of the buffer is calculated (empty or full). The 'boundary' state is transitional, so that the system sets its parameters, creates an internal output signal to notice the changes and then switch into the 'wait' state in order to wait that a change signal occurs. After receiving an !updateFR = '0' the system switches from the 'wait' state to the 'down' state and sets all its parameters to '0'. As a transitional state, the system switches to the 'wait' state after the creation of an internal output in order to notice an operative change.

3.3.2. Formal description

$$\text{'}\boldsymbol{working\ speed\ and\ accumulation\ object}\text{'} = \ < X_{WSA}, Y_{WSA}, T_{WSA}, \delta_{ext}, \delta_{int}, \lambda, t > \qquad (3)$$

Input event variable: X_{WSA} = (inputS,updateFR) where:

- 'inputS' = S where S = { f, s, $param_1$, ... , $param_k$, ... , $param_n$}
  - f ={d,u} is the flowparameter, where ?f ='d' represents amessage coming froma downstream unit while ?f = 'u' represents a message from an upstream unit;
  - s is the working speed;
  - $param_k$, for k = 1, ... ,n, is the function describing the variation of the additional parameter k over time.
- 'updateFR' = {0,1} the internal signal that is generated as output by the failures and repairs object, where ?updateFR = '0' indicates that the state of the base unit is down and ?updateFR = '1' indicates that the state of the base unit is up.

State variables: T_{WSA} = (phase, s^*, s^*_i, s^*_o, acc*, s) where:

- 'phase' ={init,down,wait,update,boundary} is a name representing the situation in the real world;
- 's*' = fs () represents a function which calculates the base unit working speed

considering the upstream and downstream base units working speed, the maximum working speed of the base unit, and the accumulation level.

- ‘s*i’ = fsi () represents a function which determines the upstream base unit working speed.
- ‘s*o’ = fso () represents a function which determines the downstream base unit working speed.
- ‘acc∗’ = facc () represents a function which determines the accumulation level considering s*, s*i, s*o and the previous accumulation level.
- $\sigma \in \mathbb{R}+0 \cup \infty$ is the life time of the current state.

Output event variable: Y_{WSA} = (updateWSA) where:

- ‘updateWSA’ = {yes} indicates a working speed and accumulation variation.

3.4. Throughput time object model

3.4.1. Informal description

By referring to Figure 5, in the initial state ‘init’, the throughput time function tr is set for the first time. This function depends on the base unit speed, accumulation and the base unit state. Later, the object waits for an input signal in order to update its function. This function can be updated only when the object receives at least one of the two input signals. These two signals are sent by an additional parameter object or by the working speed and accumulation object. Finally, the object is able to send a signal to the Interface object.

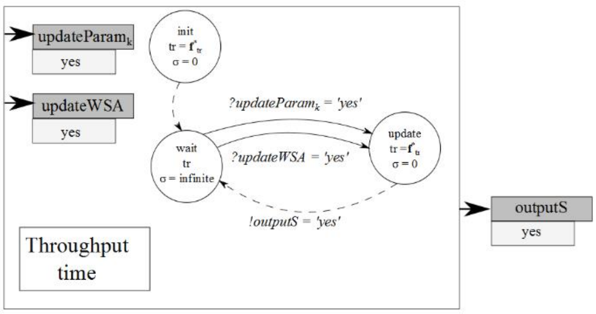


Figure 5: The Throughput Time Object.

3.4.2. Formal description

$$\text{‘throughput time object’} = \ < X_{tr}, Y_{tr}, T_{tr}, \delta_{ext}, \delta_{int}, \lambda, t > \quad (4)$$

Input event variable: X_{tr} = (updateParamk ,updateWSA) where:

- ‘updateParamk’ = {yes} indicates a variation in the k-th additional parameter function;
- ‘updateWSA’ = {yes} indicates a work speed and accumulation variation.

State variables: T_{tr} = (phase, t_r, σ) where:

- ‘phase’ = {init,wait,update} is a name representing the situation in the real world;
- ‘tr’ = f ∗tr (s∗,acc∗, state) is the function that sets the throughput time and depends on:
  - the base unit speed s*;
  - the current accumulation acc*;
  - the base unit state.
- $\sigma \in \mathbb{R}+0 \cup \infty$ is the life time of the current state.

Output event variable: Y_{tr} = (outputS) where:

- ‘outputS’ = {yes} indicates a signal to be sent to the Interface object.

3.5. Interface object model

3.5.1. Informal description

By referring to Figure 6, in the initial state ‘init’, the output signal is set to zero. When the interface receives a signal from the throughput time object, S is updated. For each input signal received, two different signals are generated: one is sent to upstream and the other one to the downstream. The first signal is S = {‘d’, s*, param*}, while the latter is S = {‘u’, s*, param*}. It means that the Interface object sends:

- ‘flow’ = {‘d′, ‘u′}, that is, the information to the upstream and the downstream with the position reference of the considered base unit;
- the speed of considered base unit;
- the additional parameter functions of the considered base unit.

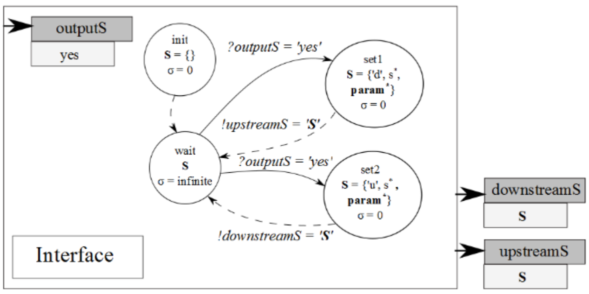


Figure 6: The Interface Object.

3.5.2. Formal description

$$\textbf{‘}\boldsymbol{interface\ object}\textbf{’} = \ < X_{int}, Y_{int}, T_{int}, \delta_{ext}, \delta_{int}, \lambda, t > \quad (5)$$

Input event variable: X_{int} = (outputS) where:

- ‘outputS’ = {yes} indicates the reception of a signal from the throughput time object.

State variables: T_{int} = (phase, S, σ) where:

- ‘phase’ = {init,wait, set1, set2} is a name representing the situation in the real world;
- S = { f low, s*, param*} is the output signal and it consists of:
  - ‘flow’ = {‘u′, ‘d′}, ‘u′ indicates that the considered base unit is the upstream of the destination base unit, while ‘d′ indicates that the considered base unit is the downstream of the destination base unit;
  - s*, the speed of the considered unit;
  - param*, the additional parameter functions of the considered base unit.
- $\sigma \in \mathbb{R}\ +0 \cup \infty$ is the life time of the current state.

Output event variable: Y_{int} = (upstreamS, downstreamS) where:

- ‘upstreamS’ = {yes} indicates a signal to be sent to the upstream base unit;
- ‘downstreamS’ = {yes} indicates a signal to be sent to the downstream base unit.

3.6. Additional parameter object

3.6.1. Informal description

Looking at Figure 7, in the initial state ‘init’, the parameter of the additional function ‘tparam’ is set for the first time. This function depends on the base unit speed, accumulation, the base unit state and on the additional parameter itself. Later, the object waits for an input signal in order to update its function. This function can be updated only when the object receives a signal by the working speed and accumulation object. Finally, the object is able to send a signal to the throughput time object.

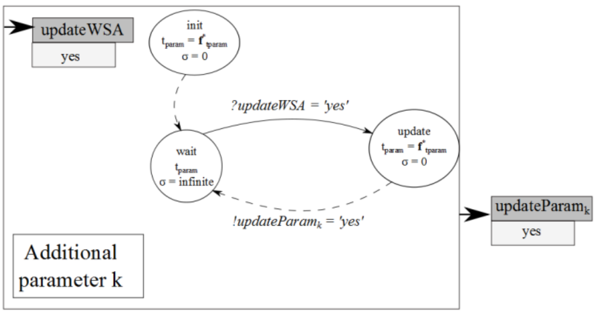


Figure 7: The generic additional parameter object.

3.6.2. Formal description

$$\text{‘}\boldsymbol{additional\ parameter\ object}\text{’} = < X_{param}, Y_{param}, T_{param}, \delta_{ext}, \delta_{int}, \lambda, t > \quad (6)$$

Input event variable: X_{param} = (updateWSA) where:

- ‘updateWSA’ = {yes} indicates a work speed and accumulation variation.

State variables: T_{param} = (phase, t, σ) where:

- ‘phase’ = {init,wait,update} is a name representing the situation in the real world;
- ‘t’ = {f^*_{tparam}} represents the function associated to the additional parameter object.
- $\sigma \in \mathbb{R}\ +0 \cup \infty$ is the life time of the current state.

Output event variable: Y_{param} = (updateParamk) where:

- ‘updateParam$_k$’ = {0,1} indicates a variation in the additional parameter k function.

4. CONCLUSIONS

A new framework for addressing simulation of continuous-time stochastic processes by exploiting a discrete event approach is presented in the paper. The core of the framework is constituted by a base unit model able to represent the minimum set of behaviors required for the modeling of a generic unit working in a real manufacturing system (i.e. workcenter, accumulator/buffer, conveyor/pipe).

The base unit is modeled by adopting the DEVS formalism and contains a set of other atomic objects whose interactions determine the sequence of events. The base concept is that events are generated by state changes in objects and propagates both in the upstream and in the downstream to notify connected objects that something has changed. In this way, the performances of each unit is determined by the staytime in states, while events are only related to state changes, thus saving a lot of computational time.

Moreover, the possibility to keep trace of additional measures of parameters of interest for the production is also added. As the model is conceived, parameters undergoing variations over time defined by continuous laws is possible without generating overheads on the event generation.

The presented work is particularly valuable for simulating manufacturing processes executed on a continuous flow of material (e.g. food and beverage industry) as well as production processes working on discrete units but characterized by a high speed flow (e.g. automated packaging lines).

Being the presented base unit model general, its implementation in simulation systems is desirable, while further extensions can be easily developed.

REFERENCES

Gebennini, E. and Gershwin, S. B., 2013. Modeling waste production into two-machine one-buffer transfer lines. *IIE Transactions* 45 (6): 591–604.

Gebennini, E., Grassi, A., Fantuzzi, C., Gershwin, S. B. and Schick, I. C., 2011. Discrete time model for two-machine one-buffer transfer lines with restart policy. *Annals of Operations Research*: 1–25.

Gershwin, S. B. and Burman, M. H., 2000. A decomposition method for analyzing inhomogeneous assembly/disassembly systems. *Annals of Operations Research* 93:91–115.

Gershwin, S. B. and Schick I. C., 1980. *Continuous Model of an Unreliable Two-Stage Material Flow System with a Finite Interstage Buffer*. Technical Report LIDS-R-1039, OSPNo. 87049, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Giambiasi, N. and Carmona J., 2006. Generalized discrete event abstraction of continuous systems: GDEVS formalism. *Simulation Modelling Practice and Theory* 14 (1): 47–70.

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. and Young, T., 2010. Simulation in manufacturing and business: A review. *European Journal of Operational Research* 203 (1): 1–13.

Levantesi, R., Matta, A. and Tolio, T., 2003. Performance evaluation of continuous production lines with machines having different processing times and multiple failure modes. *Performance Evaluation* 51:247–268.

Praehofer, H., 1991. System theoretic formalisms for combined discrete-continuous system simulation. *International Journal of General Systems* 19:226–240.

Pujo, P., Pedetti, M. and Giambiasi, N., 2006. Formal DEVS modelling and simulation of a flow-shop relocation method without interrupting the production. *Simulation Modelling Practice and Theory* 14 (7): 817–842.

Tamani, K., Boukezzoula, R. and Habchi, G., 2009. Intelligent distributed and supervised flow control methodology for production systems. *Engineering Applications of Artificial Intelligence* 22 (7): 1104– 1116.

Tan, B. and Gershwin, S. B., 2009. Analysis of a general Markovian two-stage continuous-flow production system with a finite buffer. *International Journal of Production Economics* 120 (2): 327–339.

Tan, B. and Gershwin, S. B., 2011. Modelling and analysis of Markovian continuous flow systems with a finite buffer. *Annals of Operations Research* 182 (1): 5–30.

Tan, B. and Yeralan, S., 1997. Analysis of multistation production systems with limited buffer capacity. Part II: The decomposition method. *Mathematical and Computer Modelling* 25 (11): 109–123.

Tolio, T., 2011. Performance evaluation of two-machines line with multiple up and down states and finite buffer capacity. *In Proceedings of the 8th International Conference on Stochastic Models of Manufacturing and Service Operations*, 117–127. 2011, Kusadasi, Turkey.

Yeralan, S. and Tan, B., 1997. Analysis of multistation production systems with limited buffer capacity. Part I: The subsystem model. *Mathematical and Computer Modelling* 25 (7): 109–122.

Zeigler, B., 1976. *Theory of modelling and simulation*. New York: John Wiley.

Zeigler, B., 1984. *Multifacetted Modelling and Discrete Event Simulation*. London: Academic Press.

Zeigler, B., Praehofer, H. and Kim, T., 2000. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. Second ed. New York, NY: Academic Press.

Zimmern, B., 1956. ´Etudes de la propagation des arrˆets al´eatoires dans les chaˆınes de production. *Rev. Statist. Appl*. 4:85–104.

PERFORMANCE ANALYSIS OF A WATER SUPPLY SYSTEM OFA DAIRY COMPANY BY MEANS OF ADVANCED SIMULATIVE METHODS

Marchini Davide[(a)], Rinaldi Marta[(b)], Montanari Roberto[(c)], Bottani Eleonora[(c)], Solari Federico[(c)]

[(a)] Interdepartmental Center Siteia.Parma
Parco Area delle Scienze , 181/A - 43100 Parma

[(b)] Interdepartmental Center Cipack
Parco Area delle Scienze , 181/A - 43100 Parma

[(c)] Department of Industrial Engineering, Department of Industrial Engineering, University of Parma
Parco Area delle Scienze , 181/A - 43100 Parma

[(a)]davide.marchini@unipr.it

ABSTRACT
This following work presents the result of a research project, that concerns the development of a simulation model of the water supply system of a dairy company, located in Parma, Italy. The reduction of water consumption is a very topical issue in many industrial fields. The approach developed aims to investigate, through process simulation, the areas of the plant where the efficiency of the water supply system can be significantly improved by means of simple modifications.
At first, the simulation model was used to reproduce the current system, so as to reach a precise knowledge of the water flows in the plant. In the second part of the work, a series of alternative scenarios was investigated, and the related performance was assessed, thus identifying the best plant configuration. The process simulator was designed under Microsoft Excel, using the potential of VBA (Visual Basic for Applications). Thanks to the study implemented, a final scenario of the water supply system was identified, which allows savings up to 30% of water compared to the original configuration.

Keywords: water consumption savings, process simulation, process optimization, layout modeling

1. INTRODUCTION

The amount of water consumption in industrialized countries is continuously increasing: it has doubled in about the last two decades, and in several countries the depletion of underground sources and/or their increasing level of contamination has become a central question. Therefore, rational use of water resources is a key issue for sustainable growth. Water consumption, especially nowadays, faces significant competing forces for change from decreasing water resource availability, stricter water quality regulations, decreased federal subsidies, increased public scrutiny, decreased financial health, and increased infrastructure replacement costs (Rogers and Luois, 2008). Consequently, the reuse of water has become an important issue within industry. Process water is used for many purposes in the food industry, i.e., as an ingredient, as part of the manufacturing process and in direct contact with the foodstuff, or in any indirect contact with the food product (Poretti, 1990).

In several industries, large amounts of water are used in cleaning and process applications. Recycling this water can be an opportunity to combine a reduction in the costs of industrial water with improved control of water management and a better environmental impact on natural resources (Centi and Perathoner, 1999). For instance, consider the cleaning-in-place (CIP) operations in a food industry. Generally speaking, cleaning is a key process in the food industry to assure safe products of high quality. However, the cleaning operations, besides the an economic impact, have environmental impacts.

Solutions for water savings always generate economic benefits related to lower procurement and energy cost, as well as to the lower cost for the treatment and discharge of waste water. For all the above reasons, the rational use of water resources has been an important research topic for many years and in different contexts. Among them, the main ones are the service processes, such as the cleaning operations (Centi and Perathoner, 1999) and the heat exchange processes (Lee and Cheng, 2012; Rezaei et al., 2010). But the topic was deepened also for the food industry, in general terms (Casani and Knochel, 2002; Casani et al., 2005), and for more specific food sectors, such as the sugar industry (Bogliolo et al., 1996), the dairy industry, etc. Among all these fields, for sure the dairy sector is one of those that could receive more contributions from the research.

Looking, for instance, at the cleaning operations, in the dairy industry the time of nonproduction dedicated to CIP is very high, ranging from 4 to 6 h per day. Moreover, the cleaning operation leads from 50 to 95% of the waste volume sent to the purification station (Marty, 2001; Sage, 2005). At present, the industrial CIP procedure mainly lies on practical experience imposed with relevant safety margins in terms of duration and chemical consumption (Alvarez et al., 2010). A further large amount of discharge water, affecting industries operating in the Parmigiano-Reggiano production, comes from the buttermilk concentration, the first operating step for obtaining the whey proteins.

All these factors lead easily to the conclusion that the exploitation of water resources by dairy industry can be greatly improved. According to Pagella et al. (2000), a strategic approach to water reuse must be based on a systematic analysis and on the principle that water users must not use more water of a higher quality than that strictly needed. A punctual and precise knowledge of the flows that run in a production plant is obviously of great help to identify those areas, devices or processes that offer the greatest optimization potentials in terms of water savings. This paper aligns precisely with this need. Indeed, its objective is to optimize the layout and the piping of an existing water supply line, in order to avoid wastes and recover and reuse the water that has been already used in the process, but is still poorly contaminated.

The first part of the study, therefore, focused on the analysis of the piping of the whole plant of the dairy company chosen as case study. The aim of this phase was to achieve a precise knowledge of the existing system and to reproduce it exploiting advanced design and simulation tools. A numerical, discrete events model of the plant was created, exploiting Microsoft Excel and Visual Basic for Applications (VBA). Once the system was re-created, we proceed with the evaluation of different innovative scenarios ("WHAT IF" analysis), obtained by varying the basic layout, to optimize the plant and water consumption. In particular, two different "WHAT IF" analysis were performed and compared to the "AS IS" configuration of the plant. From the simulations, it emerged that, through some changes to the water distribution and storage system, it is possible to get to significant savings both in terms of water demand from wells, and of the amount of waste water discharged.

2. MATERIALS AND METHODS

2.1. Initial layout of the water supply line

The work presented in this paper refers to a dairy company located near Parma, Italy. This company has been active for decades in the production of Parmigiano-Reggiano cheese and butter and is one of the greatest Parmigiano-Reggiano producer of the area, with a production volume of about 60 wheels of cheese and more than 10 tons of butter per day.

The approach followed in the study has focused initially on the detailed understanding of the operating conditions of the plant lines, as they were ("AS IS" configuration). The "AS IS" scenario includes two independent water circuits, one for the cheese and the other one for the butter production line. The layout of the whole factory is shown in Figure 1. In the left side of the picture, the water supply system is represented, with three wells. After extraction, the water is purified and stored in the main tank, named TK901. From this reservoir, all the processes of dairy and creamery plants will be fed. The two distinct sections of the factory are highlighted, in Figure 1, by the two dashed rectangles: the upper one delimits the water service facilities of the cheese production plants, while the lower one delimits the facilities of the butter production plants.

Regard the "cheese section", detailed in Figure 2, water withdrawals are basically associated with two main processes: the fermenters and the washing cycles. The fermenters installed are three, but, when simulating the system, they were considered as a single machine. Such a choice does not modify the water flows management, since the three fermenters always work in parallel.

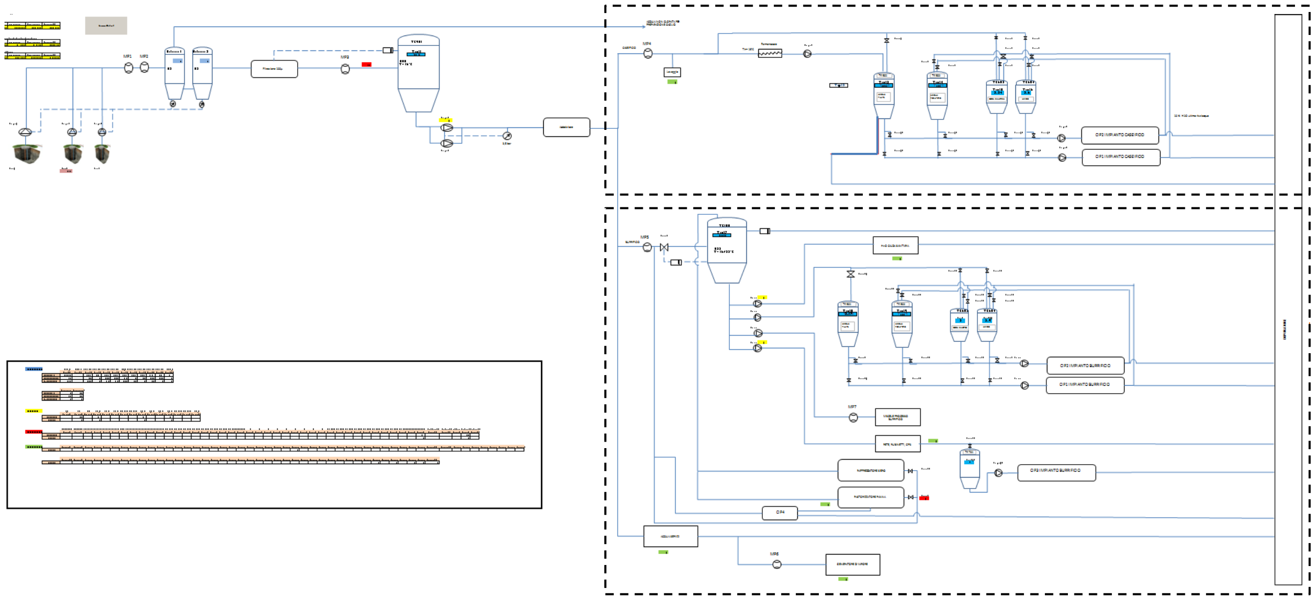


Figure 1 - Schematic layout of the water circuit of the two parts of the plant

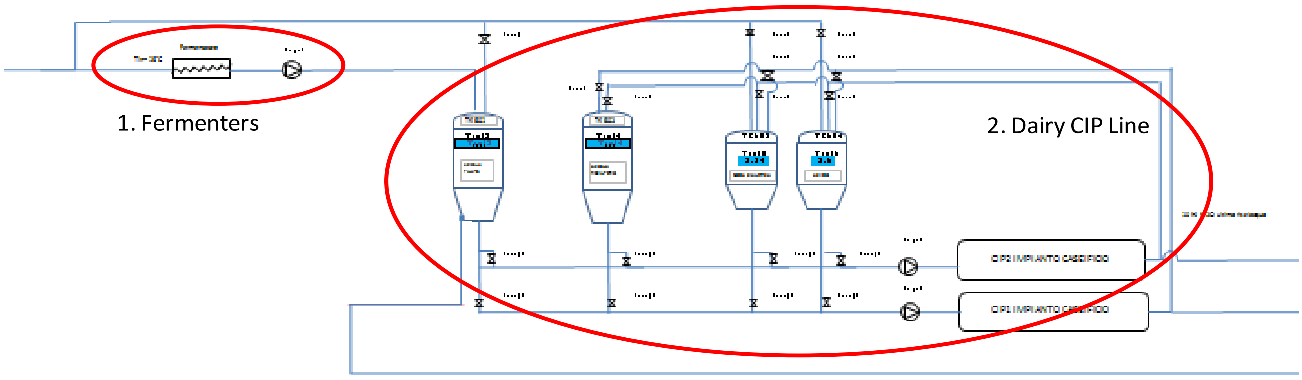


Figure 2 – Detail of Figure 1: layout of the service water line of the cheese production section of the factory

Within the fermenters, the plant reprocesses part of the whey left from the processes of the previous day; this is done because the re-fermented whey has the property of promoting the formation of cheese. The water absorbed by the fermenters is used to cool the whey.

The further relevant amount of water consumption in this part of the factory is generated by the washing cycles. The dairy section consists of 2 different CIP circuits, that operate on different elements of the plant. Specifically, they can operate simultaneously because they are equipped of two different and independent lines. In general, each washing is composed of different sequential phases, which can be combined in different way: (1) first rinse, (2) basic phase, (3) middle rinse, (4) acid phase and (5) final rinse. Washing can be carried out by means, for instance, of simple rinse cycles, or cycles of washing with soda, or combined cycles of soda and acid. Moreover, each washing cycle is characterised by a particular receipt in terms of duration and flow rate of each phase. The four tanks shown in Figure 2 are therefore essential for providing water and other cleaning solutions to both lines of the cleaning-in-place circuit. Each tank has a specific function, only the first one on the left, named TK605, contains pure water. The other ones store different kinds of cleaning solutions. The tank TK605 is fed by two sources: the general tank TK901, or the cooling water, from the fermenters. Overall, the global volume of water absorbed daily by the cheese production line of the company accounts for about 50 m^3. This data refers to the measured daily water discharge in the sewerage system.

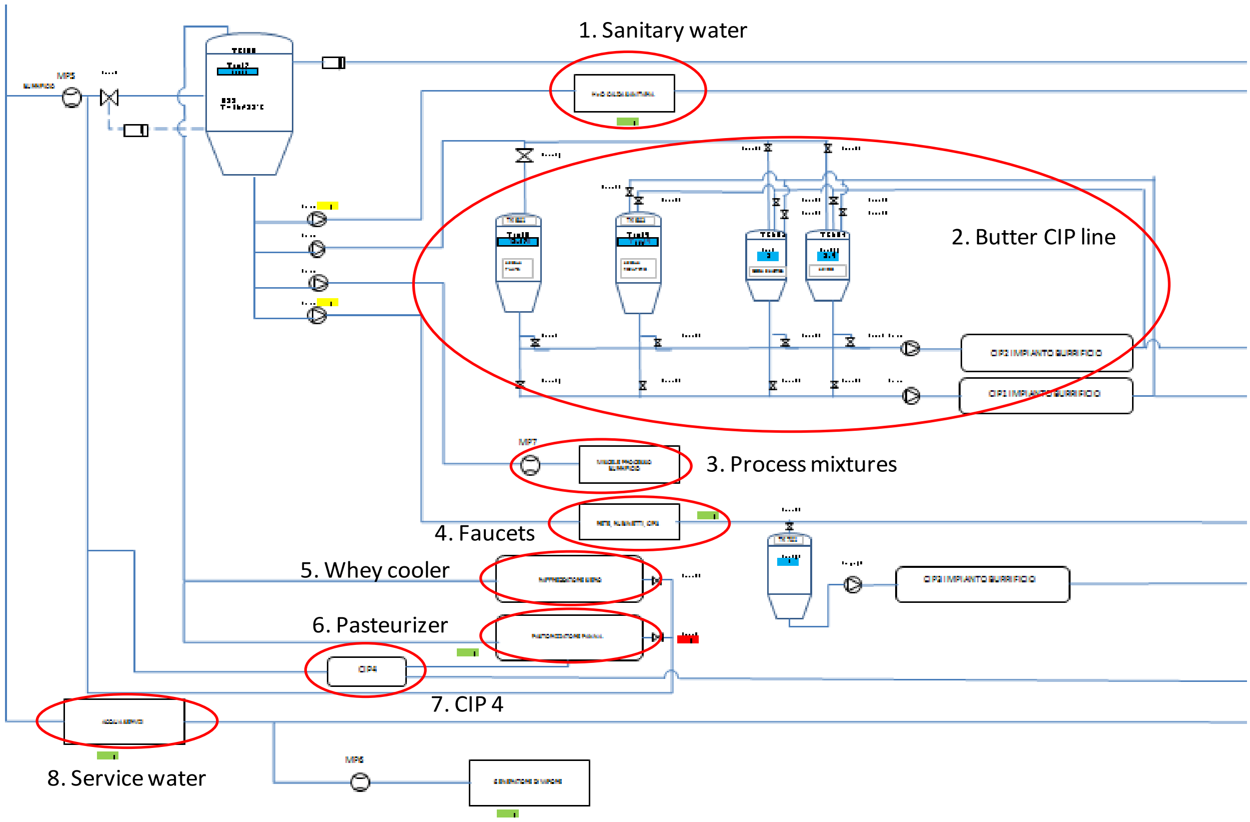


Figure 3 - Detail of Figure 1: layout of the service water line of the butter production section of the factory

The situation is slightly different with respect to the "butter section" of the factory, as can be seen in Figure 3. In fact, in this case there are eight main water withdrawals.

1. The first, small claim, is sanitary water.
2. The second withdrawal is due to the CIP circuit of the butter plants. Even in this case, the CIP circuit consists of 4 tanks: it is structured and works exactly as the CIP system of the "cheese section".
3. There is a certain amount of water that is used as an ingredient to be added to the raw materials, to create appropriate mixtures in the first phases of the process of buttermaking.
4. A further water consumption is related to the use of faucets, for rinsing and washing of floors and surfaces. This supply line feeds also a small secondary CIP circuit, named CIP3, responsible for the cleaning of the churning machines, that require a particular washing cycle.
5. Another point where there is a call for water is the whey cooling system. In fact, the whey comes out from the butter production plants, after being separated from the milk cream, with a temperature higher than 50°C. Before being concentrated and stored, it must be cooled up to 16°C.
6. The milk cream, before the process of churning, must be pasteurized. This process brings the milk cream up to a temperature of 90°C. To avoid organoleptic damage to the product, however, this temperature must be lowered rapidly after the treatment.
7. The pasteurizer machine is washed by means of CIP4, which is equipped of an independent line and requires clean water from tank TK901.
8. A certain quantity of water is finally used as service water, i.e. to feed all the service devices of the factory, such as, for instance, the steam generator.

All the lines of the butter production section are supplied by a single central tank, the TK105 (at the top left in Figure 3), except for CIP4, service water, cooler and pasteurizer. The water which exits from the pasteurizer and cooler is not contaminated at all; it only has a temperature slightly higher than the initial one. For this reason it can be reused and is sent back to the tank TK105. The global daily water consumption of the butter production line of the company amounts to approx. 160 m^3. 110 m^3 pass through the tank TK105 and are destined to the users connected to it, while the remaining 50 m^3 are taken directly from the wells and mainly converted in service water.

2.2. Simulation settings

As previously mentioned, the overall water consumption of the dairy company was studied using a simulation model, developed under MS Excel^{TM}.

Specifically, discrete event simulation has been used to reproduce the flows within the water supply system, in order to develop a useful tool to analyze the current performance of the company. The system is represented as a chronological sequence of events, each of them modifying the state of the whole system.

A main parameter of a discrete events simulation is the "time step" between two subsequent calculation steps of the simulator (clock setting). Obviously, the clock setting implemented in the simulator must be selected as a compromise between the accuracy of the results and the computational time. Moreover, in the current scenario, two particular CIP cycles are launched only once every 2 weeks. That forced us to set the overall simulation duration to 14 complete days (i.e., two complete weeks, from Monday to Sunday), and thus 336 hours. After a series of attempts, a time unit of 1 minute turned out to be appropriate for the analysis. Consequently, for the different configurations of the line, the simulator has computed the state of the system for 20160 steps overall.

The simulation model consists of 4 MS Excel^{TM} files, that reproduce: (1) the layout of the whole system, (2) the input setting of the model, in terms of the relevant raw data collected from the company, (3) the elaborated input data, and (4) the final database with the simulation results.

After the definition of the system layout, for each element, a specific cell has been identified as representative of a single object, and easily recognized through the VBA code. After that, the raw data have been grouped in the second spreadsheet: for instance for the tanks, the on/off levels have been derived from the logics of the real system. The "ON" status reflects the call of water from the upstream line, while the "OFF" status interrupts the operation.

For the various components of the system, such as pumps and valves, the punctual flow rate has been measured using a flow meter (model Krohne Optisonic 6300), depicted below (Figure 4). It exploits ultrasonic wave frequencies to determine the rate at which the fluid is moving inside the pipe.

Figure 4 - Ultrasonic clamp-on flow meter

We analyzed the processes and washings cycle listed, with the collection of their relative flow rates, starting times and durations. Moreover, these pieces of information has been inserted in the second spreadsheet, to be then elaborated in the third one by means of a VBA macro, able to reproduce the logic of the system.

The last spreadsheet contains the final results obtained from the simulations: for each time step, all the data related to the water volume inside each tank and the on/off state and the flow rate passing through the various pumps and valves.

However, even though the use of VBA allows great flexibility in programming and writing a specific code, some simplifications had to be introduced to model the original system configuration. The main simplification regard the pumps and valves functioning, as follows:

- The flow rate provided by each pump is equal to the sum of the flow rates required from any process or washing fed by the pump itself (a great operative flexibility of the pumps was assumed). Therefore, to simplify the model, the behavior of the centrifugal pumps has been assumed as comparable to the one of volumetric pumps.
- On the other hand, the behavior of the valves was considered similar to that of gate valves: they are assumed to be unable to act on flow regulation, but only to act as on/off systems.

The remaining simplifications have a really negligible impact on the whole system. An example could be the truck washings: in the real system, trucks are washed in sequence, so the washing cycle phases (called, for simplicity, P1, P2 and P3) are repeated 3 times. In the simulation model, conversely, only one single washing cycle is launched. This means that the washing phases of the trucks are grouped, and their durations are summed up: 3 generic "macro-phases" are carried out. The Figure 5 below can well clarify such change.

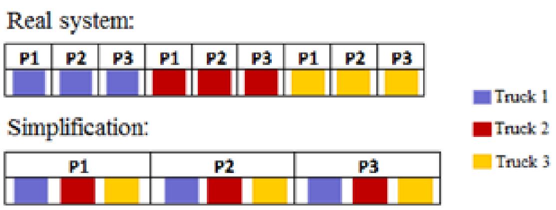


Figure 5 – Truck washing simplification

3. SIMULATION RESULTS

3.1. Limitations of the initial water supply line

Once defined the layout, the logical connection and the system simplifications, it was possible to launch the simulation process. The most interesting outcomes from the simulation concern the performance of the tanks of the plant. In particular the limitations emerged were two:

1. The tank TK605, with a volume of 9.5 m^3.
2. The tank TK105, with a volume of 32 m^3.

From the simulations, we found that the capacity of the two tanks is not sufficient. This is mainly due to the scheduling of the production and washing phases of the company. There is a deep difference between the time distribution of the processes, which the filling and emptying of the two tanks depend upon. By the very nature of the business activities, it cannot be guaranteed the simultaneity of the operations, and this implies the need of higher buffering volumes.

3.1.1. Cheese factory analysis

Considering the tank TK605, it can be fed by both wells water and by the water in output from the fermenters. Therefore, the two water users of the cheese producing plants (the fermenters and the CIP line), could be reduced to only one (the tanks of the CIP cycle), since the water that exits the fermenters could be entirely reused. The working cycle of the fermenters is spread over the whole day, with a variable flow rate of cooling water, shown in Figure 6. As a result, by these machines, an almost constant stream of water arrives to the tank. The overall daily amount of water that passes through the fermenters is 20.2 m^3, equal to 20200 liters.

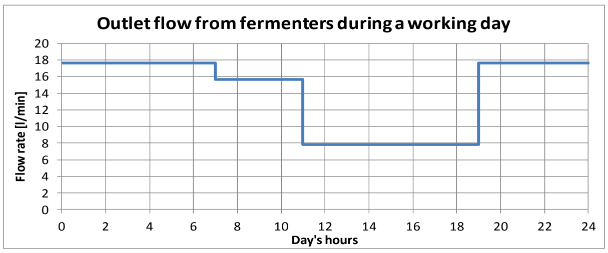


Figure 6 - Outlet flow from fermenters (inlet flow of the tank TK 605) during a working day

A different evaluation has to be done for the outlet of the tank TK605, which feeds the rinses of the CIP cycles of the dairy plants. The washing cycles change from day to day. The absorption of water from the tank TK605 will, therefore, be variable as a function of the day of the week, with a cyclicality of all the various CIP operations every 2 weeks, as explained in the previous paragraph. Moreover, in evaluating the outputs from the tank, it should be taken into account that, during the final rinse of each CIP cycle, the water used during the last 10% of the time, is recycled and used for the following cycle. Therefore, this amount of water should not be considered when quantifying the global water consumption. As an example, Figure 7 shows two graphs, with the water withdrawals from the tank TK605 during the days of Monday (in blue) and Tuesday (in red). By adding up the data of punctual consumption, the global daily quantity of water in outlet from the tank can be derived. On Monday, we obtained a total consumption of about 37000 liters (36925), on Tuesday 39500 liters (39414).

It is evident that the processes of cleaning of the cheese factory are concentrated in rather limited time windows, with large flow rates required in a short time. Considering the whole week, day by day the differences in the absorption cycle from the tank are not so deep, but there is nonetheless a variability, which had to be taken into account. By integrating the output flow rates from the tank TK605 during 24 hours, the volume of water daily emitted by the tank can be computed. The data are collected in the graph of Figure 8.

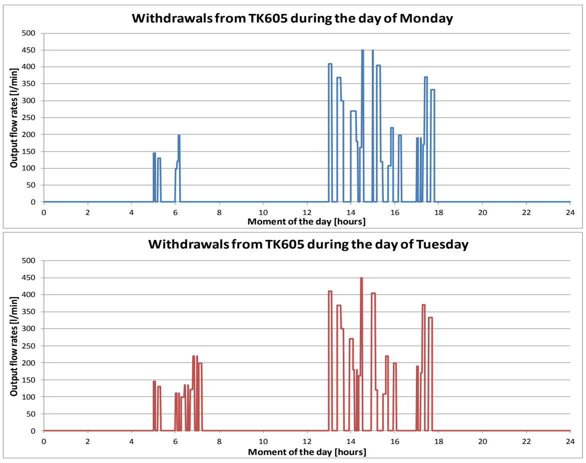


Figure 7 - Water output from the tank TK605 during the days of Monday and Tuesday for the CIP cycles

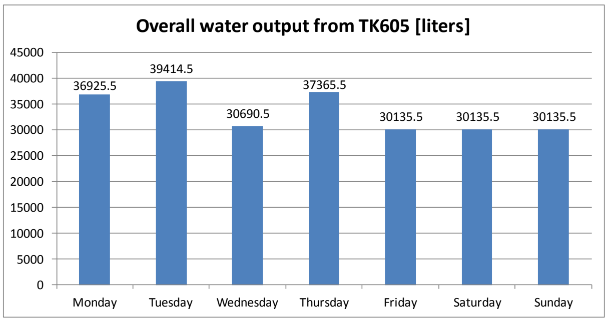


Figure 8 - Daily water output from tank TK605, used for the rinse phase of the CIP cycles

By computing the average consumption over the 7 days, it is found that approximately 33,500 liters (33543) of water are processed every day by TK605. This is more than the 20200 liters arriving each day to the tank from the fermenters, that, consequently, could be entirely reused, avoiding to take it from the wells. The only limitation to the reuse of this water is related, as anticipated, to the current volume of the tank (9.5 m^3), that is not sufficient to this purpose. Therefore, during the night, when the CIP are off, the tank reaches the filling level and the water that continues to arrive from the fermenters is discharged to the ground. This is evident from the graph in Figure 9.

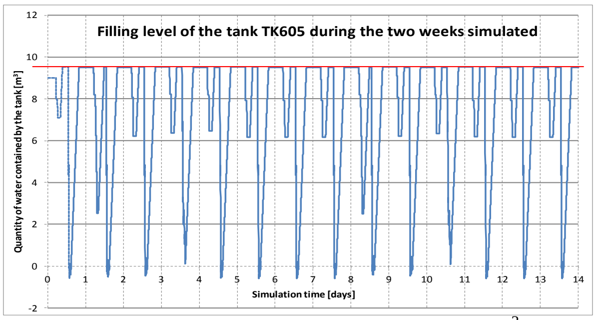


Figure 9 - Water contained in the tank TK605 (in m^3) during the two weeks simulated

In the picture, the red line represents the tank capacity. Once this limit is reached, the tank cannot accept further water and is forced to discharge it. This is even more evident analyzing the behavior of the drain valve of the tank, which opens when the limit capacity is reached (Figure 10).

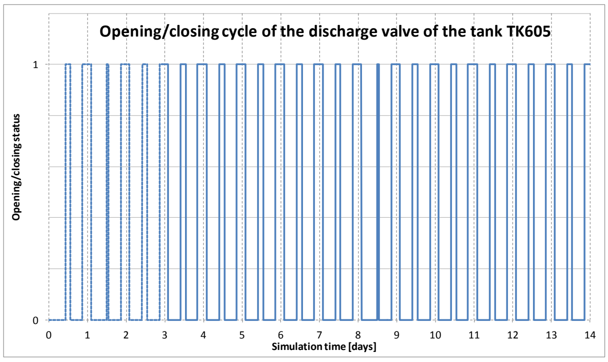


Figure 10 - Cycle of opening and closing of the discharge valve of the tank TK605. When the status is on 1, the valve is open and vice versa.

From the AS IS simulation, it emerged that the valve remains open for 34.36% of the global time. Averaging the results on the fourteen simulated days, this generates 10.57 m^3 (10570 liters) of water wasted daily, that, conversely, could be saved. This account for 52.33% of the water potentially recoverable (i.e. the 20200 liters in output from the fermenters' cooling). Moreover, by comparing the quantity of wasted water with the daily global absorption of water of the cheese factory (50 m^3), this quota represents the 21.14%.

It is obvious that avoiding to discharge more than 10000 liters per day of potentially reusable water would result in a significant economic savings. This, in turn, would justify an investment to increase the storage capacity of the tank TK605.

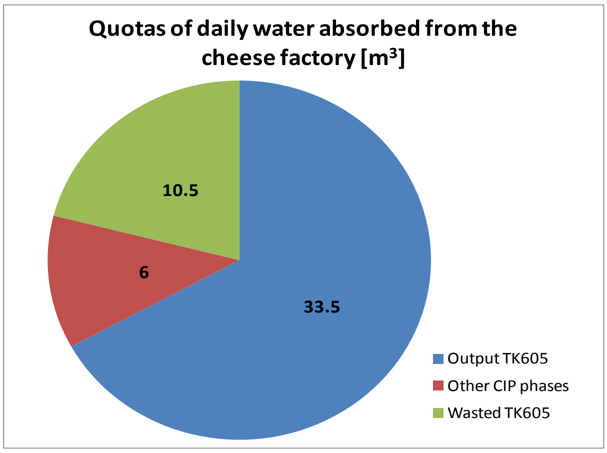


Figure 11 - Pie chart with the use distribution of the total daily amount of water absorbed from the "cheese section" of the company (50 m^3)

3.1.2. Butter factory analysis

As previously mentioned, from the analysis of the butter production line, we were able to identify that the critical point is the tank TK105, i.e. the central tank that feeds the line of the service water, the CIP circuits, the faucets and the line that brings water directly to the product to generate the correct mixtures. Also in this case, the tank can be fed in two ways: directly from the wells, or by the water in output from the pasteurizer and cooler.

Looking at the cooler and pasteurizer, their working cycles are shown in Figure 12, that shows the water absorbed (and then transferred to the tank TK105) by the two machines.

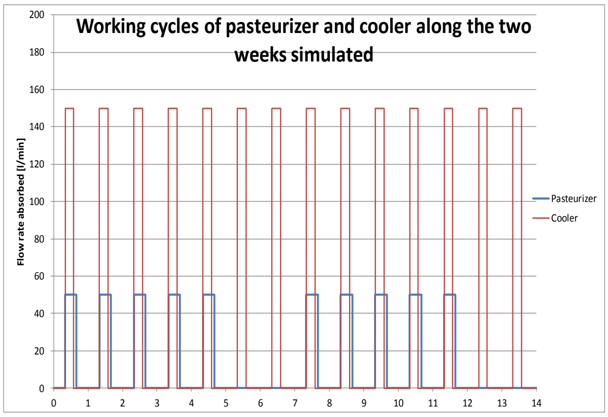


Figure 12 - Working cycles of cooler and pasteurizer of the butter section of the plant factory

While the duty cycle of the cooler is distributed across the fourteen days, the one of the pasteurizer is limited from Monday to Friday, while the machine remains turned off during weekends. Both machines start working at 8 a.m. The cooler works for the next 6 hours absorbing a water flow of 150 liters per minute, while the pasteurizer continues to work for 8 hours absorbing 50 liters of water per minute. It should be remarked that the production cycle of the butter factory is active from Monday to Friday: indeed, during the weekend the company does not manufacture butter, and the water absorption is, therefore, minimal. Conversely, when the production of butter is active, the water requirements of the various lines is considerably higher (Figure 13). In particular the daily water needs of the users connected with the tank TK105 amounts to 110 m^3 on average. Therefore, in these days the total water output from pasteurizer and cooler (72200 liters, equal to approximately 72 m^3), is fully used from the downstream processes, and an additional withdrawal from the wells is generated.

The main problems are concentrated in the weekends. The water needs, on Saturday and Sunday, are respectively of 37.8 and 41 m^3, the 34% and the 37% compared with the average of the other days of the week. Therefore, even if the pasteurizer does not operate, the incoming water from the cooler cannot be fully exploited, and must be discharged to the ground. In particular, Figure 14 depicts the filling level of the tank TK105 during the two weeks simulated. The red line represents the maximum capacity of the tank (32 m^3). When the maximum capacity of the tank is reached, the discharge valve opens, and the inlet water is discharged to the ground. The state of opening and closing of the valve is depicted in Figure 15. The valve remains open for a total of 318 of the total 20160 minutes simulated, i.e. for the 1.58% of the time. This data may seem to be very small, but, nonetheless, it results in a significant loss of water, reaching 63.6 m^3 during the two weeks, with a maximum value of 23.2 m^3 registered on the first Sunday. These value are even more significant when compared with the 110 m^3 of water that are absorbed daily from the butter production line.

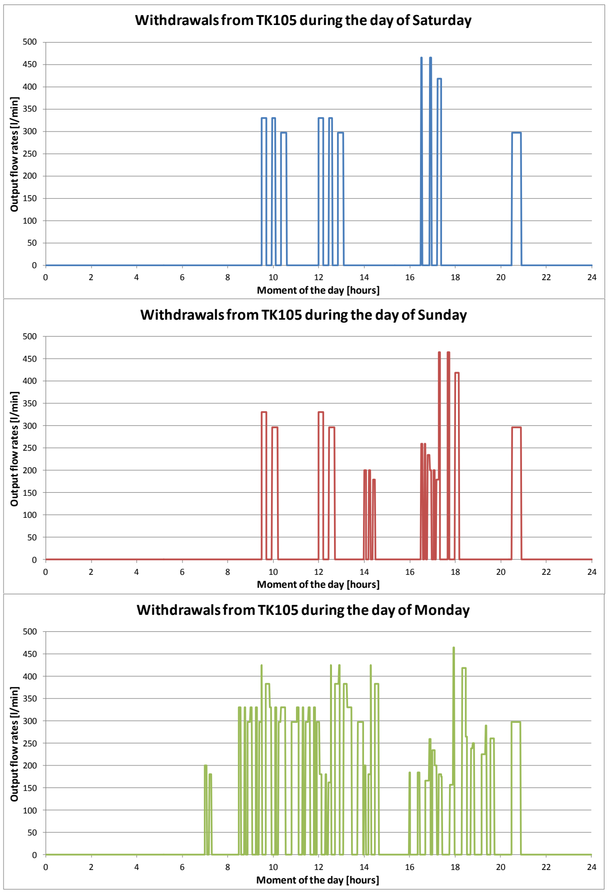


Figure 13 - Water absorbed from TK105 during the days of Saturday, Sunday and Monday

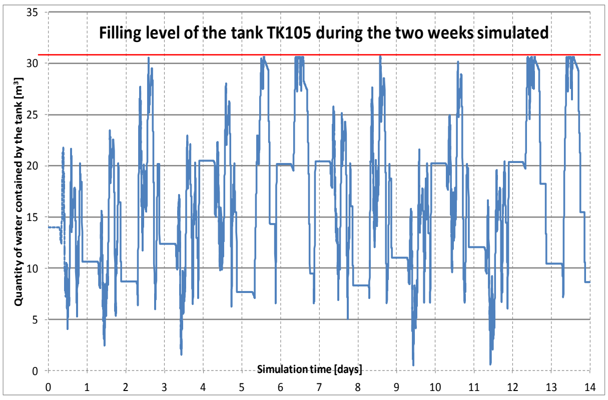


Figure 14 - Filling level of the tank TK 105 along the simulated two weeks

The valve remains open for a total of 318 of the total 20160 minutes simulated, i.e. for the 1.58% of the time. This data may seem to be very small, but, nonetheless, it results in a significant loss of water, reaching 63.6 m^3 during the two weeks, with a

maximum value of 23.2 m^3 registered on the first Sunday. These value are even more significant when compared with the 110 m^3 of water that are absorbed daily from the butter production line.

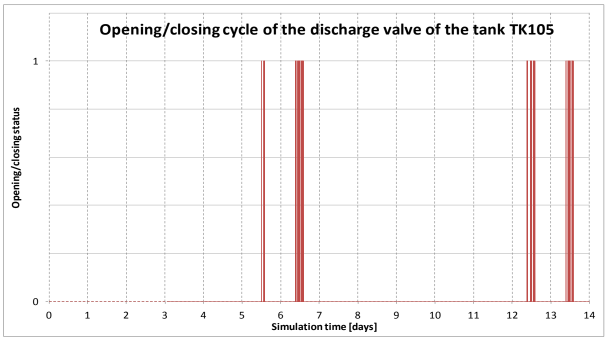


Figure 15 - Cycle of opening and closing of the discharge valve of the tank TK105. When the status is on 1, the valve is open and vice versa.

3.2. What if analysis: layout variations

After identifying the critical points of the AS IS system, we tried to identify some corrective actions, with the aim of decreasing (or eliminating) the waste identified. In this section, two different alternatives to the current layout of the lines of water supply will be analyzed, exploiting the simulator already prepared for the AS IS analysis.

3.2.1. CASE 1: Enlargement TK605 and introduction of cooling tower

In this scenario, we tried to optimize individually the two productive sections of the factory, keeping them separated. In particular, with respect to the "cheese section", we evaluated the effect of an enlargement of the tank TK605, which feeds the rinse phases of the CIPs, and which receives water from the fermenters. Looking at the "butter section", we evaluated the introduction of a cooling tower that would feed the cold water circuit of both the pasteurizer and the cooler. The tower would provide the two machine with a colder water than that withdrawn from the wells, thus generating interesting water savings.

Analyzing separately the two changes, replacing the tank TK605 (volume 9.5 m^3) with a tank of 30m^3, would allow to recover all of the water discharged on the ground by the cheese production lines in the current configuration (i.e., more than 10 m^3 per day).

The introduction of the cooling tower, instead, implies more complex considerations. First of all, the inclusion of such a system has been suggested to the company by a plant engineer, who has proposed a tower of a defined potential, namely 500 kW. The manager of the company wanted to understand if such investment was actually profitable. The behavior of the manufacturing system, in fact, changes depending on the seasons, as the performance of the tower depends on the environmental temperature. Beginning from the analysis of the AS IS scenario, the pasteurizer absorbs 240 m^3 of water over the two weeks simulated, while the cooler 756 m^3. The cooling tower would reduce these requirements. In particular, during Summer, considering an average temperature of 25°C, the tower would let to save about 100 m^3 of water every 2 weeks. In fact, the pasteurizer could be entirely fed with the tower water, and, as the water coming from the tower would be cooler than that of the wells, it would be sufficient to use 140 m^3 instead of 240.

During Winter, considering an average temperature of -1°C, the efficiency of the plant would increase, and water savings could reach 728 m^3, out of the 996 currently required. The pasteurizer and the cooler could be both fed with 268 m^3 of water from the tower.

Averaging over the year, the pasteurizer and the cooler would have their water needs reduced by approximately 414 m^3 every two weeks, equal to 41.6%. Hence, the amount of water saved would be relevant, but it must be considered that all the wells water used by pasteurizer and cooler can be reused, while it is not possible to recover the tower water. Consequently, with respect to the butter production line, the overall water absorption of the system would actually increase rather than decreasing (Figure 16).

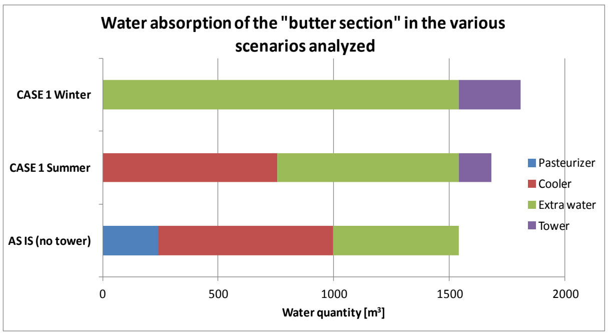


Figure 16 - Water absorbed by the butter factory during the two weeks simulated in the different cases analyzed: the total quantity are subdivided according to the users

3.2.2. CASE 2: Introduction of a new tank, parallel to the TK901

In this scenario, we assume to increment the total storage capacity of the central tank of the whole plant, the TK901. In the current configuration, the tank has a capacity of approximately 10 m^3, that has been kept constant in this analysis. Instead of enlarging it, we pondered the introduction of a new reservoir, named TK902, that would work in a complementary manner to the first one.

This tank would allow to avoid water waste, recovering the outputs from the fermenters of the "cheese section", and the discharge water downloaded by the TK105 (in the "butter section") when it reaches its maximum capacity. Therefore, it would work as an inter-operational buffer, with the purpose to decouple water recovery and the water withdrawals from the various lines. A schematic representation of the line with the insertion of the tank TK902 is depicted in Figure 17. The new part of the line is highlighted in red. The red continuous lines represent the piping in output from TK902, while the dashed lines are the inlets, that allow the tank to fill, recovering water from other processes. By introducing the new tank, the operating conditions of the plant would not be altered, but the two sections would have a preliminary point of contact, exactly the tank itself. In fact, inside the TK902, the recirculation water of both the cheese and the butter productive lines would be recovered. The volume of the tank was not defined *a priori*; rather, it has been derived from simulations, identifying the maximum filling levels achieved. The volume of water that the tank TK902 would contain along the two weeks is represented in the diagram of Figure 18. As can be seen, the maximum level is reached on Sundays, when the cheese production lines work normally, while the production of butter is off. This leads to an accumulation of the water coming from the cooler of the serum, that increases the level of water within the tank TK902.

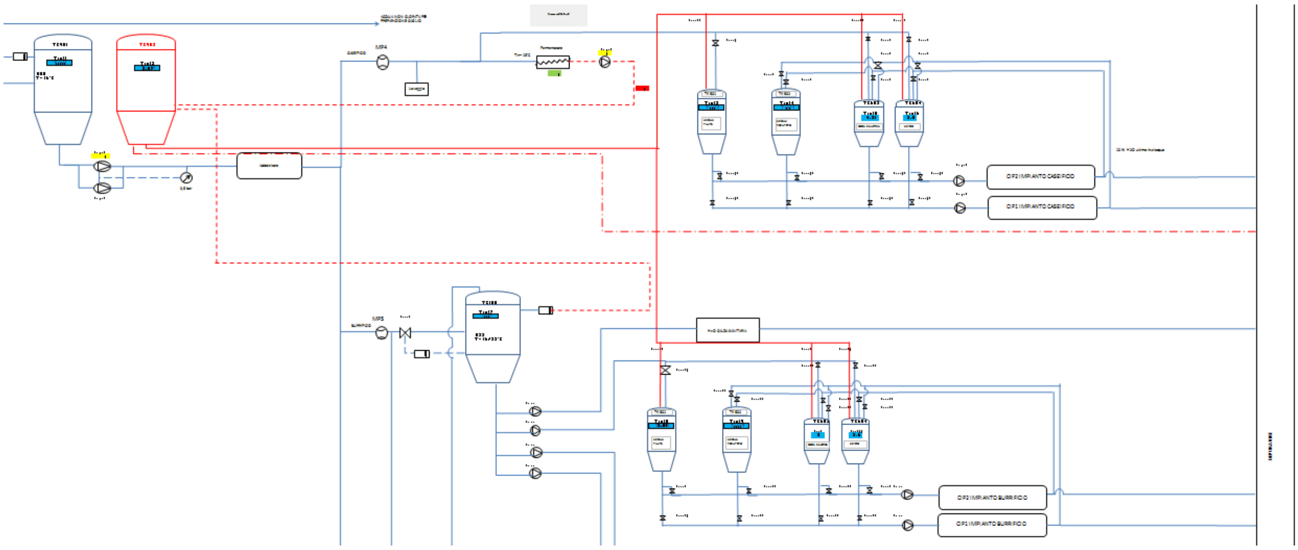


Figure 17 - New configuration of the water distribution system of the factory after the introduction of the tank TK902

It is evident that the tank level has a tendency to grow at night and in the early morning, when the productive and CIP processes are off. Conversely, during the day, the level drops until the tank is empty, when the various CIP cycles and the other utilities connected to the tank require water. However, a tank with a capacity of 30 m^3 seems to be appropriate to face the normal water demand and also in presence of potential peaks of request.

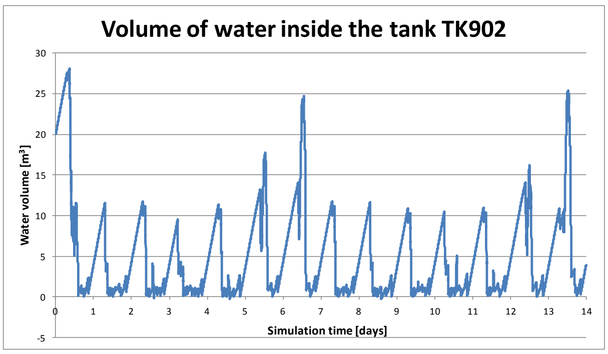


Figure 18 - Volume of water contained by the TK902 along the two weeks

The tank, if correctly dimensioned, could completely avoid the waste of water. The water saved would therefore amount to a total of 211.58 m^3 all along the two weeks. This volume of water is very relevant, even more if compared with the daily water consumption of the whole plant. In fact, the set of cheese and butter production lines and the other services, do absorb every day about 210 m^3 of water. Therefore, the installation of the tank TK902, would bring to save a volume of water equal to that necessary for one day of operation of the entire factory, every two weeks (or, in terms of percentage, the 7.2%).

4. CONCLUSIONS

Due to the increasing costs of water and water discharge, reuse of water in the food industry is becoming more and more important. The work carried out in this paper has allowed, first of all, to identify and quantify the water wastes of a real company, operating in the dairy industry. After the analysis of the current system layout, two alternative configurations of the water supply lines were evaluated, with the purpose of understanding which one would ensure the most performing results in terms of water saving. In this respect, the first plant configuration did not give satisfactory results, so the company management rejected the possibility of installing a cooling tower. Conversely, the second scenario provided interesting results in terms of water saving, by eliminating almost all the water wastes identified in the AS IS scenario. The introduction of a new tank (TK902, with a capacity of 30 m^3), which would work in a complementary way to existing TK901, would reduce the company's water consumption by 7.2%.

This study is a typical example of how discrete events simulation can be a very useful tool from many points of view. In fact, it allows to predict the performance of a new system configuration, before its physical implementation. This is an important advantage for companies, as it can guide strategic decisions through a "WHAT IF" analysis. Once set the basic logic of the simulation model, it is, indeed, very fast and cheap to change some parts of the layout and to assess the corresponding impact on the system performance. Through discrete events simulation, in fact, it is possible to monitor in detail the operating conditions of a water supply line, taking into account of the circulating flow rates, the on/off periods of pumping systems, the filling levels of the storage tanks. This allows a precise knowledge of what is happening on the entire system, as well as to identify those areas where interventions should be planned to improve the system performance.

Discrete events simulation is not an innovative tool by itself, but becomes so when it is applied to contexts hitherto unconsidered. The company with whom we collaborated has realized the benefits potentially arising from the method, and has managed to avoid an unnecessary investment on one hand and to understand how to reduce their water waste on the other.

REFERENCES

J.W. Rogers and G.E. Louis, Conceptualization of a robust performance assessment and evaluation model for consolidating community water systems. Journal of Environmental Management 90 (2009) 786–797.

M. Poretti, Quality control of water as raw material in the food industry. Food Control 1 (1990) 79–83.

G. Centi and S. Perathoner, Recycle rinse water: problems and opportunities. Catalysis Today 53 (1999) 11-21.

K.P. Lee and T.A. Cheng, A simulation approach for energy efficiency of chilled water system. Energy and Buildings 54 (2012) 290-296.

E. Rezaei, S. Shafiei and A. Abdollahnezhad, Reducing water consumption of an industrial plant cooling unit using hybrid cooling tower. Energy Conversion and Management 51 (2010) 311-319.

S. Casani and S. Knochel, Application of HACCP to water reuse in the food industry. Food Control 13 (2002) 315-327.

S. Casani, M. Rouhany and S. Knochel, A discussion paper on challenges and limitations to water reuse and hygiene in the food industry. Water Research 39 (2005) 1134-1146.

M. Bogliolo, A. Bottino, G. Capannelli, M. De Pedro, A. Servida, G. pezzi and G. Vallini, Clean water recycle in sugar extraction process: Performance analysis of reverse osmosis in the treatment of sugar beet press water. Desalination 108 (1996) 261-271.

Y.P. Marty, Maitrise de la consommation d'eau dans les industries agro-alimentaires. Ind. Alim. Agric. 118 (2001) 35–38.

M. Sage, Valorisation des eaux de procedes de l'industrie laitiere en methanisation et denitrification. PhD Thesis. Ecole Nationale Superieure d'Agronomie de Rennes, 2005, Rennes, France.

N. Alvarez, G. Daufin, G. Gésan-Guiziou, Recommendations for rationalizing cleaning-in-place in the dairy industry: Case study of an ultra-high temperature heat exchanger. Journal of Dairy Science 93 (2010) 808-821.

C. Pagella, R. Galli and D.M. De Faveri, Water reuse in industrial food processing. Journal of Food Technology Africa 5 (2000) 25-29.

AUTHORS BIOGRAPHY

Davide MARCHINI is a scholarship holder in Industrial Engineering at the University of Parma (Interdepartmental Center Siteia.Parma). He got a master degree in Mechanical Engineering for the Food Industry, discussing a thesis titled: "Advanced design of a UV reactor for water treatment using computational fluid dynamics". He attended the 11th International Conference on Modeling and Applied Simulation, presenting the paper titled "Advanced design of industrial mixers for fluid foods using computational fluid dynamics" (awarded as Best Paper), and the 2011 EFFoST Annual Meeting (Berlin 9-11 November 2011), presenting the paper titled "Advanced design of a UV reactor for water treatment using computational fluid dynamics".

Federico SOLARI is a PhD Student in Industrial Engineering at the University of Parma; master degree in Mechanical Engineering of the Food Industry, dissertation of the thesis: "Analysis and design of a plant for the extraction of volatile compounds from aqueous matrix"; Attending many international conferences, he's author or coauthor of 7 international papers. His research activities mainly concern industrial plants and food process modeling and simulation, with a particular focus on the CFD simulation for the advanced design of food plants.

Roberto MONTANARI is full professor of Mechanical Plants at the University of Parma (Italy). 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. 60 papers, which appear in qualified international journals and conferences. He acts as a referee for several scientific journals and is editorial board member of 2 international scientific journals.

Marta RINALDI is a PhD Student at the University of Parma, and Scholarship Holder in Industrial Engineering at the Interdepartmental Center CIPACK. She got a master degree in Management Engineering with a thesis titled "Analysis and evaluation of energy efficiency of a shrinkwrap-packer". Her main fields of research are discrete event simulation and simulation of industrial plants.

Eleonora BOTTANI is Lecturer (with tenure) in Mechanical industrial plants at the Department of Industrial Engineering of the University of Parma (Italy). She graduated in 2002 in Industrial Engineering and Management at the University of Parma, and got her Ph.D. in Industrial Engineering in 2006. Her research activities concern logistics and supply chain management issues, encompassing intermodal transportation, development of methodologies for supplier selection, and, recently, the impact of RFID technology on the optimization of logistics processes and supply chain dynamics. She is the author or coauthor of approx. 100 scientific papers, referee for more than 50 international scientific journals, editorial board member of 3 scientific journals.

A DATA COLLECTION METHODOLOGY TO PERFORM DHMS-BASED ERGONOMIC ANALYSIS OF MANUFACTURING TASKS

Nadia Rego-Monteil[(a)], Mariangela Suriano[(b)], Diego Crespo Pereira[(c)], David del Rio Vilas[(d)], Rosa Rios Prado[(e)], Francesco Longo[(f)]

[(a), (c), (d), (e)] Integrated Group for Engineering Research, University of Coruña, Calle Mendizábal, s/n, Ferrol, 15403, Spain
[(b) (f)] MSC-LES, University of Calabria, Mechanical Department, Ponte P. Bucci, 87036, Rende (CS), Italy

[(a)]nadia.rego@udc.es, [(b)]suriano.m@unical.it, [(c)]dcrespo@udc.es, [(d)]daviddelrio@udc.es, [(e)]rrios@udc.es
[(f)]f.longo@unical.it

ABSTRACT
A key stage in any simulation project is the data collection process. In most simulation-based ergonomic analysis, data from the movements performed by the operator during the task are obtained either by direct observation, video observation or by means of a Motion Capture (MoCap) System. Direct or video observation methods are quick and inexpensive, however less accurate and objective than commercial Mocap Systems. MoCap Systems are much more expensive and limited to the use in laboratories. The objective of this paper is to present a data collection methodology based on the joint employment of the per-pixel depth technology and a free open source video analysis software. The proposed system is portable, low-cost and suitable for fieldwork. A real case study is presented to explain how objective posture data of a user performing a task can be obtained and used for its workstation analysis and improvement.

Keywords: ergonomics, workstation design, work measurement, digital human modelling and simulation, motion capture systems, data collection methodology

1. INTRODUCTION
Despite the contribution of Digital Human Modelling and Simulation (DHMS) methodologies and tools in increasing both the consideration and the efficiency in the treatment of the ergonomic aspects among production engineers, the effective design –and most often, re-design- of industrial workstations heavily relies on the quality and reliability of an efficient data collection phase (del Rio et al. 2012). Over the years several data collection methodologies have been used by researchers and scientists working in the field of workstation effective design. Among the used for acquiring motion data, the following two have to be regarded as the most used ones, i.e., (i) observation based methods (video based systems, walking-through observation, video capture and playback technology) and (ii) Motion Capture Technology (MoCap).

Observation based methods consist in observing the worker while performing the manufacturing operations and collecting information about the work methods. MoCap systems can be based on sensor located on the subject like accelerometers, acoustic transmitters, electromagnetic sensors, or inertial sensors. Optical MoCap systems based on stereophotogrammetry are by far the most widespread in digital human modelling for ergonomic simulation (Ausejo and Wang 2010). Optical reflective markers are attached to a test subject, who is then digitally filmed with several infrared cameras. Then, from the 2D marker position recorded at each frame, the 3D marker positions can be estimated using stereophotogrammetric methods. Some commercial examples are Viconpeak, Phase-Space, Optotrak, or Motion Analysis Corporation. It is generally accepted that all postures from motion capture technology are realistic and accurate (Stephens and Jones 2009). However, the MoCap technology increases the cost and it is usually difficult to use in real-world applications due to complexity, bulk and space requirements (Best and Begg 2006)

When observation-based methods are used to collect data from body postures, designers must use their experience and judgment to manually posture the body and hands of human figures to simulate the tasks (Zhou et al. 2009). This method can result a limitation in the accuracy of the model as it introduces subjectivity. On the other hand, MoCap driven models may help to overcome this limitation, but they present other drawbacks, such as the cost and the intrusive technology.

This paper presents a combined methodology of observational data and motion capture system in order to develop a DHM ergonomic analysis to improve workstation design. The proposed methodology uses two simultaneous different motion capture systems; a per-pixel depth sensing camera (ASUS Xtion) and a free software tool to perform video analysis (Tracker).

Although not originally created for this purpose, the per-pixel depth sensing technology (Patent US 2010/0118123 A1 of May 2010) is capable of tracking the orientation and position of the body segments of a user at a frame rate of up to 30 fps. It is commercially available in Microsoft Kinect and Asus Xtion PRO (see Figure 1).

The depth cameras have been established in recent research studies as an inexpensive, portable and markless 3D data acquisition device for anatomical position measurement (Dutta 2012, Clark et al. 2012). In the experiments of Dutta (2012) it was concluded that the Kinect is able to capture 3D posture coordinates with acceptable errors (RMS < 1.5 cm in the three directions) when the device is placed between 1 and 3 m from the camera. As a drawback, it is stated that Kinect has trouble detecting dark, shiny or rough surfaces.

Figure 1. Depth Sensing Device. (Source. Wikipedia Commons file "The Microsoft Kinect Peripheral for the Xbox 360")

Clark et al (2012) compared its accuracy to a commercial 3D commercial motion capture system in an experimental setup where different subjects performed postural tests. They concluded that commercial systems and the depth camera have similar reliability for the majority of the measurement of postural coordinates. Two important limitations have been noticed in their experiments, the presence of biases for outcome measures in pelvis and sternum and the inability to assess internal/external joint rotations in the limbs.

Tracker is a free and open source video analysis and modelling tool built on Java, able to track position, velocity and acceleration of point mass particles and two-body systems, intended for aiding in Physics education. It has been built as a part of the awarded Open Source Physics (OSP) Project sponsored by the National Science Foundation and Davidson College (comPADRE 2012). The strategy of most of the video analysis tools, including Tracker, consists in tracking discrete objects within the field of view of the camera and accounting for changes in position through time. Although the accuracy depends on the analyst's individual skills, the quality of the video and the size of the objects being tracked, it is generally considered acceptable (Bryan 2004; Hedrick 2008). The tracker interface to automatically follow (track) a mass point is presented in Figure 6.

In sections 2 and 3 the data collection methodology to develop Digital Human Modelling and Simulation (DHMS) is explained. In section 4, an illustrative real case study in the slate industry in which the data collection has been used is then introduced.

2. THE PROPOSED METHODOLOGY

The proposed methodology for data collection has been designed to fit a general DHMS improvement procedure. Usually in this analysis schema, the workstation improvement approach is divided into two phases. The descriptive analysis has the purpose of obtaining an ergonomic diagnosis of the task as it is in the present. The predictive stage implies the proposal of modifications (if needed) to the present workstation configuration, and the assessment of the level of improvement achieved by them. The data collection system described in this paper has been applied to both stages –descriptive and predictive- at the laboratory. However, in principle it could be easily used in real manufacturing environments because it is portable and it does not need to attach markers to the skin or clothes of the user being tracked.

The following steps are included:

1. **Observation of the real task.** By means of videos or in situ observation, general information can be determined such as work content, objects and layout main dimensions, and worker characteristics.
2. **Definition of work elements.** From the task observation, the main activity can be divided in fundamental tasks ("subtasks") that are repeated into the work cycle. The main plane of movement in which the subtask occurs should be taken into consideration to the division.
3. **Motion Data.** For each subtask, position of the MoCap system has to be adapted, so that postural data can be best recorded.
    a. Per-pixel depth camera. Usually the best location for the device is in front of the operator.
    b. Conventional camera. Placed on the main movement plane. If the movement mainly includes arms abduction or trunk lateral bending, the best camera position is in front of the subject. If the movement mainly implies neck, arms or trunk flexion over the sagittal plane, the camera is better placed 90˚ on the side.
4. **Information analysis and posture modelling** Raw data coming from the depth device has to be filtered and combined to obtain available data to perform the manikin training in the DHMS software. The following section describes with more detail this procedure. Tracker post-processing of the videos has to be performed to obtain the necessary data to validate and complete the angles characterized with the depth device. Finally, the information recorded (positions along the time) are the input data for the posture modelling of the manikin, and the basis to perform the biomechanical and postural analysis of the task.

The motivation for using both data collection systems is derived from the fact that some limitations of the depth camera have been detected apart from the ones above mentioned. Some of them, such as the inaccuracy to detect abduction movements of the extended arm or the trunk bending movements when the subject is turned 90˚ can be solved by breaking the task in fundamental elements in which the movements occur in the same plane (and so the device can be always placed oppositely). Some other limitations, such as the inaccuracy to detect arm positions in movements of extreme flexion (more than 80˚) when the subject stands in front of the device, can be solved by using a conventional camera placed perpendicularly to the depth device to check the movements on the Tracker post-process.

In the following section, the filtering techniques and the parameters configurations are discussed. In section 4, an illustrative case study is presented.

3. DATA CONFIGURATION

There exist several approaches to obtain kinematic data from the sensing technology. The developers of the depth sensor made available the Natural Interaction (OpenNI) middleware which provides position and orientation data of 20 body joints of the body. Windows also provided freely the Kinect for Windows Software Development Kit (Windows SDK version 1.6) with similar tracking features. MS Kinect or Asus Xtion can recognize six people appearing at a time and can track two users simultaneously. It is optimized to track standing or sitting users facing the device. (MSDN 2013). The visualization of the user being tracked is presented in Figure 2 (right).

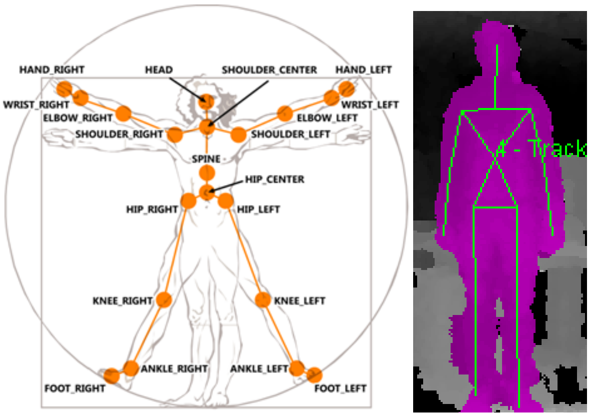


Figure 2. Skeleton Positions that are Tracked by the Depth Sensor (Source: Microsoft Kinect http://msdn.microsoft.com/)

Our experimental setup includes the use of the ASUS Xtion Pro with the OpenNI middleware (estimated price, 130 €) and Tracker 4.7x, Copyright (c) 2012 Douglas Brown and Wolfgang Christian, freely available at www.cabrillo.edu/~dbrown/tracker/.

The hip center joint provides the absolute orientation of the user. The position and orientation of each segment (child) is relative to the position and orientation of its parents. A Java application has been developed to store and combine the raw data to calculate the relative angles of the body. To do so, we had in mind the parameters that are the basis of the posture definition in Delmia V5R20, a commercial CAM software to develop DHMS. A simplified set of the degrees of freedom (without the fingers movement definition or the legs) available to set a posture is shown in Table 1.

Table 1. Degrees of Freedom that Define a Posture in Delmia V5R20 (A reduced set)

	DOF 1	DOF 2	DOF 3
Neck (N)	Flexion / Extension	Lateral R/L	Rotation R/L
Full Spine (FS)	Flexion / Extension	Lateral R/L	Medial / Lat. Rot
Right Arm (RA)	Flexion / Extension	Abduction / Adduction	Medial / Lat. Rot
(RF)Right Forearm	Flexion / Extension	Pronation / Supination	-
Right Hand (RH)	Flexion / Extension	Radial / Ulnar Deviation	-
Left Arm (LA)	Flexion / Extension	Abduction / Adduction	Medial / Lat. Rot
(LF) Left Forearm	Flexion / Extension	Pronation / Supination	-
Left Hand (LH)	Flexion / Extension	Radial / Ulnar Deviation	-
Clavicular (C)	Flexion / Extension	Elevation / Depression	-

The tracking joint information can be smoothed across different frames to minimize noise and stabilize the joint positions over time. One approach for smoothing the time series is to replace each value of the series with a new value which is obtained from an n-order polynomial fit to the 2n+1 neighbouring points (including the point to be smoothed).

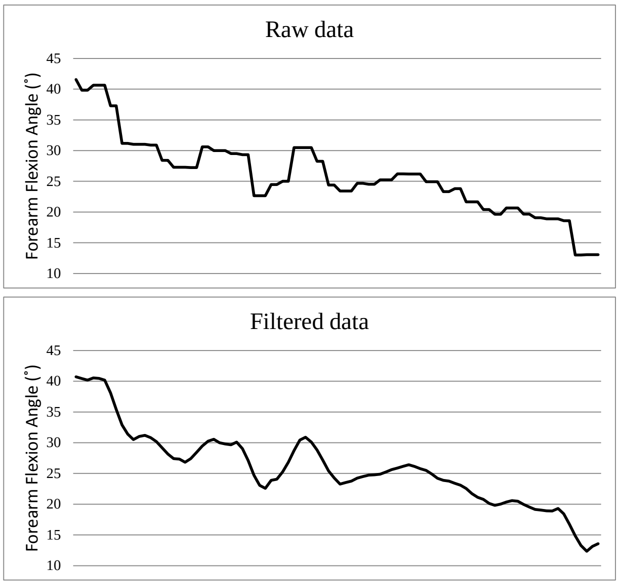


Figure 3. Forearm Flexion Angle Before and After the Savitzky and Golay Filter Application

This technique, known as the Savitzky-Golay Filter was proposed in 1964 and it has the advantage of preserving features of the distribution such as relative maxima, minima and width (Savitzky and Golay, 1964). Figure 3 presents the original right forearm flexion angle variation obtained from the depth device during some trials and the same data treated with a second order filter over seven points.

Once the different limbs of the body have been characterized, this information can be used as input into the postures modelling (by using the angles defined in Table 1). Although the angles variation is very detailed (data are acquired every 0.033 seconds), for the ergonomic analysis it is not of practical interest modelling 30 postures per second. Instead, we assumed that an input modelling of 3 postures per second was more appropriate to perform the modelling in reasonable time, as well as to conserve a realistic visual quality of the movement sequences.

4. CASE STUDY

This section presents the data collection methodology applied to a real case study of ergonomic analysis in a Spanish company devoted to the extraction and manufacturing of roofing slates. This ergonomic analysis is a part of a general improvement project which applied Lean Manufacturing, Modelling and Simulation to the proposal of standardization procedures for statistical control (del Rio, 2008), redefinition of the layout (Crespo et al. 2011) and the ergonomic analysis of other tasks (Rego-Monteil 2010).

The slate process is highly dependent on manual operations. One of them, at the end of the process is the packing station. Their work implies lifting, bending, performing repetitive work and working under pressure. These activities are extremely undesirable in terms of ergonomic impact as they will most likely provoke musculoskeletal disorders (MSD) in the long term. In addition, packers sometimes become the bottleneck of the process, blocking the exit of the final product.

4.1. Data collection methodology

The four steps of the proposed methodology to collect process data in the packers' case will be described in the following sections.

4.1.1. Step 1. General information

The first step is to characterize the work, the environment, and the worker.

The packers are responsible for performing one by one inspection of every single tile produced and packing them on crate pallets. Quality is assessed in terms of roughness, colour homogeneity, thickness and presence of imperfections - mainly quartzite lines and waving. Crate pallets are directly put on the ground so classifiers have to bend their back and extend their arms every time they place a lot of tiles. The operation is repeated until the pallet is full (around 2,300 tiles in three layers). The cycle starts with the classifiers taking a pile of tiles (a number between 10 and 12 tiles, depending on their thickness and geometry, and weighing circa 10 kilos) and walking to place them inside crate pallets.

In this company, packers usually are medium-aged men or women in the same proportion. They have been represented by the P50 of French population for proximity with their physical conditions. Their physical workstations include a table from which slates are picked up and several pallets for different thickness and qualities of slate. Geometrical information has been acquired in situ.

Figure 4. Packing Task in the Slate Process. (Source. Asociación Gallega de Pizarristas)

4.1.2. Step 2. Work elements

The second step of the data collection methodology is to define the work elements or subtasks. In particular, we can divide the entire classifier task in three subareas, to facilitate the analysis of each part:

1. Inspect, pick up and lift a group of slates
2. Transportation (turn around and walk up to the pallet)
3. Placement into the pallet and counting process. The placement of the piles occurs around 65 times per layer. For the purpose of this analysis, the placement operation has been divided in type I (worst ergonomic conditions, place tiles in the center of the first layer of the pallet), type II (all the rest of the placements in the first layer or any placement on the second layer), or type III (third layer)

4.1.3. Step 3. Motion data acquisition

The next step is the motion data acquisition, performed with the proposed system in our laboratory. A subject has been recorded performing the operations of pick up, transportation of slates and placement into a crate pallet. The experimental setup for the placement operation (type II) is shown in Figure 5. In this case, the disposition of the system cameras has been frontal to the user for the depth device and lateral (90˚) for the conventional camera. Both devices have been placed at a horizontal distance of 2 m and 1 m height.

Figure 5. Experimental Setup for Data Collection during Placement Operation

4.1.4. Step 4. Data Post-processing and postures definition

The depth camera opposite the subject performed a good control of each subtask movement. The lateral conventional camera provided a second set of angles to compare and complete the angles from the depth camera (especially trunk, arm and neck flexion, as shown in Figure 6).

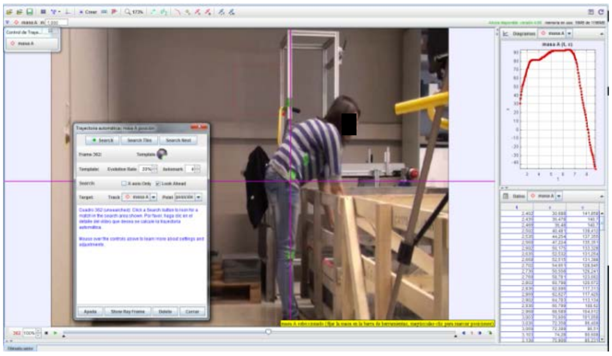


Figure 6. Tracker GUI over a Task Video during the Experimentation

Motion data have been used to perform the set of models that are presented in next section.

4.2. Task analysis

The workplace physical layout has been modelled with the geometrical information obtained in the first step. The model has been reproduced for male and female manikin when performing each of the subtasks defined on the second step. Each subtask is made up from the different postures. The posture definition included the motion data obtained in step three.

The descriptive analysis of the task includes the modelling of the inspection and pick up, the transportation and the placement from different types. The key performance indicators included to perform the ergonomic evaluations have been the RULA score and the L4/L5 vertebrae moment on the spine. RULA is a well-known ergonomic index, proposed in 1993 (McAtamney 1993). It has been especially designed for the assessment of tasks that mainly imply the upper limbs as is the case with the packers. The Spine Compression value is a complementary measure of risk of MSDs. According to NIOSH guidelines, compression force on the intervertebral disk over 3.4 kN may eventually lead to injuries. Both RULA scores and L4/L5 values are provided by Delmia V5R20.

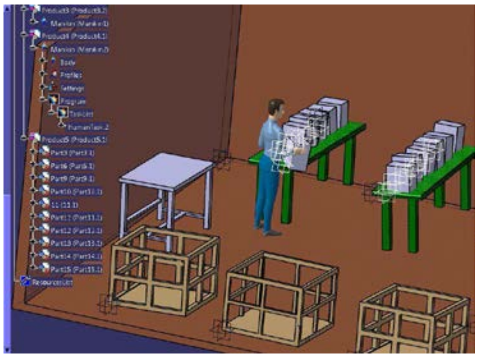


Figure 7. Model Capture of the Inspection and Pick up Operation

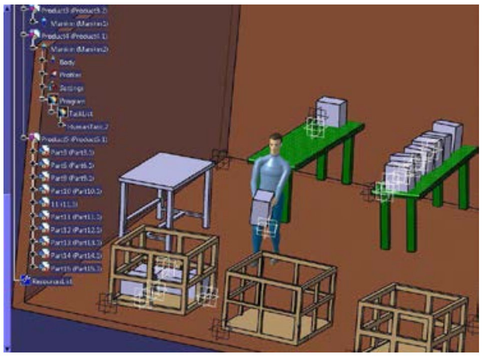


Figure 8. Model Capture of the Transportation of a Group of Slates

The pickup operation model is shown in Figure 7. This task is performed at the beginning of the cycle. Each tile is inspected to assess its quality and the group of tiles (around 10 or 12) is then picked up. The transportation is presented in Figure 8. The placement operation has been modelled for the worst (type I), intermediate (type II) and best conditions (type III), as it can be seen in Figure 9, Figure 10 and Figure 11 respectively.

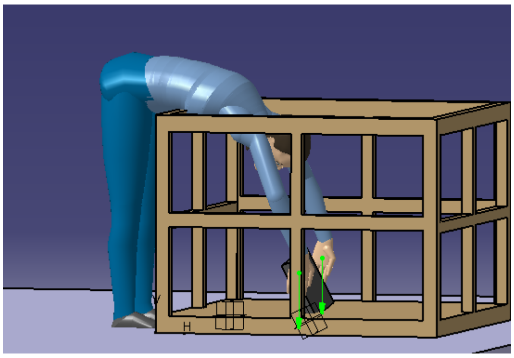


Figure 9. Model Capture of the Placement Operation in Worst Conditions (Type I)

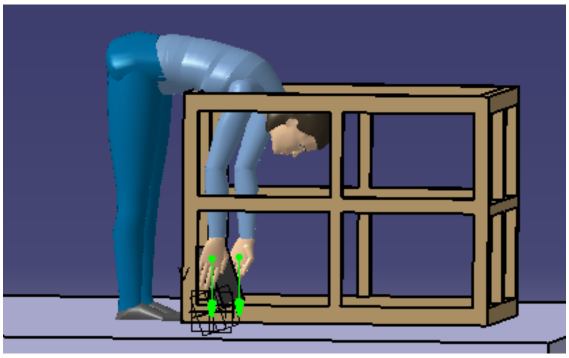


Figure 10. Model Capture of the Placement Operation, Intermediate Conditions (Type II)

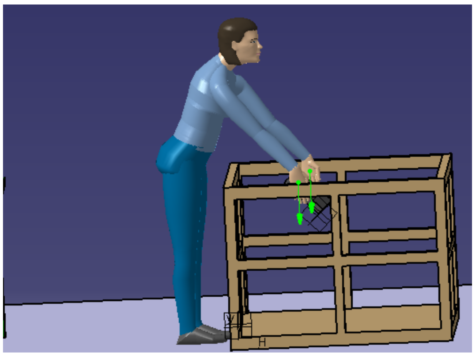


Figure 11. Model Capture of the Placement Operation, Best Conditions (Type III)

Ergonomic analysis results are summarized in Table 2. The postural risk is measured by the RULA index, which classifies the risk of developing MSD's from 1 (no risk at all) to 7 (urgent need of change). The L4/L5 compression on the spine accounts for the spine stress during the task. The NIOSH limit of 3,400 N has to be regarded as a reference for safe or unsafe task. Both maximum values and average are useful to describe the risk profile of the task. On the one hand, the maximum gives an idea of the pick forces achieved during the subtask. On the other hand, the average is a measure of cumulative stress.

Table 2. Ergonomic Results of the Packer's Task

	RULA		L4/L5 Compr. (N)	
	Average	Max	Average	Max
Pick up	3.54	6.00	748.92	1511.11
Transport	2.58	5.00	1168.96	1328.26
Placement (type I)	7.00	7.00	2852.00	4647.00
Placement (type II)	6.42	7.00	1693.63	2798.00
Placement (type III)	3.14	5.00	1168.63	1328.00

The pickup, transport and type III placement operations remain in relatively safe levels of postural risk and spine compression. Types I and II placements show a high risk profile in terms of RULA score. In the case of type I placement, the task is also very hard when considering the spine compression (that exceeds the NIOSH safe limit).

For the packers, trunk and arms are very likely to suffer MSD and therefore, these results suggest the need to redesign the task to alleviate their work conditions. The overall placement risk has been obtained by calculating a weighted average, in which the weigh is the frequency of each type of placement (Table 3)

Table 3. Overall Assessment of the Placement Operation

Type I	Type II	Type III
12%	55%	33%
Overall RULA = 5.34		
Overall L4/L5 = 1640.51 N		

Therefore, the objective of the workplace improvement is to reduce the overall RULA and L4/L5 values of the present workstation configuration.

4.3. Improvement measures

The proposed improvement for this task implies to use a modified crate pallet with an accesible lateral part to help its fill. It is initially placed on an automatic lifting platform at the corresponding ergonomic standard height avoiding problems to the worker (70 cm from the ground in our case). As the classifier fills the pallet with a row of plates, he can operate the automatic mechanism of the platform so that it is lowered and he can continue to fill the pallet with the slates always at the same height, which would not involve any physical effort that might cause ergonomic risks in the long time.

The analysis of this task implies predicting the postures that the new workstation would produce. The experimental setup has been also used to obtain motion data of the proposed modification.

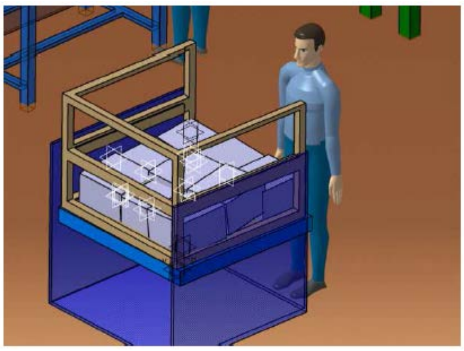


Figure 12. Proposed Workstation to Reduce the Ergonomic Risk of the Placement Operation.

In this case the filling process of each layer is equivalent. However, there still exists a difference in the placement of the slate blocks closer to the free part of the operator or the farthest possible. The ergonomic analysis of this workstation proposal is provided in Table 4.

Table 4. Ergonomic Results of the Modified Placement Workstation

	RULA		L4/L5 Compr. (N)	
	Average	Max	Average	Max
Placement (closest)	2.47	5.00	627	2495
Placement (farthest)	4.56	7.00	1153	2650
Overall	3.52	-	890	-

As it can be noticed in Table 4, the maximum value of the global score in RULA is still high. However, the average value in the improvement is much lower and so, better. For this reason it can be said that the proposed improvement would actually reduce the ergonomic stress of the operator.

5. CONCLUSIONS

A general methodological description for the motion data collection to perform ergonomic analysis and improvement projects of manufacturing tasks has been presented. The proposed methodology is portable, markerless and very affordable, especially when compared to other MOCAP systems. An illustrative case study has also been described.

During the experiments, it became clear that the error of the depth device was motion dependent and for large range of motion the errors were larger. Although there are several references that had studied the accuracy of the device to track 3D measurements, these practical considerations seem to be systematically not considered in the studies yet. More research has to be performed to clearly establish the practical limitations to obtain reliable posture information in the field.

This work supposes an improvement compared to the traditional observation of videos for estimation of postures when performing a task. However, a limitation of this approach is that, until now, the posture input modelling into the DHMS tool is still a manual task, and therefore there is a limitation on the number of postures per second to be introduced for practical reasons. We are currently working on the automation of the postures definition to solve this problem. To do so, a different but supplementary issue has to be regarded. The postural information needed to simulate a task is dependent on the speed of the movement. Variation between consecutive data should be considered as a factor to determine the appropriate number of postures that have to be introduced on the model. This is directly related to the tracking frequency of the depth device. The optimal solution would be an adjustable recording in which frequency is established according to the motion observed. Our future work lines also include this approach to the problem.

Another important advantage of the use of this methodology compared with the traditional visualization of videos is the possibility to obtain other useful information to define the task. The tracked data can be used to obtain information of the speed, acceleration, cycle time, frequency, rest time, within-minute range variation, between-minute variation, etc. However, dynamic aspects of the task are often disregarded (Wells et al. 2007). Probably this is related to two main factors. On the one hand, there used to be a need of more sophisticated tools to obtain speeds and accelerations. On the other hand, because of the lack of standardized criteria (such as speeds, work pace or repeatability) of the maximums to establish whether the risk exists in the task or not. The potential of a data collection methodology such as the one proposed is also linked to the development of dynamic ergonomic standards.

BIBLIOGRAPHY

Ausejo S., Wang X., 2009. Motion Capture and Human Motion Reconstruction. In: V. G. Duffy, ed. *Handbook of Digital Human Modelling*. Boca Raton, Florida: CRC Press, Taylor & Francis Group, 38.1-38.13.

Best, R., Begg, R., 2006. Overview of movement analysis and gait features. In: R. Begg, M. Palaniswami, eds. *Computational Intelligence for Movement Sciences.*. Hershey, PA: Idea Group Publishing, 1-69.

Bryan, J., 2004. Video Analysis Software and the Investigation of the Conservation of Mechanical Energy. *Contemporary Issues in Technology and Teacher Education*, 4(3), 284–298.

Clark, R., Pua, Y., Fortin, K., Ritchie, C., Webster, K.E., Denehy, L., Bryant , A.L., 2012. Validity of the Microsoft Kinect for assessment of postural control. *Gait & posture*, 36(3), 372–377.

comPADRE, 2012. *Tracker Video Analysis and Modeling Tool for Physics Education*. Available from: http://www.cabrillo.edu/~dbrown/tracker/ [Accessed December 17, 2012]

Crespo Pereira D., del Rio Vilas D., Rego-Monteil N, Rios Prado R., 2012. Simulation and Highly Variable Environments: A Case Study in a Natural Roofing Slates Manufacturing Plant. In: S. Bangsow, ed. *Use Cases of Discrete Event Simulation*. Heidelberg, Germany: Springer-Verlag, 147-178.

del Rio Vilas D., Crespo Pereira D., Crespo Mariño J.L., Garcia del Valle A., 2009. Modelling and Simulation of a Natural Roofing Slates Manufacturing Plant. *Proceedings of The International Workshop on Modelling and Applied Simulation*, pp. 232-239. September 23-25, Puerto de la Cruz (Tenerife, Spain).

del Rio Vilas, D., Longo, F., Rego-Monteil, N., 2012. A general framework for the manufacturing workstation design optimization: a combined ergonomic and operational approach. *Simulation Transactions of the Society for Modeling and Simulation International*, 89, 306-329.

Dutta, T., 2012. Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied ergonomics*, 43(4), 645–9.

MSDN, 2013. *Library for developers using Microsoft Tools*, *Skeletal Tracking Overview*. Available from: http://msdn.microsoft.com/en- us/library/hh973074.aspx [Accessed December 17, 2012]

Hedrick, T.L., 2008. Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspiration & biomimetics*, 3(3), p.034001.

Rego-Monteil N., del Rio Vilas D., Crespo Pereira D., Rios Prado R., 2011. An Overall DHM-Based Ergonomic and Operational Assessment of a Manufacturing Task: A Case Study. *Proceedings of the 10th International Conference on Modelling and Applied Simulation*, pp. 375-382. September 12-14, Rome (Italy).

Rego-Monteil N., del Rio Vilas D., Crespo Pereira D., Rios Prado R., 2010. A Simulation-Based Ergonomic Evaluation for the Operational Improvement of the Slate Splitters Work. *Proceedings of the 22nd European Modeling and Simulation Symposium*, pp. 191-200. October 13-15, Fes (Morocco).

Savitzky A.; Golay M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry,* 36 (8), 1627–1639.

Stephens, A., Jones, M., 2009. Workplace Methods and Use of Digital Human Models. In: V. G. Duffy, ed. *Handbook of Digital Human Modelling*. Boca Raton, Florida: CRC Press, Taylor & Francis Group, 1-11.

Wells, R., Mathiassen, S.E., Medbo, L., Winkel , J., 2007. Time--a key issue for musculoskeletal health and manufacturing. *Applied ergonomics*, 38(6), 733–44.

Zhou, W., Armstrong, T.J., Reed, M.P., Hoffman, S.G., Wegner, D., 2009. Simulating Complex Automotive Assembly Tasks using the HUMOSIM Framework. *Proceedings of the 2009 SAE Digital Human Modeling for Design and Engineering Conference*. p. 2009-01-2279. June 9-11, Göteborg (Sweden).

AUTHORS BIOGRAPHY

Nadia Rego Monteil obtained her MSc in Industrial Engineering in 2010 from the University of A Coruna (UDC). She has been working as a research engineer at the Integrated Group for Engineering Research (GII) and has also worked as a research Industrial Engineer for the Galician Naval Technological Centre (CETNAGA). Her areas of major interest are in the fields of Ergonomics, Process Optimization and Production Planning.

Mariangela Suriano obtained her Master Degree in Management Engineering from the University of Calabria (Italy) in February 2013. She collaborated with the GII of the UDC (Spain) during her last year of degree. Since March 2013 she has been collaborating with the Modelling & Simulation Center – Laboratory of Enterprise Solutions (MSC- LES).

Diego Crespo Pereira, MSc Industrial Engineer**,** PhD. Diego is Assistant Professor of the Department of Economic Analysis and Company Management of the UDC. He also works in the GII of the UDC as a research engineer since 2008. He is mainly involved in the development of R&D projects related to industrial and logistical processes optimization. He has also worked in the field of human factors affecting manufacturing processes.

Rosa Rios Prado works as a research engineer in the GII of the UDC since 2009. She holds an MSc in Industrial Engineering. She has previous professional experience as an Industrial Engineer in an installations engineering company. Rosa`s areas of main expertise relate with the development of transportation and logistical models for the assessment of multimodal networks and infrastructures.

David del Rio Vilas holds an MSc in Industrial Engineering from the University of A Coruna, where he is Adjunct Professor of the Department of Economic Analysis and Company Management. Since 2007 he has been working in the GII mainly involved in R&D projects related to industrial and logistical processes optimization. Since 2010 he is R&D Manager in Proyfe S.L., a Spanish company within the Civil Engineering sector.

Francesco Longo received his Ph.D. in Mechanical Engineering from University of Calabria in January 2006. He is currently Assistant Professor at the Mechanical Department of University of Calabria and Director of the Modelling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES). He has published more than 80 papers on international journals and conferences. He is Associate Editor of the "Simulation: Transaction of the society for Modeling & Simulation International".

COMPARING MATERIAL FLOW CONTROL MECHANISMS USING SIMULATION OPTIMIZATION

Manuela André[(a)], Luís Dias[(b)], Guilherme Pereira[(c)], José Oliveira[(d)], Nuno Fernandes[(e)], Sílvio Carmo-Silva[(f)]

[(a) (b) (c) (d) (f)] University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal.
[(e)] Polytechnic Institute of Castelo Branco, Av. do Empresário 6000-767, Castelo Branco, Portugal
[(b) (c) (d)] Algoritmi Research Center

[(a)] manuela.andre@ics.uminho.pt, [(b)] lsd@dps.uminho.pt, [(c)] gui@dps.uminho.pt, [(d)] zan@dps.uminho.pt, [(e)] nogf@ipcb.pt, [(f)] scarmo@dps.uminho.pt

ABSTRACT
In this study, discrete event simulation is used for comparing the performance of three material flow control mechanisms: push-MRP, Generic Kanban System (GKS) and generic Paired-cell Overlapping Loops of Cards with Authorization (GPOLCA). The former does not impose restriction to the number of jobs that are released into the supply chain. The latter two are card-based control mechanisms, where the number of jobs in the supply chain is restricted. The simulation models of these mechanisms are developed in Arena® and optimized using OptQuest®. The average total work in process and the average system throughput are used to evaluate the performance of the mechanisms. We found that GKS outperforms GPOLCA and MRP for high levels of throughput.

Keywords: simulation optimization, GPOLCA, GKS, MRP, ARENA, OptQuest

1. INTRODUCTION

Global competition has forced manufacturers to seek innovative ways to manufacture and control the flow of materials. The success of just-in-time (JIT) production, along with their pull production and logistic control methods, has led to considerable interest in the study of control mechanisms for manufacturing systems (Krishnamurthy and Suri 2009). These mechanisms can be classified into push, pull or hybrid push-pull (Hopp and Spearman 2004).

Push systems are typically associated with material requirements planning (MRP). Pull systems are also called kanban control systems (Krishnamurthy and Suri 2009). The Toyota Kanban System (TKS) (Sugimori 1977) is a card-based pull system that has attracted the attention of many companies. However, it was created to fulfil the specific needs of a company (Toyota). It is not suitable in situations with unstable demand, processing time instability, non-standardised operations, long setup times, great variety of items, and raw material supply uncertainty (Junior and Filho 2010). Thus, variations (or adaptations) were created to adapt TKS to the companies' specific reality. For example the Generic Kanban System (GKS) was proposed by Chang and Yih (1994) as an adaptation of TKS for non-repetitive production environments. Junior and Filho (2010) review the literature regarding variations of TKS.

Given the numerous control mechanisms introduced in recent years, it is not an easy task to evaluate which is the best approach for a specific situation. Thus it is an important issue to address the problem of how to compare these mechanisms.

In this study we use simulation optimization to: first, optimize and then, compare material flow control mechanisms. The simulation models were developed in Arena® (Kelton and Sadowsky 1998, Dias, Pereira, Vik and Oliveira 2011) and the control mechanisms were optimized using the OptQuest® tool (Bapat and Sturrock 2003, Rogers 2002).

A case study was developed to demonstrate the methodology by evaluating the performance of a recently introduced mechanism called generic Paired-cell Overlapping Loops of Cards with Authorization (GPOLCA) (Fernandes and Carmo-Silva 2006) along with existing control mechanisms, such as GKS and push-MRP (which is used as a benchmark mechanism) in the context of a small supply chain with multi-products and stochastic operation times.

This paper has two purposes: one is to demonstrate how simulation optimization can be used to solve complex design and control problems; the other is to evaluate the performance of three material flow control mechanisms, namely GKS, GPOLCA and push- MRP.

The paper is organised as follows. In the following section, section 2, a description of the studied material flow control mechanisms is provided. In section 3 the case study and the research design and experimental setup are described. Results and findings from the experiments are provided in section 4. Finally section 5 summarizes the study conclusions and includes directions for future research.

2. MATERIAL FLOW CONTROL MECHANISMS

Each material flow control mechanism coordinates the release of jobs to the supply chain and its progress from one stage to another in different ways. Thus, this section describes how GKS, GPOLCA and push-MRP operate as illustrated in Figure 1.

The GKS mechanism, introduced by Chang and Yih (1994), is a material flow control mechanism suited to make-to-order MTO non-repetitive and dynamic production environments. One characteristic inherited from the TKS is that it uses a well-defined number of cards for each stage of the supply chain, as a means of controlling the work-in-process (WIP). These cards (or kanbans) are not specific of any particular product, contrarily to what happens in the TKS, and thus can be attributed to any job waiting to be released into the system. A job cannot be released unless it acquires all the cards that it needs, from each supply chain stage, according processing requirements, as long as there are cards available, i.e., not yet allocated to jobs. The number of cards allocated to a job, for each processing stage in the supply chain, depends on its workload, since each card represents a specific amount of workload. After release jobs are pushed through the supply chain and, as they are completed at each stage, the attached cards are dropped to the respective cards' boxes, becoming available for new requests, i.e. for allocation to new jobs waiting to be released. Figure 1a) illustrates the operation and the elements of the GKS material flow control mechanism.

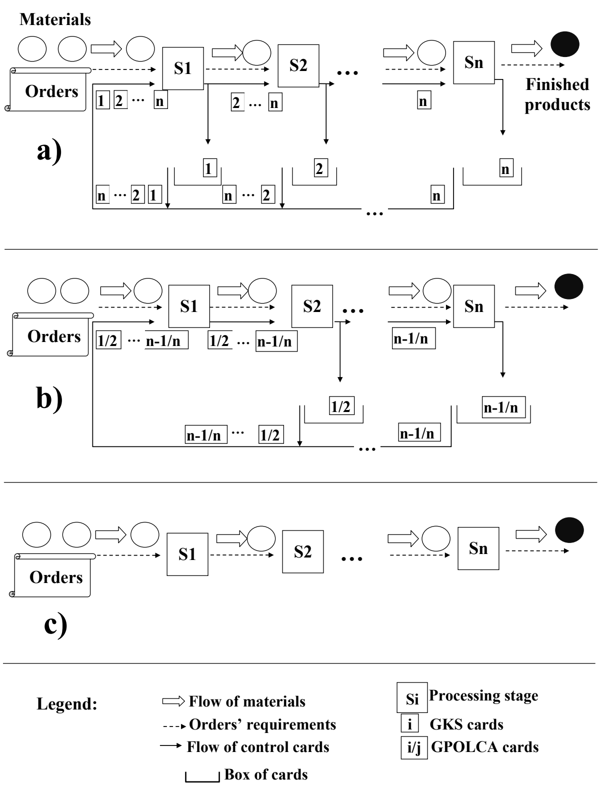


Figure 1: Material flow control mechanisms: a) GKS, b) GPOLCA, c) Push-MRP.

GPOLCA was introduced by Fernandes and Carmo-Silva (2006). Figure 1b) illustrates the GPOLCA mechanism. It is also suited for MTO environments. GPOLCA adapts the POLCA (Suri 1998) mechanism to production environments with high routing diversity, by controlling job release through a combination of planned release dates and production authorization cards, which are allocated to pairs of stages in a supply chain. Moreover, the second stage of a pair in a card is the first in the card that must follow, according to job routing. A clear operating advantage of GPOLCA in relation to GKS, resulting from this, is that the cards have routing information, i.e. it is always possible to know, solely based on the GPOLCA cards to which stage to send a job after it has been processed in another. Nevertheless GPOLCA mechanism controls the release and the flow of work in a manner similar to GKS. Thus, GPOLCA also uses generic cards, i.e. cards that are not specific of any particular product, to control the number of jobs or the workload in the supply chain, and a job to be released must seize the required number of cards. However, cards only become available after the processing is carried out on the second stage of the pair of stages of the card.

In the push-MRP mechanism job release is not controlled by production authorization cards, which means that no limit is imposed to the number of jobs in the shop floor neither to the flow of work between the supply chain stages. Job release is based only in planned release dates and jobs start processing at stages according the dispatching mechanism adopted. Figure 1c) illustrates the operation and the elements of the push-MRP mechanism.

3. SIMULATION MODEL AND EXPERIMENTAL DESIGN

3.1. Simulation model

A discrete-event simulation model was developed using Arena® software. We consider a supply chain with two products types and three stages, identical to that used by Krishnamurthy et al. (2004) (see Figure 1 and Figure 2). A stage represents a major processing function in the supply chain, e.g., procurement of a raw material, the fabrication of a part or the subassembly of a component, test of a finished product and the transportation of goods from a central distribution centre to a regional warehouse.

The physical and operational configuration of the production system includes:

- The same routing for all jobs (products): first they are processed at stage 1, then at stage 2 and last in stage 3;
- A Poisson process for the demand arrival rate;
- Exponentially distributed job processing times at each stage with mean equal to one time unit;
- Set-up times considered as part of the operation times;
- First-come-first-served (FCFS) priority dispatching at all stages of the supply chain.

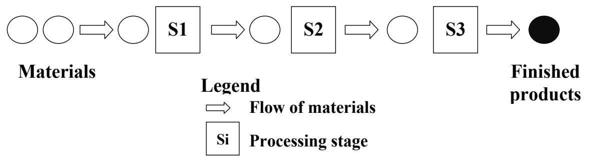


Figure 2: Simulated three-stage supply chain

We also consider the presence of short-term product mix imbalances caused by changes in demand. These changes take place over a short period of time, i.e. during a day, at the seventh day of every week.

The two types of jobs have an equal probability of arriving to the system, except at the seventh day. In this day the total demand for the two products, λ1+λ2, is kept unchanged, acting upon the demand rate ratio λ1/λ2 to achieve a product mix change. This ratio was set to 1/2. To ensure that workload stays balanced, the average service time μ of each job in each supply chain stage is kept unchanged and, therefore equal to one. Thus, the following equation applies:

$$\frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2} = 1 \qquad (1)$$

Assuming a service time ratio μ1/μ2 of 2 the expected mean service time of each job is determined by equation (1).

3.2. Experimental Design

The experimental factors and simulated levels considered in this study are summarized in Table 1. The material flow control mechanism is tested at three levels (MRP, GPOLCA and GKS), job arrival rate is tested at 12 levels, varying arrival from 0.5 to 0.95 jobs per time unit. This results in a full factorial design with thirty-six, i.e. 3x12, test cases.

Table 1: Experimental factors and corresponding levels

Factors	Levels		
Mechanism	MRP	GPOLCA	GKS
Release rate	0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.875, 0.9, 0.925, 0.95		

The main objective of applying a material flow control mechanisms in supply chains is reducing work in progress (WIP). To achieve this objective, we optimize the number of GKS and GPOLCA cards using OptQuest. Thus the release rate at these mechanisms results from the minimum number of cards that leads to the desired system throughput (TP).

For all material flow control mechanisms the performance measures monitored are TP, WIP and Flow time. TP is measured as the mean number of completed jobs per time unit. WIP is defined as the number of jobs released into the production system, but not yet finished. Flow time is measured as the time between the job release and its completion.

During simulation experiments, data is collected under steady state. Each simulation was run for 50 independent replications of 96000 time units with a warm-up period of 9600 time units to ensure that steady-state condition was reached.

3.3. Simulation optimization for card-based systems

To compare material flow control mechanisms first they must be optimized. Since material flow control mechanisms, such as GKS and GPOLCA, enforce WIP limits by restricting the number of cards that can be in the system, and consequently the buffer capacity at stages, this number is a critical parameter that must be optimized.

For determining the minimum number of cards needed for each throughput rate, we used the tool OptQuest for Arena.

Although the maximum throughput can be achieved with any number of cards above the minimum, the objective is to minimize the number of cards that achieves the desired throughput. This, in turn, reduces the WIP for each card-based material flow control mechanism.

The minimum number of cards, obtained from the best solution given by OptQuest, is then used to run the simulation in order to evaluate and compare the material flow control mechanisms.

We exemplify the optimization procedure adopted focusing on GPOLCA. In this case, Num_Card_1_2 and Num_Card_2_3 are variables in the Arena model, identified as user specified Controls in OptQuest. For these controls we specified the range to be used by the optimizer, through low and high boundaries.

In the Responses option we have to select items that we want to use in the constraint, in this case the Arena variable TRP (throughput).

A constraint was defined in order to ensure a TRP larger than 99.98% of the release rate, as shown in Figure 3. For instance, for the average release rate of 0.50 jobs per time unit in the Arena model, we use in OptQuest the Constraint: *[TRP] > (4999/10000).*

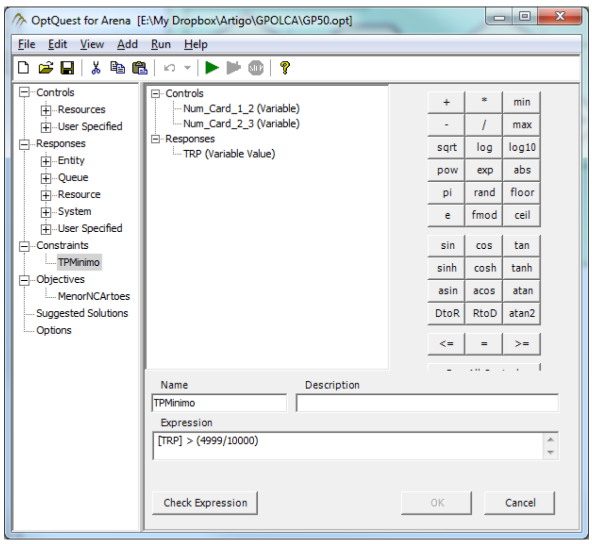


Figure 3: OptQuest constraint

The Objective function is: *minimize [Num_Card_1_2] + [Num_Card_2_3]*. The cards number boundaries and the constraint value varies according to each release rate. This was evaluated for all the release rate values tested.

The OptQuest optimization run, with the above parameterization, is shown in Figure 4. In the graphic of this figure, we can follow the best objective function value evolution. For each iteration OptQuest performed a simulation using different values for the number of cards, keeping the best solution founded until it runs all the possible combinations. The OptQuest best feasible solution, i.e. number of cards that ensures minimum WIP for meeting the required TRP is also shown in Figure 4. For the case where TRP is 0.5 and GPOLCA is used, the best solution funded by OptQuest was 2 and 3 cards respectively for the card type 1_2 (pair 1 and 2 of processing stages) and type 2_3.

The amount of time needed to complete each optimization can be between ten minutes to one hour, depending on the high boundary value defined to the cards number, in a core 2 Duo 3.16GHz processor.

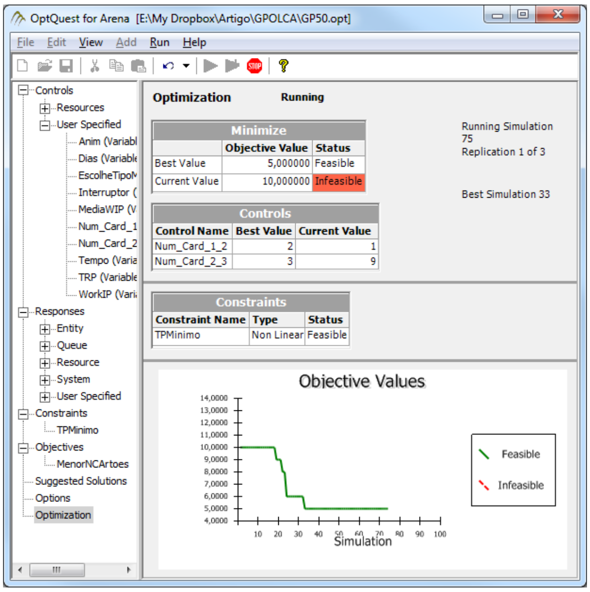


Figure 4 : OptQuest running

4. RESULTS OF THE EXPERIMENTS

In this section we report the results of the set of experiments conducted in order to compare the material flow control mechanism performance. The 95% confidence intervals were developed for the mean value of each performance measure, namely WIP, Throughput (TP) and Flow Time.

Results are shown in Figure 5 and Table 2. The former shows a production logistic curve for each of the control mechanisms studied: push-MRP, GPOLCA and GKS. These curves show the average WIP as a function of the system throughput. A point on a curve corresponds to the best cards configuration of the mechanism that achieves that throughput with the lowest WIP. A particular control mechanism is superior to another if, for a given TP it shows a lower WIP.

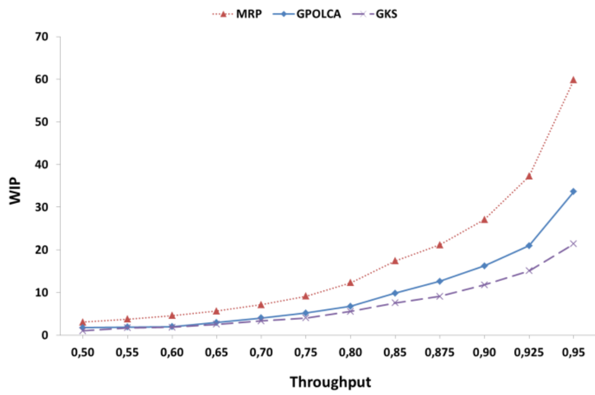


Figure 5: Mechanisms performance

Table 2: Mechanisms' performance for a TP of 0.95

Mechanism	WIP	Flow time
MRP	59.8 ± 2.19	62.6 ± 2.07
GPOLCA	33.6 ± 0.23	36.4 ± 0.14
GKS	21.4 ± 0.15	23.3 ± 0.11

It is known that as throughput approaches the system capacity, i.e. the maximum possible throughput, WIP tends to infinite. This can be observed in Figure 5.

Table 2 shows performance results for a throughput of 0.95. A paired t-test revealed that, the difference between mechanisms, for WIP and flow time, is significant at 95% confidence level.

Analysing the above results the following observations can be made. The total WIP required to meet a particular TP is higher under MRP, particularly for high levels of TP. Results under GPOLCA are clearly better than under MRP and improve as the system TP increases. For a TP of 0.95, changing from MRP to GPOLCA produced a reduction in WIP of 44%. This means that the same TP can be obtained under GPOLCA with less WIP, as shown in Table 2, and also lower flow time, as could be expected by the Little's law. The better performing mechanism is GKS having a reduction of WIP in relation to MRP in about 64% for a TP of 0.95. Over a large range of TP values, GKS and GPOLCA have practically identical performance rates. However, at the highest levels of TP, GKS performs better than GPOLCA. In fact, for the stated TP objective of 0.95, changing from GPOLCA to GKS reduces WIP in about 36%.

From these results we may conclude that the strategy adopted by GKS of controlling the release of jobs to every stage of a serial supply chain performs better than the strategy used by GPOLCA of controlling the release of jobs to pairs of stages.

5. CONCLUSIONS

This paper considers the material flow control within a supply chain with stochastic operations times and highly variable product demand. Three material flow control mechanisms were compared by simulation, namely: push-MRP, GPOLCA and GKS, all suitable for MTO environments. Before being compared, the latter two were first optimized in relation to the number of production authorization cards (or kanbans). These card-based mechanisms showed to successfully reduce work-in-process, with GKS performing better than GPOLCA.

In the study the FCFS dispatching rule was used. An important avenue for future research would be to investigate how the dispatching rule interacts with the materials flow control mechanisms. It would also be interesting to investigate which method yield robust performance in the presence of bottlenecks.

Since several other material flow control mechanisms have emerged recently it is important to identify the suitability of them to the demand environment studied and confront them with GKS. This will help to improve knowledge about the mechanisms that should be recommended for supply chains of high variable product demand.

ACKNOWLEDGMENTS

This work had financial support of FCT-Fundação para a Ciência e Tecnologia of Portugal under the project PEst- OE/EME/UI0252/2011.

This work has been supported by the "Programa Operacional Fatores de Competitividade - COMPETE" and by the FCT - Fundação para a Ciência e Tecnologia in the scope of the project: FCOMP-01-0124-FEDER-022674.

REFERENCES

Bapat, Vivek and David T. Sturrock. 2003. The arena product family: enterprise modeling solutions. In *Proceedings of the 35th conference on Winter simulation: driving innovation* (WSC '03), New Orleans, Louisiana, USA, December 7-10, 2003, 210-217,

Chang, T.M., Yih, Y., 1994. Generic kanban systems for dynamic environments, *International Journal of Production Research*, 32 (4), 889-902.

Dias, L.S., Pereira, G.B., Vik, P., Oliveira, J., 2011. Discrete Simulation Tools Ranking – a Commercial Software Packages comparison based on popularity, *ISC 2011 - Industrial Simulation Conference*, 5-11, June 6-8 2011, Centro Culturale Don Orione, Venice, Italy.

Fernandes, N.O., Carmo-Silva, S., 2006. Generic POLCA—A production and materials flow control mechanism for quick response manufacturing, *International Journal of Production Economics,* 104, 257-262.

Hopp, W.J., Spearman, M.L., 2004. To Pull or Not to Pull: What Is the Question?, *Manufacturing & Service Operations Management*, 6, 133-148.

Junior, L.M., and Filho, G.M., 2010. Variations of the Kanban system: Literature review and classification. International Journal of Production Economics, 125(1), 13-21.

Kelton, W.D., Sadowsky, R.P., Sadowsky, D.A., 1998. Simulation with Arena, Boston-USA, McGraw-Hill.

Krishnamurthy, A. and Suri, R. 2009. Planning and Implementing POLCA: A Card-Based Control System for High Variety or Custom Engineered Products, *Production Planning and Control*, 20 (7), 596-610.

Krishnamurthy, A., Suri, R., Vernon, M., 2004. Re-examining the performance of MRP and Kanban material control strategies for multi-product flexible manufacturing systems, *International Journal of Flexible Manufacturing Systems* 16, 123-150.

Rogers, P. 2002. Simulation of manufacturing operations: optimum-seeking simulation in the design and control of manufacturing systems: experience with optquest for arena. In *Proceedings of the 34th conference on Winter simulation: exploring new frontiers* (WSC '02), San Diego, California, USA, December 8-11, 2002, 1142-1150,

Sugimori, Y. Kusunoki, K., Cho, F. and. Uchikawa, S., 1977. Toyota production system and Kanban system Materialization of just-in-time and respect-for-human system *International Journal of Production Research, 5 (6), 553-564.*

Suri, R., 1998. *Quick Response Manufacturing: A Company-Wide Approach to Lead Time Reduction*. Portland-USA, Productivity Press.

AUTHORS BIOGRAPHY

Manuela André was born in 1971 in Cabinda, Angola. She graduated in Information Systems and Technologies in the University of Minho, Portugal and is finishing the Master's in Education. Her main interests are computer science, education and photography.

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 Modelling.

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.

Nuno O. Fernandes is currently working as a professor in the School of technology of Polytechnic Institute of Castelo Branco, Portugal. He graduated in Production Engineering and received a PhD in Industrial Management in 2007 from University of Minho. His research interests are in workload control, card-based control systems and production planning and control in general.

Sílvio Carmo-Silva, is an Associate Professor of the University of Minho and research coordinator of the Industrial Management and Systems Research Group (IMS) of the Centre for Industrial and Technology Management (CITM). He obtained a master degree in Management and Technology from UWIST, UK, in 1980, and a PhD in Manufacturing Engineering, from Loughborough University of Technology, in UK, in the year 1988. His main research work is on production systems organization and methods and mechanisms for production and materials flow control. Presently is responsible for Production Activity Control and Production Systems Organization curricular units, among others, within post-graduate courses of the University of Minho.

THE USE OF SIMULATION AS A TOOL TO SUPPORT DECISION-MAKING IN A PISTON MANUFACTURING SYSTEM

Alexandre Augusto Massote[(a)], Bruno Alvarez Ferreira Ignácio[(b)]

[(a)] Centro Universitário da Fei
[(b)] Fundação Armando Alvares Penteado - FAAP

[(a)]massote@fei.edu.br, [(b)]bruno.alvarez@outlook.com

ABSTRACT
The business scenario is characterized by a high complexity, which influences various aspects of the organizations, such as the decision-making. In this way, companies can use several techniques to reduce the risks and uncertainties of this process; among them is the computer simulation. Considering that, this article aims to demonstrate the effectiveness of this technique. In order to do so, it will simulate several situations of a piston manufacturing system linked with the expansion of the manufacturing cell. In summary, the method will be used as a support for the manager to decide which is the best option to expand the manufactory cell.

Keywords: Computer Simulation; Decision-making; Manufacturing System

1. INTRODUCTION

This work presents the use of simulation as a tool for decision support in manufacturing environments, to evaluate different alternatives for improving productivity and therefore, various situations that can occur in a manufacturing system are simulated. These conditions, in turn, were based on a manufacturing cell of a piston factory located in the state of São Paulo - Brazil.

In this sense, it is the perception that the business setting faced by organizations in recent decades is characterized by increasing complexity, so that companies are dealing with situations that make decision making increasingly difficult in their daily routines, such as: the time available for making the decision, the risks and uncertainties inherent in the organizational environment (BATEMAN, SNELL, 2008), the decision-makers and conflicts of interests (LACHTERMACHER, 2002).

Given this context, organizations can use several techniques that assist in the decision making, including computer simulation, a tool of Operational Research that has been gaining prominence in recent years due to the improvement of hardware and software and awareness among companies about the applicability tool, which allows a process of decision making wider, in which managers can simulate various scenarios representing different market variables (SARGENT, 2004).

In its most basic definition, computer simulation can be considered to be a technique based on queuing theory, which reproduces a real system by means of equations or a mathematical model to evaluate and improve performance. Thus, trials of predefined scenarios can be made from its use, verifying how they affect system performance (HARRELL; GHOSH; BOWDEN, 2000; ANDRADE, 2004; HILLIER; LIEBERMAN, 2005). It is still necessary to emphasize that the trials only occur in a controlled environment allowing a substantial reduction of costs, since working in this environment is cheaper than a real system (LAW; KELTON, 2000).

In this sense, one can say that the simulation is a tool that can assist in decision making in different situations, enabling a reduction of the time of the decision process (BARTON, 2004). Furthermore, through it, it is possible to analyze the situation faced in more detail, exploring all possibilities for its resolution thus reducing uncertainty and risk decisions (LIMA; BARBOSA; BEAL, 2003; SLACK et al., 2002).

In this study, the simulation will be used to represent a piston manufacturing system. In this sense, it is understood that the car market has shown a steady growth in the last decade, reaching in 2012 a total of 3,415,486 million vehicles produced, twice the amount produced in 2000 - 1,691,240 million (ANFAVEA, 2013). This increase resulted in an expansion in the amounts of parts made, thereby manufacturing systems have to be expanded or optimized to meet the growing demand.

Thus, the simulation will be used to assist in decision making regarding the expansion of the manufacturing system of pistons. Three different scenarios are simulated using the software Promodel, which involve: increasing the amount and better distribution of inputs received, the increase in the number of employees, and the purchase of new machinery.

In general, the goal is to demonstrate the operational and financial results of these scenarios, facilitating the process of decision making of those responsible for this manufacturing system, since they can both design costs, as the results and consequences of each action.

2. LITERATURE REVIEW

2.1. Organization Decisions

A decision is the act of choosing one alternative among all possible. This option is, according to the decision maker, the best way to achieve the goal, already pondering the consequences and risks (ANDRADE, 2004). Thus, the act of deciding involves at least six elements: the decision maker, the goals the decision maker seeks to achieve; preferences; strategy, the situation, and the result (SIMON, 1997).

When analyzed under the organizational scope, decisions become even more important. That happens because they become more frequent and the majority of cases, more complex having characteristics such as risk and uncertainty. In general, most of the decisions in a company have different characteristics from personal ones. They are complex and extensive, they lack information about the problem faced and involve situations where one can't predict and evaluate the results of the alternatives, thus hindering the taking of decision.

Accordingly, Bateman and Snell (2008) reported that the majority of decisions within an organization lacks structure and entails risk, uncertainty and conflict, which can slow decision making, as decision makers hesitate to act because there are certain difficulties to be faced.

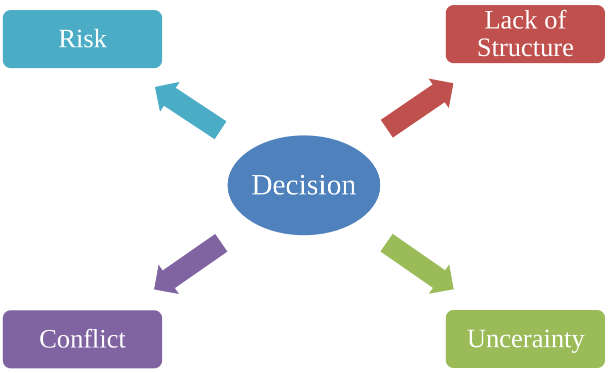


Figure 1: Characteristics of administrative decisions

Risk and uncertainty are present situations in decision-making of any modern organization, involving both the difficulty of predicting the outcome of a particular decision, and the difficulty to estimate and quantify the consequences of the decision. This situation tends to be avoided by decision makers who try to act to anticipate, minimize and control risk (SHIMIZU, 2001; BATEMAN, SNELL, 2008).

The decisions made in organizations can be classified in two types of criteria: the decision level and the degree of structuring of the problem (ANDRADE, 2004; SHIMIZU, 2001).

According to the first criterion, the decisions made within an organization are not restricted to an organizational level and can occur at three different levels: the strategic, the tactical and the operational. The second criterion divides decisions into scheduled and non-scheduled. The routine scheduled are those that are taken in an environment of certainty and have easy solutions. However those unscheduled refer to those natural and difficult to solve (SIMON, 1997; SHIMIZU, 2001).

The organizational decision-making process has evolved over the last century due to the emergence of new management techniques and assimilation of quantitative techniques such as computer simulation, which can minimize the risk and uncertainty inherent in the decisions. Through it, the decision maker can experience different situations, so as to find the optimal solution for the situation, which is against the process of rational decision (SIMON, 1997).

2.2 Computer Simulation

Computer simulation is a technique of Operational Research, which seeks to represent a real system by a computer model (VIEIRA, 2006; HARRELL; GHOSH; BOWDEN, 2000). Mathematical models are developed in order to realize these representations. They are abstractions of reality trying to imitate the main features of it (TURBAN, 1995). From the use of these models it is possible to experiment on the system without major encumbrances (HARRELL; GHOSH; BOWDEN, 2000; ANDRADE, 2004).

Therefore the use of models can provide several benefits for organizations such as: a decrease in the expenditures made in the analysis, a reduction in expenses for research and development, a reduction in analysis time, the use of a large amount of variables, ease of handling models, etc. (TURBAN, 1995; RASGDALE, 2004).

These benefits inherent in modeling and, consequently, the simulation started to draw the attention of companies. It was realized that they could play an important role in the business world, being used by organizations to support the process of decision making (HILLIER; LIEBERMAN, 2005).

In this sense, computer simulation has become a management tool that has gained prominence in recent years. This is because these days most organizations are involved in projects, large or small, which in turn encompass the use of resource-machines, persons, tools, and equipment to be optimized, since they involve considerable investment needs that need to be optimized (VIEIRA, 2006). In this sense, the simulation has gained prominence since it allows companies to represent a real situation in a controlled system, in which it is possible to experiment, in order to find the best alternative to solve the situation without adding cost to the project (VIEIRA, 2006; HARRELL; GHOSH; BOWDEN, 2000).

In this sense, briefly, we can say that the simulation is a tool that can assist in decision making in different situations, enabling a reduction of the uncertainty of decisions. For this, it seeks: to represent a real system; hypotheses and scenarios to be tested, and; to reproduce and analyze the results derived from such conjecture (LIMA; BARBOSA; BEAL, 2003). However, it should be emphasized that the simulation does not replace the decision maker. It only helps him find alternatives to solve the situation faced (DUARTE, 2003).

Even as to its basic objectives, the simulation has four distinct purposes: a) to evaluate the behavior of a system over time, b) to predict how the system simulated will perform under certain circumstances; c) to investigate and understand the system dynamics, and d) to compare the results obtained (VACCARO, 1999).

The computer simulation is an ancient technique, since the first models emerged in the 1950s to assist in solving problems, improving the utilization of business resources. However, only with the improvement of the hardware, which now supports more complex models, and software development, with more friendly interfaces to decision makers (VIEIRA, 2006).

Despite its current reputation, before implementing the use of computer simulation a company has to weigh its advantages, as well as their disadvantages. Among the advantages, it can be noted that: a) it represents complex systems that cannot be modeled mathematically; b) it enables the creation and testing of various situations which are always under the control of the user; c) it enables the study of long term situations; and d) it is cheaper than a real system experimentation (LAW; KELTON, 2000).

The main disadvantage is related to the fact that it is a representation of reality, i.e., that it uses modeling, which is the process of representing reality in a simplified manner (TURBAN, 1995). This simplification can affect the outcome of the simulation, since the model does not accurately represent realit, the results obtained by it are useless. Besides, it is also necessary to point out other disadvantages, such as: the need to empower the decision maker to operate the software simulation, the time required to model the system, which can take months (DUARTE, 2003), and because the simulation is not an optimizing technique, it only allows the user to test various scenarios (LAW; KELTON, 2000).

Finally, it is necessary to mention as a disadvantage the initial costs involved with the implementation of the simulation in a company, which is often enhanced due to the difficulties of designing the future economic returns that will be provided by the use of this technique of Operations Research (HARRELL; GHOSH; BOWDEN, 2000).

In general, despite the potential drawbacks, there is a growing use of simulation in organizational scope, which is applied in different areas such as logistics and supply chain management, project management, systems manufacturers, among others. In this study, we intend to implement the simulation in a piston manufacturing system, which due to increased demand needs to be expanded. Thus the technique will help the manager to find an alternative that optimizes the financial and operating system returns.

3. SIMULATION OF A PISTON MANUFACTURING SYSTEM

In this work, the simulation is used to represent a piston manufacturing system of a plant located in the state of São Paulo. It should be noted that in recent years the Brazilian economy has shown considerable growth, and the auto industry follows this trend. According to the National Association of Automobile Manufacturers a total of 3,415,486 million vehicles were produced in 2012, a number that must be overcome in 2013, since this year's production, considering the first five months, exceeds the previous year by 18.6% (ANFAVEA, 2013).

This increase in production is directly related to an increase in domestic sales. In 2012, 3,802,071 million new vehicles were licensed in Brazil. This growth, in turn, can be explained both by the income improvement of the Brazilian people, and the ease of getting credit. Over the past decade the dealerships have been offering various facilities in the allocation of loans, which in addition to propelling car sales, also increased demand for accessories such as power steering, electric trio, air conditioning, among others (ANFAVEA, 2013).

Among the accessories with the highest demand is air conditioning. In 2005, 58% of the vehicles sold were equipped with factory air conditioning. It is projected that this figure will increase each year, reaching 74% of vehicles in 2012, approaching the number of European markets, where 75% of cars now have factory air conditioning (COUTTO; PROVATTI, 2005).

The piston is an essential component in the compressor used in the refrigeration circuit of the car’s HVAC system, so that the article will verify alternatives to increase production in order to meet the growing demand for the product.

The evaluated piston manufacturing cell produces 17,000 pistons daily with a 0.534 unit cost. However, about 18,000 inputs are available daily. After fabrication, the finished products are transported to another location within the factory where they are used in the production of hermetic compressors.

The manufacturing system is distributed in a space of ten feet long and seven meters wide. In the layout used, there are two part inputs and one output, the machines are positioned facing each other. Currently, there are seventeen machines with capacities and different production times: five machines of type A, eight type B, one type C, a D-type, plus a washer and quality control.

Table 1: Production times of the piston manufacturing system

Product	Machine	Units	Pçs per cycle	Time of manual labor	Time of operation (Total)
Piston	A	5	8	16	152
	C	1	4	6	86
	D	1	4	6	102
	B	8	1		30

The operation is performed by three workers, which can be divided into three distinct processes. In the first there are the five machines A, the machine C and D machine. In it, a worker takes parts of the two inputs and leads them to another machine. There he will start the setup of the machining of the parts, which are released on sequence.

The second process starts after machining. Another employee takes the initial machined parts and leads them to type B, where another machining process is initiated. In the latter case, a third employee collects the pieces from machine B and leads them to the washer and quality control, where their specifications are checked.

This production process was reproduced in this study using the software Promodel. All features of the real system were respected, such as the distances between machines, the number of employees, the amount of inputs, the lead time of the machines, among others.

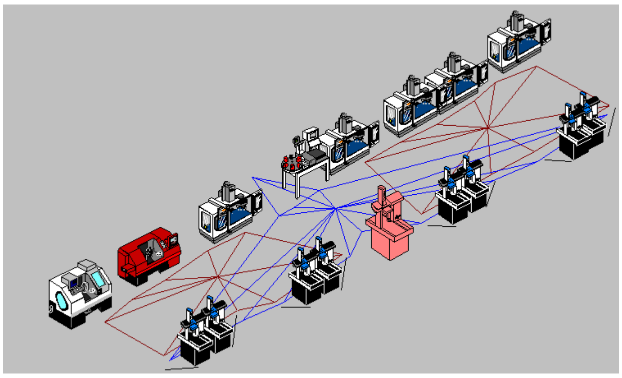


Figure 2: A representation of the piston manufacturing system

The simulation model obtained a production of 17,288 units, a value close to 17,000 pistons manufactured by the actual system. Still from the computational model it was possible to notice some system problems, such as the disparity in the level of use of machinery compared to the others.

It was also observed that employees of manufacturing cell had a major downtime. These two problems, concerning the disparity of machine utilization and idle work, were the points that initially guided the four trials made by the simulation, which involved the following changes: a) increasing the amount and the better distribution of the inputs received, b) increasing the number of employees, and c) purchasing new machinery and increased inputs.

Table 2: Machines utilization of the real system

Machines	% Utilization	% Operation
A	51,51	49,57
A	51,48	49,57
A	51,92	49,57
A	51,23	49,57
A	90,03	86,41
C	66,23	61,27
D	63,20	59,97
B	78,06	78,06
B	78,06	78,06
B	78,06	78,06
B	78,06	78,06
B	77,43	77,43
B	77,27	77,27
B	85,69	85,69
B	70,56	70,56

Table 3: Working and idle time of the workers of the real system

Workers	Number Times Used	% In Use	% Travel To Use	% Idle
1	3372	27,32	10,67	62,01
2	4844	26,51	16,98	56,51
3	8695	24,22	25,34	50,44

3.1. Simulation of Scenario I

The first simulated scenario involved two distinct strategies. Firstly, it was tried to increase the amount of incoming input from 18,000 to 20,000. It was intended to increase the level of use of machinery which, in most cases, was below 80%.

The second strategy involved modifying the number of entry pieces, from two to four. In the real system the company used only two incoming places, creating a great disparity in the level of utilization of the machines of the first process (machine A - machines C and D).

The simulation pondering these two changes had a positive result, the total number of pistons produced could grow to 19,352, an amount that would be totally absorbed by demand. Furthermore, the level of use of machinery significantly increased, exceeding 20% more than the use of other instruments used in the first process. Finally, it is noticed that there was also a considerable drop in the idleness of workers, due to the increased amount produced performing more operations related to transportation of goods and the setup of the machines. The results of this scenario are presented on tables 4 and 5.

Table 4: Machines utilization of the scenario 1

Machines	% Utilization	% Operation
A	82,64	66,15
A	82,09	66,15
A	78,97	66,15
A	78,83	66,15
A	82,75	79,38
C	64,52	59,88
D	61,52	58,67
B	99,79	99,79
B	99,76	99,76
B	99,75	99,75
B	99,71	99,71
B	71,25	71,25
B	71,25	71,25
B	83,75	83,75
B	69,03	69,03

Table 5: Working and idle time of the workers of the scenario 1

Workers	Number Times Used	% In Use	% Travel To Use	% Idle
1	4444	36,21	14,65	49,14
2	4650	25,19	16,34	58,46
3	9732	27,15	28,95	43,91

3.2. Simulation of Scenario II

The second simulation encompassed changes from the previous scenario, so the number of inputs was 20,000 pieces per day, being divided into four entries. Furthermore, it was also experienced an increase in the number of workers.

Despite the level of idleness found in previous simulations, it was intended with this trial to ascertain whether the addition of a worker influences the quantity produced. Thus, it was added another employee in C, since this worker was the one who performed most operations, a total of 9,732.

The results of this experiment were unsatisfactory, the total number of pistons produced was 19,360, only 8 more than the previous scenario. The level of machine utilization was practically not changed and the level of idleness of worker C nearly doubled from 43.91% in the previous scenario to 78.76% in the current one. Thus, it is clear that the number of employees has no influence on the total produced, so the company should consider new alternatives to manufacture more products. The results of this scenario are presented on tables 6 and 7.

Table 6: Machines utilization of the scenario 2

Machines	% Utilization	% Operation
A	82,64	66,15
A	82,09	66,15
A	78,97	66,15
A	78,83	66,15
A	82,75	79,38
C	64,52	59,88
D	61,52	58,67
B	99,79	99,79
B	99,76	99,76
B	99,75	99,75
B	99,71	99,71
B	71,25	71,25
B	71,25	71,25
B	83,75	83,75
B	69,03	69,03

Table 7: Working and idle time of the workers of the scenario 2

Workers	Number Times Used	% In Use	% Travel To Use	% Idle
1	4444	36,21	14,65	49,14
2	4650	25,19	16,34	58,46
3A	4870	13,60	14,50	78,73
3B	4862	13,55	14,45	78,76
3	9732	27,15	28,95	78,76

3.3. Simulation of Scenario III

The third and final simulation involved the purchase of new machinery. Analyzing the results of the previous experiments it was noted that the level of use of certain machines B reached almost 100%.

Accordingly, in this last experiment, two more machines B were added. Despite the investments required, it was believed that the use of new machinery would increase considerably the total of products. In this sense, the amount of inputs received daily was also increased to 22,000 units.

The simulation showed that with these changes the total production of pistons would increase to 20,928 units. However, it is clear that there is a significant drop in the level of machine B use. The results of this scenario are presented on tables 8 and 9.

Table 8: Machines utilization of the scenario 3

Machines	% Utilization	% Operation
A	75,22	72,85
A	75,26	72,67
A	91,02	67,38
A	90,40	67,35
A	89,66	86,08
C	71,35	65,24
D	67,99	65,28
B	57,36	57,36
B	57,36	57,36
B	57,36	57,36
B	57,22	57,22
B	99,75	99,75
B	99,71	99,71
B	77,15	77,15
B	76,96	76,96
B	91,25	91,25
B	76,81	76,81

Table 9: Working and idle time of the workers of the scenario 3

Workers	Number Times Used	% In Use	% Travel To Use	% Idle
1	4725	38,43	16,60	44,97
2	5096	27,54	17,88	54,59
3	10525	29,57	31,37	39,07

4. DISCUSSION OF RESULTS AND FINAL CONSIDERATIONS

The aim of this study was to demonstrate that the use of simulation as a technique of Operational Research can assist the process of decision-making by businesses, which are increasingly complex due to uncertainties, risks, conflicts and lack of infrastructure decisions.

To achieve this goal we simulated a piston manufacturing cell, which represented an auto parts factory located in the state of São Paulo. Pistons are components used in air-conditioned cars, having a steady increase in demand. In this sense, the goal of the simulations was to test alternatives to increase the daily output of the production system that was 17,000 units.

The results in all scenarios were satisfactory with respect to the main objective of the trial. In all scenarios, the daily production was increased by at least 2,000 units.

However, strategies have simulated different investments, so comparing them is necessary to determine both the investment required and the unit cost of the pistons. For this, it is estimated that the unit cost is $0.534, the hourly cost of the cell is $291.98, the cost of direct labor per worker is $8.18 and the cost of labor indirectly, with the share of overhead, is $ 40.91. Table 10 presents the summary of the costs.

Table 10: Costs of the real system

Unit Cost	0.534
Cost/hour	291.88
Direct labor costs / worker	8.18
Indirect labor cost + General expenses	40.91

In the first alternative, modifications involved only a change in the input quantity and inputs received, i.e., no considerable investment. It was obtained production of 19,352 units, thus the product unit cost was $ 0.4430.

In the second scenario, the modifications involved changes experienced in the first stage, and the addition of an employee. The total production of 19,360 was obtained, so this unit cost was $ 0.4530.

Finally, in the latter scenario, the number of inputs was increased to 22,000 units. In addition, we simulated the purchase of two machines B, which required an investment of $ 200,000. The unit cost was 0.423, and the total production quantity reached 20,928 units. Table 11 presents the summary of scenarios.

Table 11: Summary of scenarios outputs

Simulation	Production	Workers	Input	Cost/piston
Scenario I	19352	3	20000	0,4430
Scenario II	19360	4	20000	0,4530
Scenario III	20928	3	22000	0,4230

It can be seen then that there are two viable options to increase daily production of pistons; Alternative 3 has a lower unit cost than the others. However, it requires an initial investment of $ 200,000. On the other hand, strategy 1 requires no investment. However, it has a unit cost of $ 0.4430.

Thus, the company should decide on the best way to face the situation, since, as mentioned earlier, the simulation is not an optimizing technique.

In general, by analyzing the results obtained by simulations and the concepts presented, it is clear that this technique can aid the decision-making process of a company making it less complex, since it allows experimentation strategies that could solve the situation faced.

REFERENCES

Anfavea. Associação Nacional Dos Fabricantes De Veículos Automotores. Available On: <Http://Www.Anfavea.Com.Br/>. Access On: 14 Apr 2013.

Andrade, E. L. Introdução A Pesquisa Operacional: Métodos E Modelos Para Análise De Decisão. 3. Ed. Rio De Janeiro: Ltc, 2004.

Barton, R. R. Designing Simulation Experiments. Proceedings Of The Winter Simulation Conference, Usa, 2004.
Bateman, Thomas; Snell, Scott. Management: Leading & Collaborating In The Competitive World. 8.Ed. Mcgraw-Hill, 2008.
Couto, Marcelo; Provatti, Roberta. O Contínuo Crescimento Do Ar Condicionado Automotivo. Abrava Informação & Negócios. Ed. 231, Novembro/2005.
Duarte, R.N. Simulação Computacional: Análise De Uma Célula De Manufatura Em Lotes Do Setor De Auto-Peças. 2003. 168.F. Dissertação (Mestrado). Universidade Federal De Itajubá.
Harrel, C; Ghosh, B.; Bowden, R. Simulation Using Promodel. Mcgraw-Hill, 2000.
Hillier, F. S.; Lieberman, G. J. Introduction To Operations Research. 8.Ed. Mcgraw-Hill, 2005.
Lachtermacher, Gerson. Pesquisa Operacional Na Tomada De Decisões. 3. Ed. Rio De Janeiro: Editora Campos Ltda., 2002.
Law, A. M.; Kelton, W. D. Simulation Modeling And Analysis. Mcgraw-Hill, 2000.
Lima, M; Barbosa, G.; Beal, C. Otimização De Transferência E Estocagem Utilizando Ferramentas De Modelagem E Simulação. 2003
Ragsdale, Cliff T. Spreadsheet Modeling And Decision Analysis: A Practical Introduction To Management Science. 4. Ed.. Cincinnati: South Western College Publishing, 2004.
Sargent, R. G. Validation And Verification Of Simulation Models. Proceedings Of The 2004 Winter Simulation Conference. R .G. Ingalls, M. D. Rossetti, J. S. Smith, And B. A. Peters, 2004.
Shimizu, T. Decisões Nas Organizações: Introdução Aos Problemas De Decisão Encontrados Nas Organizações E Nos Sistemas De Apoio À Decisão. São Paulo: Atlas, 2001.
Simon, H. A. Administrative Behavior: A Study Of Decision-Making Process In Administrative Organizations. 4. Ed. New York: Free Press, 1997.
Slack, N. Et Al. Operations And Process Management: Principles And Practice For Strategic Impact. Prentice Hall, 2006.
Turban, E. Decision Support And Expert Systems: Management Support Systems. 4. Ed. Englewood Cliffs, N.J: Prentice-Hall, 1995.
Vaccaro, G. L. R. Modelagem E Análise De Dados Em Simulação. Porto Alegre: Instituto De. Informática/Ufrgs, 1999.
Vieira, G. E. Uma Revisão Sobre A Aplicação De Simulação Computacional Em Processos Industriais. Xiii Simpep, 2006.

SIMULATING AN INTERNET PRODUCT DELIVERY SUPPLY CHAIN WITH MULTI-ITEM ORDERS

Oliverio Cruz-Mejia [(a)], Richard Eglese [(b)]

[(a)] Faculty of Engineering, Universidad Anáhuac Mexico Norte
Av. Universidad Anáhuac 46, Huixquilucan, Estado de México, 52786, MEXICO
[(b)]Department of Management Science, Lancaster University, Lancaster, LA1 4YX UNITED KINGDOM

[(a)] oliverio.cruz@anahuac.mx, [(b)] r.eglese@lancaster.ac.uk

ABSTRACT
The increase of e-commerce has developed new challenges for online retailing companies delivering product orders to customers. With this challenge, a new type of supply chains have been developed aiming to cope with a strict control of lead time and associated costs. In this paper a model of an internet product delivery supply chain with multi-item orders is simulated. We address specifically the mismatch between supply and demand when retailers for any reason are unable to estimate the configuration of multi-item orders or single item orders. Three scenarios of demand configuration are simulated (demand as expected, lower than expected and higher than expected) using discrete-event simulation to look at the effect on lead time. A detailed numerical analysis is used to draw conclusions.

Keywords: internet retailing, product delivery, merge-in-transit, discrete event simulation

1. INTRODUCTION

Distribution channels in many industries have experienced major changes in recent years in terms of their structure, collaborative partnership, operational practices and performance requirements. These channel transformations arise from several factors that have altered the rules for providing competitive delivery services. Changes in product delivery management can be attributed to three main driving forces:

Customers are raising their service expectations. Customer demands for quick response and customized products are propagating along supply networks. Changes in life style of people require manufacturers and service providers to adjust to the new circumstances.

Information technologies are providing more timely and detailed supply chain data. Advances in information technologies in both connectivity and reach increase the potential for information sharing and enable tighter integration among supply chain partners.

Partnerships with logistics service providers allow manufacturers to focus on their core competences while taking advantage of the distribution efficiency and expertise of dedicated distributors. In turn, distributors are offering their services beyond the traditional warehousing and transportation functions to include value-added activities e.g., repackaging, labeling, light assembly, and non-inventory distribution services of which cross-docking and merge-in-transit distribution are examples.

Merge-in-transit distribution (MiT) is a logistics process introduced in practice in the late 1990s.

Merge-in-transit is defined as a distribution process that brings together at a consolidation centre multi-product order components, coming from different origins, consolidates them into a single order, and then ships it for final delivery to the end customers.

Some of the advantages obtained with MiT are:

Higher customer satisfaction is obtained by delivering multi-product orders in one event instead of making more than one delivery, one for each component or partial group of them.

Savings are achieved by not keeping inventories in the distribution process; since merge-in-transit centers just hold order components for a short time (usually less than 24 hours) so the order is all the way in transit to its final delivery point. Holding costs associated with warehousing operations are avoided or at least minimized.

Savings also arise by avoiding the risk of keeping obsolete inventories. MiT is normally applied to distribute orders where sometimes one component has been made-to-order. Those tailored components have been made for a specific need and are never kept in stock so there is no risk of keeping obsolete components (Ala-Risku, Karkkaainen and Holmstrom 2003).

2. SUPPPLY CHIAN DESCRIPTION

In this section it is described a prototypical supply chain that will represent a generalization of normal operation of MiT supply chains. It is considered a customer that is online at home or office and makes the selection of items that he wants to buy in the same transaction. The information is sent to the retailer and the retailer sends the multi-item purchase order to the order consolidation center. The order consolidation center collects the items needed and a single multi-item package is assembled for the specific customer order. It may happen that some items required are not in stock because they are in transit to the consolidation center. Having products out of stock obviously causes delay in the delivery process.

A graphic explanation of the supply chain can be seen in Figure 1.

The stock of items at the consolidation center is replenished by a continuous review policy. The following logic is applied: if the stock level at the consolidation center goes below the reorder point, then place an order that replenishes the stock at the consolidation center. The replenishment shipments have an implicit transportation time. Finally, when all the items required for a multi-item order are available, a single shipment is transported and delivered at the customer location.

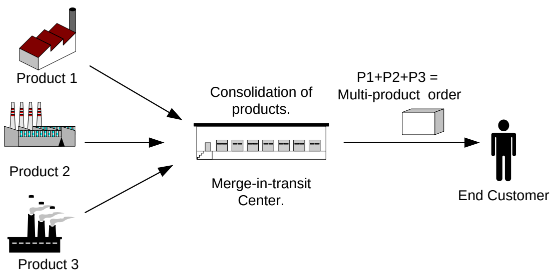


Figure 1: Supply Chain Model

3. CONCEPTUAL MODEL

The supply chain described in section 3 can be translated in a conceptual model (Pidd 1998) to approach the construction of the simulation model required for the analysis. Figure 2 shows a conceptual model with four basic operations:

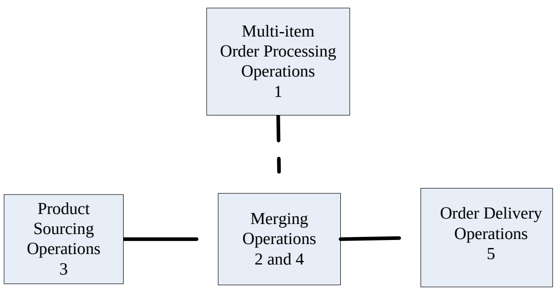


Figure 2: Conceptual model of the MiT supply chain

3.1. Multi-item Order Processing Operations

Multi-item order processing operations include the activities involved in the communication and initial information processing of the order placed by the customer. Communication of the order placed by the customer means how the MiT Supply Chain gets the preferences of the customer into their system. It can be a website portal in which the customer selects from a catalogue of products. Web internet ordering systems have the advantage of communicating the information instantaneously.

Initial information processing means the initial classification of information that MiT Supply Chains may do to batch orders with similarities that represent an advantage to the system. It also includes setting priorities to special orders or customers as another example. Three main issues can be found to be relevant in Multi-item Order Processing Operations: first is the order size. How many items are included in the order? Large order sizes may have implicit long consolidation times due to the number of items required to have available to have a complete order. Second is the mix of products Assemble-to-Order (ATO) and Make-to-Stock (MTS). Coordination of items with different natures (ATO versus MTS) represents a challenge as usually they have different holding costs and processing times. ATO items are assembled just as an order has been placed, so its assembly time and transportation is critical to have lead time under control. Making right decisions in the sourcing of items needed in order to have the required mix of items can be a challenge with implications on performance and cost. Third are the order processing decisions.

3.2. Product Sourcing Operations

Product sourcing operations include the activities required to have available at the consolidation point the items to be merged to integrate a multi-item order. In the case of MiT Supply Chains that merge items ATO and MTS, the Sourcing Policy decisions and the Transportation are relevant decisions. Sourcing Policy means what type of Inventory Management logic will be used to bring in the required items to the consolidation centre. It cannot be said that a MiT Supply Chain holds inventory as this would be contradictory to its principle but it certainly can hold items that are waiting to be merged. The more desynchronized is the merge of multi-item orders, the more the process turns from merging operations to holding inventory operations. Sourcing Policies can operate under periodic review logic where time is the variable that triggers the transportation process or it can be reorder point systems where the level of in hand items triggers the transportation of items. Transportation is another significant decision in the Product Sourcing Operation. Distance of Transportation and Speed of the transportation media can be important factors in the costing and service levels reached in the operation. In the case of ATO items, Postponement can be a relevant strategic decision to be included. If the assembly time of ATO items is one of the critical elements in the whole order lead time, then it can be understood that if is possible to pre-assemble some parts of the ATO items before the order is placed into the system, then some time can be saved to cut lead times.

3.3. Merging Operations

The orders taken online are transferred in the sequence they arrive to the point of consolidation of items. There is no order batching. Based on the items' availability, customer orders are scheduled for consolidation following a first-in-first-out rule. The point of consolidation maintains a minimum level of stock and every time the stock level goes below the level a replenishment order is scheduled following a fixed replenishment point/fixed replenishment quantity inventory policy (Q, R). The replenishment batch size is fixed. Customer orders that require items out of stock at a given time are put away temporarily and as soon as the item required is available in the consolidation point, the waiting order is scheduled to be processed. The consolidation process is modelled as a quick operation, very similar to the sortation in material handling systems. Once an order is consolidated, a delivery truck delivers a batch of orders. The time taken for the delivery truck to complete its deliveries may vary according to a given probability distribution. The delivery truck is dispatched regularly regardless of the number of orders to be loaded but having a maximum capacity. The base model includes the consolidation of up to 3 items, each of them being sourced from different origins. Figure 3 shows the flowchart that describes the logic for the model.

3.4. Order Delivery Operations

Order delivery operations include the transportation of the MiT order once it leaves the merge centre. Order Delivery Operations is what Boyer, Frohlich and Hult 2005; Bowersox, Closs and Cooper, 2007) discuss in the extended supply chain concept. From the operations perspective, Order Delivery includes two typical decisions to consider. One is the consideration of Time Windows on the final delivery trip. It can happen that delivery staff arrives to drop off the order and customers are just unable to receive it. It can be also the case that companies are offering a higher service level to customers by allowing customers to set a day and time range for their delivery operations. A second issue to consider is the design of an Optimal Delivery Network. This means looking at finding the best delivery network for the MiT supply chain that as example can set as its objective to minimize the cost of operations given a set of constraints. Alternatively, the objective may be to minimise the delivery time given a set of constraints.

4. PROBLEM DESCRIPTION

The supply chain model utilized is able to deliver up to three items for the same customer order and the supply chain structure and operating principles remain the same.

For this case we use three demand scenarios, each of them having different proportions of dependent orders. The three scenarios are: a) Demand with a high proportion of dependent orders, this type of demand will be called higher than expected, b) Demand with a medium proportion of dependent orders, we will call this type of demand as expected, c) Demand with a low proportion of dependent orders and this will be called lower than expected.

The key concept to explore in this simulation scenario is whether the supply chain operation is set to supply and cope with the delivery of orders to customers with an expected level of demand per individual item, which may be different to the demand actually received. Based on the fact that orders are not single-item and that the multi-item orders depend on customer choice, the real demand per item may be different from what was expected and consequently the capability of the supply chain to fulfil the delivery orders on time may be affected. By order configuration based on customer choice we mean the group of items requested per multi item order.

Three order types will be defined: type A represents an order for item 3 only, type B represents an order for items 2 and 3 and type C represents an order for items 1, 2 and 3. We assume a situation where we expect an equal proportion of orders from customers of types A, B and C.

In this paper, we simulate three customer demand configurations, each one representing levels of demand: higher than expected, as expected, and lower than expected. In the higher than expected configuration, we assume that a higher proportion of orders are of type C and a lower proportion are for type A. In the lower than expected configuration, we assume that a lower proportion of orders are of type C and a higher proportion are for type A. In this condition it is expected that the system will operate with excess of capacity.

The objective of this three scenario experiment is to quantify the implications in the delivery supply chain when the demand for some items is significantly more or less than expected, due to different proportions of orders combining orders for different items.

The values used to generate each of the demand scenarios are in tables 1,2 and 3:

Table 1: a) Demand higher than expected

Order Type	Item 1	Item 2	Item 3	% of orders per type
A	0	0	1	15.00%
B	0	1	1	30.00%
C	1	1	1	55.00 %

Table 2: b) Demand as expected

Order Type	Item 1	Item 2	Item 3	% of orders per type
A	0	0	1	33.33%
B	0	1	1	33.33%
C	1	1	1	33.33 %

Table 3: c) Demand lower than expected

Order Type	Item 1	Item 2	Item 3	% of orders per type
A	0	0	1	55.00%
B	0	1	1	30.00%
C	1	1	1	15.00 %

5. DATA AND INPUT PARAMETERS

The simulation model was run for 50 replications of 3 months of continuous operation, each 2196 hr. The conditions of the model were: 50 % of stock out risk, inbound and outbound transportation times were modelled following a Normal distribution with means of 24 hrs and 72 hrs respectively. Standard deviations for the transportation times were 2.4 hr and 7.2 respectively. Table 4 summarize data and input parameters used for the simulation runs.

Table 4: Input parameters, control variables and experimental variables.

Section of the model	Variable name	Variable value
Order taking	Order inter arrival time	Exp (0.15)
	Maximum number of different type of products	3
	Order configuration (independent & dependent)	2
Inventory	Excess of Supply factor	1.25
	Reorder policy	(R, Q)
	Induced stock-out probability	30%
	Inbound transportation time	N(24, 2.4)
	Inbound transportation vehicle size	2MLTD
Consolidation	Consolidation operation time	0.005/item
Outbound transportation	Outbound transportation time	N(72, 7.2)

6. METHODOLOGY

Discrete event simulation (DES) was used as the modeling methodology for the analysis of the supply chains under study. DES is a well-established technique for the study of operational scenarios in real world situations (Banks, Carson, Nelsona and Nicol 1995; Pidd, 1998; Law and Kelton 2000). DES is a suitable analysis tool for the research objectives set for this thesis because of the following advantages:

- DES allows a high level of detail to be modeled for the operating scenarios under study while mathematic analytic models would only allow the simplified representations of real world scenarios (low level of detail).
- Using DES can easily model alternative scenarios of operation (experimentation) of the supply chains under study and allow practical conclusions to be drawn.
- One of the main research aims in this thesis is the study of the composition of customers' multi-item orders when buying using the Internet. This order composition is a behavioral element that can be nicely modeled and experimented upon with DES. Supply chain problems involving behavioral issues use predominantly simulation over analytical methods as the primary research tool, since the complexity of human interaction with complex systems precludes analytical methods for examining customer election issues (McCreery, Krajewski, Leong, G. K. and Ward, P. T. 2004 (McCreey et al, 2004).
- DES is a methodology that allows the dynamic analysis of operations. As the name suggests, a simulation run is a sequence of events for which the model can be can be stopped at any event in the run. This feature is very useful in the verification of the model, as it allows to carefully checking that the operating conditions in the computer model are accurately executed as in the conceptual model.

RESULT AND ANALYSIS

Table 5 shows the results for the three scenarios of demand being higher than expected, as expected and lower than expected. The table includes the segregated values for orders delayed and non-delayed as well as all the orders

Table 5: Input parameters, control variables and experimental variables.

	Demand Pattern Condition		
	Higher	As expected	Lower
% orders of type A, B and C	15,30,55	33,33,33	55,30,15
Avg. Time in System (1)	279.00	110.41	110.17
Minimum Time in System (1)	66.99	66.06	70.57
Max. Time in System (1)	1660.09	157.97	152.12
St Dev of (1)	214.64	21.94	21.01
CV (1)	0.77	0.20	0.19
Items Entered (delayed)	6727.66	804.14	390.26
Avg. Time in System (2)	108.81	108.53	108.52
Min. Time in System (2)	57.67	57.36	57.36
Max. Time in System (2)	164.07	164.38	164.39
St Dev of (2)	22.59	22.59	22.59
CV	0.21	0.21	0.21

The average time in system for the case of orders with demand higher than expected is almost double (195.50) that for the as expected (108.63) and lower than expected (108.56) cases. The reason for this is the higher number of out of stocks registered under the higher than expected demand. The high coefficient of variation (CV) of time in system for orders with higher demand than expected (0.90) compared to the ones from as expected (0.21) and lower than expected (0.21) confirms the higher variability in delivery time of the system with higher demand than expected caused by the lack of stock. As orders with demand as expected or lower than expected have the same CV, we can argue that they expect the same degree of variation in the delivery time. The percentage of orders delayed for higher than expected, as expected and lower than expected demand is 52.8%, 5.5%, 2.7% respectively.

More than half of the orders are delayed when demand is higher than expected. This value is ten times higher than when demand is as expected and almost twenty times higher than when demand is lower than expected. So this proportion can be very high and implies a potential high impact on the delivery time of the order and customer satisfaction. Again, the high proportion of orders delayed when demand is higher than expected is related to the lack of items to fulfill orders.

The average difference in lead time between delayed and non-delayed orders for the different scenarios show that when demand is higher than expected, the orders (170.19 hr) have around a 100 % longer delays compared to when demand is as expected (1.88 hr) and lower than expected (1.64 hr). This means that delays registered when demand is higher than expected are hundred percent longer on average than for the other types of demand.

The percentage of difference in lead time between delayed and non-delayed orders for runs with the same demand values show that when the demand is higher than expected a delayed order takes 156.4% more time to be fulfilled than a corresponding order that did not register delay. In the case when demand is as expected, the difference between delayed and non-delayed is only 1.7 % more time, while when the demand is lower than expected, the delay is 1.5% time in excess.

Table 5: Cont. input parameters, control variables and experimental variables

	Demand Pattern Condition		
	Higher	As expected	Lower
% orders of type A, B and C	15,30,55	33,33,33	55,30,15
Items Entered (non-delayed)	6019.84	13847.4	14260.64
Avg. Time in System	195.5	108.63	108.56
Min. Time in System	57.67	57.36	57.36
Max. Time in System	1660.09	164.68	164.43
St.Dev(1+2)	175.95	22.58	22.57
CV	0.9	0.21	0.21
Number Completed (all)	12763.8	14552.9	14553.4
% of orders delayed	52.8%	5.5%	2.7%
Ave. Diff. in LT(Delayed vs. Non-delayed)	170.19	1.88	1.64
% of Diff in LT(Delayed vs. Non-delayed)	156.4%	1.7%	1.5%

For orders non-delayed the lead time in systems is the same for the three types of demand (108.53, 108.81, and 108.52). This data confirms the correct operation of the simulation model including that the model is not blocking at the merge operation. The standard deviations for delayed orders are 214.64, 21.94, 21.01 for high, medium and low mix respectively. We can understand from this that orders delayed in when demand is higher than expected have a 10 times larger standard deviation in the delivery time.

7. CONCLUSIONS

The higher value on average time in system for the demand pattern defined as higher than expected is not counterintuitive. However, it can help managers to quantify the delay on the delivery systems with consolidation if there is mismatch between supply and demand on at least one item of the order. Figures for the scenario evaluated show that the delivery time can go from 4.58 to 11.62 days. The proportion is considerable and having a delivery system with delays double than base delivery time can have considerable implication on customer satisfaction. The system can have a better performance if the stock-out probability or reorder policy for items is adjusted for products likely to be ordered in a consolidated order. The average difference in lead time between orders delayed and not-delayed with figures of (170.19, 1.88 and 1.64) can provide mangers an idea on how “ long or deep” is the delay of an order. This figure is also linked to the inbound transportation time. The smaller the inbound transportation time is the smaller the impact on the delivery of items consolidated will be. An alternative for managers is sourcing from suppliers that offer shorter lead time.

The systems evaluated in this paper help to profile the expected behavior of the logistic system with implications with customer satisfaction and indirectly the cost of running the logistics system if improvements on the system want to be realized.

ACKNOWLEDGMENTS

This research was funded by the National Council of Science and Technology of Mexico (CONACYT) under the grant No. 117263. We also thank SIMUL8 Corporation for providing a license of their software Simul8 Professional.

REFERENCES

Ala-Risku, T, Karkkaainen, M, Holmstrom, J, 2003. Evaluating the Applicability of Merge-in-Transit: A Step by Step Process for Supply Chain Managers. *International Journal of Logistics Management* 14(2): 67-81.

Banks, J, Carson, J, S, Nelson, B, L, Nicol, D, M, 1995. *Discrete-Event System Simulation*. New Jersey: Prentice Hall.

Bowersox, D, J, Closs, D, J, Cooper, M,B, 2007. *Supply Chain Logistics Management*. New York: McGraw-Hill.

Boyer K, K, Frohlich. T M, Hult M, T, 2005. *Extending the Supply Chain: How Cutting-Edge Companies Bridge the Critical Last Mile into Customers' Homes*. New York, USA: American Management Association.

Cole, M, H, Parthasarathy, M, 1998. *Optimal Design of Merge-in-Transit Distribution Networks*. Research Report.University of Arkansas,USA.

Croxton, K, L, Gendron, B, Magnanti, T, L, 2003. Models and Methods for Merge in Transit Operations. *Transportation Science* 37(1):1-22.

Hoffman, K, C, 1998. *Luncent Revamps Order Deliveries with help of Skyway Freight*. Available from:<http://www.glscs.com/archives/casestudies.htm> [Accessed 3 January 2008].

Karkkainen, M, Ala-Risku, T, Holmstrom, J, 2003. Increasing Customer Value and Decreasing Distribution Costs with Merge-in-Transit. *International Journal of Physical Distribution and Logistics Management* 33(2):132-148.

Kopczak, L, R, 1995. *Logistics Partnership and Supply Chain Restructuring*. Ph D Thesis. Stanford University.

Law, A, M,, Kelton, W, D, 2000. *Simulation Modeling and Analysis*. New York City: McGraw-Hill.

McCreery, J, K, Krajewski, L, J, Leong, G, K, Ward, P, T, 2004. Performance Applications of Assembly Work Teams. *Journal of Operations Management* 22(4):387-412.

O'Leary, D, E, 2000. Reengineering Assembly, Warehousing and Billing Processes, *Information Systems Frontiers* 1:379-387.

Pidd, M, 1998. *Computer Simulation in Management Science*. West Sussex: Wiley.

AUTHORS BIOGRAPHY

OLIVERIO CRUZ-MEJIA is an Associate Professor of Industrial Engineering at Universidad Anáhuac, Mexico. He obtained a Ph D. in Management Science from the Lancaster University Management School. He holds an M. S. in Manufacturing Engineering from Syracuse University, USA and a B. S. in Mechanical Engineering from the National Polytechnic Institute, Mexico. He worked previously in the automobile industry for San Luis Rassini, Federal Mogul and Kostal GmbH in areas of product design, production control and project management. He obtained a specialization in Oil Hydraulics from the Japan Cooperation Agency, Kitakyushu, Japan. He serves as board review member for the Journal of Supply Chain Management (JSCM) and as ad-hoc reviewer for the Journal of Operations Management (JOM). His research interest is in the area of discrete-event simulation applied for internet operations.

RICHARD EGLESE is a Full Professor in the Department of Management Science at Lancaster University Management School. He has published widely in the areas of Vehicle Routing and Scheduling in journals such as the European Journal of Operational Research and the Journal of the Operational Research Society. He has been a program committee member for the International Federation of Operational Research Societies (IFORS) Conference and President of the Operational Research Society (OR Society).

A STUDY OF DEVS-BASED PROCESS SCHEDULING ON MULTI-USER SEMICONDUCTOR TEST EQUIPMENT

Soonchul Lim[(a),(b)], Youngsin Han[(a)], Chilgee Lee[(a)]

[(a)] College of Information and Communication Engineering, Sungkyunkwan University
[(b)] Development and Evaluation Group, Memory Business, Samsung Electronics

[(a)]scandth1@naver.com, yshan@skku.edu, cslee@skku.edu

ABSTRACT

Single processor semiconductor test equipment inevitably experiences idle time between tests. This idle time is increased when multiple operators use the same equipment. An increase in idle time is considered a loss factor and produces low equipment efficiency; therefore, it is important to decrease it. In this paper, we offer two methods to effectively decrease idle time in multi-user test equipment. The methods proposed here were developed using an atomic model and were coupled with discrete event system specification methodology. Features of our model include idle time that can be decreased more than the typical sequence and equipment status that can be monitored beforehand without adding extra time.

Keywords: equipment efficiency, idle time, process scheduling, DEVS

1. INTRODUCTION

The market for semiconductor memory devices is rapidly changing. The demand for desktop PC memory is stable or slightly decreasing, but the demand for memory for mobile devices is rapidly increasing. These changes require an increased investment and reduce the life of certain equipment. The effective use of limited resources is essential in the semiconductor industry, which already requires heavy investment.

Equipment efficiency can be expressed numerically and most manufacturing processes apply various methods to increase equipment efficiency. Semiconductor test equipment is used to screen devices and evaluate their characteristics. Test equipment efficiency needs to be improved by optimizing test items, decreasing test times, and reducing unnecessary idle time. Overall equipment efficiency (OEE) is one of the methods used to illustrate how effectively equipment and resources are utilized. The overall performance of a piece of equipment or a factory is always governed by the cumulative impact of three OEE factors: availability, performance rate, and quality rate (A.J. De, Ron, 2006). The OEE is defined as:

$$\%OEE = (\%Availability) \times (\%Performance) \times (\%Quality) \quad (1)$$

Our proposal relates to the availability of the OEE factors.

2. BACKGROUND AND METHODOLOGY

2.1. Background

The test equipment process sequence applied by multiple users is expressed as shown in Figure 1. User *A*, occupying the test processor, needs idle time (*I1*) to set up the test environment. After the test program is executed, idle time (*I2*) is generated from the end of the program, or by user interruption, before starting the next program. After user *A* completes the procedure on the test processor, idle time (*I3*) is generated until user *B* occupies the test processor. The process sequence hereafter is identical to the previously described sequence of user *A*. Thus, idle time (*Itp*) between users occupying the test processor and idle time (*Is*) between the test programs, are generated. The total idle time is defined by the formula below:

$$I_{total} = \sum I_{tp} + \sum I_s \quad (2)$$

The total idle time is one of the loss items in an availability factor; the higher the total idle time, the lower the system's efficiency. Process scheduling is a method used to decrease the waiting time and a common representation of process scheduling is a queuing system (Silberschatz and Galvin, 1994; Hopp, 2008). We propose the queuing system using discrete event system specification (DEVS) methodology to minimize the total idle time on multi-user equipment.

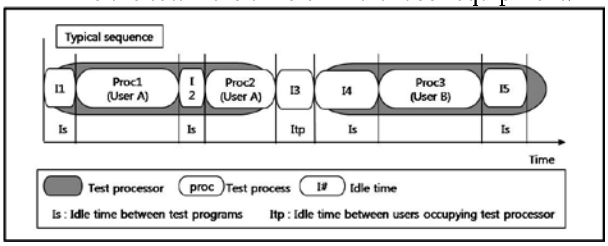


Figure 1: Typical sequence for multi-user equipment

2.2. Methodology

The DEVS is a modeling methodology used in a system that operates discrete events in linear time (Hill, 1996). This methodology is based on a set theory to extract the structure and behaviors of the model. It is easily expressed in hierarchical and modular systems. To provide these features, a DEVS has an atomic model and a coupled model. An atomic model (*M*) is organized to express the dynamic characteristic, as follows:

$$M = < X, Y, S, \delta_{ext}, \delta_{int}, \lambda, ta >$$

The coupled model (*CM*) expresses the interaction between the components of the system and the hierarchical structure of the system, as follows:

$$CM = < X, Y, M, EIC, EOC, IC, SELECT >$$

Detailed descriptions and modeling methods of DEVS can be found in Zeigler, Praehofer and Kim (2000), Kim (2007), and Han and Song (2012).

3. SYSTEM MODELING

We propose a queuing system that consists of the InQueue model, the processor model, and the InQueuehandler model using the DEVS to produce a system with minimal idle time.

3.1. InQueue model

A state diagram of the InQueue model is shown in Figure 2. There are three input messages and two output messages. The initial state of the InQueue model is the WaitForPgm state with a time of infinity. The InQueue model accumulates the program in a queue if receiving a PgmRaw message. The model produces PgmRaw and Qval messages if receiving ReqPgmRaw and ReqQval messages, respectively.

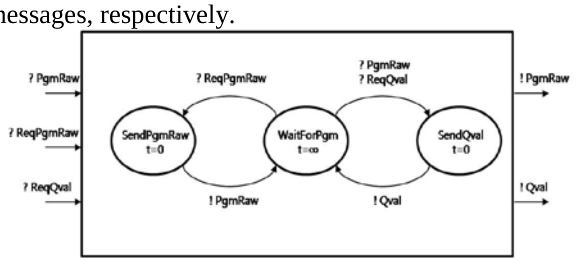


Figure 2: State diagram for the InQueue model

3.2. Processor model

The processor model's behavior is depicted in Figure 3. The processor model has two states, with an input message and an output message. The initial state of the processor model is the WaitForPgm with infinity time. If receiving a PgmRaw message, the state of the model is changed to BUSY which has a random *ta*. After executing the PgmRaw, the model generates a TestDat message and returns to its initial state.

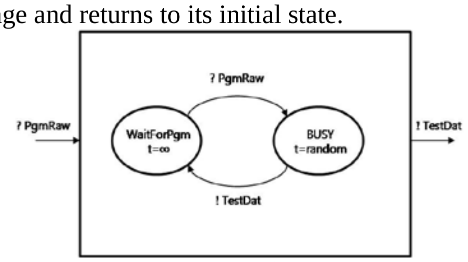


Figure 3: State diagram for the processor model

3.3. InQueuehandler model

A state diagram for the InQueuehandler model is provided in Figure 4. In the figure, there are three input messages and four output messages. The model has a WaitForQval state with infinity time at the initial state. If Qval is not 0, the model moves to the CheckTP state. If the test processor state is free, the model goes to the ReqForPgmRaw state. If it is not free, the model goes to the ReqForPreemptiveTP state and waits until the specified time. A ReqForPgmRaw message is produced and the state of the model is changed to WaitForPgmRaw. When receiving a GetPgmRaw message, the model goes to the SendPgmRaw state and immediately produces a SendPgmRaw message. Subsequently, the model goes to WaitForProcessDone with infinity time. After receiving the GetDoneSig message, the state of the model is changed to SendDoneSig state and the model generates a SendPgmRaw message. The CheckInQ state produces a Qval message and the model returns to the WaitForQval state.

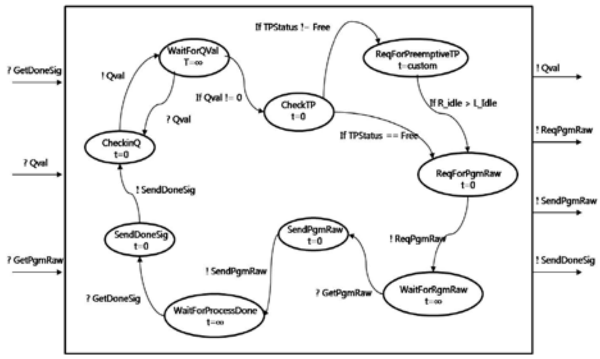


Figure 4: State diagram for the InQueuehandler model

3.4. Coupled model

Figure 5 represents the overall coupled model of the queuing system to apply idle time to the effective process. This coupled model consists of three atomic models and combines the InQueue, processor, and InQueuehandler models. In the InQueuehandler, the WaitForPgmRaw state changes to the SendPgmRaw state when receiving PgmRaw from the inGetPgmRaw port. PgmRaw is transferred from the InQueuehandler output port to the processor model and the Qval is changed. The inputted PgmRaw is executed by the defined environment process and the model produces TestDat messages for the user that initially provided the PgmRaw to the InQueue model.

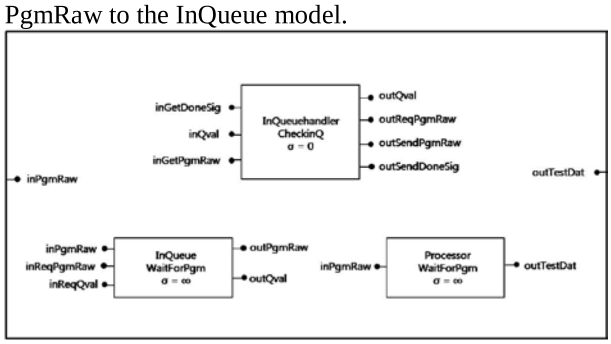


Figure 5: Overall view of the coupled model

4. RESULTS

4.1. Queuing system using the DEVS

Our queuing system using the DEVS is shown in Figure 6. This method can be executed without the loss of time. The upper sequence in Figure 6 is a typical user sequence (*A*, *B*, *C*); it is necessary for each user to occupy the test processor. In Figure 1, as described above, idle time *(I1 ~ I4*) occurs due to latency times (*Itp* and *Is*). The queuing system using the DEVS is executed with non-preemptive processes, as in the lower sequence in Figure 6. This method has features for minimizing *Itp* and *Is*; however, it has an assumption that requires the same hardware infrastructure for proc1, proc2, and proc3. In the experimental results, as shown in Figure 7, this method works more efficiently for multi-user equipment than the previous process, but the data for the single-user equipment does not seem to be effective.

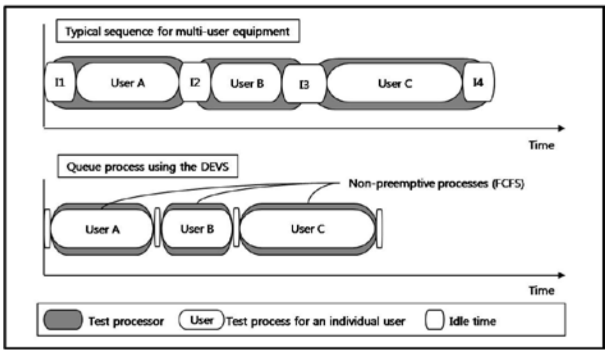


Figure 6: Queue process using the DEVS

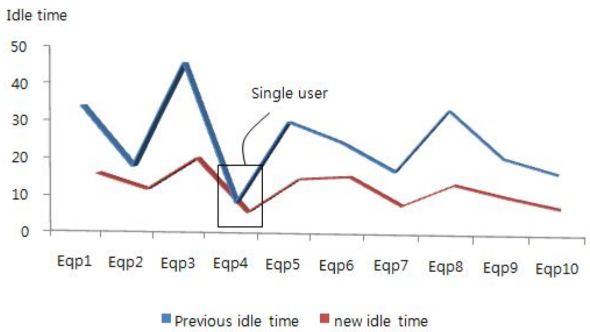


Figure 7: Comparison of the idle time

4.2. Self-diagnostic process using the DEVS

The upper image in Figure 8 shows a typical sequence. Idle time occurs between the test programs. The length of idle time is different according to the type of work the user is performing. The lower image in Figure 8 shows our method with a self-diagnostic process, applying a preemptive process to the user's idle time. If the idle time (*I3*) exceeds the limited time defined by the system user, the user's test processor is terminated and changed to the system user. Subsequently, a self-diagnostic process (SDP) is executed by the queuing system. For effective execution, the running time of the SDP should be considered, depending on the type of work. Once the SDP ends, the test processor is free for the next user's process.

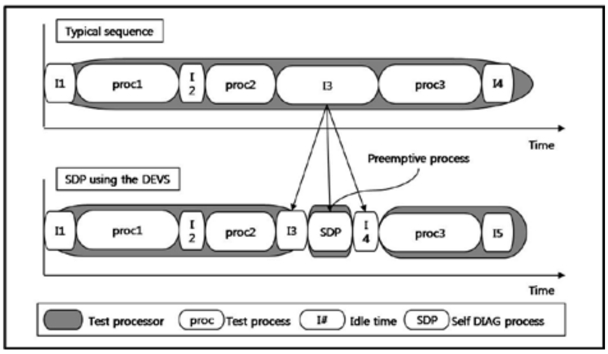


Figure 8: Self-diagnostic process using the DEVS

5. CONCLUSION

Using limited resources and minimizing loss of time are important factors in test equipment efficiency. In this paper, we proposed two methods to effectively utilize idle time during test processes. One of the proposals is a queuing system with non-preemptive scheduling to minimize idle time on multi-user equipment. The other is a self-diagnostic process with preemptive scheduling, applying the user's idle time. As shown in the results, idle time using the DEVS-based queuing system is effective for multi-user equipment but its effectiveness in single-user systems is negligible. We also confirmed that self-diagnostic processes with the DEVS, applying the user's idle time, is available for such systems.

REFERENCES

A.J.De, Ron, 2006. OEE and Equipment effectiveness: an evaluation. *International journal of production Research*, Vol 44, No. 23, 4987-5003.

Avi Silberschatz, Peter Galvin, 1994. Operaing system (8th Edition), *Addison-Wesley*.

Wallace J. Hopp, 2008. Single server queuing models. *International Series in Operations Research & Management Scienceh*, Vol 115, pp 51-79.

David R. Hill, 1996. Object-Oriented Analysis and Simulation. *Addison-Wesley*

Bernard P. Zeigler, Herbert Praehofer and Tag Gon Kim, 2000. Theory of Modeling and Simulation (2nd Edition), *Academic Press*.

Tag Gon Kim, 2007. EE612 Lecture Note, EECS, KAIST, http://smslab.kaist.ac.kr/Course/EE612/.

Tag Gon Kim, 2007 DEVSim++ User's Manual, http://smslab.kaist.ac.kr

Youngsin Han, Hae Sang Song, 2011. Co-modeling Methodology for Semiconductor Manufacturing and Automobile. *Communications in Computer and Information Science*, Vol 341, pp 7-14

ADDRESSING ROBUST BERTH PLANNING UNDER UNCERTAINTY VIA SIMULATION BASED OPTIMIZATION

Pasquale Legato[(a)] and Rina Mary Mazza[(b)]

[(a), (b)]Department of Informatics, Modeling, Electronics and System Engineering
University of Calabria
Via P. Bucci 42C
87036 Rende (CS), Italy

[(a)]legato@dimes.unical.it, [(b)]rmazza@dimes.unical.it

ABSTRACT
Decisions of allocating berth segments to incoming vessels, in maritime container terminals, has been extensively modeled in the scientific literature by resorting to formulations of mathematical programming with integer variables. Both vessel arrival times and processing times are usually considered as a deterministic input to the mathematical model despite of the uncertainty affecting berth decisions at the operational level, when several unpredictable events and operation delays occur and require to be managed. In this paper, we propose to apply the methodology of simulation based optimization to cope with uncertainty: a constructive algorithm is used to obtain a weekly plan at the tactical level; the allocation decisions are then adjusted at the operational level. Randomness in events and operations is taken into account by Monte Carlo simulation, while moving-average sample mean estimators are used to reduce the number of simulation runs required. Preliminary numerical results are also given.

Keywords: Simulation Optimization, statistical selection, port logistics, berth planning

1. INTRODUCTION

Many modern day systems providing products and services in the fields of logistics, manufacturing, transportation, network-centric computing, etc., are event-driven and, thus, can be modeled as discrete-event systems with the objective of carrying-out performance analysis and optimization. When pursuing decision integration and performance optimization in similar complex systems, the idea of inserting a simulation engine in an optimization algorithm is often the only practical solution method available in order to deal with difficult-to-solve combinatorial problems, embedded in realistic and dynamic processes characterized by several elements of randomness. The optimization algorithm is aimed at first generating an initial feasible solution and then exploring the whole feasible region (search process) until no further improvements of the performance results are obtained or until computation time is exhausted. The use of the simulation engine is required (evaluation process) since an estimate of the objective function cannot be returned by simply fitting a set of possible decision variables into a simple closed-form formula.

In the resulting methodology, known as Simulation Optimization (SO) (Andradottir 2007), the trade-off between the amount of computational time needed to find improved alternative solutions on the optimization side versus the effort in estimating via simulation the performance of a particular solution has always been a key issue in most SO techniques and general frameworks (Lee et al. 2006). (Fu 2001) divides these techniques in the following main categories:

- statistical procedures (e.g. ranking & selection procedures and multiple comparison for the comparison of two or more alternative system configurations);
- metaheuristics (methods directly adopted from deterministic optimization search strategies such as simulated annealing);
- stochastic optimization (random search, stochastic optimization);
- other, including ordinal optimization and sample path optimization.

In this paper we propose a Simulation Optimization scheme to manage both tactical and operational planning issues when facing the berth allocation problem (BAP) in port logistics. The object of the scheme is to enable the tuning of the tactical solution returned for the above problem when unforeseen and/or unwanted conditions overcome in the operational stage. In doing so, the SO scheme may benefit of a minor computational effort by employing a moving-average estimator for the sample mean in the procedures used to compare alternative solutions for the problem. The scheme, in fact, is based on a procedure belonging to the first category (i.e. ranking & selection) to estimate the best among a set of alternative berth allocation solutions, as well as metaheuristics that take care of solution generation and improvement at both the tactical and operational level.

The paper is organized as follows. In Section 2 an integrated tactical-operational view of the berth allocation problem is given, followed by considerations on how this problem is dealt with at the real container terminal of our interest. In Section 3 the Simulation Optimization scheme is described by focusing on the constructive algorithm used to find the tactical solution and the simulation approach used to find the operational solution. In particular, in the latter case a low-variance sample mean is proposed for use in ranking & selection techniques when selecting the best berth allocation solution. Some numerical examples that explain the concept of robustness of the final berth allocation solution provided via Simulation Optimization are presented in Section 4, while conclusions are drawn in Section 5.

2. THE BERTH ALLOCATION PROBLEM

Deciding which berthing position to assign to an incoming vessel is certainly the first and most important step of the overall resource and activity planning process in a maritime container terminal. Besides the (obvious) physical constraints that must be satisfied with respect to vessel size, draft and security measures, the assignment must take into account the distance lying between the candidate berthing position and the yard area where containers for the incoming vessel are to be stacked/retrieved. The ideal berthing position, meaning the position which minimizes the above distance, is also known as “home berthing”. As one may observe from Figure 1, with respect to the yard block in the dotted lines, berthing option n°1 is a more suitable choice than berthing option n°2. Indeed, the first option leads to shorter container transfer times and, thus, to greater throughput and higher revenues for the maritime terminal and improved customer satisfaction for the shipping companies calling the port.

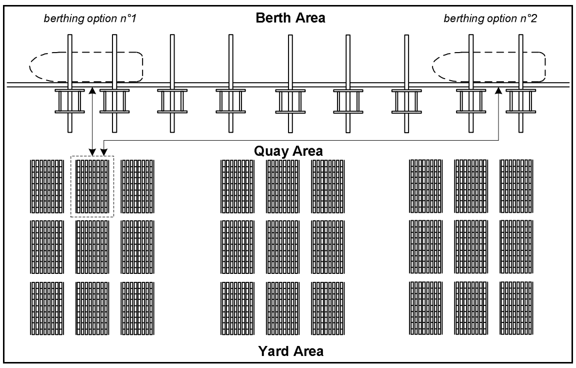


Figure 1: An example of how different positions yield different distances to be covered when performing vessel discharge/loading

A great number of factors may affect the vessel’s sojourn time in the assigned berthing position, among which resource availability (i.e. cranes, transfer vehicles and manpower), number of container moves to be performed, congestion due to other traffic, weather conditions, equipment failure, lack of synchronization in operations performed across bordering terminal areas involved in the D/L operations and so on. Therefore, when dealing with the BAP, both deterministic and stochastic models of Operations Research can give a valuable support. Decision models pertaining to the BAP are often integrated in solution methodologies designed for wider logistic processes (Steenken et al. 2004; Meisel and Bierwirth 2006; Stahlbock and Voß 2008). Moreover, BAP decisions have also been included in discrete-event simulation models of port logistics according to global outer views to evaluate the global performance of the terminal in terms of productivity and vessel turnaround time (Yun and Choi 1999; Legato and Mazza 2001; Bielli et al. 2006).

As was well pointed out by (Moorty and Teo 2006), the BAP may be suitably viewed at both a tactical and operational level. In the former case it pertains to the definition of a “weekly plan”, i.e. a berth template where arriving vessels are expected to be moored at some preferable berth segments, under the assumption that (1) incoming vessels enter the port at a forecasted, but deterministic time instant and (2) a specific but fixed, average service rate should be guaranteed in discharging-loading each berthed vessel, whatever be the availability of the quay cranes when the operations start. Vice versa, at the operational level, the terminal manager is asked to face delays on vessel arrival time, actual availability in time of each crane and manpower gang to be assigned, plus delays within operations, physical obstacles such as draft and work in progress that restrict mooring locations and so on. This calls for adjusting the tactical weekly plan in real time. At this level of the planning process, a finer representation of the berth segment and handling equipment (e.g. position of quay cranes, shift constraints for manpower, etc.), together with a finer reproduction of the complex discharge/loading process by means of a discrete-event simulator is well appreciated. To this purpose (Legato et al. 2010) developed a simulator where the effect of container transfer from the yard to the berth and vice versa is also highlighted when simulating container discharge/loading processes under a given assignment profile of some cranes to a vessel and a given schedule of container moves.

Coming back to the tactical level of decisions, the whole berth may be viewed as a discrete set of small berthing segments or a continuous, unique long segment and each vessel is represented as a space-time rectangle to reflect it space-time occupancy within the berth template. Whatever the berth representation, the planning goal is to achieve a good matching between container storage positions on the yard versus container discharge/loading positions on the berth. This is a prerequisite before organizing the container picking, transfer and delivery process back and forth between the yard and quay. In previous literature both the discrete and the continuous approaches to berth modeling at a tactical level have been pursued (Lim 1998; Imai et al.

2001; Kim and Moon 2003; Guan and Cheung 2004; Imai et al. 2005, Cordeau et al. 2005) under the common assumption that both the arrival time and processing time per ship are known without any uncertainty.

More recent papers concentrate on improving computational performance of the heuristic methods proposed for problem solution (Wang and Lim 2007; Hansen et al. 2008; Lee and Chen 2009; Buhrkal et al. 2009), while others are devoted to integrate berth allocation with the subsequent decision of assigning the right number of cranes, hour by hour, to each vessel during discharge-loading operations (see Bierwith and Meisel 2010 for an extensive survey). Observe that just couple of papers (Zhou and Kang 2008; Hendriks et al 2010) focus on the problem of managing the uncertainty in vessel arrival times and service times. Precisely, (Zhou and Kang 2008) adopt a discrete berth representation, with a given number of service points along the quay, and propose an integrated berth & quay-crane model by using a stochastic 0-1 programming model aiming to minimize the waiting time for both berth and crane assignment. Vice versa, (Hendriks et al. 2010) remark that container terminal operators and shipping lines agree upon arrival time windows and develop a planning model that explicitly takes this arrival time window into account. The concept of robustness in berth planning is implemented as the capability of returning a feasible solution for each arrival scenario where all vessels arrive within their arrival time window. This window is obtained by simply shifting the arrival time of each vessel with the goal of minimizing the maximal crane capacity reservation that would result from adopting a plan based on fixed arrival times.

The major limit of the stochastic programming approach pursued in recent literature for managing uncertainty lies in the prohibitive computational costs of representing and analyzing all possible scenarios arising from the joint variations of both vessel arrivals and time duration of discharge/loading operations. In such a case, a simulation based approach to the optimization of berthing decisions is preferable.

In the work at hand, we refer to the container terminal at the port of Gioia Tauro in Southern Italy. The generation of their weekly plan is supported by CaLeMa: a simulation environment designed to reproduce ship berthing, with a particular focus on contention of the entrance channel and the management of berthing points (Canonaco et al. 2007). Besides being used in scenario analysis, CaLeMa is under further development to include quay crane management by taking into account the uncertainty in quay crane availability and durations of discharge/loading operations.

3. THE SIMULATION-OPTIMIZATION SCHEME

To model and solve the berth allocation problem we propose the Simulation Optimization scheme illustrated in Figure 2 that bridges the natural gap between the related tactical and operational solution methodologies.

As for the tactical level, the scheme responds to the quest of producing a berth template by applying a constructive algorithm. This template specifies the position in time and space of single berth windows by taking into account both the contractual agreements defined between the terminal and the shipping companies and the physical constraints imposed by vessel drafts. Although this solution is obtained in a fast and accurate way, it still corresponds to a static representation that is unable to embody the uncertainty of the major activities taking place in the terminal facility such as the vessel arrival process and the container discharge/loading (D/L) process, as well as the actual availability of the resources required to carry out the above processes.

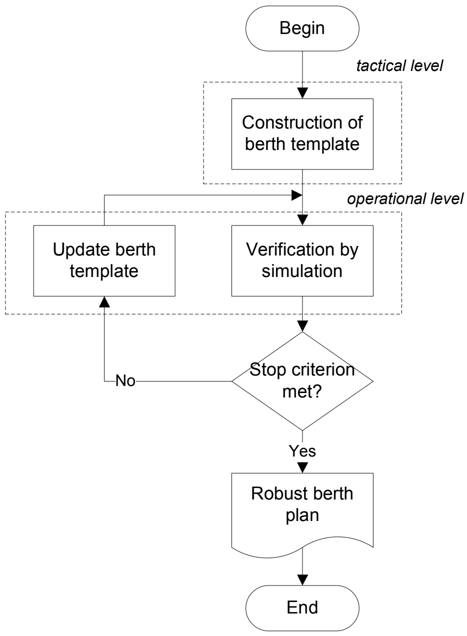


Figure 2: The SO scheme for robust berth planning

As a result, the performance of this initial solution needs to be tested at the operational level with respect to a wide range of additional conditions that, in our view, offer a measure of the so-called robustness of the solution found. In other words, in the SO scheme the goodness of the solution found at the tactical level is later assessed and compared with alternative berthing plans on the operational level generated by a heuristic algorithm for neighborhood exploration. The overall aim of the scheme consists in minimizing the waiting time suffered by vessels due to untimely arrivals, non-deterministic service times and/or unavailable resources.

3.1. View at the tactical level

Constructing a solution for the BAP at the tactical level may certainly vary from one facility to another, although the information used to do so is practically the same (e.g. vessel size and draft, expected vessel arrival/departure, workload, resource availability). As reported in Section 2, mathematical programming models are often used for this purpose. In particular, the work by (Kim and Moon 2003), which minimizes the cost for berthing a vessel far from its home berthing and the cost for delaying vessel departure, may be considered the starting point for most models based on the continuous location space approach to the BAP.

Unfortunately, despite it being very useful, a few practical requirements prevent us from applying this model. First of all, the commercial solvers normally used for this purpose can only solve small instances. Secondly, whatever the dimension of the problem, commercial solvers cannot be embedded in the software applications already in use at most container terminals. As a consequence, the SO scheme in Figure 2 has been designed to use a constructive algorithm in order to provide a tactical solution for the BAP. In companion papers we experimented and analyzed the properties of metaheuristics used to cope with similar complex logistic problems (Legato et al. 2008; Legato et al. 2010). However, we recognize that in this particular case obtaining a fast and accurate solution at the tactical level calls for the use of a constructive approach, rather than feeding randomly-generated initial solutions that almost certainly do not resemble those provided by the terminal operators in real-life planning.

This stated, we designed a constructive solution algorithm that extends the model proposed by (Kim and Moon 2003) in order to include restrictions on vessel berthing along certain segments due to the lack of compliance between vessel draft and berth depth. The algorithm, which minimizes the additional cost sustained by the terminal operator when vessels are berthed in non-optimal conditions (i.e. delay in berthing and far from its ideal berth position), is described by the pseudo-code given below.

Initialization
1: *Parameter setting (Δ, m)*
2: *Generation vessel and order by arrival time*
Berth definition for current vessel
3: *Extract vessel with smallest arrival time*
4: *Define all feasible berth segments*
5: *Determine all vessel berthing positions*
Berth definition for next Δ vessels
6: *Define all feasible berth segments for the next Δ vessels*
7: *Select best berthing position for each of Δ vessels*
Selection
8: *Evaluate objective function* f_Δ
9: *Order solution and select of top m*
10: *Eliminate of non-selected solutions*
Exit condition
11: *Return to step 3 if berthing of all incoming vessels is not completed*

In the above algorithm, just as in real-life company practice, all vessels arriving to the port are bound to be berthed, so the problem is highly combinatorial. However, a natural pruning stage is delivered by the limited number of feasible berthing positions (step 4) and, thus, the number of possible combinations is reduced. As a matter of fact, a vessel can be berthed along a segment only if the size of the segment matches the length of the vessel measured in bollards, followed by an extra bollard for security matters. Furthermore, as previously mentioned, water depth in free berth areas need to comply with vessel draft. If both of these conditions are met, then one of the following two rules applies: the vessel can be berthed in the upper or lower angle of a free area, as illustrated for berthing options 1 to 4 for vessel 4 in Figure 3; in contrast, the vessel can be berthed in the same positions as previously berthed vessels once these have completed their D/L operations and have been unberthed, as illustrated by berthing option 5 in Figure 3.

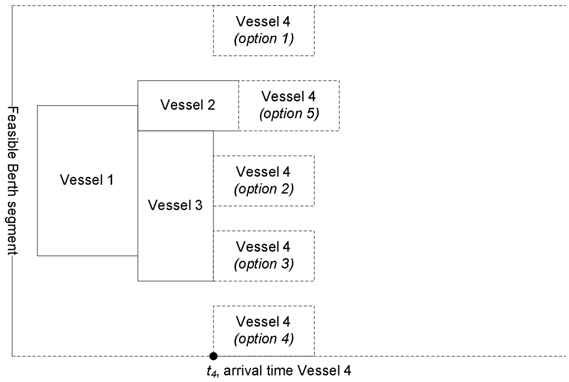


Figure 3: Five possible berthing solutions for Vessel 4

The estimation of the objective function f_Δ associated with any of the above berth allocation decisions is performed in the *Selection* section by taking into account two contributions. One term represents the cost of the first i vessels already berthed and, thus, it returns an immediate evaluation based on the previously made berth assignments. The other is an estimate of the cost required to berth the next Δ vessels ($i+1, i+2, ..., i+\Delta$) based on a greedy-operating logic (steps 6 and 7). Thanks to this estimation, node sampling aimed at selecting the best partial solutions to be fed as input to the next solution-building iterations of the algorithm is performed according to the classic top m criterion (step 9). The entire mechanism is cycled until all incoming vessels are berthed and, thus, the algorithm returns a final solution for the BAP.

3.2. View at the operational level

A discrete-event simulator has been designed to test the solution returned by the tactical level with respect to its so-called robustness at the operational level. As a result, the weekly template may not hold because of the randomness featured by the terminal activities in which process initiation and duration change over time. In our experience, the major sources of randomness, for which the simulator must account for, are given by:

- vessel arrival times;
- quay crane availability and deployment;
- D/L service times.

Vessel arrival generally occurs within a fixed time window in a week for oceanic vessels or on the basis of a probabilistic profile as in the case of common feeders. An example of a similar profile is given in Figure 4, according to which the real interarrival times of 1030 common feeder vessels (in one year) can be suitably modeled by an exponential law with mean value equal to 505 minutes. So, while the periodic arrivals of the oceanic vessels are of limited impact on the berth planning activities, the simulator is necessary to account for the random arrivals of the common feeders.

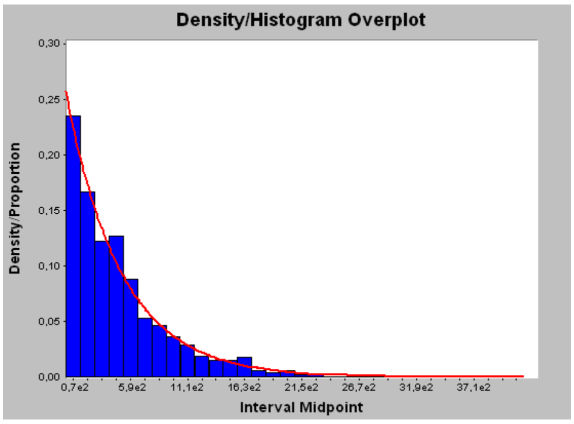


Figure 4: Profile of common feeder interarrival times

The use of quay cranes is another source of randomness for which the simulator is meant to cope with. As a matter of fact, the terminal's operations manager must first verify the overall availability of the cranes and then provide for assigning specific cranes to a specific vessel and deploying these cranes along the quay according to an hourly profile.

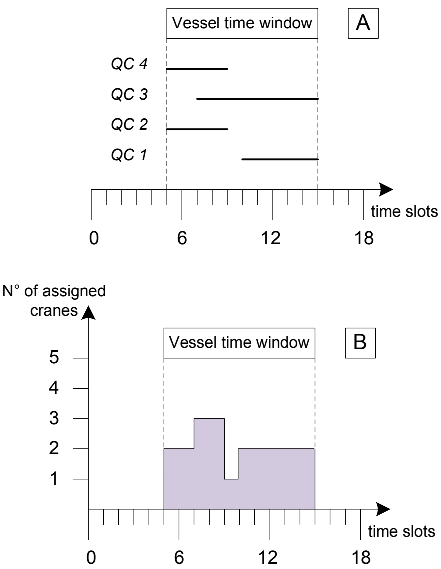


Figure 5: Crane availability and crane intensity

Frame A of Figure 5 illustrates both the availability of 4 quay cranes and the length of the time window during which a vessel requires crane assignment. Specifically, the average number of cranes to be assigned to a vessel during its time window, a.k.a. *crane intensity* (*CI*), is usually fixed by contractual agreements. The actual value of the crane intensity for a given vessel, which ought to match the target value of the crane intensity, can be determined from the corresponding quay crane hourly deployment profile as, for example, the one illustrated in frame B of Figure 5, and computed according to the following

$$CI = \frac{1}{tw_f - tw_s} \sum_{i=1}^{n} aQC_i \qquad (1)$$

where tw_f, tw_s, n and aQC_i are the time window's start and finish times, the number of quay cranes deployed and the availability (in hours) of each quay crane, respectively. As a result, since the berth template is built according to the target value of the crane intensity for each vessel bearing a time window, it should be clear how any kind of change in quay crane availability and/or deployment at the operational level may affect the goodness of the entire template.

As for the final, yet most important source of randomness, the simulator must account for the D/L service times by considering eventual disruptions due to *i*) failure in the container handling and/or transfer equipment and *ii*) congestion and/or starvation phenomena arising from the lack of synchronization among the equipment involved in container transfer from the quay area to the yard area and vice versa. In other words, if the transfer activity carried-out by transfer vehicles from the quay to the yard is too slow, then the quay crane discharge activity is prone to be affected by *blocking* during operations due to container space that will be likely unavailable in the buffer area under the crane. Vice versa, because of an empty buffer area, crane *starvation* is likely to occur during container loading operations on the vessel if the transfer vehicles from the yard to the quay are too slow.

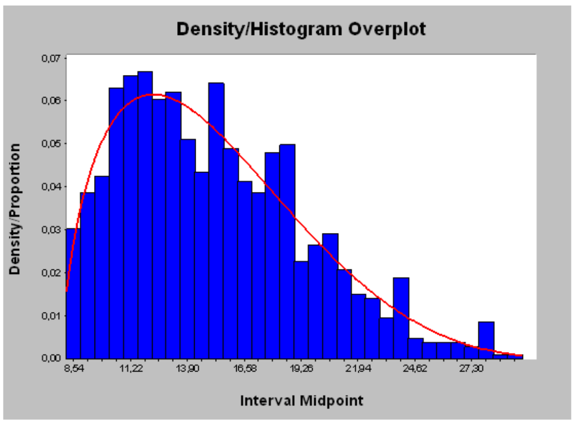


Figure 6: Profile of common feeder overall D/L times

An example of the randomness in the overall D/L times of the same real 1030 common feeder vessels previously mentioned is represented by the Beta-like profile in Figure 6 bearing mean=15.12, variance=18.96 and skewness=0.66.

This stated, at every iteration of the scheme in Figure 2, the simulator is meant to play its role by feeding to the SO procedure a sample mean that represents an estimate of the expected value of the objective function for the current solution (berth template). In turn, the SO procedure is asked to compare alternative competing solutions. To guarantee the correct selection of the "best" sample mean under a fixed level of confidence, a great computational effort may be required in terms of number of observations upon which each sample mean is defined: the greater the variance of the sample mean, the greater the number of observations required. The most common procedures of ranking and selection (R&S) work with the standard sample mean (Kim and Nelson 2006), which, in our context is computed across multiple simulation replications (sample size). Here a moving-window based logic, inspired by Welch's procedure for estimating the length of the transient in simulation, is adopted. Thus, we first organize n independent, simulated output observations into b groups and then compute the average value of the *ith* observation across these groups according to the width w of the moving window. Let Y_{ji} be the *ith* observation within group b, then

$$\overline{Y}_i = \frac{1}{b}\sum_{j=1}^{b} Y_{ji}. \tag{2}$$

Hence, the set of values, $\overline{Y}_1, \overline{Y}_2, ..., \overline{Y}_m$, is used to define the moving average (MA) $\overline{Y}_i(w)$ with a window length of w as follows:

$$\overline{Y}_i(w) = \frac{\sum_{s=-w}^{w} \overline{Y}_{i+s}}{2w+1} \qquad i = w+1, ..., m-w. \tag{3}$$

Observe that the neighboring moving averages (say $\overline{Y}_i(w)$ and $\overline{Y}_{i+1}(w)$) are still unbiased estimators of the mean of the output observations but they are (positively) correlated due to those common observations shared when averaging over the fixed w values. As a result, the variance of the moving-average estimator $\overline{Y}_i(w)$ is smaller than the variance associated to the standard estimator (with $w = 0$) and, thus, less simulation effort is expected to be required. Clearly, the MA estimator can be used in any type of R&S procedure whether it be one-stage (Bechhofer 1954), two-stage (Rinott 1978) or *n*-stage (Goldsman et al. 2002).

Turning the attention to the optimization part of the SO procedure, observe that the main component at the operational level is based on a neighborhood structure that allows to move from one BAP solution to another. In particular, a vessel is selected from the BAP solution currently proposed and meant to be swapped with that of another vessel. The swap is considered feasible if, on one side, the vessel's future position is compliant with vessel size and draft and, on the other, if the vessel's arrival and departure is covered by that of the other vessel. Obviously, for the other nearby vessels the swapping activity may require an "adjustment" along the berth. All the vessels fulfilling the above conditions are inserted in a set of so-called swappable vessels and the corresponding BAP solutions represent new neighboring solutions for the current berth plan. One (or more than one) neighbor will be chosen from this set and, then, verified via simulation. When deciding which solution to choose between the current and the new BAP solutions, a simulated annealing (SA) metaheuristics (Kim and Moon 2003) is used.

The pseudo-code describing this part of the SO scheme is given below.

Initialization
1: *Parameter setting (T, α, threshold, n)*
2: *Assign tactical template to current BAP solution*
Definition of swappable vessels for current vessel
3: *Select vessel from current BAP solution*
4: *Create a set of swappable vessels and select a vessel*
5: *Swap vessels and perform adjustments*
Solution comparison
6: *Compare new solution with current solution*
7: *Accept new solution with probability p=1 if value of objective function is best or with probability* $p = e^{\Delta/T}$ *if value of objective function is worst*
8: *Decrease T according to α*
Exit condition
9: *T<threshold or no improvements in last n iterations (else return to step 3)*

At this point, it is possible to discuss how solution robustness is conceived and accomplished by the above SA-based search for the optimal operational BAP template. When the tactical template is simulated, the stochastic operational conditions unavoidably affect the value of the objective function returned by the tactical planning phase. In particular, the delays in the vessel handling time highlighted by the simulation must be recovered by rearranging the space-time windows pertaining to every single vessel, at the price of settling for a new berthing position that is distant from the home berthing. As a result, the value of the objective function is expected to deteriorate. Therefore, the aim of the overall SO scheme consists in keeping deterioration within a limited range from the initial value corresponding to the tactical template. This stated, a solution is said to be robust when, under the uncertainty of the operational level, it is able to limit its deviation from optimality - for instance, given a final operational template, in 90% of the cases vessels already berthed are able to complete their operations without triggering delay propagations on the incoming vessels waiting to be berthed.

4. NUMERICAL EXAMPLES

The aim of this section is twofold: on one hand, we expect to show how under the previously described conditions of uncertainty the tactical solution returned for the BAP requires tuning at the operational level; on the other, we wish to verify to what extent the SO scheme proposed may benefit of a minor computational expense when comparing simulated solutions by means of a moving-average estimator for the sample means within R&S procedures, rather than the straightforward sample means based on independent observations.

As for the first aim, preliminary experiments mainly devoted to illustrative purposes focus on a few vessels (i.e. 2 oceanic vessels and 3 feeders which both share a common and dedicated berth segment) belonging to a major service for which real data is provided by the company that runs the container terminal located at the port of Gioia Tauro in Southern Italy. Considering the small number of vessels to be berthed and under the hypothesis that all problem data is deterministic, the corresponding integer programming based formulation of the BAP has been solved under Excel, thus obtaining the tactical solution required as initial step of the SO scheme. A graphical representation of the tactical solution for the problem instance at hand is given in the left side of Figure 7.

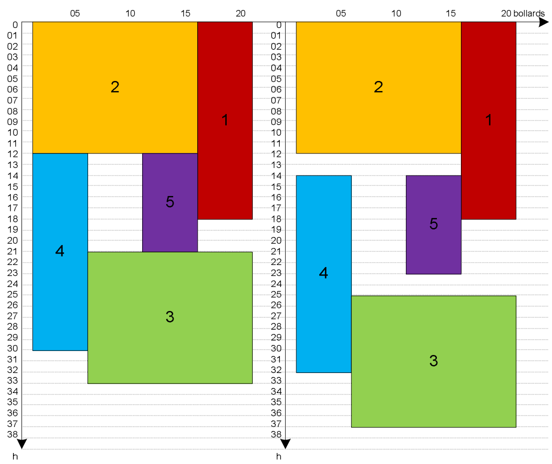


Figure 7: Tactical vs operational berth template

Now, according to company records, the delay in vessel operations is distributed according to the Pearson Type VI profile illustrated in Figure 8 (scale I, shape 1.945) for oceanic vessels and the Log-Logistic profile illustrated in Figure 9 (scale 0.586, shape 0.262) for feeder vessels. Unfortunately, as one may observe from Figure 7, the template on the left does not tolerate any kind of delay on oceanic vessel 2 and feeder vessel 5. Vice versa, since the time gap of two hours left in the right side template in Figure 7 for vessels 2 and 5 is sufficient to prevent delay propagation in 72% and 85% of the cases, respectively, then the operational template on the right in Figure 7 may be viewed as the robust counterpart of the tactical template. In other words, the degree of robustness lies in the time gap located after every single window which is delay-tolerant, i.e. unpredictable delays in operations do not immediately affect the operations scheduled on the vessels planned to be berthed afterwards. A robust template is also valuable for its managerial implications: the number of vessels delayed beyond their time windows is reduced and, therefore, the senior management reduces the payment of extra charges to shipping companies for not achieving the level of service stipulated in their contracts.

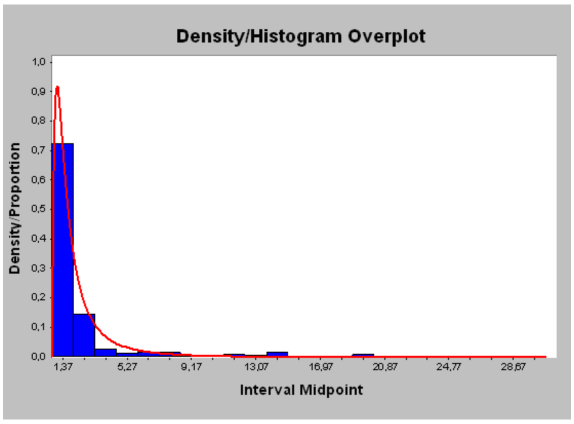


Figure 8: Profile of delays for oceanic vessels

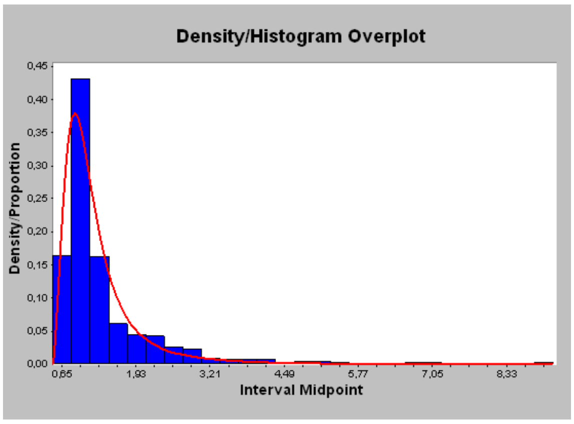


Figure 9: Profile of delays for feeder vessels

As for the second aim of the preliminary numerical experiments, to investigate the effect of using R&S procedures based on a moving-average, rather than standard estimator for the sample mean, we organized the n simulated output observations into b groups, each of size m, and then computed the average value of the *ith* observation across these groups according to the width *w* of the moving window. We then focused on the computational results returned when increasing the value of *w* from 1 to 5. In Table 1 for five selected instances we recorded the number of simulated observations (runs) required by the moving-average

(MA) sample mean (3) against the standard sample mean (S) estimator for the expected value of the objective function associated with each berth allocation plan.

Table 1: Average simulation runs required by different estimators for the sample mean in R&S procedures

W	E/I	#1	#2	#3	#4	#5
1	MA	110	283	459	195	145
	S	828	3771	6835	2571	1952
2	MA	104	192	233	133	117
	S	847	3840	7026	2448	1952
3	MA	104	142	195	119	103
	S	834	3836	7025	2544	1964
4	MA	100	139	184	119	101
	S	832	3925	7124	2641	1963
5	MA	100	118	141	105	100
	S	829	3765	7107	2507	1975
indifference zone parameter					δ=1‰	

In the above examples, both estimators were used to determine the number of additional runs required according to a classic two-stage indifference-zone ranking and selection procedure, under a fixed probability of correct selection $1-\alpha = 0.95$ and $\delta = 1$‰. To this purpose, we remark the effectiveness of the MA estimator since it reduces the sample size from intolerable to tolerable, especially if one considers the number of runs cumulated over all the neighbor solutions to be compared (consider that in the above five instances neighborhood size may reach 100). Furthermore, MA also delivers a growing variance reduction effect as the width of the window increases. In Table 1, this may be appreciated by reading the MA vertical values for each instance.

5. CONCLUSIONS

Simulation Optimization has been shown to be well suitable in addressing the solution of the berth allocation problem under uncertainty. It has allowed to integrate the proper deterministic model formulation at the tactical level with discrete-event simulation, used at the operational level, to cope with uncertainty in the duration of vessel discharge/loading activities and other sources of random occurring events and availability of resources in time. Real-size BAP instances may be solved by combining constructive heuristics and a simulated annealing based search process, where a discrete-event simulator is called to evaluate competing solutions. The computational burden due to the number of simulation runs required for the ranking and selection of solutions is kept tolerable by resorting to window-based moving sample means within a classic two-stage procedure. The SO scheme proposed has returned "robust" BAP solutions that well contain delay propagation at the operational level. Results of extensive numerical experiments on large-size instances will be presented in a companion paper.

REFERENCES

Andradóttir, S., 2007. Simulation Optimization. In J. Banks, ed. *Handbook of simulation: principles, methodology, advances, applications, and practice*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Bechhofer, R.E., 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics* 25(1), 16-39.

Bielli, M., Boulmakoul, A., and Rida, M., 2006. Object oriented model for container terminal distributed simulation. *European Journal of Operational Research,* 175(3), 1731-1751.

Bierwirth, C. and Meisel, F., 2010. A survey of berth allocation and quay crane scheduling problems in container terminals. *European Journal of Operational Research,* 202(3), 615-627.

Buhrkal, K., Zuglian, S., Ropke, S., Larsen, J., and Lusby, R., 2009. *Models for the discrete berth allocation problem: a computational comparison.* DTU Technical Report 14-2009. Delft, Denmark.

Canonaco, P., Legato, P., and Mazza, R.M., 2007. An integrated simulation model for channel contention and berth management at a maritime container terminal. *Proceedings of the 21st European Conference on Modelling and Simulation*, pp. 353-362. June 04-06, Prague (Czech Republic).

Cordeau, J.-F., Laporte, G., Legato, P., and Moccia, L., 2005. Models and tabu search heuristics for the berth-allocation problem. *Transportation Science* 39(4), 526-538.

Fu, M.C., 2001. Simulation optimization in *Proceedings of the 2001 Winter Simulation Conference*, pp. 53-61. December 09-12, Arlington, (Virginia, USA).

Goldsman, D.M., Kim, S.-H., Marshall, W.S., and Nelson, B.L., 2002. Ranking and selection for steady-state simulation: procedures and perspectives. *INFORMS Journal on Computing* 14(1), 2-19.

Guan, Y. and Cheung, R.K., 2004. The berth allocation problem: models and solution methods. *OR Spectrum* 26, 75-92.

Hansen, P., Oguz, C., and Mladenovic, N., 2008. Variable neighborhood search for minimum cost berth allocation. *European Journal of Operational Research* 191, 636-649.

Hendriks, M., Laumanns, M., Lefeber, E., and Udding, J.T., 2010. Robust cyclic berth planning of container vessels. *OR Spectrum* 32, 501-517.

Imai, A., Nishimura, E., and Papadimitriou, S., 2001. The dynamic berth allocation problem for a container port. *Transportation Research Part B* 35, 401-417.

Imai, A., Sun, X., Nishimura, E., and Papadimitriou, S., 2005. Berth allocation in a container port: using a continuous location space approach. *Transportation Research Part B* 39(3), 199-221.

Kim, K.H. and Moon, K.C., 2003. Berth scheduling by simulated annealing. *Transportation Research Part B* 37, 541–560.

Kim, S.H. and Nelson, B.L., 2003. Selecting the Best System: Theory and Methods. *Proceedings of the 2003 Winter Simulation Conference*, pp. 101-112. December 07-10, New Orleans, (Louisiana, USA).

Lee, Y. and Chen, C.Y., 2009. An optimization heuristic for the berth scheduling problem. *European Journal of Operational Research* 196(2), 500-508.

Lee, L.H., Chew, E.P., and Manikam, P., 2006. A general framework on the simulation-based optimization under fixed computing budget. *European Journal of Operational Research* 174(3), 1828-1841.

Legato, P. and Mazza, R.M., 2001. Berth planning and resources optimisation at a container terminal via discrete event simulation. *European Journal of Operational Research* 133(3), 537-547.

Legato, P., Mazza, R.M., and Trunfio, R., 2008. Simulation-based optimization for the quay crane scheduling problem. *Proceedings of the 2008 Winter Simulation Conference*, pp. 2717-2725. December 07-10, Miami (Florida, USA).

Legato, P., Mazza, R.M., and Trunfio, R., 2010. Simulation-based optimization for discharge/loading operations at a maritime container terminal. *OR Spectrum* 32, 543-567.

Lim, A., 1998. The berth planning problem. *Operations Research Letters* 22, 105-110.

Meisel, F. and Bierwith, C., 2006. Integration of berth allocation and crane assignment to improve the resource utilization at a seaport container terminal. *Operations Research Proceedings 2005*, pp. 105-110. September 07-09, Bremen (Germany).

Moorthy, R. and Teo, C.P., 2006. Berth management in container terminal: the template design problem. *OR Spectrum* 28, 495-518.

Rinott, Y., 1978. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics - Theory and Methods* A7(8), 799-811.

Stahlbock, R. and Voß, S., 2008. Operations research at container terminals: a literature update. *OR Spectrum* 30, 1-52.

Steenken, D., Voß, S., and Stahlbock R., 2004. Container terminal operation and operations research – a classification and literature review. *OR Spectrum* 26, 3-49.

Wang, F. and Lim, A., 2007. A stochastic beam search for the berth allocation problem. *Decision Support Systems* 42(4), 2186-2196.

Yun, W.Y. and Choi, Y.S., 1999. A simulation model for container-terminal operation analysis using an object-oriented approach. *International Journal of Production Economics* 59, 221-230.

Zhou, P.F. and Kang, H.G., 2008. Study on Berth and Quay-crane Allocation under Stochastic Environments in Container Terminal. *Systems Engineering — Theory & Practice* 28(1) 161-169.

AUTHORS BIOGRAPHY

Pasquale Legato is an Associate Professor of Operations Research at the University of Calabria, Rende (CS, Italy), where he teaches courses on simulation for system performance evaluation. He has published on queuing network models for job shop and logistic systems, as well as on integer programming models. He has been involved in several national and international applied research projects and is serving as a reviewer for many international journals. His current research activities focus on the development and analysis of queuing network models for logistic systems, discrete-event simulation and the integration of simulation output analysis techniques with combinatorial optimization algorithms for real-life applications in transportation and logistics. His home-page is www.deis.unical.it/legato.

Rina Mary Mazza received her Laurea degree in Management Engineering and obtained a Ph.D. degree in Operations Research at the University of Calabria, Rende (CS, Italy). She currently heads the "Research Projects Office" at the Department of Informatics, Modeling, Electronics and System Engineering (DIMES) of the above University. She has a seven-year working experience on knowledge management and quality assurance in research centers. She is also a consultant for operations modeling and simulation in container terminals. Her current research interests include discrete-event simulation and optimum-seeking by simulation in complex logistic systems.

MODELLING AND MULTI-OBJECTIVE OPTIMIZATION OF THE VHP STERILIZATION PROCESS OF POUCH PACKAGING

Matteo Folezzani[(a)], Michele Manfredi[(b)], Giuseppe Vignali[(c)]

[(a), (b)] CIPACK Interdepartmental Centre, University of Parma, viale G.P. Usberti 181/A, 43124 Parma (Italy)
[(c)] Department of Industrial Engineering, University of Parma, viale G.P. Usberti 181/A, 43124 (Italy)

[(a)]matteo.folezzani@unipr.it, [(b)] michele.manfredi@unipr.it, [(c)] giuseppe.vignali@unipr.it

ABSTRACT

Aseptic filling technology of Spouted Pouch packaging, even more used in the beverages sector, requires a very complex sterilization and rinsing of the packaging before filling.

This work aims to optimize the sterilization process of a pouch packaging used in aseptic technology. Starting from a fixed position of the nozzle inside the pouch, this work goals to varying the flow rate through the nozzle and the process time, in order to ensure the contact between H_2O_2 solution and all zone of the packaging wall, so to optimize the sterilization process. The analysed sterilization process adopts a mixture composed by vaporized hydrogen peroxide and hot sterile air.

This problem has been resolved using a multi-objective optimisation with the mains objectives to minimize the consumption of H_2O_2 and the relatives costs and maximize the sterilization efficacy on the packaging volume in a defined time.

Keywords: pouch packaging, VHP sterilization, multi-objective optimization, ModeFrontier.

1. INTRODUCTION

The Sterilization of food packaging is one of the most critical phases of the aseptic processing (Ferretti et al., 2006). During the last ten years many experiments and studies have been performed in order to identify the best sterilizing agent for each combination of product-packaging (Robertson, 2006; Ansari and Datta, 2003). In particular, polymeric-based packaging has assumed a main role in aseptic packaging thanks to its economic convenience, an easy stock and transport. On these kinds of packaging materials, both Chemicals (wet peroxide acetic acid, wet and vaporized hydrogen peroxide, etc..) and physical methods (pulsed light, β and γ radiation) have been adopted as sterilization principle (Abreu and Faria, 2004; Riganakosa et al., 1999; Farkas, 1998).

Regarding chemical methods, the main problem is caused by the temperature of the chemical agent reached during the polymeric packaging treatment. Hydrogen Peroxide reaches a good efficacy only at a temperature higher than 70°C. So, if the packaging material to treat has a lower point of glass transition (i.e. for PET 69°C), it is not possible to use liquid solution, with the risk of damaging the packaging. For this reason an addition of peroxide acetic acid at a percentage of about 1% has been frequently used, in order to decrease the temperature of action of the hydrogen peroxide solution. However this addition increase the cost of the solution, so many companies tried to use only Hydrogen peroxide in a vaporized form (Yun and Sastry, 2007).

The two techniques primarily adopting Hydrogen Peroxide in vapour condition were the “Vaporized Hydrogen Peroxide” (VHP) and “Condensed Hydrogen Peroxide” (CHP) (Cerny, 1992). The first method does not produce vapour condensation on the inner side of the packaging, thanks to the use of a pre-heating section; instead CHP method wants produce this condensation in order to be more powerful on microbial reduction, using a packaging not pre-heated. Both these methods have been experimentally tested by several authors for many food packaging in order to optimize the sterilization process (Klapes and Vesley, 1990; Cerny, 1992).

The application of these latter sterilization techniques on flexible containers shows, also, additional problems about the removal process of the sterilizing agent, which is complicated by the small size of the exit hole and by the type of material (Castle et al, 1995; Abdul Ghani, 2001). For this reason only VHP method is currently adopted in the food sector, according to its better removal property of hydrogen peroxide from the packaging surface.

In order to simulate the behaviour of this latter technology, numerical simulations could be able to predict the velocity and the concentration of the chemical agent inside the pouch (Patrick Kirchner, 2013). In some cases Computational Fluid Dynamics (CFD) analysis is used to investigate and to optimize industrial beverage processes, i.e. filling, rinsing and empting. Instead, in the case of sterilization using vaporized solutions only one author recently try to approach the problem using CFD methods, mainly due to the complexity of simulate two different vapour component acting simultaneously (Qian Zou, 2006, Shao-Ping Wang et al, ,2004).

Based on these premises, this work aims to optimize the sterilization process of a pouch packaging used in aseptic technology. Starting from a fixed position of the nozzle inside the pouch, this work goals to simulate the process, varying the flow rate through the nozzle and the process time, in order to ensure the contact between H_2O_2 solution and all zone of the packaging wall, so to optimize the sterilization process.

This work then is divided in two parts: the first one concerns the simulation and modelling the physic condition of the sterilization process on the pouch, and the second one regards the optimization of the process with find the bests configuration.

In the first part of this work, a multicomponent CFD models were used to simulate the Vaporized Hydrogen Peroxide (VHP) sterilization process of spouted pouches. The CFD simulations were in a transient state, being necessary to show the behaviour during the initial phase and because it is important to determinate the correct sterilization time. The use of transient simulations allows testing combinations between variables in order to optimize also the time of treatment. For the first part ANSYS CFX software was used.

In the second part of the work the CFX software was implemented in an optimization tool (ModeFrontier software). Before any optimization, the problem must first be modelled. A multi-objective optimization task considers several conflicting objectives simultaneously. In such a case, there is usually no single optimal solution, but a set of alternatives with different trade-offs, called Pareto optimal solutions, or non-dominated solutions. Despite the existence of multiple Pareto optimal solutions, in practice, usually only one of these solutions is to be chosen. Thus, compared to single-objective optimization problems, in multi-objective optimization, there are at least two equally important tasks: an optimization task for finding Pareto optimal solutions (involving a computer-based procedure) and a decision-making task for choosing a single most preferred solution. All the optimization analysis were performed using Mode Frontier software. In this software we designed a workflow, where is necessary to specify the logic node and the algorithm which will generate all possible allowed combinations. In this case, the logic node was connected to ANSYS CFX.

2. MATERIAL AND METHODS

2.1 Materials

The sterilization is obtained from a mixture of air and hydrogen peroxide. The hydrogen peroxide at 30% of concentration is vaporized in a plate maintained at 200°C. This solution is subsequently mixed with a flow of sterile hot air that has a rate of 2,000 Nl/h. The concentration of hydrogen peroxide which should arrive in each envelope is 5,000 ppm. The temperature of the whole system is maintained at 55°C to avoid the problem of condensation. The condensation of the sterilizing mixture would make it impossible to remove and obtains a residual value in compliance with the regulations. The nozzles have an overall height of 200 mm and an outlet section with a diameter of 2.5 mm. This form can be modified as required. The pouch pack has a very complex geometry, with a total height of 170 mm, a maximum width of 90 mm, a maximum depth of 53 mm and an overall volume of 340 cm^3.

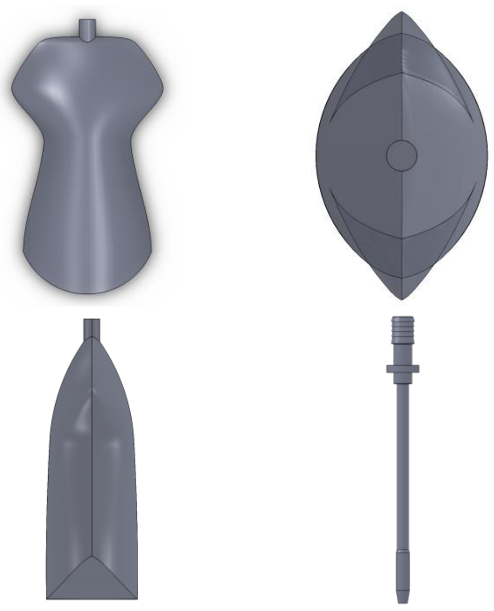


Figure 1. Different view of pouch packaging and sterilization nozzle

2.2 Modelling

2.2.1 CFD Modelling

The CFD simulations have been performed using ANSYS CFX 14.5 version, which allows simulating multi component fluids and to analyse this problem in a more reliable way. One component in each phase must be calculated using a constraint component in the same way as for single-phase multicomponent flow.

For a multicomponent fluid, additional equations must be solved to determine how the components of the fluid are transported within the fluid.

The variation of the properties for a single component influence the global system. Each component has its own equation for conservation of mass. After Reynolds-averaging can be expressed in tensor notation as:

$$\frac{\delta \tilde{\rho}_i}{\delta t} + \frac{\delta\,(\tilde{\rho}_i \tilde{U}_j)}{\delta x_j} = -\frac{\delta}{\delta x_j}\left[\rho_i\left(\tilde{U}_{ij} - \tilde{U}_j\right) - \overline{\rho_i'' U_f''}\right] + S_i \qquad (1)$$

where:

$\tilde{\rho}_i$ is the mass-average density of fluid component i in the mixture

$\tilde{U}_j = \sum \frac{(\tilde{\rho}_i \tilde{U}_{ij})}{\tilde{\rho}}$ is the mass-average velocity field,

$\tilde{U}_{ij}$ is the mass-average velocity of fluid component i,

$\rho_i(\tilde{U}_{ij} - \tilde{U}_j)$ is the relative mass flux,

S_i is the source term for component i which includes the effects of chemical reactions.

The physical properties of general multicomponent mixtures are treated using the assumption that the components form an ideal mixture, i.e. a mixture of components such that the properties of the mixture can be calculated directly from the properties of the components and their proportions in the mixture.

Thus, the mixture density ρ may be calculated from the mass fractions Yi and the thermodynamic density of each component ρi, which may require knowledge of the mixture temperature and pressure, as well as an appropriate equation of state for each component.

$$\frac{1}{\rho} = \sum_{i=A,B\ldots}^{N_C} \frac{Y_i}{\langle\rho_i\rangle} \qquad (2)$$

Extending the Reynolds-averaged conservation equation for energy of a single component fluid for multicomponent fluids involves adding an additional diffusion term. In the special case that all species diffusivities are the same and equal to thermal conductivity divided by specific heat capacity, the energy equation is the following:

$$\frac{\delta}{\delta t}(\rho H) - \frac{\delta P}{\delta t} + \frac{\delta}{\delta x_j}(\rho U_J H) = \frac{\delta}{\delta x_j}\left[\left(\frac{\lambda}{C_p} + \frac{\mu}{Pr_t}\right)\frac{\delta h}{\delta x_j}\right] + S_E \qquad (3)$$

This equation has the advantage that only a single diffusion term needs to be assembled, rather than one for each component plus one for heat conduction. This can significantly reduce numerical cost, in particular when the fluid consists of a large number of components. The turbulence model adopted was the Standard k-ε. The k-ε turbulence models is one of the most used to solve this kind of problems (Ferziger and Peric, 2002; Margaris and Ghiaus, 2006; Bottani et al, 2008). It is part of the Reynolds Averaged Navier-Stokes models (RANS), which consider the average time of the speed to which add terms of fluctuation. In particular, the k-ε is a model with two equations, which means that it includes two additional equations to the classical ones to represent the properties of the turbulent flow.

The fluid domain of the pouch was obtained using ICEM CFD, the modeller associated with ANSYS CFX. The volume of the pouch is divided into a finite number of volumes on which the analysis is carried out. Figure 2 shows the generated mesh used for the calculation: an unstructured tetrahedral meshing scheme was used.

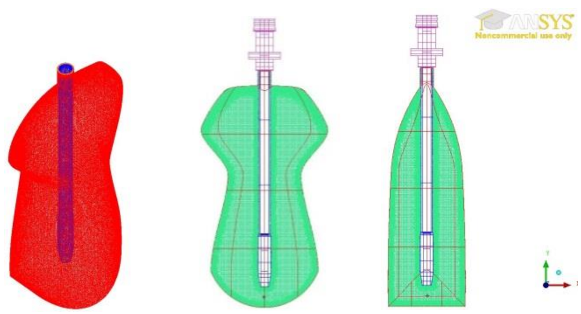


Figure 2. Pouch mesh (for the surface and the volume)

The number of cells used in the simulations was determined starting from a coarse meshing gradually refined, evaluating the changes in the results. In particular, a finer mesh was used near the outlet section of the nozzle, where it is foreseeable that the shear rates would be higher and close to the wall of the pouch in order to simulate accurately the flow boundary layer. The final mesh was determined when increasing the fineness of the mesh there were no significant improvements in the results. The overall number of cells created is about 4,100,000.

2.2.2 Optimisation Modelling

The optimization modelling is the most important part of the process, because it is necessary to specify the optimization strategy that will be used. ModeFrontier software (version 4.4.2) has been adopted for this study.

Optimization process starts with the workflow creation (Figure 3).

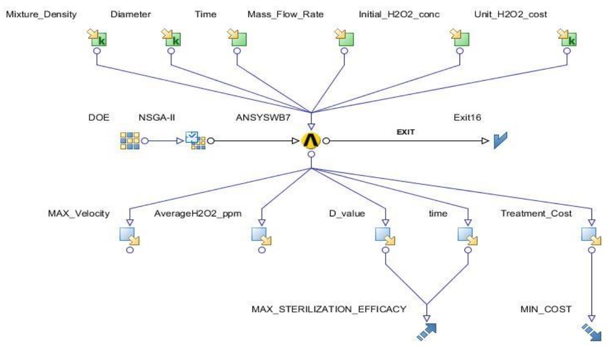


Figure 3. ModeFrontier workflow

The workflow is composed by a process flow and a data flow. The process flow describes the sequences of actions and the data flow describes which data should be moved from one step to another.

First of all, the input variables of the system were defined. In particular, six input were individuated: three of them like a constant and the other three like a variable. In this work, input like the mass flow rate, the initial concentration of H_2O_2 and the time of the sterilization process were kept variable in order to assess the optimal configurations in according to the variation of these particular three inputs.

All the input parameters are defined by dedicated nodes, which specify their range of variation, and they are all linked to the application node. In this work, the Ansys Workbench node was implemented. Table 1 shows the input name and the relative range of values.

Input variables	Type	Range of Values
Mixture density	Constant	1.18 [kg/m^3]
Diameter nozzle	Constant	0.005 [m]
Unit cost H_2O_2	Constant	0.57 [€/kg]
Mass flow rate	Variable	[0.1 l/s; 0.5 l/s]
Initial H_2O_2 conc	Variable	[0.0015; 0.0050]
Time	Variable	[5 s; 10 s]

Table 1. Input variables

Hence the outputs of the system were set. In the model, the outputs have been selected in order to optimize some of them. In particular, the maximum velocity, the final concentration of H_2O_2 expressed in ppm, the treatment cost for each bottle and the treatment efficacy were calculated (Table 2).

Output variables	
Max velocity	Mass flow rate) / (density · surface)
Final average PPM	Ave(H_2O_2.Mass Fraction)·1000000
Treatment cost	Unit Cost · Mass flow rate · Time
Decimal reduction value (D)	4901,5·(AveragePPM)^-1.069
Efficacy treatment	Time/6D

Table 2. Output variables

About the microbial inactivation some studies in literature were already performed (Wang and Toledo, 1986, Malik et al, 2013). With reference to these studies, the Weibull model provided the best fit, and its use was extended to produce a correlation yielding D values over a range from 10 to almost 4000 ppm (D.J. Malik et al, 2013).

In this work, the D parameter was fixed using the Weibull model. In this context, the model accounts for biological variation with respect to inactivation times. The following form of the model was applied here:

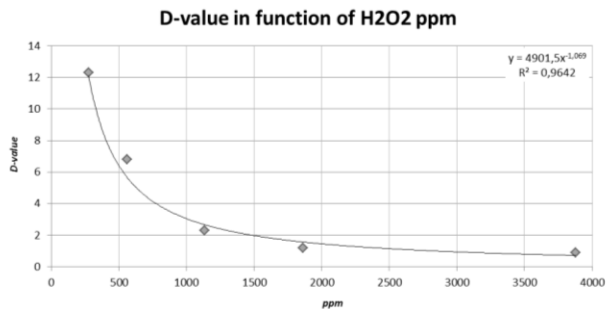
$$\frac{C}{C_0} = 10^{-(\frac{t}{D})^p} \qquad (4)$$

where the parameter p is commonly referred to as the 'shape parameter', and D is the decimal reduction value. Wang and Toledo (1986) had examined the inactivation of *B. Subtilis* spores by hydrogen peroxide vapour at concentrations in the range 275-3879 ppm. Table 3 reports D and p values for the Weibull model at high concentration of VHP.

Weibull		
ppm	***p***	***D-value***
275	3.36	12.3
558	1.95	6.8
1131	2.60	2.3
1859	3.33	1.2
3879	3.62	0.9

Table 3. Weibull model inactivation

A power-law regression model describes the hydrogen peroxide concentration dependency of the decimal reduction values as show Graph 1.



Graph 1. D-value in function of H2O2 ppm for the Weibull model

On the basis of these studies it was possible to determinate the D-value in function of the H_2O_2 concentration and to obtain the efficacy treatment parameter.

As already specified, between the input and the output variables, it is necessary to insert the Workbench application node. This part connects the input and output variables. To complete the process flow, the objectives were set. In this work the main objectives to be achieved are two:

- Maximize the efficacy sterilization treatment to exceed six decimal reductions.
- Minimize the costs consumption of H_2O_2.

For the first one, the ratio between the simulation time and the relative D-value was calculated. For the second one, the multiplication among the unit cost of H_2O_2, the mass flow rate of H_2O_2 and the time of the sterilization process was calculated. To complete the data flow, and the entire workflow, ModeFrontier offers many types of modern optimization algorithms, whose specificities are adapted to different optimization strategies.

The correct number of design and the appropriate scheduler for the simulation were individuated in according with the number of the input and output variables and the number of the objectives (Table 4).

DOE method		Optimization algorithm	
Sobol	9 designs	*NSGA-II*	10 generations

Table 4. ModeFrontier scheduler

Sobol DOE method creates sequences of "n-tuples" that fill the "n-space" more uniformly than random sequence. This type of sequence are called quasi-random sequences (Chi et al, 2005). That term is somewhat of a misnomer, since there is nothing random in this algorithm, but filling in a uniform manner the design space.

NSGA-II optimization algorithm is a fast and selective multi-objective (Deb, 2002). A fast non-dominated sorting procedure is implemented. Sorting the individuals of a given population according to the level of non-domination is a complex task: non-dominated sorting algorithms are in general computationally expensive for large population sizes. The adopted solution performs a clever sorting strategy.

Furthermore, NSGA-II implements elitism for multi-objective search, using an elitism-preserving approach. Selectiveness is introduced storing all non-dominated solutions discovered so far, beginning from the initial population. Selectiveness enhances the convergence properties towards the true Pareto-optimal set.

3. RESULTS

3.1 Pareto frontier

ModeFrontier software simulated 90 configurations. In the figure and table below, the Pareto frontier, obtained thanks to all of them, is shown. Also Scatter charts have been created.

A Scatter chart is a two by two quantity chart, it reveals relationships or associations between variables. The values of the variables selected specify the X and Y coordinates of the design.

The first scatter chart (Figure 4) correlates the two objectives: maximize the efficacy sterilization treatment [x] with minimize the costs consumption of H_2O_2 [y].

This case considers two conflicting objectives. From figure 4 it can be observed two different areas: the lower right area shows the maximum in terms of efficacy of treatment but not in terms of the minimum treatment cost. The second area, located in the lower left part of the graph, shows the configurations that reach the minimum cost but not the maximum treatment efficacy. In all the following figures, the green line represents the regression line for the all ID designs.

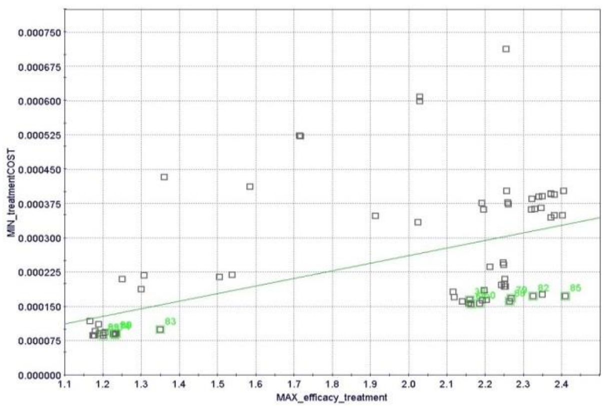


Figure 4. Scatter chart between objectives

The second scatter chart (Figure 5) correlates the average H_2O_2 concentration [x] with the D-value [y].

Increasing the average H_2O_2 concentration, the D-value decrease.

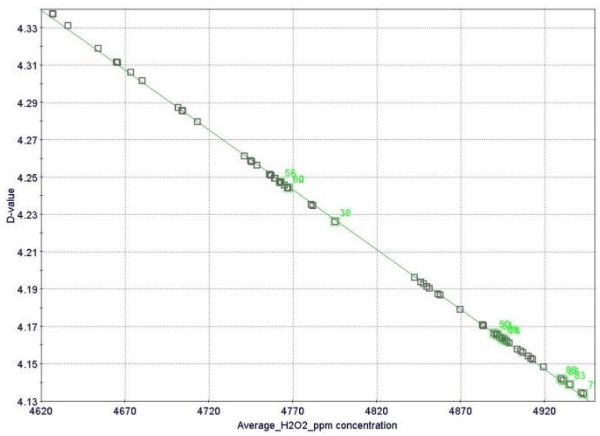


Figure 5. Scatter chats between average H2O2 concentration and D-value

A Scatter 3D chart displays points at the locations specified by the 3-dimensional vectors X, Y, and Z. It reveals relationships or associations between variables.

In the Figure 6, the two objectives and the time of the process were analyzed.

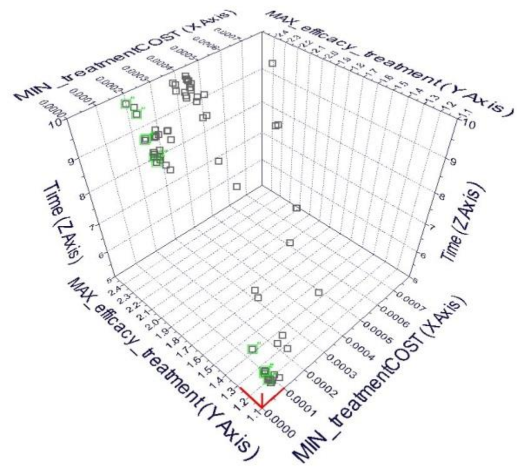


Figure 6. Scatter 3D between two objectives and time

Also the scatter 3D shows the two areas previously described. Furthermore, this 3D graph reports how the two objectives are influenced from the time variable.

The parallel coordinates chart is a method of displaying multivariate data. In this type of chart, a set of parallel axes are drawn for each variable. Then a given set of data is represented by connecting the value of each data on each corresponding axis. The parallel chart is useful to quickly evaluate designs whose variables are in a particular range.

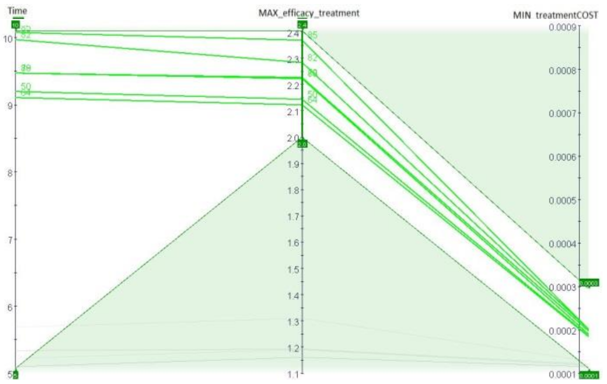


Figure 7. Parallel history of the 13 Pareto designs

Figure 7 shows the Pareto designs with efficacy treatment higher than 2 and treatment cost below 0.00030 € per bottle.

When the sterilizing effect is equal to 1, six decimal reductions are achieved. For some applications this value can be considered a sufficient sterilizing effect. Though, this work purposes to maximize the sterilization level consistently with the other variables. For this case a MCDM method was applied.

MCDM (Multiple Criteria Decision Making) is a post-processing tool which helps the user to make selections of best designs from a family of Pareto solutions. It is extremely useful when dealing with several conflicting objectives.

3.2 Optimal design and MCDM application

The previous results define a clear Pareto frontier and will be used to determine the optimal sterilization process. Among the 90 designs, 11 of them are true Pareto designs, meaning that for these 11 designs one cannot decrease the cost of treatment without decreasing the efficacy of treatment and vice versa. In other words, without performing a multi-objective analysis, one can only treat these designs as equally optimal. Table 5 summarizes the 11 optimal designs including the cost of treatment, the efficacy of treatment, the average H_2O_2 ppm, the D-value, the time and the design ID.

Design ID	Efficacy treatment	Treatment cost	Time	D-value	Average H_2O_2 ppm concentration
50	2.19	1.56E-04	9.10	4.166	4890
56	1.23	0.91E-04	5.23	4.247	4762
64	2.16	1.54E-04	9.01	4.165	4892
65	1.20	0.85E-04	5.00	4.163	4895
74	1.23	0.88E-04	5.13	4.163	4896
79	2.27	1.68E-04	9.37	4.134	4943
80	1.23	0.91E-04	5.23	4.244	4767
82	2.32	1.71E-04	9.87	4.244	4767
83	1.35	0.99E-04	5.59	4.139	4935
85	2.41	1.72E-04	9.98	4.141	4931
88	2.26	1.60E-04	9.37	4.142	4930

Table 5. Resume of 14 optimal designs

These 11 designs are all optimal designs because for each of them it is not possible to reduce treatment cost and increase the sterilization efficacy simultaneously.

A possible way is to apply an MCDM tool that enables one to find the best solution among a set of reasonable alternatives. Table 6 and Figure 8 show the relationships between the variables.

Attribute 1	Type	Weight	Attribute 2
Average H_2O_2 PPM	>	2.0	TreatmentCOST
Average H_2O_2 PPM	=	1.0	Time

Table 6. Relationship between variables

As described before, this case provides the optimum configuration that ensures the optimal sterilizing effect.

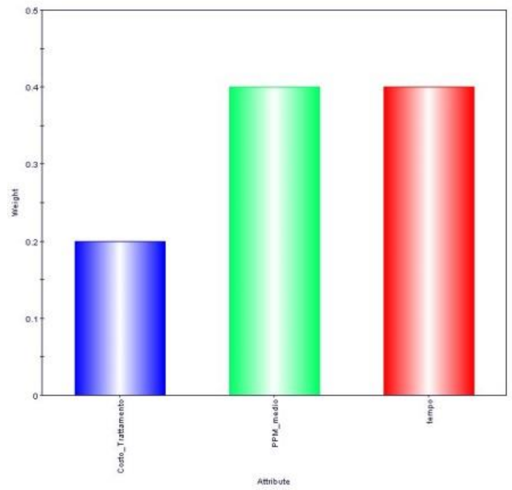


Figure 8. Attributes and weight chart

In respect of the weights previously set, the optimum design can be seen in Figure 9.

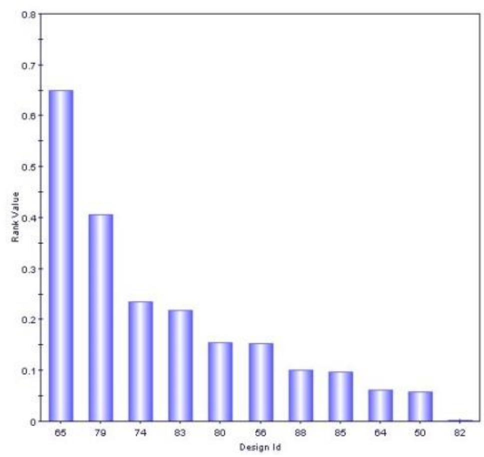


Figure 9. Rank value and Design ID charts

The greatest rank value represents the optimal ID configuration (Table 7).

ID	Treatment cost	Average H_2O_2 ppm concentration	Time	Rank Value
65	0.85E-04	4895	5.00	0.649
79	1.68E-04	4943	9.37	0.406
74	0.88E-04	4896	5.13	0.235
83	0.99E-04	4935	5.57	0.218
80	0.91E-04	4767	5.23	0.154
56	0.91E-04	4762	5.23	0.153
88	1.60E-04	4930	9.37	0.100
85	1.72E-04	4931	9.98	0.096
64	1.54E-04	4892	9.01	0.061
50	1.56E-04	4890	9.10	0.058
82	1.72E-04	4867	9.87	0.002

Table 7. Ranking Value

Among the Pareto frontier, the design 65 represents the best configuration able to reach the optimal treatment efficacy possible according to the other objective and the others variables.

4. CONCLUSIONS

Aseptic filling technology of Spouted Pouch packaging requires a very complex sterilization and rinsing of the packaging before filling.

The aim of this work was to optimize the sterilization process of pouch packaging used in aseptic technology and it has been resolved using a multi-objective optimisation.

ModeFrontier is a multi-objective optimization and design environment which combines a comprehensive process integration platform with sophisticated, optimization algorithms, and powerful post-processing capabilities.

To fully utilize the performance simulation and increase the efficiency of the design, we can introduce optimization algorithms and integrate the simulation program to automatically generate and simulate new designs.

This study goals to optimize the sterilization treatment and its relative cost. A contradiction between these two objectives occurs. By using the MCDM tool, the designer can express his preferences and let the software rank all the optimal designs for selection.

The present study purposes to individuate the optimal sterilization effect able to achieved at least 6 decimal reductions. Decimal reductions pair to 6 means that the sterilizing effect is equal to 1. Basing on the previous considerations, the optimum configuration is the design 65. The parameters of the optimal configuration are: high average H_2O_2 concentration equal to 4895 ppm; short time of the process whose value is 5.00 seconds and low treatment cost equal to 0.85E-04 Euro per bottle.

Future researches could be related to improve the optimization process, adding a SolidWorks node in ModeFrontier in order to change also the nozzle position inside the pouch packaging.

REFERENCES

Abdul Ghani, A.G., Farid, M.M., Chen, X.D., Richards, P., 2001. Thermal sterilization of canned food in a 3-D pouch using computational fluid dynamics. *Journal of Food Engineering 48 (2)*, pp. 147-156.

Abreu L.F., Faria J.A.F., 2004. Evaluation of a system for chemical sterilization of packages. *Packaging Technology and Science, 17, (1)*, pp. 37-42.

Ansari I.A., Datta A.K., 2003. An overview of sterilization methods for packaging materials used in aseptic packaging systems. *Food and Bioproducts Processing, 81, (1)*, pp. 57–65.

Bottani E., Rizzo R., Vignali G. (2008). Numerical Simulation of Turbulent Air Flows in Aseptic Clean Rooms. *Recent Advances in Modelling and Simulation*. pp. 633-650.

Castle, L., Mercer, A.J. and Gilbert, J. 1995. Chemical migration from polyproylene and polyethylene aseptic food packaging as affected by hydrogen peroxide sterilization. *J. Food Prot. 58*, pp. 170–174.

Cerny G., Testing of aseptic machines for efficiency of sterilization of packaging materials by means of hydrogen peroxide, *Pack. Technol. Sci. 5 (1992)*, pp. 77–81.

Chi, H., Beerli, P., Evans, D.W., Mascagni, M., 2005. On the Scrambled Sobol Sequence. *Lecture Notes in Computer Science 3516, Springer*, pp. 775-782.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, vol. 6, n°. 2, pp. 182-197.

Farkas, J. 1998. Irradiation as a method for decontaminating food: A review. *International Journal of Food Microbiology, 44 (3)*, pp. 189-204.

Ferretti, G., Montanari, R., Rizzo, R., Vignali, G., 2006. Ionising Radiation for Food Packaging Sterilisation. *Proceding of 3r Central European Congress on Food*, May 22-24, Sofia, Bulgaria.

Ferziger J.H. and Peric M., 2002. Computational methods for fluid dynamics, *Springer-Verlag*, Berlin/Heidelberg, Germany.

Jun, S., Sastry, S., 2007. Reusable pouch development for long term space missions: A 3D ohmic model for verification of sterilization efficacy, *Journal of Food Engineering, 80,* pp. 1199–1205.

Kirchner, P., Oberländer, J., Suso, H.P., Rysstad, G., Keusgen, M., Schöning, M.J., 2013. Monitoring the microbicidal effectiveness of gaseous hydrogen peroxide in sterilisation process by means of a calorimetric gas sensor. *Food Control 31*, pp. 530-538.

Klapes, N.A., Vesley, D., 1990. Vapor-phase hydrogen peroxide as a surface decontaminant and sterilant.

Applied and Environmental Microbiology, 56 (2), pp. 503-506.
Malik, D.J., Shaw, C.M., Rielly, C.D., Shama, G., 2013. The inactivation of Bacillus subtilis spores at low concentrations of hydrogen peroxide vapour, *Journal of Food Engineering 114*, pp. 391-396.
Margaris D.P. and Ghiaus A.G., 2006. Dried product quality improvement by air flow manipulation in tray dryers, *Journal of Food Engineering 75*, pp. 542–550.
Qian Zou, Linus, U., Opara, McKibbin, R., 2006. A CFD modelling system for airflow and heat transfer in ventilated packaging for fresh foods: I. Initial analysis and development of mathematical models. *Journal of Food Engineering 77*, pp. 1037-1047.
Riganakosa K.A., Kollerb W.D., Ehlermannb D.A.E., Bauerb B. and Kontominasa M.G., 1999. Effects of ionizing radiation on properties of monolayer and multilayer flexible food packaging materials. *Radiation Physics and Chemistry, 54, (5)*, pp. 527-540.
Robertson G.L., 2006. Food Packaging: Principles and Practice, *Second Edition CRC Press*, Boca Raton FL (USA).
Shao-Ping Wang, Mark, H., Anderson, Jason, G., Oakley, Michael, L., Corradini, Bonazza, R., 2004. A thermodynamically consistent and fully conservative treatment of contact discontinuities for compressible multicomponent flows. *Journal of Computational Physics 195*, pp. 528-559.
Wang, J., Toledo, R.T., 1986. Sporicidal properties of mixtures of hydrogen peroxide vapour and hot air. *Food Technology 40*, pp. 60-67.

AUTHORS BIOGRAPHY

Matteo FOLEZZANI is scholarship holder in Industrial Engineering at the Interdepartmental Center CIPACK. In 2011 he has achieved a master degree in Mechanical Engineering for the Food Industry, discussing a thesis titled: "Analysis and simulation of innovation system of sugar dissolution for the food industry". He attended the 2011 EFFoST Annual Meeting (Berlin 9-11 November 2011), with a poster presentation titled “Analysis and simulation of a powders dissolution system based on the hydrodynamic controlled cavitation.”. His main fields of research concern food process modeling and simulation, with a particular focus on the CFD simulation for the advanced design of food plants.

Michele MANFREDI is a PhD Student at the University of Parma, and Scholarship Holder in Industrial Engineering at the Interdepartmental Center CIPACK. He has achieved a master degree in mechanical engineering of food industry, discussing a thesis titled: “Analysis of the sterilization process of pouch packaging in an aseptic line”. His main fields of research concern analysis LCA (Life Cycle Assessment) of food process and food products.

Giuseppe VIGNALI graduated in 2004 in Mechanical Engineering at the University of Parma. In 2009, he received his PhD in Industrial Engineering at the same university, related to the analysis and optimization of food processes. Since August 2007, he works as a Lecturer at the Department of Industrial Engineering of the University of Parma, and, since the employment at the university, he has been teaching materials, technologies and equipment for food packaging to the food industry engineering class. His research activities concern food processing and packaging issues and safety/security of industrial plant. Results of his studies related to the above topics have been published in more than 50 scientific papers, some of which appear both in national and international journals, as well in national and international conferences.

MULTI-PERSPECTIVE VISUALIZATION WITHIN SIMULATION-BASED DESIGN ENHANCED SHARED UNDERSTANDING

Bart van Zaalen[a], Rens Kortmann[b], Wijnand W. Veeneman[c] Alexander Verbraeck [d] Marten Busstra[e]

[a,b,c,d]Delft University of Technology - Faculty Technology, Policy and Management
[e] NedTrain B.V.

[a]b.van.zaalen@gmail.com, [b]L.J.Kortmann@tudelft.nl , [c]W.W.Veeneman@tudelft.nl,
[d] A.Verbraeck@tudelft.nl, [e]Marten.Busstra@nedtrain.nl

ABSTRACT

Within complex design projects the methodology of Simulation-Based Design (SBD) is regularly used. In the literature critical remarks have been made on the lack of multiple perspectives within such simulations, which creates inefficiency and dissatisfying results of the design. The use of multi-perspective visualization during an SBD process may give participants a better insight in the impact of the design on their own and others' interests. As a result it may create a higher shared understanding (SU) among participants and improve the design. We examined this hypothesis during a design case for a new shunting plan on a marshalling yard, which led to the following conclusion: the addition of multi-perspective visualization enhanced SU among the participants significantly and contributed to a better design result. For similar design cases this approach is expected to be successful too, although small adjustments to the approach and other types of visualizations will be required.

Keywords: Simulation-Based Design, Shared Understanding, multi-perspective visualizations, marshalling yard

1. INTRODUCTION

Simulation-Based Design (SBD) is used as a method to support the design of complex systems. This method is experienced to be successful because it enhances shared understanding among actors involved in the design process. Shared Understanding (SU) is of great importance since large technological systems within an environment with a lot of actors are hard to manage due to its complexity, interaction between those actors and their uncertain behaviour (Xia and Lee 2005). Designing a new system in these kind of environments requires SU, defining the problem and solving it through a process of finding a satisficing solution (Simon 1996). In order to increase the quality of design it is therefore also very important to create a high level of SU (Piirainen, Kolfschoten and Lukosch 2000).

Within an SBD process, the tool of simulation is used to solve challenges and meet requirements in a multi-actor environment concerning a complex system (Hengst, Vreede and Maghnouji 2007). Within this methodology the two systems approaches are combined. Hard systems thinking is the approach for the simulation of systems of which a current and desired state are taken for granted and the problem of the system to be designed is structured. Soft systems thinking is the approach for ill-defined and unstructured problems and of which the design process is not goal oriented (Robinson 2001). Simulation is used as a tool to combine these approaches and creates a lot of opportunities; higher acceptance of outcomes, increased shared understanding, better stakeholder involvement, higher quality of the model and its use (Fumarola 2011, Hengst, Vreede and Maghnouji 2007).

To make advantage of these opportunities several frameworks have been developed to structure an SBD process in which the multi-actor design processes leave more room for negotiation, mutual learning and aim for the creation of a higher shared understanding (Huang, Seck and Fumarola 2012).

However, the SBD process still have its limitations. Evaluation of the design process led to the discussion of the actual contribution to a higher level of SU and in the end a higher quality of design (Fumarola et al. 2010). Fumarola et al. conclude that a lack of multiple perspectives exists within SBD processes which can lead to unintended results of the design and opportunities are not used optimally. Just simulating and visualizing from a single perspective reduces important information about the reality since actors try to have intuitions from a single perspective simulation. Important information can be neglected, which is critical to get a better understanding of the system (Bürgi and Roos 2003).

The simulation within an SBD process should therefore be developed with multiple perspectives, so each actor can identify himself with the system and to resolve the limitation encountered within SBD. The enhancement of SU is the main objective in this case.

An experiment has been executed to test whether or not the addition of multiple perspectives within an SBD process creates a higher level of SU. The main research question for this experiment was: *To what extent does the addition of multi-perspective visualization contribute to an enhanced shared understanding in the multi-actor simulation based design process for a logistic process design on a marshalling yard?*

A design for a new shunting plan at a marshalling yard of NedTrain has been used as a case to examine the effect of multi-perspective visualization on the level of SU. The method of case study research has been used because this method gives the opportunity to develop and test new theories within a realistic environment (Yin 2003).

In the next section the construct of SU is explained. The experimental setup, creating multiple perspectives within the simulation and description of the case are discussed in the third section. Results of this research are given in section four, followed by the conclusions and discussion in the final section.

2. SHARED UNDERSTANDING

Shared Understanding (SU) is a conjoined term for the mutual knowledge, beliefs and assumptions by a group of actors. The amount of overlap in understanding and concepts of the particular system of study among actors can be seen as the level of SU (Mulder, Swaak and Kessels 2002). Different actors state that the creation of SU will lead to a better performance of business processes within a multi-actor environment (Bondar, Katzy and Mason 2012; Zhao, Yunfeng and Xiaochun 2009). As Mulder denotes; "..shared understanding facilitates working and interacting effectively and efficiently. Interacting effectively and efficiently is possible when the group members use the same symbols and assign the same meanings to those symbols in their interaction processes." (Mulder 1999, p. 1).

Through interaction between actors the SU is affected. During interaction actors exchange information which can be used to create SU. Therefore SU is not on a fixed level, but is always on-going through the interactions between actors (Mulder 1999).

Mulder denotes three aspects of SU; social relation, content and process. During interaction, so also during design processes, actors should have SU on these three aspects (Mulder 1999). Together, this creates an overall level of SU which is important in interaction processes like for example a design process. The aspect of social relation is about who is communicating messages and in what way. Messages from different persons can be the same, but the interpretation by others can differ a lot because of non-verbal behaviour. Interaction about the content should frame the problem so all group members have the same meaning of the problem and the problem area; 'what' are they working on. The third aspect is the process related understanding, for which actors should have the same way of communication, structure of interaction (protocols) and understanding of roles within an actor field. Actors should have a SU on how to work together (Mulder 1999).

Literature and quantitative tests to measure the level of SU are exceptional. However, Mulder developed a quantitative test to measure the level of SU on the three different aspects identified and on the overall level of SU. The test is a questionnaire which has to be filled in by participants of a design or decision making process. They are asked on each aspect on their perceived SU and on the perceived SU among other stakeholders. Additionally the participants can be asked whether or not they think the SU is enhanced. These questions can be used for a pre-test and post-test, before and after a design or decision making process is completed. This test is very suitable for the experiment which has been executed, since a pre-test and post-test can give insight in the enhanced level of SU.

3. RESEARCH METHOD

The research was based on the Design Science Research theory by Hevner et al. (2004). By the addition of multiple perspectives in the visualization of the simulation during an SBD process the method can be improved, as proposed by Fumarola et al. (2010) and discussed in section 1. To examine whether or not the addition of multi-perspective visualization contributes to an enhancement of SU a case study has been performed, of which the results can be evaluated and justified to draw conclusions for the methodology of SBD (Yin 2003).

First of all the environment in which the case study has been performed will be discussed. The setup for the experiment on the enhancement of SU is explained subsequently, followed by an explanation of the test methods and organization of the design workshop.

3.1. Case NedTrain

An experiment on the enhancement of SU has been performed using a design case at NedTrain, the service and maintenance company of the Dutch train operator Nederlandse Spoorwegen (NS). NedTrain is planning to reorganize its service and maintenance processes. In this process new Technical Centres (TC) are built on four locations scattered around the Netherlands, of which one in Utrecht. The marshalling yard at Utrecht, Cartesiusweg (Ctw), has a remarkable lay-out which can be described best as a bottle and is illustrated in figure 1.

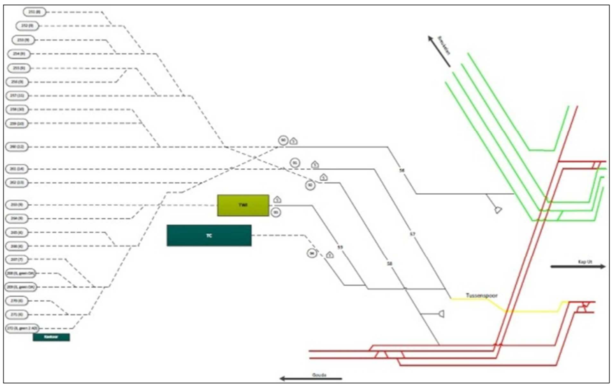


Figure 1: Lay-out marshalling yard Cartesiusweg (Ctw)

The location was already chosen and the tender for the construction project was finished, then thoughts about the risks on the logistical process arose. In figure 1 the location of the TC is the dark green box. Shunting trains from the parking tracks on the left side of the marshalling yard towards the TC creates a lot of problems and conflicts with other train movements, due to the bottleneck at the right side of the marshalling yard. At the same time the requirement for a new shunting plan is stated, which created the opportunity to adjust the shunting process in order to mitigate the risks on the logistic process by the shunting movements to the TC. The design for a new shunting plan including solutions for the accessibility of the TC can be considered as a complex design problem, since the system of Ctw is within a multi-actor environment and its processes and techniques are complex. Therefore this design problem was a good case for the research to be performed on the enhancement of SU within an SBD process.

3.2. SBD approach

To structure an SBD process several frameworks can be found in literature. Especially Fumarola et al. (2010), den Hengst et al. (2007) and Robinson (2001) have developed frameworks to structure the design process of an SBD project. Although these frameworks differ from each other, the combination of soft systems thinking with hard systems thinking is found in these frameworks. Due to practical limitations and the fact that the design approach had to align with a reference case which will be exemplified in next paragraph, not one of these frameworks can be adopted. Elements of both the framework of Fumarola et al. (2010) and of den Hengst et al. (2007) are merged into a specific SBD framework for this case study (Zaalen 2013, p. 45).

A part of this framework is focused on the preparation and execution of a design workshop (figure 2). The execution of the design workshop is marked green in figure 2.

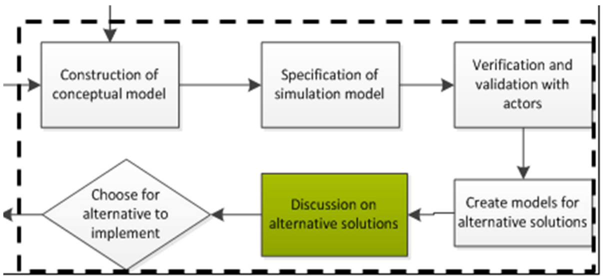


Figure 2: Part of SBD framework focused on the Design Workshop

During this workshop critical actors involved discussed on alternative solutions to implement, in order to improve the logistic process on the marshalling yard Ctw. During this workshop the influence of multi-perspective visualization on the enhancement of SU has been examined.

3.3. Tests on enhancement of SU

The enhancement of SU as a result of multiple perspectives within the simulation during an SBD process could not be measured by just the tool of Mulder. These quantitative tests can only measure the enhancement of SU. First of all the tool of Mulder has been used to draw up a pre-test and post-test as discussed in section 2 and examine whether or not there is an increased level of SU. The questions within the pre-test and post-test could be answered on a scale between 1 and 6 (1 = low level of SU, 6 = high level of SU). The post-test improvement questions could be answered on a 7-point Likert scale (1 = high decrease of SU, 4 = no improvement, 7 = high increase of SU).

To identify whether or not the addition of multiple perspectives had a clear influence on the enhancement of SU an observer of a reference case in which an SBD process has been used for the design of a new shunting plan was used additionally to the test of Mulder. Recently an SBD process has been executed on the marshalling yard Watergraafsmeer (Wgm), near Amsterdam. The environment of this design problem was similar to the case of Ctw and was therefore suitable to use as a reference case. In the SBD project of Wgm they just used a single perspective visualization just for the discussion with involved actors and validation of the design. The project manager of this design project has been invited to join and observe the design process to be able to identify whether or not the multi-perspective visualization in an SBD process lead to a higher SU.

Because there was just a single observer, a third method has been used to draw stronger conclusions on the enhancement of SU and the influence of the multiple perspectives hereon. The third method was a post survey, in which participants of the design workshop were presented a list of propositions about the influence of multiple perspectives in a simulation on the level of SU. For each aspect in the pre-test and post-test the participants were presented two propositions, a negative and a positive one on the influence of multi-perspective visualisation on the enhanced SU. An example of such a propositions is: ‘the use of multiple perspective visualization has led to a higher SU on the logistic processes and problems accordingly.’ The participants could answer on a 7-point Likert scale (1 = totally disagree, 4 = neutral, 7 = totally agree).

3.4. Design workshop

The design process was drawn up according to the design process as followed in the reference case Wgm and to the possibilities of the case Ctw. Due to human resource and time limitations there has been chosen for a one-off design workshop. Within this 3-hour workshop the critical actors discussed on alternatives for the shunting plan, which were already composed before the workshop in consultation with these actors. The workshop was supported by a simulation, since it is an SBD process. The simulation differed from other SBD processes in the amount of perspectives to be visualized. For each actor the most important KPI or KPIs were identified and visualized using the information from the simulation model.

The visualizations to provide information on actors perspectives were drawn up to the possibilities which were limited due to time constraints and development possibilities of the software used. The behaviour of the system is animated with the visualization of the lay-out of Ctw and train movements on the infrastructure. Several perspectives on KPIs have been visualized by graphs and tables. Figure 3 illustrates the setting of the workshop.

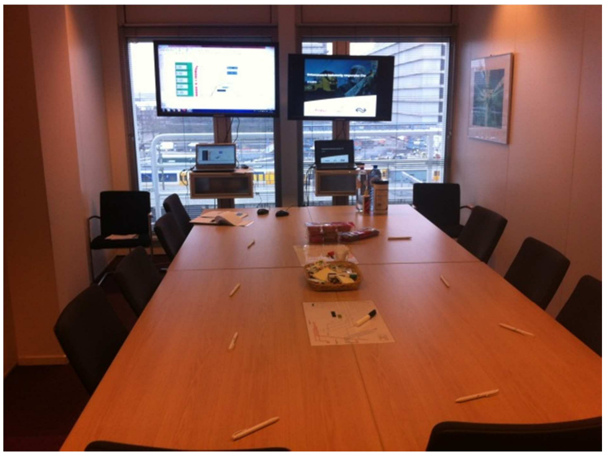


Figure 3: Setting Design Workshop

On the left screen the animation runs for the visualization of system's behaviour and some performance indicators are added for the visualization of system's performance on actors' KPIs. On the right screen a presentation was passed through with the visualization of the performance of the system on more KPIs by graphs and tables.

Within the 3-hour workshop there has been started with the discussion on the performance and behaviour of the system of Ctw in the current situation and future situation with the TC in operation. Subsequently the discussion on the design for a shunting plan including measures to improve the logistic process on Ctw have been discussed.

Upfront the workshop actors were asked to fill out the pre-test and after the workshop ended the post-test. The observer from the reference case Wgm joined the workshop and is consulted a few days after the workshop to reflect on the influence of multiple perspectives within the simulation. The post-survey was filled out by the participants a few weeks after, to identify the influence of multiple perspectives moreover.

4. RESULTS

Subsequently the results on the pre-test and post-test, the survey and the observations will be discussed as described in section 3.3.

4.1. Enhancement of SU

The pre-test and post-test have been filled out by 7 participants. In table 1 the results for both tests are given. During the post-test the participants were also asked about their perceived improvement of their SU. These results are shown in the rightmost column.

Table 1: Results Pre-test and Post-test on level of SU (SR = Social Relation)

	Results			
Question	Pre-test*	Post-test*	Wilcoxon test (p-value)	Perceived improve-ment**
Content 1	4.71	5.43	0.01	5.57
Content 2	3.00	5.00	0.01	5.43
SR 3	3.43	5.29	0.01	5.43
SR 4	3.29	5.00	0.01	5.71
Process 5	3.57	4.89	0.03	5.14
Process 6	3.43	4.43	0.03	5.00

Average scores of participants. N = 7
* range 1-6
** range 1-7

Table 1 shows that the average scores on questions in the post-test are significantly higher than in the pre-test. We used a non-parametric test to account for our small sample size. Moreover the average scores on the perceived improvement are in the range 5.00 – 5.71 out of 7. Our interpretation of these results is that the participants experienced an increased level of SU after the workshop.

4.2. Influence of multi-perspective visualization on enhanced SU

Among the participants of the design workshop an additional survey was held to identify the effect of the multi-perspective visualization on the enhancement of SU.

Table 2: Perceived influence of Multi-Perspective Visualization on SU

Aspect of SU	Average score on propositions*
Content	5.92
Process	5.42
Social Relation	5.04
Overall	5.75

*** (1 = totally disagree, 4 = neutral, 7 = totally agree)**

In table 2 the average scores for each aspect of SU as defined by Mulder (1999) are shown. These scores were calculated from the individual scores on the positive and negative propositions as discussed in section 3.3 (Zaalen 2013, p. 70).

It can be concluded the average scores for the aspects of SU are all higher than 4. This indicates that for each aspect the participants agreed that the multi-perspective visualization contributed to a higher SU.

From the participants' response we derive that the addition of the multiple perspectives in the visualization has a medium to large influence on the level of SU (5.75 out of 7). The results of the post survey show little variation in the scores of the different aspects, ranging from 5.04 to 5.92. Therefore the influence of multi-perspective visualization of the simulation within an SBD process is concluded to be substantial as indicated by the participants of the workshop.

4.3. Observed effect of multiple perspectives

The project manager of the reference case Wgm joined the workshop to observe whether or not the addition of multiple perspectives created a higher level of SU. The project manager was consulted after the design workshop and was questioned during a 2-hour interview on the behaviour of actors and the influence of the multi-perspective visualizations. The following conclusions were drawn from the interview:

Great influence on the level of SU: through the multi-perspective visualization actors got a very good insight in the behaviour and performance of the system, not just for their own interest but for the entire environment. During the discussion on the current performance and behaviour of the system just a few questions arose and all actors indicated to understand the system, its behaviour, performance and the problem to solve.

Structures discussions: Discussions on alternatives were primary based on the visualizations of KPIs. During the discussions the participants often referred to the numbers visualized for the KPIs, comparing them and using this in their argumentation. The observer indicated that actors were convinced more easily by the reference to the visualizations. As she said: 'the participants had a lot of handles to use in their argumentation'. As expected, certain dilemmas arose during the workshop, but they were solved by the insights that the visualizations gave on actors KPIs. In the discussions each actor structured his line of argumentation on the visualizations of the KPIs, pointing out the positive and negative effects for his own values, but acknowledging negative effects for other actors if present. As a final step in their argumentation the actors summarised the effects as perceived by themselves, resulting in their final opinion.

Leads to relevant discussions: The discussions were very substantive and only addressed those aspects which were useful to discuss. While discussing the alternatives, actors encountered other actors' positions in an early stage. Doing so, they made quick progress in the discussion because they could see the impact of the alternatives on other actors' KPIs: if the impact of a particular alternative was very negative to others, actors already took this into account in their argumentation and opinion about the specific alternative. This led to just very useful discussions on the details of alternatives which were acceptable for all actors.

Every type of discussion needs a specific type of visualization: During the design workshop the discussions could be split up into 2 different types; 1. on the current behaviour and performance of the system and identification of the problem and 2. on the alternative solutions. For the first type of discussion the animation of the system of Ctw was much most important. Actors got a very good insight in the behaviour of the system and understood what the problem was. During the discussion on alternatives actors referred a lot to the visualization of KPIs. The animation was not as important anymore and participants even asked if the animation could not speed up a bit because they could not identify the effect of the alternative solution from the animation as shown.

Altogether, the observations led to the conclusion that the addition of multi-perspective visualization within an SBD process creates a more structured and efficient design process and enhances SU among participants.

4.4. Practical implications

Using multi-perspective visualizations in an SBD process lead to a one-off design session in which all actors came to a consensus on the alternative to implement. It gave rise to open discussions that enabled actors to identify the full effects of the alternatives. As a result the final decision to implement a particular alternative was supported by all actors.

Moreover, the design workshop that was organized gave the actors the insight that collaboration leads to better design results. The insight that actors received about other actors' values and interests and the added value of a collaborative design process created the intention for further collaboration.

However, organising a design workshop, in which all actors together discuss on a design for a particular system, is quite difficult. All actors have their own agenda and priorities, which makes it difficult to gather all critical actors within an SBD process simultaneously. These organizational problems concerning the presence of actors can be mitigated by for example video conferencing.

5. CONCLUSION AND DISCUSSION

The experiment on the enhancement of SU by the addition of multiple perspectives within an SBD process has shown that SU is enhanced through the addition of multi-perspective visualization. By the design workshop for a new shunting plan for the marshalling yard Cartesiusweg critical actors participated in a 3-hour workshop, in which an observer from a reference case was present to identify the influence of multi-perspective visualization. Moreover the participants were asked to fill out a pre-test and post-test to identify their perceived SU and a post survey to check whether or not the enhanced SU is caused by the addition of multi-perspective visualization. From the evaluation tools the conclusion is that by the addition of multiple perspectives the SU is enhanced significantly, with a substantial influence of the multi-perspective visualization. From observations it can be concluded that for the SU on the behaviour of the system the visualization by animation has the most impact. For the SU on the performance of the system, the focus is more on the visualization of actors' KPIs. Because actors have insight in the performance of the system on other actors' KPIs the discussion is already discussed for a greater part. Actors already assess others' reactions and take this into account during the discussion.

The case used for the research study is typified by the actor field, in which all actors are dedicated to the logistic process. Furthermore the location and type of railway section, the marshalling yard of Utrecht, are typical for the case study. For similar cases the effects of the addition of multi-perspective visualization are expected to be the same. Design projects focused on logistic processes on a railway network will require other type of visualizations and possibly another software tool to simulate, but the effect of adding more perspectives is expected to be the same. For SBD projects in general the approach with multi-perspective visualization is promising, certainly because the actor field will be more diverse. However, this will require another approach of the SBD process and other types of visualization. To introduce actors with the design project and involve them in a design workshop will take a lot more effort, but will certainly contribute to a higher level of SU and a better design result in the end.

The tool of Mulder has been used in order to measure the level of SU. This tool was not a thoroughly validated tool. Moreover this case study research consists of a single case. Therefore more case studies should be performed to strengthen the conclusions of this experiment. In the end it is concluded there is a substantial influence on the enhancement of SU by the addition of multiple perspectives, however the quantitative extent of this influence is not known yet. Further research on the quantitative extent of influence should be performed to identify to what extent the multiple perspectives contribute to a higher SU.

Finally, it is experienced that the key to success is the openness and willingness to cooperate of critical actors involved. The creation of SU is crucial in whatever design project and can be enhanced significantly by the method of SBD with multi-perspective visualization.

REFERENCES

Bondar, K., Katzy, B.R., Mason, R.M. 2012. Shared understanding in networked organizations. *Proceedings of the 2012 18th International conference on Engineering, Technology and Innovation (ICE),* pp. 1-11. June 18-20, Munich (Germany).

Bürgi, P., Roos, J. 2003. Images of Strategy. *European Management Journal*, 21 (1), 69–78.

Fumarola, M., Huang, Y., Tekinay, C. & Seck, M.D. 2010. Simulation based design for infrastructure system simulation*.* *Proceedings of The 2010 European Simulation and Modelling Conference, Eurosis-ETI,* pp. 288-293. October 25-27, Hasselt (Belgium).

Fumarola, M., Klofschoten, G., Verbraeck, A. & Versteegt, C. 2011. Experimenting with the multiple worlds concept to support the design of automated container terminals. *Proceedings of the 2011 Winter Simulation Conference*, pp 3030-3036. December 11-14, Phoenix (Arizona, USA).

Hengst den, M., Vreede de, G.-J. & Maghnouji, R. 2007. Using soft OR principles for collaborative Simulation: A Case study in the Dutch Airline Industry. *The Journal of Operational Research Society*, 58(5), 669-682.

Hevner, A.R., March, S.T., Park, J. & Ram, S. 2004. Design science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105.

Huang, Y., Seck, M.D. & Fumarola, M. 2012. A simulation-based design framework for large scale infrastructure systems design. *Proceedings of The 11th International Conference on Modelling & Applied Simulation*, pp. 194-201. September 19-21, Vienna (Austria).

Mulder, I. 1999. *Understanding technology mediated interaction processes: A theoretical context*. , Enschede: Telematica Instituut.

Mulder, I., Swaak, J., Kessels, J. 2002. Assessing group learning and shared understanding in technology-mediated interaction. *Educational Technology & Science*, 5 (1), 2002.

Piirianen, K., Kolfschoten, G.L., Lukosch, S. 2010. In search for the right tools to fix the right problem: a look into the challenge of collaborative design. *43rd Hawaii International Conference on System Sciences*, pp. 1-10. January 5-8, Koloa, Kauai (Hawaii, USA).

Robinson, S. 2001. Soft with a hard centre: Discrete-Event Simulation in Facilitation. *The Journal of Operational Research Society*, 52(8), 905-915.

Simon, H.A. 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge, MA: MIT Press..

Xia, W. & Lee, G. 2005. Complexity of information systems development projects: conceptualization and measurement development. *Journal of management information systems*, 22 (1), 45-83.

Yin, R.K. 2003. *Case study research. Design and Methods.* Thousands Oaks, CA: Sage Publications.

Zaalen, van B. 2013. *Multi-actor simulation-based design supported by multi-perspective visualization*. Thesis (MSc). Delft University of Technology.

Zhao, Z., Yunfeng, W., Xiaochun, C. 2009. The Development of Shared Understanding among IS Leadership Team: A Multiple Case Study in China. *International Conference on Information Management, Innovation Management and Industrial Engineering*, pp. 356-359. December 26-27, Xi'an (China).

POLYNOMIAL CONTROLLER FOR AC-DC CONVERTER WITH POWER FACTOR CORRECTION

Jean N. RAZAFINJAKA[(a)], Jean Daniel LAN SUN LUK [(b)]

[(a)]Laboratoire d'Automatique, Ecole Supérieure Polytechnique, Université d'Antsiranana, Madagascar
[(b)]Laboratoire d'Energétique, d'Electronique et des Procédés, Faculté des Sciences et Technologies, La Réunion

[(a)]razafinjaka@yahoo.fr, [(b)]lanson@univ-reunion.fr

ABSTRACT

This paper deals with the control of a AC-DC converter with power factor correction (Boost PFC). The polynomial controller named RST by the three polynomials which constitute it, R(z), T(z) and S(z), is applied to the loop voltage and the loop current is controlled by hysteresis. This kind of controller generalizes the structure of the PID one. Simulation and experimental results show that modelling the open voltage loop by a first order system gives good results supported by a THD satisfying standard IEC even the system presents nonlinearities. Teaching and industrial applications are taken into account: it is the reason of using Labview with the peripheral NI 6009.

Keywords: Boost PFC, RST controller, hysteresis, total harmonic distortion (TDH), Labview

1. INTRODUCTION

Currently, there is apparition of the increased use of the apparatuses, primarily in the informatics fields and in the field of the electric household appliances, requiring supply provided with converter AC-DC using capacitor filters. Although of lower cost, this type of supply generates harmonics in the network. These current harmonics can generate problems for the energy distributor (Feld 2009):

- Increase of line losses
- accelerated ageing of the condensers of compensation because of their low impedance: their rated current may be exceeded
- over sizing of the transformers of distribution

The rate of re-injection of these current harmonics can be quantified by the harmonic rate of distortion TDH.
The power-factor fp is defined by:

$$f_p = \frac{P}{S} = \frac{V.I_1.\cos\varphi_1}{V.I} = \frac{I_1.\cos\varphi_1}{I} \qquad (1)$$

With
S, P, indicating respectively, apparent power, active power
I, I_1,φ_1 : the effective value of the AC current, the effective value of fundamental of current, dephasing enters the tension and the fundamental current. The effective value of current is:

$$I = \sqrt{\left(\sum_{k=1} I_k^2\right)} = \sqrt{I_1^2 + \sum_{k=2} I_k^2} \qquad (2)$$

I_k , harmonic of current of rank k
The expression of the THD is also defined as:

$$TDH = \sqrt{\left(\frac{I_2}{I_1}\right)^2 + \left(\frac{I_3}{I_1}\right)^2 \ldots} = \frac{1}{I_1}\sqrt{\sum_{k=2} I_k^2} \qquad (3)$$

So, according to these three relations:

$$f_p = \frac{\cos\varphi_1}{\sqrt{1+TDH^2}} \qquad (4)$$

The power-factor fp is thus related to the harmonic rate of distortion TDH. It means that this TDH may be an adapted parameter to quantify the harmonic degree of pollution on the network. In all that follows, it will be taken as index of comparison (in practice TDH expressed in % is used).
With a purely sinusoidal fundamental current and in phase with the voltage, a power-factor approaches the unit value (fp = 1).

Figure 1 and Figure 2 show respectively the current and voltage waveforms as well as the output voltage for a traditional rectifier (C = 470 [µF] and R = 328 [Ω]) and the resulting of current spectrum.

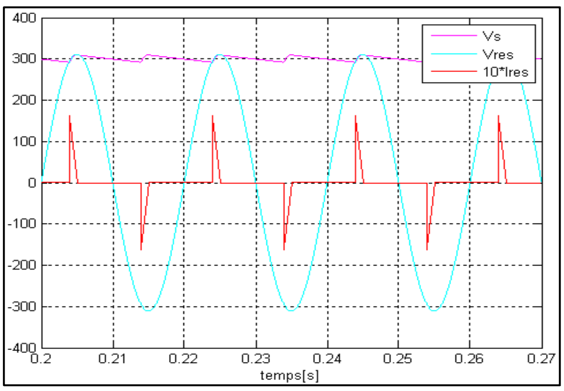


Figure1: Current and voltages waveforms

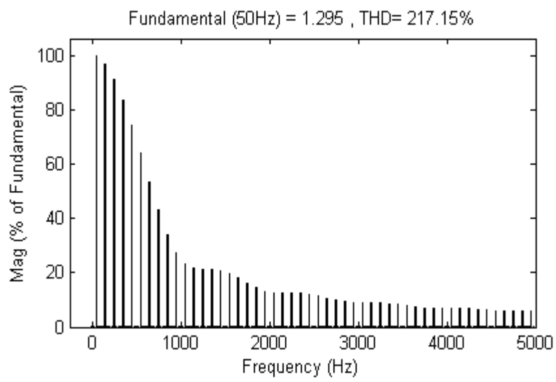


Figure 2: Current spectrum resulting

To bring solution of this problem, various strategies are proposed whose principal objectives are summarized as follows (Benaïssa 2006), (Razafinjaka 2008),(Tédjini 2008), (Singh 2003), (Keraï 2003):

- Obtaining a sinusoidal current network and in phase with the tension
- Or ensuring the smallest possible TDH in order to respect the standard normalizes: IEC-61000-3-2 for example for the systems of class D.
- Ensuring a voltage output constant

The generalized structure is shown in Figure 3.

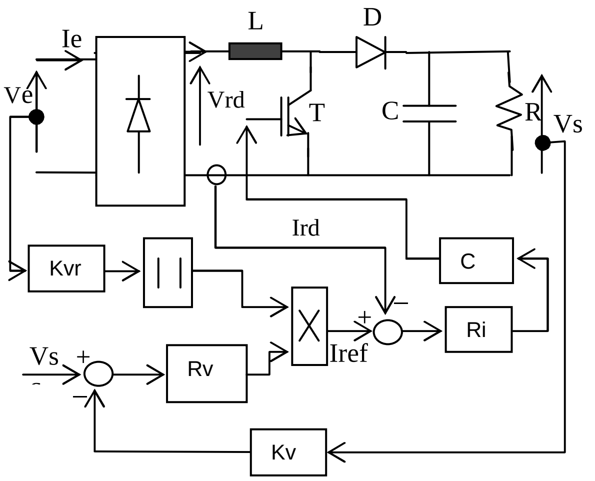


Figure 3: Structure for boost PFC

The existence of two loops is highlighted. The reference of the current Iref is obtained by multiplying the output voltage regulator by a party (K*Vrd) of rectified voltage. The output current regulator is treated in a shaping circuit CMF to obtain the command u(t) used to control the static inverter CS.

In this paper, the structure using polynomial RST for the loop tension and hysteresis control for the loop of current are used. First the loop of current is studied to obtain some conditions having a perfect loop in comparison with loop voltage. A structure using regulator PI is first used for the voltage loop to compare results. Modeling is then necessary for the synthesis of this kind of regulator. The model obtained is used when RST controller, primarily a numerical type, is applied.

2. STRUCTURE WITH PI CONTROLLER

In this case, voltage loop is controlled by a PI regulator and the current loop by a hysteresis (Figure 3).

2.1. Current loop

The control by hysteresis is selected for the current loop because the nonlinear model of the static inverter is. However, it is necessary to express the quench frequency in order to establish a dimensioning of inductance L.

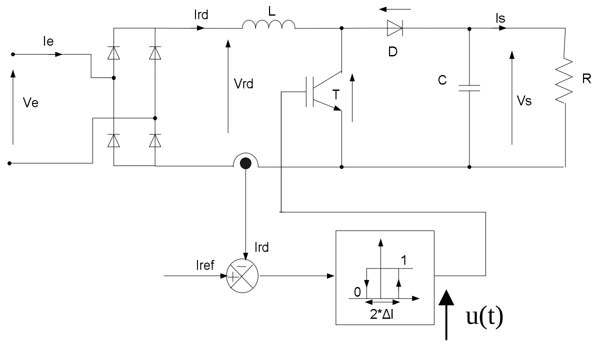


Figure 4: Control current scheme

The set value of current Iref (see Figure 4) must be in phase with the tension. It is also needed to ensure Iref ≈ Ird (Feld 2009): a fast variation of Ird around its reference Iref must be then satisfied which implies a high chopping frequency (Fd >10[kHz]). A value of inductance L according to the undulation of current ΔI must be so determined for this purpose. The value of the output voltage V_s and the effective value V_{rd} are considered as constant. When the variation of Ired around its reference Iref is supposed obtained, the output voltage Vs, and the effective value Vrd are considered as constants (Multon 2003), (Feld 2009). The expression is given by:

$$F_d = \frac{1}{T_d} = \frac{Vrd(Vs - Vrd)}{2.L.\Delta I.Vs} \tag{5}$$

The figure 5 shows curves giving Fd according to the inductance L for imposed ΔI. In this case, Vs = 400 [V], Vrd = 235 [V], ΔI = ± 0,1[A], ±0,2 [A], ± 0,3[A]

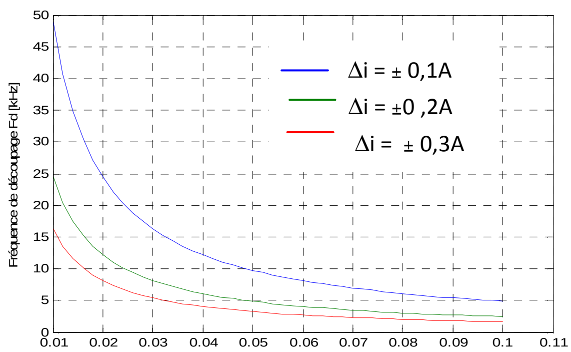


Figure 5: Curves giving Fd according L

2.2. Voltage loop

The voltage loop gives the signal which will act to the reference current Iref. This current must have an sinusoidal form conformed with the network voltage.

It is assumed that current loop is faster than the voltage one and every time Ired = Iref. It is there possible to adopt the following approximation obtained by modelling by assessment of power (Feld 2009):

$$\frac{V_s(p)}{I_{red}(p)} \simeq \frac{V_s(p)}{I_{ref}(p)} \qquad (6)$$

So

$$\frac{V_s(p)}{I_{red}(p)} = \frac{V_M}{4.V_S} \times \frac{R}{1+\frac{RC}{2}p} \qquad (7)$$

Where, VM is the peak value of the network voltage, Vs the value of the output voltage, R represents the load resistance and C the capacitor.

The opened loop is defined by a first order t function. (8)

$$G(p) = \frac{K}{1+pT}$$

With $K = \frac{V_M.R}{4.V_S}$ and $T = \frac{RC}{2}$ (9)

A PI regulator is sufficient to control such system. Its function transfer may be expressed like followed:

$$G_R(p) = \frac{1+p.AT_i}{pT_i} \qquad (10)$$

The gain A and the constant time Ti can be determined by imposing a frequency Fc for the closed loop. Assuming that the transfer function for opened loop Go(p) is:

$$Go(p) = G_R(p).G(p) = \frac{1+pAT_i}{pT_i}.\frac{K}{1+pT} \qquad (11)$$

It is possible to pose by method compensation:

$$A.Ti = T \qquad (12)$$

The transfer function for closed loop is:

$$H(p) = \frac{Go(p)}{1+Go(p)} = \frac{1}{1+p\frac{T_i}{K}} \qquad (13)$$

Imposing frequency Fc according the relation (13) gives Ti and then the gain A by relation (12).

Figure 6 shows the simulation results when frequency at closed loop is Fc= 5 [Hz] and FC=15[Hz]. Current waveform at steady state and the spectrum analysis at the frequency Fc = 5 [Hz] are presented respectively in Figures 7 and 8.

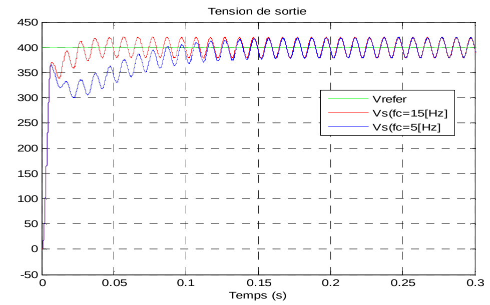


Figure 6: Step responses of Vs at Fc=5[Hz] and Fc=15[Hz]

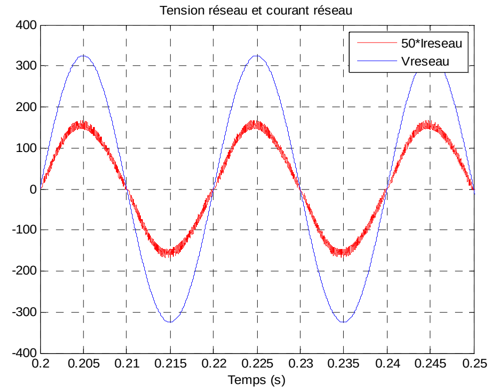


Figure 7: Current network waveform (Fc = 5[Hz])

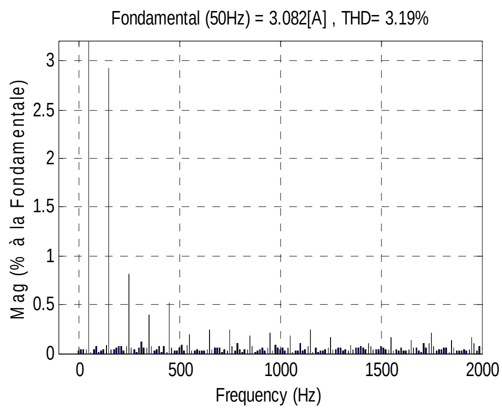


Figure 8: Current spectrum (Fc = 5[Hz])

Following conclusions can be expressed:

- The current is sinusoidal and in phase with the voltage
- More the frequency loop is higher more the regulation is faster but the harmonic rate distortion of the current is higher
- F_c= 5 [Hz] TDH = 3,19%
  F_c = 15 [Hz] TDH = 7,62 %

3. STRUCTURE WITH RST CONTROLLER

In this case, numerical regulation is applied for the loop voltage because RST controller is primarily numerical. The current loop is always controlled by hysteresis command. Using RST controller, the functional scheme is given by Figure 9. The basic idea is to search the

Proceedings of the International Conference on Modeling and Applied Simulation, 2013
978-88-97999-23-2 Affenzeller, Bruzzone, De Felice, Del Rio, Frydman, Massei, Merkuryev, Eds.

three polynomials R(z), S(z) and T(z) to have correspondence. The closed transfer function $H_m(z)$ is given: it represents the desired model.

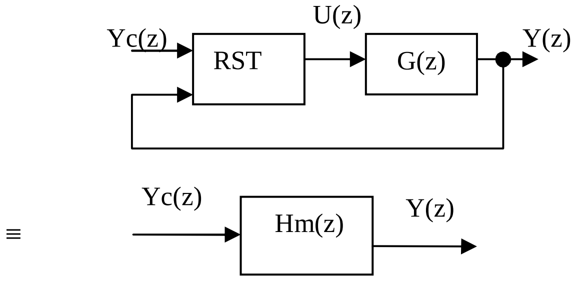


Figure 9: Basic schemes for RST calculation

The RST controller generalizes the law of command obtained by a classic one.

$$R(z).U(z) = T(z).Yc(z) - S(z).Y(z) \quad (14)$$

Where R(z) , T(z) and S(z) are polynomials and R(z) is selected as a normalized polynomial.
Having G(p), the discrete transfer function G(z) can be calculated with a sampling period h :

$$G(z) = (1 - z^{-1})Z\left\{L^{-1}[\frac{G(p)}{p}]\right\} \quad (15)$$

Where Z denotes the discretization operation and L^{-1} the inverse operation of Laplace's transformation.
Using the relation (8), G(z) is as followed:

$$G(z) = K\frac{1 - e^{-h/T}}{z - e^{-h/T}} \quad (16)$$

It is more practical to use the following form for G(z):

$$G(z) = \frac{b_o}{z - z_o} = \frac{B(z)}{A(z)} \quad (17)$$

Where bo and zo are related to K, h and T. The polynomials B(z) and A(z) have no common factor.

According Figure 9, $Y(z) = G(z).U(z)$ (18)

Using relations (14) and (17), the closed loop function transfer is given as:

$$\frac{Y(z)}{Yc(z)} = \frac{B(z).T(z)}{A(z).R(z) + B(z).S(z)} \quad (19)$$

Assume that the desired closed loop is:

$$H_m(z) = \frac{B_m(z)}{A_m(z)} \quad (20)$$

Relations (19) and (20) give:

$$\frac{B(z).T(z)}{A(z).R(z) + B(z).S(z)} = \frac{B_m(z)}{A_m(z)} \quad (21)$$

To obtain relation (21), it can be posed:

$$B(z).T(z) = Ao(z).Bm(z) \quad (22)$$

$$A(z).R(z) + B(z).S(z) = Ao(z).Am(z) \quad (23)$$

Where Ao(z) is defined as the observant polynomial.

The polynomial B(z) is defined as follow:

$$B(z) = B^+(z).\ B^-(z) \quad (24)$$

There are several methods to calculate the polynomials RST: with zero cancellation or without zero cancellation. Both methods can be completed by compensation of disturbance. All these methods lead in the resolution of Diophantine equation:

$$A_1(z).R_1(z) + B_1(z).S_1(z) = C(z) \quad (25)$$

Where the unknown polynomials are $R_1(z)$ and $S_1(z)$.

Theorem (Longchamp 1991)

There is a causal RST when these conditions are verified:

$$\deg(A_m) - \deg(B_m) \geq \deg(A) - \deg(B) \quad (26)$$

$$\deg(Ao) \geq 2.\deg(A) - \deg(A_m) - \deg(B^+) - 1 \quad (27)$$

In several cases, $H_m(z)$ is selected as followed:

$$H_m(z) = \frac{B^-(z)\frac{P(1)}{B^-(1)}}{z^d.P(z)} \quad (28)$$

Where z^d is chosen to respect (23).

The polynomial P(z) is generally as followed:

$$P(z) = z + c \quad (29)$$

$$P(z) = z^2 + c_1 z + c_2 \quad (30)$$

In these expressions, the different coefficients c, c_1, c_2 are selected to ensure the absolute and relative conditions of damping.
To ensure permanent error e_p equal to zero, $H_m(z)$ must verify:

$$H_m(1) = 1 \quad (31)$$

This condition is obtained by using the relation (28). Even integrating effects are introduced this error cannot be cancelled if this condition is not checked.

3.1. Algorithm

According the form of G(z), there is no zero to be cancelled here. So,

$$B^{+}(z) = 1 \quad \text{and} \quad B^{-}(z) = B(z) \tag{32}$$

The Diophantine equation gives R(z) and S(z). The relations (24), (28) and (32) imply that:

$$B_m(z) = B(z).B'_m(z) \tag{33}$$

With

$$B'_m(z) = \frac{P(1)}{B^{-}(1)} \tag{34}$$

The polynomial T(z) is calculated as followed:

$$T(z) = B'_m(z).Ao(z) \tag{35}$$

3.2. Application for the loop voltage

It already said that there is no zero to be cancelled. In this study, the case with no perturbation compensation is used. A first degree polynomial P(z) degree can be selected: it ensure the absolute condition of damping.
In this case, the observant polynomial Ao(z) is defined like followed:

$$\deg(Ao) = 0 \quad \text{and} \quad Ao(z) = 1 \tag{36}$$

By solving equation (26) and using relation (27), the polynomials RST verify:

$$\deg(R) = \deg(S) = \deg(T) = 0 \tag{37}$$

$$R(z) = 1 \quad S(z) = s_1 \quad T(z) = t_1 \tag{38}$$

The relation (14) gives the implemented law:

$$u(k) = t_1.yc(k) - s_1.y(k) \tag{39}$$

For the practical test, an autotransformer which can deliver variable voltage is used. To ensure a boost function, Vrd < Vs. First, the conditions are:

$$V_{rd} = 60\sqrt{2}\ [V] \qquad Vs = 90\ [V] \tag{40}$$

The figure 10 shows the front panel. The teaching applications are taken into account for this purpose.

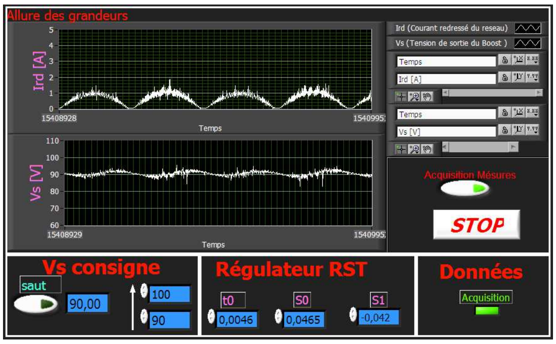


Figure 10: Front panel window

The complete stand is presented by Figure 11.

Figure 11: Complete stand for boost PFC

3.3. Results

The results obtained by simulation using Matlab and Simulink are given by Figure 12 and Figure 13. The set value of Vs is here: Vsc= 90[V]. These comments can be notified:

- The current is sinusoidal and in phase with the voltage (Figure 12).
- The THD is 3,74 %. It is more than the THD obtained by regulator PI (3,19% - Fc= 5[Hz]) but the regulation is more faster.

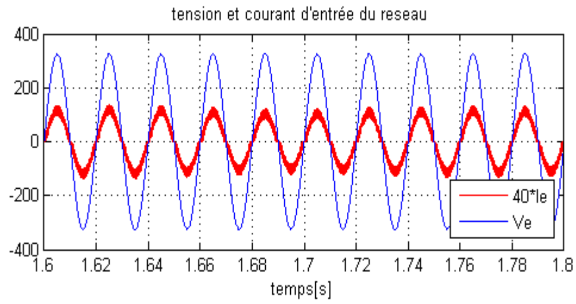


Figure 12: Voltage and current waveforms

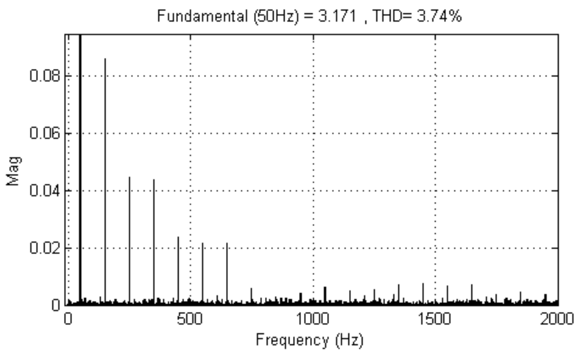


Figure 13: Current spectrum (THD=3,74%)

In Figure 14 and Figure 15, the simulations and the experimental results are given in steady states. The corresponding must be notified.

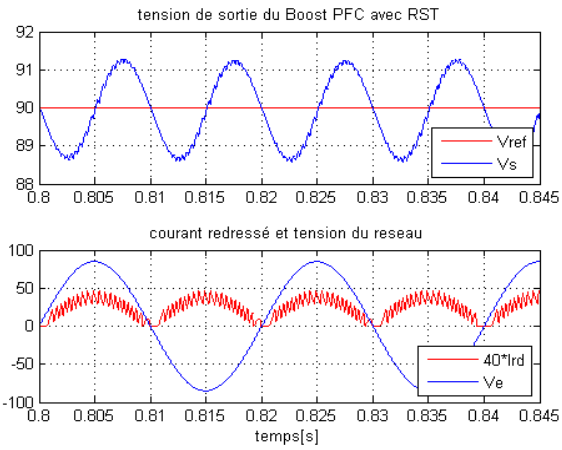


Figure 13: Steady state by simulation

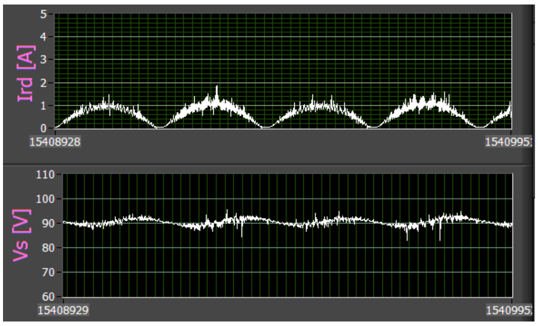


Figure 14: Experimental result (Vsc = 90 V)

Figure 15 show the result when the set value of voltage is Vsc= 100 V.

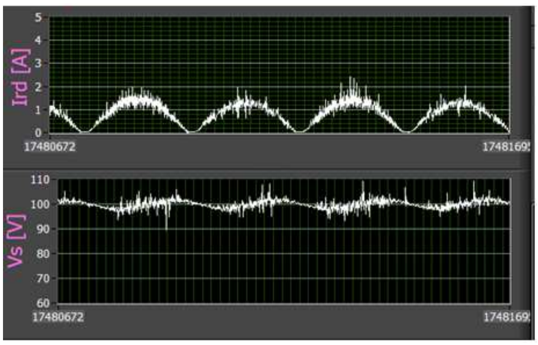


Figure 15 Experimental result (Vsc = 100 V)

In all these applications, the undulation at a frequency 100 [Hz] is present around Vs. His amplitude depends on the value of C. From the principle of the power factor correction, it cannot be eliminated. It is necessary to note that more the value of C is higher more this amplitude of Vs decreases but more the effects of harmonics current increase.

CONCLUSION

In this paper, polynomial RST controller is used for a boost PFC. Modeling the loop voltage as a first order system is sufficient even there is nonlinearity created by the static inverter. The RST controller gives good results in comparison with the classic PI one. The same TDH is obtained but the great difference is located especially at the velocity of the regulation. It must be noted that if the regulation with corrector PI is wanted to be faster, the distortion of the current waveform increases.

Labview with the peripheral NI 6009 is used for the applications. Teaching and industrial applications are aimed at same time. The implementation with Labview gives possibility to visualize the effects by varying the different coefficients of the RST controller. Applying nonlinear control like fuzzy command is also possible by using Labview.

REFERENCES

Gilles FELD, A. Cunière, 2009, Etude et simulation d'un AC/DC à absorption sinusoïdale de courant, *Publication ENS Cachan,*

S. Keraï, G. Meki, A. Fodil, K. Ghaffour, 2003, Calcul du convertisseur AC-DC avec correction de facteur de puissance, *Publication LMER,* Département d'Electronique, Université Abou Bakr Belkaïd, Algérie

R. Longchamp, 1991, *Commande adaptative*, Cours d'été, Institut d'Automatique, EPF Lausanne, Suisse.

B. Multon, N. Bernard, H. Ben Ahmed, 2003, Le redresseur MLI en absorption sinusoïdale de courant, *Revue ENS Cachan*, Antenne de Bretagne.

Razafinjaka, M. Sabata, J. D. Lan Sun Luk, 2008, Réglage par Logique Floue d'un Boost PFC, *Article présenté au Forum de la Recherche,* MENRS, Madagascar

B. Singh, 2003, *Modelling and simulation of power factor corrected AC-DC converters*, IJEEE

H. Tédjini, Y. Meslem, M. Rahli, 2008, Contrôle non linéaire avancé du redresseur MLI triphasé en absorption sinusoïdale de courant, *ACTA Electrotehnica, Vol.49, N°3 pp 231-300.*

.

AUTHOR BIOGRAPHY

Jean N. Razafinjaka was born in 1956, Madagascar. He is teaching Automatic and Optimization at the Higher Polytechnic School, Diego Suarez. His area of research concerns the advanced control for electromechanical systems. He is currently preparing his HDR.

A PRELIMINARY SUPPLY CHAIN MODEL TO HOUSING RECOVERY AFTER THE OCCURRENCE OF A NATURAL DISASTER

Diaz Rafael[(a)(b)(c)], Behr Joshua[(a)(b)], Sameer Kumar[(d)], Toba Ange-Lionel[(a)], Longo Francesco [(e)], Nicoletti Letizia[(e)]

[(a)]Virginia, Modeling, Analysis and Simulation Center, Old Dominion University, USA
[(b)]Eastern Virginia Medical School (EVMS), USA
[(c)] Department of Decision Science, College of Business and Public Administration, Old Dominion University, USA
[(d)] Department of Operations and Supply Chain Management, University of St. Thomas, USA
[(e)]MSC-LES, University of Calabria, Mechanical Department, Italy

[(a)]rdiaz@odu.edu, [(a)] jbehr@odu.edu, [(d)] skumar@stthomas.edu, [(a)] dtoba001@odu.edu, [(e)]f.longo@unical.it, [(e)]Letizia.nicoletti@unical.it

ABSTRACT
Severe storm events adversely affect housing stock and regional capacity to produce them. Rebuilding this capacity takes time while the affected region faces an unexpected surge in the demand for housing. This research presents a simulation model that considers this problem from the supply chain and production perspective. It allows characterizing capital fluctuations over time and determining bottlenecks to recovery. The model enables an understanding of the dynamics of supply and demand as it pertains to producing housing solutions as part of the recovery process. The ability to anticipate the composition of the demand as well as understanding capital fluctuations is critical to a fast recovery process.

1. INTRODUCTION
Hurricanes are major catastrophic events that cause death and suffering when they hit a populated region. Sheltering before, during, and after a severe storm is an essential activity to preserve the life of residents of the affected communities. In the aftermath of the storm, authorities and communities engage in different activities to recover from the devastating effects of the disaster. Depending on the magnitude and location of the event, different segments of the population may be affected and the regional capacity to recover may be diminished.

Flooding and damage produced by winds may significantly affect existing dwellings. The displaced population remains in shelters and may transition to temporary housing solutions until reconstruction endeavors are able to provide more permanent residential solutions. However, these reconstruction efforts are not uniform among affected communities since the damage stemming from a hurricane may have a dissimilar impact on the housing stock (Nigg, Barnshaw et al. 2006).

The spectrum of recovering the housing stock is extensive. Some houses require some minor repairs while others involve demolition and reconstruction. While government assistance rapidly flows into the affected region, supplies, materials, and labor converge at different rates. In an effort to quickly respond to the displaced population's sheltering needs for those that cannot return to their residences, decision makers may rule to establish temporary housing. However, some of these policies may cause unintended consequences that may delay the rebuilding process. For example, the longer the distance between temporary dwellings and the affected neighborhood, the longer it takes to rebuild the housing stock for underserved communities (Green, Bates et al. 2007). The examination of the recovery process as well as additional examples have been developed by Cemea (1997), Levine, Esnard et al. (2007), Kovács, Matopoulos et al. (2010), Frimpong (2011), Nejat and Damnjanovic (2012) among others.

This paper suggests an innovative application of a stock management structure suggested by Sterman (2000) in the production of housing solutions viewed from the supply chain perspective. The proposed simulation model allows examining the implications of transitioning displaced population to intermediate and long-term housing in the demand side while exploring the effects of barriers and enablers during the housing reconstruction process.

2. RESEARCH QUESTION
Literature that explores and models the progression of housing stock recovery after a catastrophic event employing a supply chain view is limited. Local, State, and Federal authorities benefit from this knowledge since it allows understanding the demand for temporal and permanent housing solutions while identifying bottlenecks that may jeopardize the recovery process. Thus, the central objective of this research is to propose a generic model that characterizes the dynamics of the

recovering housing stock process viewed as a production process. This production process may be seen as a process in which a set of structures, resources, and processes are set in place to deliver an output to the customer who is, in this case, the displaced household. As the production of housing solutions progresses, important fluctuations in labor and material may affect regional capacity to produce them. Thus, the production process assumed in this model is influenced by material and labor availabilities as well as building permits and housing reassignments rates.

3. APPROACH

This abstract suggests using a System Dynamics approach to gain knowledge and understanding of the dynamics of housing stock reconstruction as recovery from a catastrophic event progresses. Größler, Thun et al. (2008) argue in favor of using System Dynamics structures for investigating operation management issues since this approach considers feedback loops, accumulation processes, and delays that actually exist and are commonly found in complex problems. The authors employ a generic System Dynamics modeling approach suggested by Sterman (2000). The methodology includes: (1) developing a causal-loop representation of the regional housing production system; (2) formulating a theory that underlies this model; and (3) formulating a simulation model to test the dynamic hypotheses created from the production model. A brief overview of the model follows.

4. THE MODEL

The model involves characterizing displaced population when progressing through several stages during the recovery process. This is best characterized by compartments, e.g. short term, intermediate, and long-term. Likewise, on the supply side, the production of housing solutions process is subdivided into stocks that represent the different stages in which residential solutions are generated. The production process is influenced by the flow of capital to the region, fluctuations in materials and labor, and the number of dwellings completed and reassigned. These are common characteristics of housing production activities during recovery (Green, Bates et al. 2007).

The model provides a mechanism to characterize the demand and supply of housing solutions considering the composition of the demand and the interplay between the demand and the supply over time. As the stock of housing increases via repairing or reconstruction, the displaced population decreases which leads to decelerating the demand for permanent housing. Figure 1 provides a snapshot sub-model of the core production process in which material, labor, and house assignment rates contribute to adjustments in production rate, and consequently, completion rates.

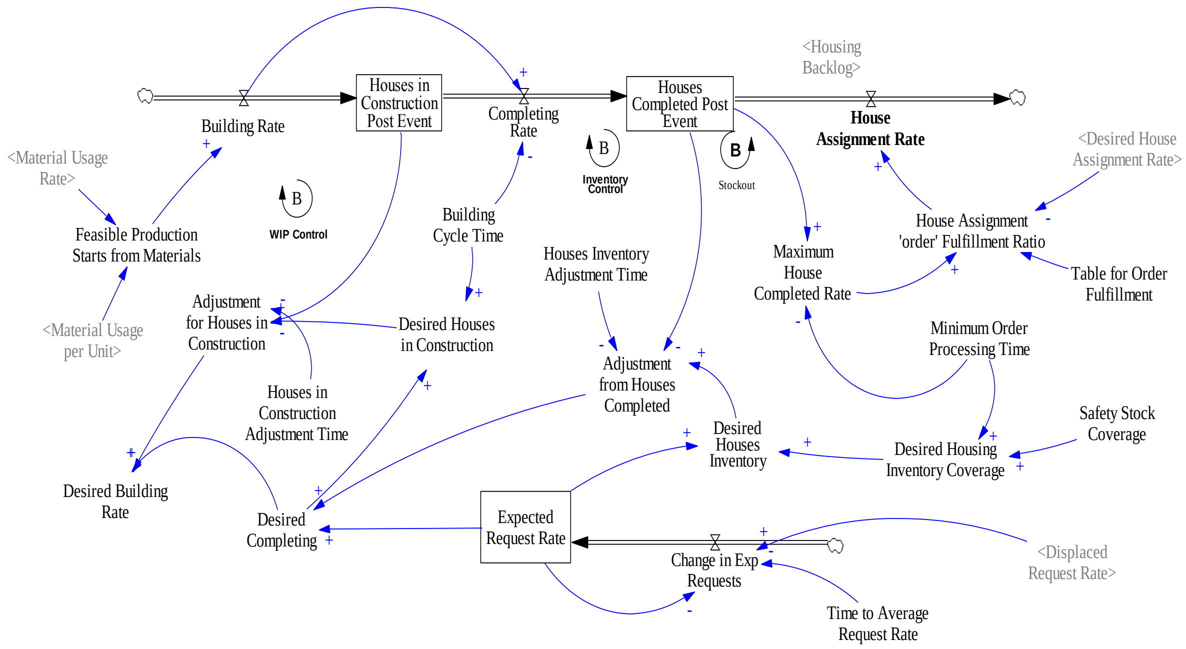


Figure 1: Production view of housing recovery

5. RESULTS

Theoretical data have been employed to mimic the behavior of a hypothetical scenario in which a hurricane hits a populated region. Figure 2 shows how displaced population stock declines as permanent housing solutions become available. Figure 3 exhibits how the number of people waiting for housing solutions first increases as more people become eligible for reconstruction, and then declines as permanent housing are produced and assigned to waiting people.

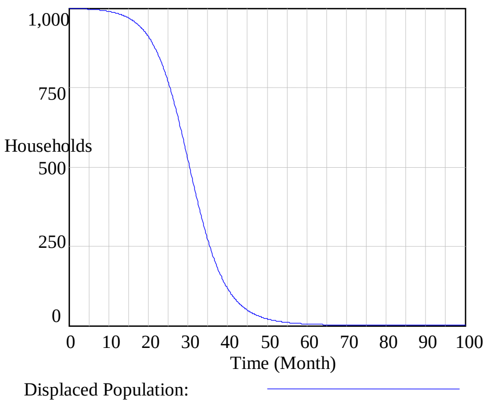


Figure 2: Displaced population volume

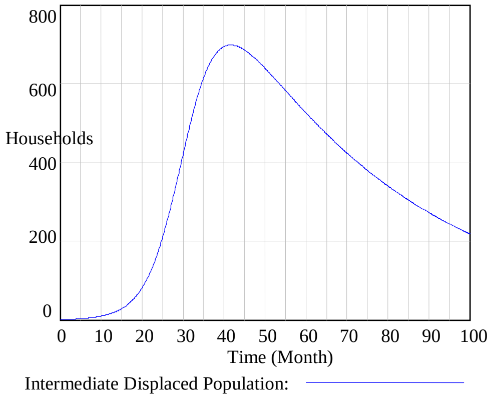


Figure 3: Intermediate displaced population

Figure 4 shows the labor flow where the hiring rate, vacancy creation, and quit rate are displayed. Right after the hurricane hits, the vacancy creation rate rises sharply since it reflects the immediate need of worker to perform the housing construction. This dictates the urgency and need to increase addition labor. Vacancies constitute the stock of potentially hired workers and is increased by the vacancy creation rate and decreased by the hiring rate. However, the hiring process takes time, e.g., considering interviews, background checks. All those activities constitute delays, necessary for the vacancies to be filled.

This delay is illustrated on the figure by the hiring rate, expectedly displaying a lag of time, compared to the vacation creation rate. The difference between the peaks of these rate curves – *vacancy creation rate* and *hiring rate* – illustrates the shortage of workers in need during this period. The quantity of workers available does not meet the need for reconstruction / repair. In the meantime, this severe lack of human resources causes a significant number of houses left without repair / reconstruction (Colten, Kates et al. 2008). Finally, this leads to the increase of the pressure on the small amount of construction builders already on duty.

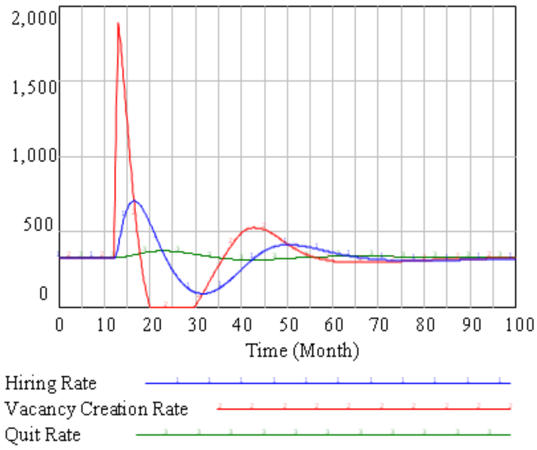


Figure 4: Labor flows

The *vacancy creation rate* remains at 0 for 10 weeks (between 20th and 30th). The 30th week corresponds to the time the *desired labor* raises from its minimum value and tries to adjust to the demand. The *vacancy creation rate* initiates its increase at the same time. An increase in *desired labor* indicates a need for workers, since the *Labor* and *hiring rate* have also gone down. The shortage having been created from those decreases need to be covered, which explains the rise of *vacancy creation rate*. The same cycle repeats until the forecasted demands match the demands.

Figure 5 shows an increase in pressure as to assure the emergency needs of population. The construction employees are pushed to work overtime in an attempt to clear the overload. Figure 5 also shows the sudden rise of the *desired labor*, similarly to the *vacancy creation rate*, as both behaviors are influenced by the demand. Thus, as the demand declines, both rates slowly regress. Given the delays involved in the hiring process, both *desired labor* and *vacancy creation rate* keep decreasing even after the demand is stabilized.

The *desired labor* depends on the *desired building start rate* (desired number of house construction starts), which is the desired *completion rate* of houses adjusted by the adequacy of the inventory of houses in construction, which depends on the demand. In other words, the behavior of *desired labor* is directed by the demand, with the presence of a lag time. In this sense, there is a constant attempt of the *Labor* to mimic the behavior of the demand.

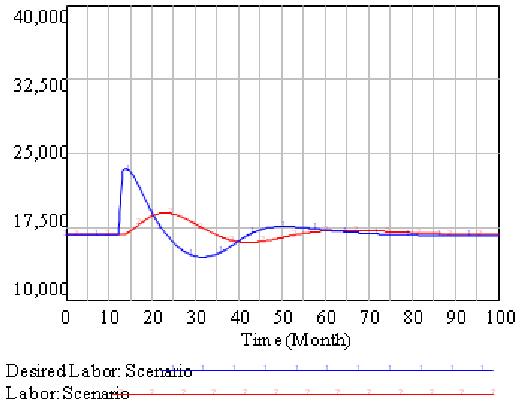


Figure 5: Desired labor

The *Desired labor* and *labor* meet at month 20, which indicates that the ideal needed number of workers is theoretically attained. There are thus no more vacancies to fill. This is illustrated by the curve of *vacancy creation rate* which takes the value of 0 at the same time. At that same time, the pressure is back to normal (value equals 1) and there is no need for overtime. The number of workers has become important enough to meet the demand. *Labor* is thus forced to stop its increase and start its decrease.

The schedule pressure is shown in Figure 6. This work schedule is also decreasing, given the diminishment of demand and the lowering of state of emergency. The vacancies are progressively being filed as the number of newly hired construction workers increases. The number of construction workers initiates its increase, but slowly as rather than local residents, mostly immigrants are the ones being hired (Petterson, Stanley et al. 2006). The *hiring rate* reaches its peak as it equals the *vacancy creation rate* and decreases.

Figure 7 shows another consequence of the labor shortage is the delivery delay, which goes up at the same time. This is due to the overwhelming demands which lead to the accumulation of orders in the backlog for some time, until they can be processed.

Figure 8 displays the behavior of housing after the shock. As expected, the *desired inventory* of houses increases right after the incident. This is happening because the *Desired Inventory is* directed by the behavior of the demand. In other words, an increase followed by a decrease of the demand causes the *desired inventory* to increase followed by a decrease. The *desired inventory* comes back down with the decline in the demand, but with a slower pace.

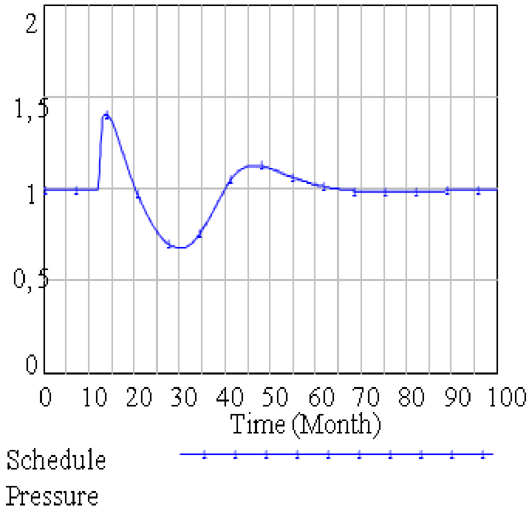


Figure 6: Schedule pressure

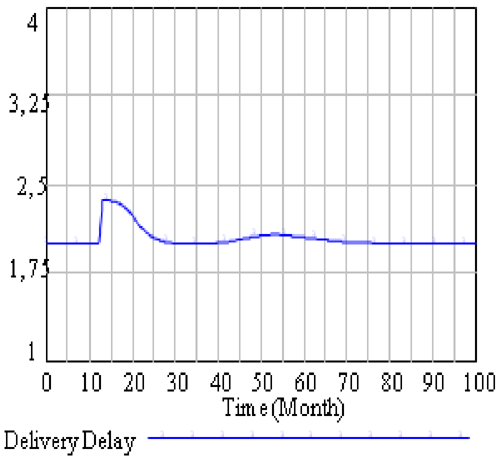


Figure 7: Delivery delay

The number of actual houses built decreases from its original value, given the unforeseen demands. The inventory stops diminishing around month 19^{th}. This time coincides with the time when the *completion rate* equals the *expected demand* rate. At this point, the number of houses newly built equals the demand so the inventory of houses already built stops being used. The stock of built houses can get back up. The *completion rate* reaches its peak around month 28th, which provides the opportunity of more time for the inventory to re-build. By the 35^{th} month, the demand takes over the building rate, which has been decreasing due to the drop in labor.

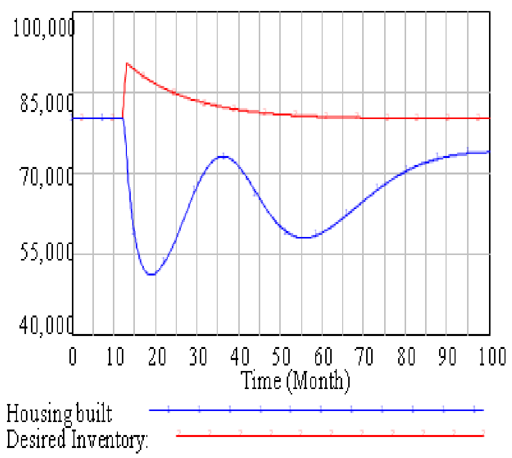


Figure 8 Housing after the shock

The houses are again in shortage and the inventory stops growing and declines back down again. This triggers the rise of the building start rate as there appears a shortage of houses to meet the demand. The inventory *housing built* is again used and the same cycle can repeat. Figure 8 also depicts the slow recovery in the area, should a hurricane event occur. The number of houses originally present does not seem to be reached anymore, given impediments in the reconstruction/recovery process. This is due to a structure in the model creating a delay for the *desired inventory*. This additional delay slows the adjustment of the *desired inventory* to the *expected demand*. That is, the *desired inventory* takes more time to follow the demand. Thus, the decrease of *desired inventory*, caused by demand declines, is more intensified. This causes the adjustment between the desired number of houses built and the actual number of those houses that take more time to occur. The *desired inventory* remains low and forces the inventory to adopt the same pattern. This behavior shows the problems of misplaced population and the struggle to recover from the disaster. From the governmental perspective, the lack of funding and the policy adjustment constitute a significant halt to post disaster recovery (Zhang and Peacock 2009).

It creates an important delay and causes negative impacts in the population and their ability to recover on the long run. From a socio-economic perspective, communities with limited resources bear most of the suffering and experience more complicated return conditions. More vulnerable victims opt for low-income rental homes, which have proved to be lengthy in construction, causing a shortage given the high demands (Dass-Brailsford 2008). The lack of shelter contributes to the slow return of the victims. The example of the locality of Lower Ninth Ward in New Orleans also confirms the behavior of the graph as it clearly shows the delay in recovery and the differences in housing after the disaster. The low and moderate income and minorities struggled the most and require more time to return (Green, Bates et al. 2007).

The presence of oscillation in all the graphs is the result of time delay and negative feedback. The structure of this model is designed to keep the inventories at their target level, which is used to compensate when faced with unexpected disturbances. The behavioral cycle described above is associated to the need for those values to pursue their target and bring the system to equilibrium.

6. CONCLUSIONS

This paper suggests using a System Dynamics modeling to understand the housing recovery process from a supply chain perspective. This model considers essential aspects of housing production systems in the presence of a highly disruptive event, namely, a severe storm. Hypothetical results indicate that the model theoretical behaves as expected, in terms of reflecting demand and supply changes as housing recovery progress. The ability to model and the housing production process when shocked by a major catastrophic event is important for decision makers. Scenario analysis provides opportunities to examine ripple effects and potential stumbling blocks that may jeopardize the recovery process. In addition, it provides a mechanism to test interventions that seek to reduce bottlenecks and increase housing production rates. Ongoing investigation involves calibrating the model and using real-world data to analyze likely scenarios for a major U.S. metropolitan region that lies in a low-lying coastal zone.

7. REFERENCES

Cemea, M. M., 1997. "The Risk and Reconstruction Model for Resettling Displaced Population." *World Development* **25**(10): 1569-1588.

Colten, C. E., et al., 2008. "Three years after Katrina: Lessons for community resilience." *Environment: Science and Policy for Sustainable Development* **50**(5): 36-47.

Dass-Brailsford, P., 2008. "After the storm: Recognition, recovery, and reconstruction." *Professional Psychology Research and Practice* **39**(1): 24-30.

Frimpong, A., 2011. "Sheltering and Housing Recovery after Disasters: Dissecting the problems of policy implementation and possible solutions." *International Journal of Humanities and Social Science* **1**(20) .

Green, R., et al., 2007. "Impediments to recovery in New Orleans' upper and lower Ninth Ward:

One year after Hurricane Katrina." *Disasters* **31**(4): 311-335.

Größler, A., et al., 2008. "System Dynamics as a Structural Theory in Operations Management." *Production and Operations Management* **17**(3): 373-384.

Kovács, G., et al., 2010. "A community-based approach to supply chain design." *International Journal of Logistics: Research and Applications* **13**(5): 411-422.

Levine, J. N., et al., 2007. "Population displacement and housing dilemmas due to catastrophic disasters." *Journal of Planning Literature* **22**(1): 3-15.

Nejat, A. and I. Damnjanovic, 2012. *Modeling dynamics of post-disaster recovery*. Construction Research Congress 2012, West Lafayette, Indiana.

Nigg, J. M., et al., 2006. "Hurricane Katrina and the flooding of New Orleans: Emergent issues in sheltering and temporary housing." *The Annals of the American Academy of Political and Social Science* **604**(1): 113-128.

Petterson, J., et al., 2006. "A preliminary assessment of social and economic impacts associated with hurricane Katrina." *American Anthropologist* **108**(4): 643–670.

Sterman, J. D., 2000. *Business Dynamics -System Dynamics for a complex world*, McGraw Hill Higher Education.

Zhang, Y. and W. G. Peacock, 2009. "Planning for housing recovery? Lessons learned from hurricane Andrew." *Journal of the American Planning Association* **76**(1): 5-24.

Biographies

Rafael Diaz graduated from the Old Dominion University with a Ph.D. in Modeling and Simulation in 2007, and became a Research Assistant Professor of Modeling and Simulation at Old Dominion University's Virginia Modeling, Analysis, and Simulation Center (VMASC). He holds an M.B.A degree in financial analysis and information technology from Old Dominion University and a B.S. in Industrial Engineering from Jose Maria Vargas University, Venezuela. His research interests include operations research, operations management, healthcare and public health policy-making, dependence modeling for stochastic simulation, and simulation-based optimization methods. He is an Adjunct Assistant Professor in the program of Public Health, at Eastern Virginia Medical School.

Joshua G. Behr received his Ph.D. training at the University of New Orleans specializing in urban and minority politics. He has taught a variety of public policy, state government, and statistical methods courses at the University of New Orleans, Southwestern Oklahoma State University, Eastern Virginia Medical School, and Old Dominion University. His is now a research professor at the Virginia Modeling, Analysis and Simulation Center (VMASC), Suffolk. He has published on a wide range of topics including presidential approval, times series methodology, minority employment patterns, public health, and emergency department utilization as well as a recent book (SUNY Press) on political redistricting and a book chapter addressing discrete event simulation. Currently, his research interests include modeling and simulating smart grid, transportation systems, and the flow of patients within a regional healthcare system.

Ange-Lionel Toba is currently a second year Ph.D. student in Modeling and Simulation at Old Dominion University. He holds a MS. in Industrial & Systems Engineering from Colorado State University and a B.S. in Electrical Engineering from Cadi Ayyad University, Morocco. His research interests are operations research, Applied Statistics and Modeling and simulation of complex systems using System Dynamics (SD) and Discrete Event Simulation (DES). He is a Graduate Research Assistant at Old Dominion University's Virginia Modeling, Analysis and Simulation Center (VMASC).

Francesco Longo received his PhD in Mechanical Engineering from the University of Calabria; he is currently an Assistant Professor and Director of the Modeling and Simulation Center, Laboratory of Enterprise Solutions (MSC-LES). His research interests include modeling and simulation for training procedures in complex environments, supply chain management and security. He has published more than 120 papers in international journals and conferences. He is Associate Editor and Guest Editor of Simulation: Transactions of the SCS. He is Guest Editor of the International Journal of Simulation and Process Modelling. He is has extensively supported the organization of international conferences as General co-Chair, Program Co-Chair and Track Chair (MAS, EMSS, I3M, SCSC, etc.).

Letizia Nicoletti is currently a PhD student in Management Engineering at the University of Calabria, Italy. Her research interests include modelling and simulation for inventory management, as well as for training in complex systems, specifically marine ports and container terminals.

A MECHANICAL APPROACH OF MULTIVARIATE DENSITY FUNCTION APPROXIMATION

László Mohácsi[(a)], Orsolya Rétallér[(b)]

[(a)] Department of Computer Science, Corvinus University of Budapest
[(b)]MTA-BCE "Lendület" Strategic Interactions Research Group and
Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest

[(a)] mohacsill@gmail.com, [(b)]retaller.orsolya@gmail.com

ABSTRACT

In this paper we are modeling multivariate density functions by going back to the roots: instead of trying to fit a well-known copula on the data, we choose to generate one. Our model approximates the two dimensional density function using physical analogy. Points on the scatter plot diagram are represented by small balls of unit mass placed on a sheet of elastic sponge, and the deformation of the surface of the sponge caused by the balls represents the density of the points on the scatter plot diagram. The elasticity of the sponge is described by Hooke's law. The distortion of the sponge can be determined by using finite element methods. The distorted surface can be approximated by functions using Fourier transformation. Hence, the model can be extended into higher dimensions.

Keywords: copulas, multivariate data analysis, finite element method, goodness of fit

1. INTRODUCTION

Modeling multivariate distributions is relatively complicated in general, even when we are aware of the marginal distributions' nature. The dependency structure among the variables is usually described by their covariance or correlation matrix, however these measures have a great disadvantage: they only measure the linear dependency, not the association in general. There are other measures, such as Spearman's rho, or Kendall's tau, which do not rely heavily on linearity, but these measures are rather used to determine how monotonic the relation is among the given variables. But we should note, that dependency doesn't imply monotonicity.

Copulas are functions that join multivariate distribution functions to their one-dimensional margins. (Nelsen 2006 and Sklar 1959) They also take it into consideration that the dependency might be different on the edges, they provide a flexible structure, and do not intend to measure linearity or monotonicity.

There are several famous copulas, which we might try to fit on the data. The problem is, that in some cases none of the famous copulas fit well – even if we take the parameter shifting into consideration. This was basically what motivated us trying to find alternatives for these situations.

In section 2 we introduce our copula generating method, for which we are using finite element methods. This is a modeling technique widely used among engineers, but as far as we're concerned it has no previous history in financial modeling. We believe, that the synergy of the different fields will provide with an interesting and promising result. In section 3 we present some calculations on financial data in two cases. Not only we would like to show how our copula fits on the data as one of the bests (compared to some of the famous copulas), but we would also like to demonstrate, that the conventional chi-square testing of the goodness-of-fit is not reliable in our case. In section 4 conclusion follows.

2. THE COPULA GENERATING METHOD

In order to understand how our copula generation works let's consider an analogy from physics. The density function of a given copula can be modeled by the distortion of an elastic lattice (sponge) when small balls of unit mass are placed on it. The more balls we place on a given point, the more the sponge sinks. Each site , where we decide not to put any balls the value of this function is zero, or – due to the interaction of the balls – close to zero. If we would like to obtain a function that has the characteristics of a density function (e.g. non-negativity) this procedure has to be reversed: instead of placing weights on the sponge, and pushing it down, we prefer to pull it up. Hence, the surface looks more like the surface of a density function, and fulfills the non-negativity criteria.

The distortion of the lattice can be determined by using finite element methods. In engineering, these methods are typical modeling applications of the distortion of statically loaded machine parts. Another typical application is determining the deformation of machine parts and temperature distribution in them with a given thermal conductivity and thermal expansion coefficient. In this application, temperature dependency is omitted , however it could lead to an extension of modeling in 3D.

For generating the 3D model, we used our own program written in C#. The Z88 Aurora is a program, which is a user interface based on the Z88 engine (http://www.z88.de/).

2.1. The parameters of the simulation

During the simulation the behavior of the sponge can be characterized with the following parameters:

1. The *Young-modulus* (E, N/mm^2), also known as the elastic modulus , which is a constant in our model. It describes the relationship between the force effecting on the lattice and its' distortion, which is also known as Hooke's law. However, it should be noted that not all materials are behaving according to Hooke's law's : e.g. amorphous materials rubber the distortion and the force on it does not have a linear dependency.
2. The *Poisson's ratio* is a dimensionless constant. It provides the negative ratio of transverse to axial strain.
3. The next parameter of the material is the *density* (ρ, kg/m^3). If we place the material in the gravitational field, it will be distorted by its own weight.
4. *Thermal conductivity* (λ, W/(m·K)) could also be taken into consideration. According to the second law of thermodynamics, an isolated system, if not already in its state of thermodynamic equilibrium, spontaneously evolves towards it. For non-evolving materials, this phenomenon is characterized by the Fourier-law.
5. There are a few materials which have a *negative coefficient of thermal expansion*, which means that cooling them leads to expansion. An example for this phenomenon is water between 0 C° and 4 C°.

2.2. The simulation steps

The simulation comprises the following steps. (Based on Deák 1990 and Ross 1997.)

2.2.1. Building the model

As a first step we have to build the model of the shape that we would like to observe. The basis of the model relies on the nodes, which are given by their coordinates in 3D. These nodes determine spatial elements, which are usually constructed of hexahedrons or tetrahedrons. Eight nodes could determine a hexahedron, while six of them lead to a tetrahedron.

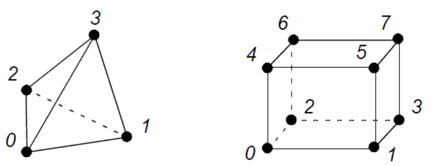


Figure 1: A tetrahedron and a hexahedron given by their nodes.

In engineering there is a common practice regarding the model building: on those areas, where the forces are more concentrated there should be more dense sampling.

Most finite element softwares are capable of decomposing models, which originate from CAD programs. In our case it was more desirable to generate the nodes and elements from the program.

2.2.2. Boundary conditions

Before running the simulation we have to provide the boundary conditions. At least one node should be designated, which has a fixed position. Without this step a static examination is unconceivable. Also, for each node the attacking force vectors can be given.

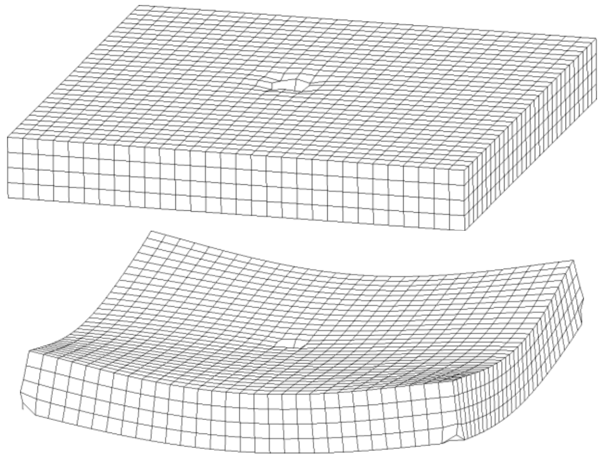


Figure 2: A loaded rubber sheet, which is supported in every point and in the middle only.

2.2.3. Running the simulation

As we expected the value of the Young modulus had no significant effect on our simulations, and it was also irrelevant, whether we pushed or pulled the material. We found however, that the thickness of the material is relevant. Also, we had an interesting side-effect, which we decided to call the sponge-effect.

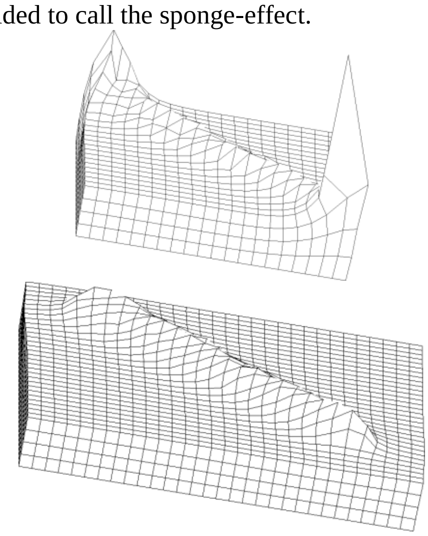


Figure 3: The sponge-effect.

The sponge-effect can be explained again with an analogy deriving from physics. Consider a mattress, and imagine if we sit on it. If we sit in the middle, there will be a big distortion in the middle, some distortion in the surrounding area, but almost no distortion on the edges. However, if we decided to sit on the edge, the distortion would be much bigger, as the surrounding area is missing, and there is no support from it. Therefore, if we use the same force to pull the sponge in the middle and on the edges, the distortion looks completely different.

As the sponge-effect distorted our simulations a lot, we decided to obtain some corrections. Figure 3 shows that we decided to put some extra elements outside our model as well, which gives some support to the nodes on the edges as well.

3. EXAMPLES AND COMPARISON

We applied the previously described methodology on two datasets, both of them are modeling dependency among certain financial indicators. In the first case we observed the stock exchange indices of London's and Paris's stock exchange market, namely the FTSE and FCHI. The second case is about the dependency between two financial assets: gold and real estate.

3.1. The FTSE and FCHI dependency

The data we used to observe the dependency between the English and the French market were the daily closing values of the indices between 01.01.2000 and 12.31.2009, but we have only taken into consideration those days, where both markets were open. As a result we obtained 2412 pairs of data.

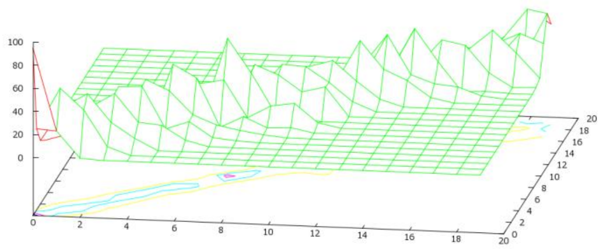


Figure 4: The relationship between ranking numbers on a 20x20 crosstabulation

As we originally expected, the FTSE and FCHI indices have a very high positive correlation, namely 0.968. Figure 4 represents the relationship of the ranking numbers on a joint histogram on a 20x20 crosstab. There is not much difference if we use the original data or the ranking number. This also means that the Pearson's correlation and Spearman's rho both provide us with similar results and these measurements provide enough information about the dependency structure. Still, let's observe which is the best fitting copula on the data.

Both in this case and the next one we fitted the independence copula along with Clayton's and Gumbel's copulas (the formulas can be found in table 1), and of course our one. For those copulas that require a parameter estimation (Clayton's and Gumbel's) we applied the maximum likelihood estimation using Excel Solver. As a result we ended up with a theta value of 7.540 for Clayton's copula, and 4.331 for Gumbel's.

Table 1: CDF of the fitted copulas

Copula name	Bivariate formula of the CDF
Independence	uv
Clayton's	$\left(max\{u^{-\theta}+v^{-\theta}-1;0\}\right)^{-1/\theta}$
Gumbel's	$exp\left(-\left[\{-log(u)\}^{\theta}+\{-log(v)\}^{\theta}\right]^{\frac{1}{\theta}}\right)$

It should be noted, that instead of providing a figure of the cumulative distribution function of the copulas we decided to represent the expected values. This way it is easier to compare the results with the original data.

To calculate the expected value of each cell we used the following formula (i denotes the number of row, and j stands for the number of column, k is dimension of the crosstab, n denotes the number of data pairs, and F stands for the cumulative distribution function of the given copula):

$$n\cdot\left(F\left(\frac{i}{k},\frac{j}{k}\right)-F\left(\frac{i-1}{k},\frac{j}{k}\right)-F\left(\frac{i}{k},\frac{j-1}{k}\right)+F\left(\frac{i}{k},\frac{j}{k}\right)\right)$$

Figure 5-8 show the results of the fitting.

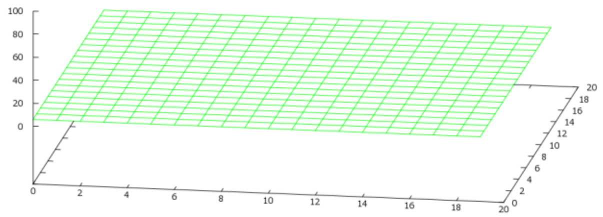


Figure 5: The expected values based on the independence copula

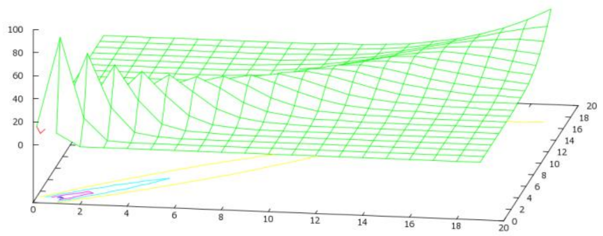


Figure 6: The expected values based on Clayton's copula (θ = 7.540)

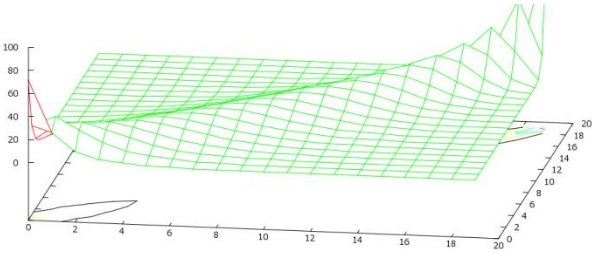


Figure 7: The expected values based on Gumbel's copula (θ = 4.331)

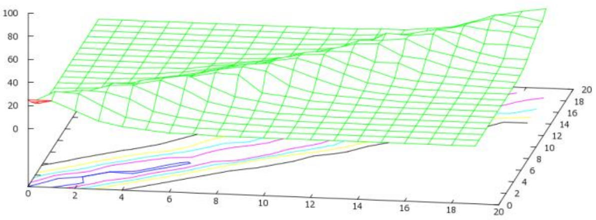


Figure 8: The expected values based on our copula

It is not easy to decide by looks which is the best fitting copula on these data. However it is quite visible, that the independence copula's case seems to be the worst one.

With the statistical testing we encountered a major difficulty, which was the fact, that the conventional chi-square method requires an expected value of at least 5 in each cell to be able to perform the test. In general if this criterion isn't met, the suggestion is to merge some of the cells, until the test can be calculated. In this case however – as the data are basically in the diagonal – even the 3x3 representation is inadequate for the statistical testing.

The goodness-of-fit testing of copulas is still a hot topic, and there is not yet an obvious answer how it should be done. (As an example see Berg 2007, Dobric and Schmid 2005, Lowin 2001, Patton 2006, Quessy 2005). This paper does not have an agenda to provide an answer to this question hence we only calculated the chi-square test (if it was possible) and the average sum of squares of the errors. In the previous case it should also be noted, that we disregarded the test's requirement regarding the minimal expected value (because of the previously described phenomenon), but this caused major distortion in the test value in those cases where the expected value was close to zero. If it was even zero, the test couldn't even be calculated. This was basically the reason why we decided to observe the average sum of square of the errors instead. Table 2-7 represent the results, and the best ones are highlighted.

It should also be noted, that the results vary if we change the numbers of cells in the crosstabs. In this case we did the calculations on a 3x3, a 5x5 and a 20x20 crosstab.

Table 2: Chi-square results (3x3 crosstab)

Fitted copula	Chi-square	df	p-value
Independence	3 720,0	8	0,000
Clayton's	140,2	7	0,000
Gumbel's	239,6	7	0,000
Our one	2 078,7	7	0,000

Table 3: Chi-square results (5x5 crosstab)

Fitted copula	Chi-square	df	p-value
Independence	5 957,5	24	0,000
Clayton's	348,3	23	0,000
Gumbel's	384,0	23	0,000
Our one	N/A	N/A	N/A

Table 4: Chi-square results (20x20 crosstab)

Fitted copula	Chi-square	df	p-value
Independence	11 923,3	399	0,000
Clayton's	1 616,0	398	0,000
Gumbel's	2 939,0	398	0,000
Our one	N/A	N/A	N/A

Table 5: Average Sum of Squares (3x3 crosstab)

Fitted copula	Average Sum of Squares
Independence	110 773,1
Clayton's	3 462,6
Gumbel's	6 259,7
Our one	65 873,2

Table 6: Average Sum of Squares (5x5 crosstab)

Fitted copula	Average Sum of Squares
Independence	22 991,0
Clayton's	2 230,3
Gumbel's	2 319,8
Our one	2 922,5

Table 7: Average Sum of Squares (20x20 crosstab)

Fitted copula	Average Sum of Squares
Independence	179,7
Clayton's	52,2
Gumbel's	84,5
Our one	75,3

This case was a difficult one, as none of the fitted copulas were good enough to fit according to the chi-square test results. However, the figures we presented can be quite persuasive, that the expected values generated based on Clayton's, Gumbel's and our copula, are all pretty close to the original data.

3.2. The gold and real estate dependency

In this case 2043 pairs of data have been observed for a 10 year period again. This case the dependency was not that obvious, the Pearson's correlation coefficient was only -0.442. But if we look at the data we can find certain groups on the scatter-dot diagrams. Figure 9 represents the original data and the rankings as well, whereas figure 10 presents the joint histogram on a 20x20 crosstabulation.

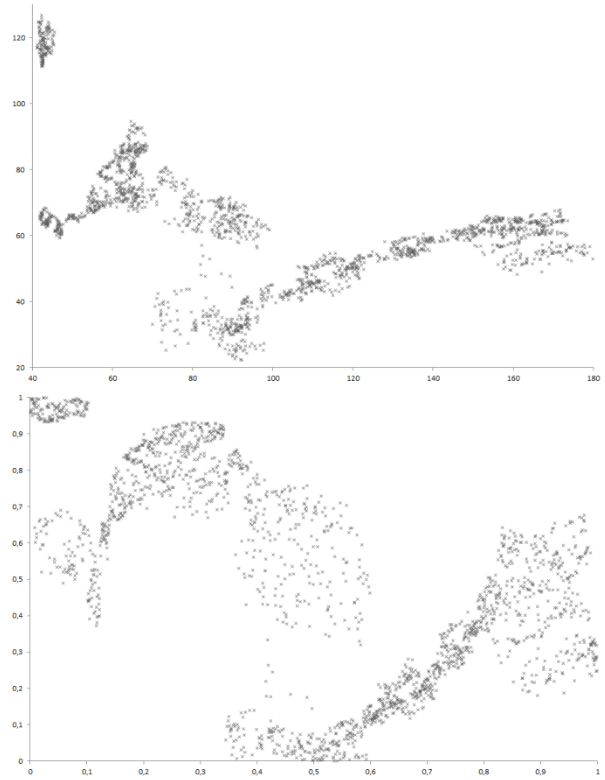


Figure 9: The relationship between the original data and the ranking numbers

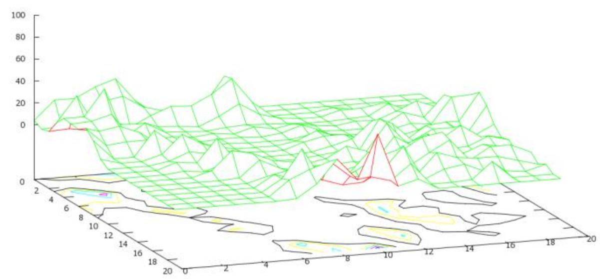


Figure 10: The relationship between ranking numbers on a 20x20 crosstabulation

As the maximum likelihood estimation didn't bring any results for neither Clayton's nor Gumbel's theta, we decided to use -0.2 for the first, and 1 for the second one (these estimations had pretty good results for the sum of the log-likelihoods). This resulted, that the Gumbel copula was identical to the independence one. The fitted copulas can be seen on figure 11-14.

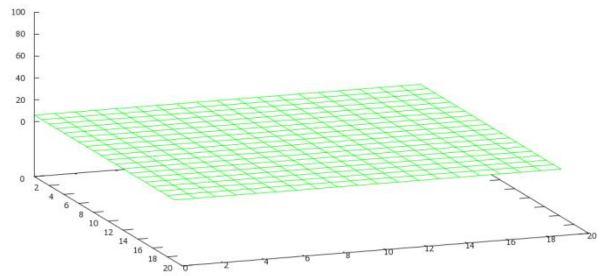


Figure 11: The expected values based on the independence copula

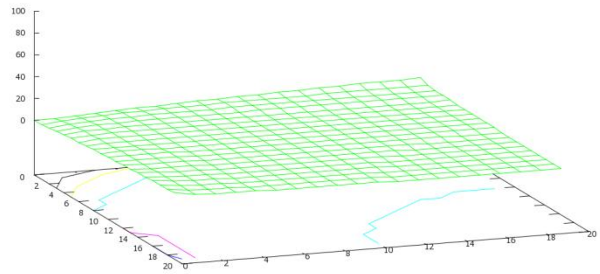


Figure 12: The expected values based on Clayton's copula (θ = -0.2)

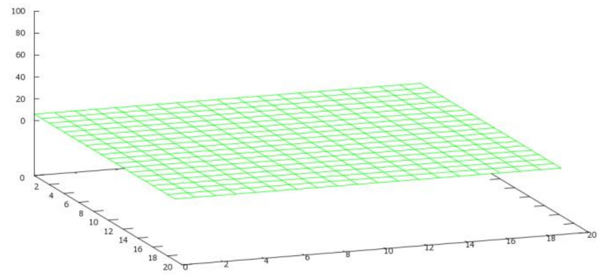


Figure 13: The expected values based on Gumbel's copula (θ = 1)

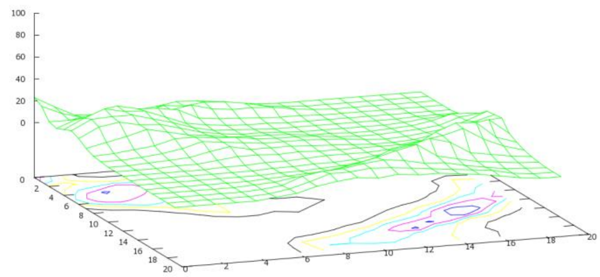


Figure 14: The expected values based on our copula

It is quite visible, that in this case our copula is the closest to the original data. Once again, we tried to obtain the conventional chi-square test to verify this statement, but we found that if we want to follow the rules, a 3x3 crosstab is the biggest that we can use. In this case however the test could really be applied, as the expected value in each cell was over 5 in all cases. Still, none of the fitted copulas were good enough to accept the null hypothesis. Table 8-13 represent the results.

Table 8: Chi-square results (3x3 crosstab)

Fitted copula	Chi-square	df	p-value
Independence	796,8	8	0,000
Clayton’s	1 291,7	7	0,000
Gumbel’s	796,8	7	0,000
Our one	1 344,3	7	0,000

Table 9: Chi-square results (5x5 crosstab)

Fitted copula	Chi-square	df	p-value
Independence	1 605,3	24	0,000
Clayton’s	2 278,3	23	0,000
Gumbel’s	1 605,3	23	0,000
Our one	1 652,6	N/A	N/A

Table 10: Chi-square results (20x20 crosstab)

Fitted copula	Chi-square	df	p-value
Independence	5 470,1	399	0,000
Clayton’s	8 717,2	398	0,000
Gumbel’s	5 470,1	398	0,000
Our one	N/A	N/A	N/A

Table 11: Average Sum of Squares (3x3 crosstab)

Fitted copula	Average Sum of Squares
Independence	23 725,8
Clayton’s	35 301,4
Gumbel’s	23 725,8
Our one	4 969,3

Table 12: Average Sum of Squares (5x5 crosstab)

Fitted copula	Average Sum of Squares
Independence	6 195,0
Clayton’s	8 039,9
Gumbel’s	6 195,0
Our one	3 443,8

Table 13: Average Sum of Squares (20x20 crosstab)

Fitted copula	Average Sum of Squares
Independence	82,5
Clayton’s	91,2
Gumbel’s	82,5
Our one	80,0

As the chi-square testing is very unreliable, we prefer to make a decision on the average sum of squares of errors. In this case our copula seems to provide with the best solution. However, the goodness-of-fit testing for copulas is still an interesting topic with a lot of open questions.

4. CONCLUSION

Even though the results are not conclusive (because of the unreliability of the testing), we believe that the copula generation method we presented could be widely applied. It is flexible enough to fit even on those data, where other copulas cannot find any relationship among the data. Also, as we have referred to it, it could be extended into higher dimensions.

ACKNOWLEDGMENTS

The authors would like to thank Erzsébet Kovács for encouraging them to work together, even though they have very different fields of expertise. They would also like to acknowledge Zsolt Benczúr for his helpful hints, and Fruzsina Gajdos for her help with the translation of the professional expressions.

REFERENCES

Berg, D., 2007. Copula goodness-of-fit testing: an overview and power comparison. Available from: https://www.duo.uio.no/bitstream/handle/10852/10361/stat-res-05-07.pdf?sequence=1

Deák, I., 1990. *Random number generators and simulation (Mathematical Methods of Operations Research)*. Budapest: Akadémiai Kiadó.

Dobric, J., Schmid, F., 2005. Testing Goodness of Fit for Parametric Families of Copulas – Application to Financial Data. *Communications in Statistics – Simulation and Computation*, 34: 1053-1068.

Klubman S.A., Panjer, H.H., Willmot G.E., 2008. *Loss Models: From Data to Decisions*. Hooken, New Jersey: John Wiley & Sons.

Lowin, J.L., 2001. The Fourier Copula: Theory & Applications. Harvard University.

Nelsen, R.B., 2006. *An Introduction to Copulas*. New York: Springer

Panjer, H.H., 2007. *Operational Risk: Modelling Analytics*. Hooken, New Jersey: John Wiley & Sons.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47 (2): 527-556.

Quessy, J.F., 2005. *Méthodologie et Application des Copules: Tests d’Adéquation, Tests d’Indépendance, et Bornes pour la Valeur-à-Risque*. Thesis (PhD). Université Laval.

Ross, S.M., 1997. *Simulation*. Boston: Academic Press.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8: 229–231.

ORCHESTRATING THE INTEROPERABILITY WORKFLOW WITHIN A TRANSPORT SIMULATION PLATFORM

Judicaël Ribault, Gregory Zacharewicz

Univ. Bordeaux, IMS, UMR 5251, F-33400 Talence, France

judicael.ribault@u-bordeaux1.fr, gregory.zacharewicz@u-bordeaux1.fr

ABSTRACT

The domain of logistics and transport is now gaining with the use of the web, geo positioning and RFID to improve the tracking and decision making for the product more appropriate routing in order to save time, cost and reduce impact on the environment. The combination of these software and hardware devices faces interoperability problems. This paper proposes to introduce a new simulation platform that will mix interaction with real world including sensor and human interfacing and simulation world. In detail, the proposition of this paper is to combine the Taverna Workflow, which handles and triggers the call of web services proposed by a platform, with several simulation models. In particular one drawback of several workflows orchestrator tools is that they do not provide time management facilities to handle time and to rhythm simulation run. This paper introduces a message clock ordering solution defined by G-DEVS models to give the beat to the transport simulation workflow system. The imbrication of G-DEVS modelling and simulation with the workflow Taverna shows the possibility of the interoperability and complementarity of these approaches.

Keywords: Workflow, Taverna, Interoperability, Discrete event simulation, G-DEVS

1. INTRODUCTION

The effectiveness of enterprise information technology system (IS) depends not only on its internal interconnectivity of its inner software components, but also on its ability to exchange data, so to collaborate, with every day new tools developed and updated in the environing digital world. This necessity led to the development of the concept called interoperability that allows improving collaborations between enterprises IS. No doubt, in such context where more and more networked enterprises are developed; that enterprise interoperability is seen as one of the most wanted solutions in the development of an enterprise IS. Also the data treatment calls both human processing and automatic treatments. The sequencing of these actions is desired to be controlled or orchestrated by a high level application that can decide the human resource and/or component to solicit.

From a research point of view, several works has been launched since the beginning of 90's in the domain of Workflow. Workflow was first designed to formalize and improve enterprise business process. A product workflow is a set of linked steps required for developing a product until it gets into market [Weske 2012]. The workflow steps are based on observing a number of steps that are usually enchained manually and formalizing them. The research on the Workflow initiated by the Workflow Management Coalition [Zacharewicz2008] was a premise to workflow modelling (e.g. with BMPN) and it permits the development of recent ERP systems in the enterprises. Nevertheless a clear distinction appeared in the late 90's between the theoretical approaches in this domain and applied approaches. In the theoretical approach, Modelling and Simulation (M&S) is a main consideration, while in the applied approaches, execution is the core problem. Few approaches compose efficiently M&S and real executions in the transport domain. Main reasons are the slowing for synchronization of the simulation engine, that is usually constrained by causality [Chandy, 1979] between real and simulated time, and the interoperability barriers that are faced between hardware and software [Chen, 2003].

Recent improvements in web-based development propose new facilities to connect the applications in a more convenient way. For instance the web services can support that question of interoperability. We can classify the Web services into two categories:

- Web services of type "REpresentational State Transfer" (REST) [Richardson, 2007] whose main purpose is to manipulate XML representations of Web resources using a uniform set of HTTP operations (GET, PUT, POST, DELETE) and URI.
- Arbitrary Web services, which expose an arbitrary set of operations that can be executed remotely by using SOAP and WSDL standards that facilitate interoperability.

We propose to use web services and workflow for interoperability among simulation and real-world application. Web services enable the integration of applications or data from heterogeneous sources (i.e. Mash-up). This paper is proposing to apply the use of workflow Web services and simulation to the PRODIGE application.

Section 2 describes the necessary background needed to understand how workflows of services and simulation can drive real application. Section 3 presents the scientific contribution while section 4 put it into practice in a real framework.

2. BACKGROUND

In this section, we first present the enterprise interoperability concept. Then we briefly present the PRODIGE system and how workflow can be used for experimentation. Then we present the DEVS formalism and its interoperability through web services. Finally we present the Taverna workflow management system to orchestrate the experimentation.

2.1. Interoperability

Enterprise Interoperability [Chen, 2003] refers to the ability of interaction between enterprise systems. The interoperability is considered as significant if the interactions can take place at least at the three different levels: data, services and process, with a semantic defined in a given business context.

Interoperability extends beyond the boundaries of any single system, and involves at least two entities. Consequently establishing interoperability means to relate two systems together and remove incompatibilities. Incompatibility is defined as the fundamental concept of interoperability [Zacharewicz 2011a]. It is the obstacle to establish seamless interoperation. The concept 'incompatibility' has a broad sense and is not only limited to 'technical' aspect as usually considered in software engineering, but also 'information' and "organization", and concerns all levels of the enterprise.. Basic concepts relating to enterprise interoperability are classified into three main dimensions as described in the cube proposed in [Chen, 2003]. The integrated approach is demanding to all partners to have the same description of information. The unified approach is asking partners just to prepare data to exchange to be compliant with a Meta model but local description can be kept. The third dimension is federated. Here, interoperability must be accommodated on the fly between partners without considering a pre-existing meta model.

Our goal is to tackle interoperability problems through the identification of barriers (incompatibilities) which prevent interoperability to happen

The first kind of barrier concerns the nonexistence of commonly recognized paradigms and data structure, for that, clarification is required to propose a sound paradigm. The second requirement not addressed at the enterprise modelling level is the synchronization of data. The right order of data exchanged is important, ignoring this can lead to misunderstanding and wrong functioning of the model. Finally the enterprise modelling must consider the confidentiality management of data. The interoperability can be considered between concurrent enterprises in that context, a strategy of data sharing/not sharing between these must be defined.

Today, most of the approaches developed are unified ones. For example, in the domain of enterprise modelling, we can mention UEML (Unified Enterprise Modelling Language) [Roque, 2008] and PSL (Process Specification Language) [NIST, 2003] which aim at supporting the interoperability between enterprise models and tools. Using the "federated approach" to develop enterprise interoperability appears to be the most challenging and few activities have been performed in this direction. The federated approach aims to develop full interoperability and is particularly suitable for an inter-organisational environment (such as networked enterprises, virtual enterprises, etc.). In the enterprise interoperability roadmap published by the European Commission in 2006, developing "federated approach" for interoperability was considered as one of the research challenges for the years to come.

From the state of the art of the enterprise interoperability domain and some implementations experiences to be presented in next points, we will introduce in the next section some propositions to address these compatibility challenges.

2.2. PRODIGE

The PRODIGE project aims to prepare the future of physical products transportation, placing the reflection at the organizational level that control the flow of commodities in order to provide a technical and organizational solution helping the reduction of the travelled distance, optimization of the tours, volumes transported and taking into account new issues related to sustainable development.

The base of the work proposed in this paper, start from a transportation Web application released in the project. This platform is composed of a server where several trucks users are remotely contacted to display their positions thanks to GPS and GSM communication. The server is proposing algorithm to optimize truck routing. It is exposing its methods through the use of SOAP Web services in order to promote interoperability (set a tour, view the results, etc.). The idea is to test the function of the tool regarding a sequence of calls in dynamic. For that purpose a simulation tool for making alive the workflow is required if you don't want to launch all the trucks on the roads for each test.

2.3. Workflow

Workflows can quickly orchestrate several experiments (and optionally simultaneously) of the PRODIGE application. Indeed, computer experimentation has no time constraints which must face the real experiment: a tour of several hours can be simulated in a few seconds. Among the possibilities offered by computer experimentations, we can mention the possibility to verify and debug the PRODIGE application during its development. This parallelism of tasks saves time and resources allocated to the development of the PRODIGE application. Computer experimentation also allows to quickly test new features. Once all the features established and verified, computer experimentation can create scenarios of use mimicking the behaviour of different actors (manager, customers, and drivers in the case of PRODIGE). A scenario can have several objectives:

- Quantitative: calculating and comparing several variables such as the number of kilometres travelled by products or the amount of CO2 emissions produced for a set of delivery
- Qualitative: following the different steps of the delivery of a product (e.g. respect of delivery times, compliance with cold chain, etc.)
- Analytics: observing a special case not understood, difficult or impossible to reproduce with the real system, often for scientific purposes.

To this are added demonstrations scenarios, to explain PRODIGE to public audience and track the movement of vehicles depending on the scenario chosen.

2.4. DEVS M&S

Discrete EVent Specification (DEVS) was introduced by [Zeigler 00]. This Moore based language describes a dynamic system with a discrete event approach using some typical concepts. In particular it represents a state lifetime. When a lifetime is elapsed an internal transition occurs that change the state of the model. The model also takes also into account the elapsed time while firing an external state transition triggered by an event received from outside the considered model.

The behavioural models are encapsulated in atomic models that are completed with input and output ports. Then, these models can be composed with others by connecting inputs and outputs. The composed models are called coupled models.

Generalized DEVS (G-DEVS) emerged with the drawback that most classical discrete event abstraction formalisms (e.g. DEVS) face: they approximate observed input–output signals as piecewise constant trajectories. G-DEVS defines abstractions of signals with piecewise polynomial trajectories [Giambiasi 00]. Thus, G-DEVS defines the coefficient-event as a list of values representing the polynomial coefficients that approximate the input–output trajectory. Therefore, a initial DEVS model is a zero order G-DEVS model (the input–output trajectories are piecewise constants).

G-DEVS keeps the concept of the coupled model introduced in DEVS [Zeigler 00]. Each basic model of a coupled model interacts with the others to produce a global behaviour. The basic models are either atomic or coupled models that are already stored in the library. The model coupling is done with a hierarchical approach (due to the closure under coupling of G-DEVS, models can be defined in a hierarchical way).

On the simulation side, G-DEVS models employ an abstract simulator [Zeigler 00] that defines the simulation semantics of the formalism. The architecture of the simulator is derived from the hierarchical model structure. Processors involved in a hierarchical simulation are Simulators which implement the simulation of atomic models, Coordinators, which implement the routing of messages between coupled models, and the Root Coordinator, which implement global simulation management. The simulation runs by sending different kind of messages between components. The specificity of G-DEVS model simulation is that the definition of an event is a list of coefficient values as opposed to a unique value in DEVS.

2.5. Services Simulation

We use discrete-event simulation results to mimic the behaviour of certain elements of the PRODIGE system. In [Al-Zoubi and Wainer 2010a] the authors discussed the advantages and disadvantages of several modelling and simulation environments, including the High Level Architecture (HLA) [Kuhl et al., 2000], CORBA, SOAP-based Web-services, etc. As discussed there, most of these distributed simulation middleware still lack of plug-and-play interoperability, dynamicity, and composition scalability. Based on this conclusion, they designed the first existing RESTful Interoperability Simulation Environment (RISE) middleware [Al-Zoubi and Wainer 2010b].

The main goal of RISE is providing simulation interoperability and mash-up regardless of their formalism, theory or implementation. Access to RISE is done through Web resources (URIs like a classic website URL) and XML messages using HTTP channels: GET (to read a resource), PUT (to create new resource or update existing data), POST (to append new data to a resource), and DELETE (to remove a resource). RISE allows modellers to run any number of experiment instances, whose settings and resources (URIs) are persistent and repeatable (unless deliberately removed or updated). An interface between RISE and CD++ [Wainer, 2002] allows running distributed simulations using the CD++ simulation engine.

We need to orchestrate various services to simulate the use of the PRODIGE application. We want to use the results of simulations to drive the PRODIGE application through formalized scenarios. This formalization and orchestration of services corresponds to the use of workflow of services. Workflows of services can be useful for computer experimentation by promoting replayability, sharing and interoperability [Ribault and Wainer 2012a].

2.6. Workflows of services

In [Tan et al., 2009], the authors compare the service discovery, service composition, workflow execution, and workflow result analysis between BPEL and a workflow management system (Taverna) in the use of scientific workflows. They determine that Taverna provides a more compact set of primitives than BPEL and a functional programming model that eases data flow modelling. Due to our needs, we identify that a workflow management system such as Taverna would be a better alternative than BPEL to illustrate the feasibility of our approach.

Taverna [Hull et al. 2006] is an application that facilitates the use and integration of a number of tools and databases available on the web, in particular Web services. It allows users who are not necessarily programmers to design, execute, and share workflows. These workflows can integrate many different resources in a single experiment.

Taverna workflow can contain services including:

- A service capable of running Java code directly within Taverna.
- A service to run a remote application via the REST protocol.
- A service to run a remote application via the SOAP/WSDL protocol.

A Taverna service can take inputs and produce outputs. The value of an entry can be part of the workflow (hardcoded) or a parameter to provide information during the execution of the workflow. A REST service returns systematically 2 outputs predefined: the return value of the Web service (404 if the resource is not found, 200 if everything went well, etc...), and the contents of the response (XML, HTML, ZIP, etc.). Figure 1 represents a REST service in Taverna. The number of input arguments is variable and chosen by the developer of the workflow. The number of output arguments is fixed.

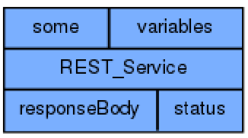


Figure 1: Taverna REST service.

In contrast, a WSDL service will find automatically, thanks to the WSDL file, the number and type of input and output. Figure 2 represents a Taverna workflow with a WSDL service in green in the middle of the figure. The service is available in Taverna after the addition of the URL of the WSDL file (such as http://xxx.xxx.xxx.xxx:8080/WS-PRODIGE/services/Identification?wsdl). Taverna offers the possibility to automatically format the input and output based on the type of parameters required by the Web service. In this example, the Web service "identificationChauffeur" that allows a driver to identify within the PRODIGE application takes as input a data type 'identificationChauffeur_input' that encapsulates 'id', 'imei', and 'pwd' input. The Web service "identificationChauffeur" produces as output a data type 'identificationChauffeur_return' that encapsulates various data such as firstName, lastName, login, etc.

Workflows are particularly suited to automate experiments, but all necessary parameters cannot always be specified in advance. In these cases, it is desirable to interact with users for decision making. Taverna offers several graphical interfaces for interacting with the user. A Taverna workflow can also contain nested workflows in a hierarchical manner. In this way, a set of simple workflows easily allow to design more complex workflows. These workflows can then be shared, reused, and adapted to new needs.

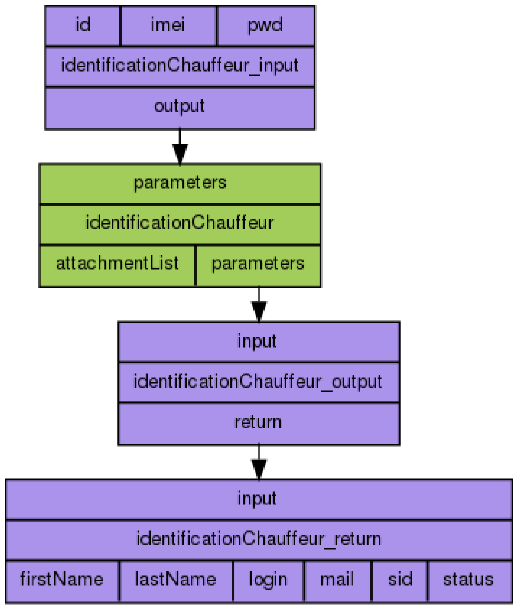


Figure 2: Taverna WSDL service.

3. CONTRIBUTION

We propose to use workflow of services as the interoperability layer among several services. In addition, we propose to integrate the G-DEVS engine as a specific workflow engine. G-DEVS is a formalism based on a state machine automaton. Workflows differ from state machines as state machine can be cyclic graphs while workflows are usually acyclic. Workflow proceeds down different branches until done. Thus, using G-DEVS coupled to another workflow engine to process a workflow could benefit from the DEVS formalism while keeping the top to bottom behaviour of the main workflow manager. Interoperability among workflow engines and applications are done using web services.

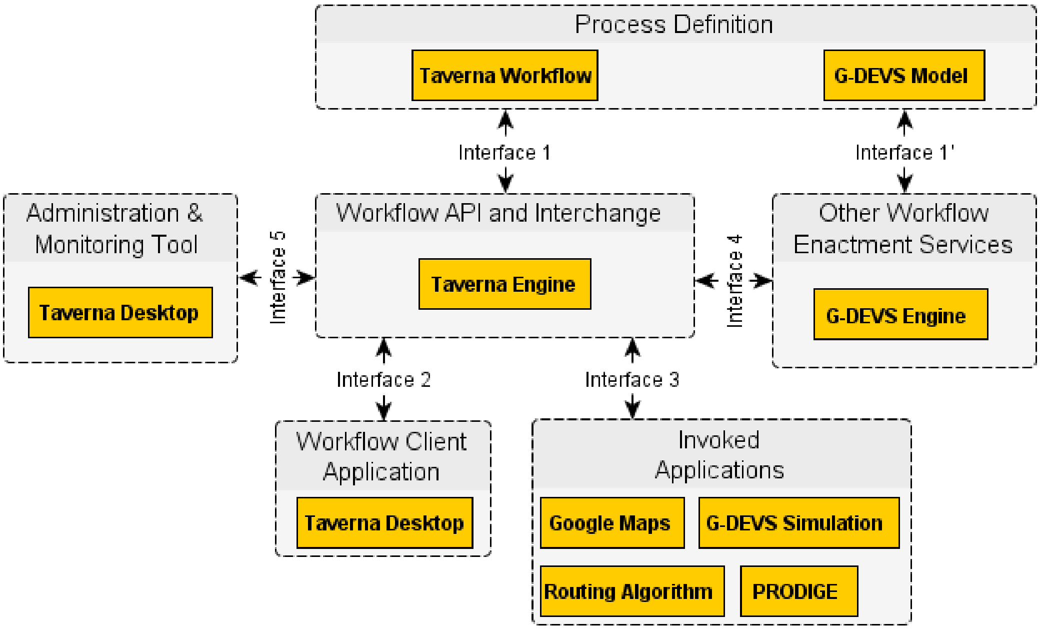


Figure 3: Workflow orchestration architecture.

3.1. Workflow Orchestration Architecture

The Figure 3 presents the orchestration architecture based on the workflow architecture by the WfMC.

We propose to use Taverna and G-DEVS as the process definition formalism to express workflows. Taverna workflow represents the main workflow that organizes all tasks and enables interoperability between services. Taverna workflow process definition will be executed by the Taverna Engine (Interface 1). G-DEVS process definition will be executed by the G-DEVS Engine (Interface 1') as an others workflow enactment services. Communication between both engines (Interface 4) will be granted by web services thanks to RISE. Taverna interprets G-DEVS workflow event and enables the interoperability with other services using RESTful or SOAP/WSDL Web services protocols. Taverna ensures interoperability between workflows (Interface 4) and among invoked applications (Interface 3) such as Google Maps, G-DEVS Simulation, Routing Algorithm and PRODIGE. Interface 2 allows Taverna workflow to interact with users through the use of the Taverna Desktop.

3.2. Taverna Workflow Model

We want to test the PRODIGE application before moving to a phase of real experimentation. Then, we want to be able to quickly test algorithm, compare studies without having to drive trucks and monopolize drivers. Taverna is used to create scenarios using the PRODIGE application through workflow showing the behaviour of the users involved in the scenario such as customer who will apply for delivery of a point to another, managers who will validate and create the tour, and the drivers who will drive trucks.

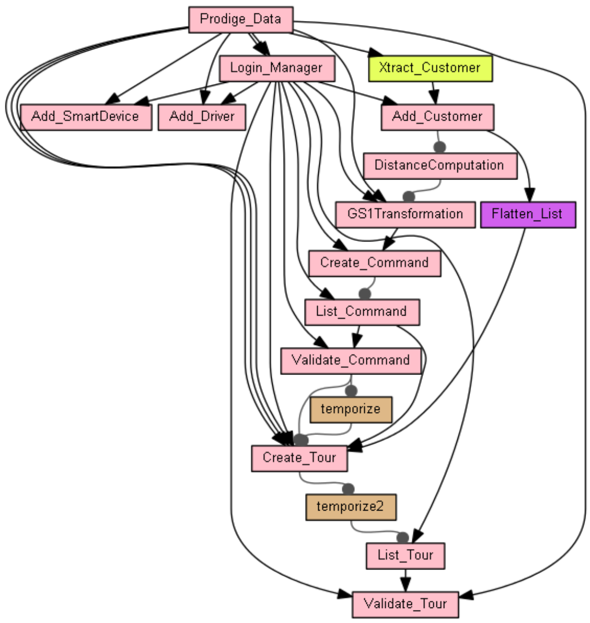


Figure 4: Taverna workflow to setup the PRODIGE system.

3.3. G-DEVS Workflow Model

In a previous work [Zacharewicz, 2011b] several G-DEVS models were introduced to represent the behaviour of the various actors of the PRODIGE system.

The main components of the PRODIGE workflow have been proposed in G-DEVS models For instance the smartphone has been described. It detail the behaviour of the smartphone and in particular it précises how this device is reacting from its environment. In this approach the synchronization was given by an HAL RTI.

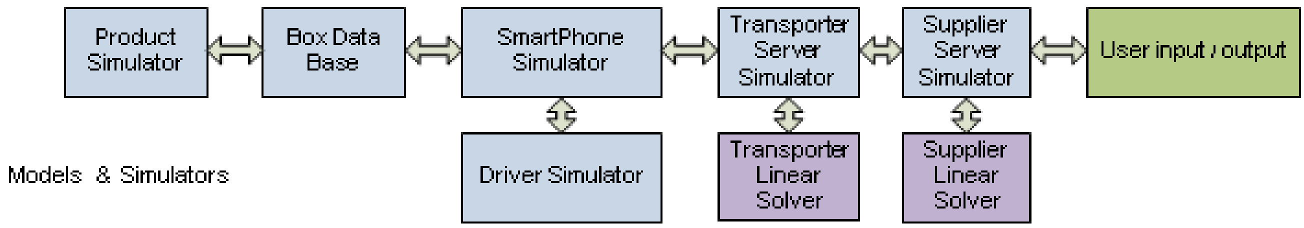


Figure 5 Workflow main components

3.4. G-DEVS Clock and Sorting Model

In this paper the interoperability is assumed by the Taverna engine that calls the services and links the different applications. Nevertheless this tool does not provide time synchronization. Two options have been envisaged. The first was using a RTI to build an HLA federation [Al-Zoubi, 2011]. This option requires reusing an existing RTI that can set up a simulation rapidly but this kind of configuration can cause overheads in the communication like discussed earlier.

Because [Zacharewicz, 2011] already uses G-DEVS models and simulators to simulate the behaviour of several components in the PRODIGE environment, the idea proposed in this research is to define a G-DEVS model dedicated to be the clock of the PRODIGE workflow. This model will define the ordering of the actions regarding their time. It also can be considered as the time driver of the simulation. In other terms G-DEVS, that is originally designed to run event driven simulation, is used in that case to run a time driven simulation.

In detail, in this paper we propose a G-DEVS model that collects messages, sorts them and triggers right on time the services call to the PRODIGE server or forward the message to the G-DEVS models that simulate the behaviour of the PRODIGE components recalled in the Figure 5. This model can receive messages both from the server as a service answer or from a G-DEVS model that send an output message as a simulation result of a local behaviour. The messages received from the server are service answers. They possess time stamp information to be used by the clock model to add the message at the right place in the queue. Then depending on the execution state of the clock it will sort the message and direct it to the proper receiver. The state of the clock can be processing a message or being available. In the first case, the approach is inspired from the conservative algorithm of [Chandy 79]. It is based on the DEVS/HLA algorithm, proposed in [Zacharewicz 2008], in particular if a message is arriving late. The message temporary blocks the simulation but will not be ignored. Then simulation is unblocked to process the next message. The receiver can be the server. In that case it prepares an output message. This output message is addressed to Taverna that transforms it to service call and then triggers the PRODIGE server. If the message is addressed to a G-DEVS model to trigger component behaviour, the message is directly sent to the appropriate G-DEVS component using the coupled model structure. In the second case (no input event to be treated) the state is transient and after a definite life time it automatically goes to another state. During transition to this state, an output message is generated in order to give the order to refresh the positioning of the trucks and product to the server according to the roadmap and geographical information extracted from Google maps. During the setting of the simulation the pace can be tuned in order to accelerate the simulation execution. Also at any simulation time the execution can be stopped to show a particular case.

3.5. Interoperability

The interoperability between G-DEVS workflow model and invoked application such as PRODIGE are ensured by the Taverna workflow. Figure 6 presents the sequence diagram of the Taverna workflow, G-DEVS workflow and PRODIGE application. The Taverna workflow represents an experimentation scenario that is executed automatically by the Taverna Engine to test the PRODIGE application and is represented by the first column of the sequence diagram. The G-DEVS workflow model represents the workflow of a smart device sending every 30 seconds a couple of GPS coordinates to simulate truck movement. The G-DEVS simulation is represented by the last two columns of the sequence diagram. Finally, PRODIGE is represented by the second column.

The sequence is expressed as follow:

1. The Taverna workflow scenario invokes the PRODIGE application to setup a new round.
2. The Taverna workflow scenario invokes and initializes the G-DEVS simulation that will create in turn the workflow model.
3. The Taverna workflow scenario executes the G-DEVS workflow. The G-DEVS simulation engine interacts with the workflow model (sendEvent).
4. The Taverna workflow scenario gets in return what is needed by the G-DEVS workflow to continue its execution (i.e. the next destination of the truck).
5. The Taverna workflow scenario invokes the PRODIGE application to request the next destination of the truck associated with the smart device simulated by the G-DEVS workflow.
6. The Taverna workflow scenario continues the execution of the G-DEVS workflow by passing the next destination event to the G-DEVS simulation.

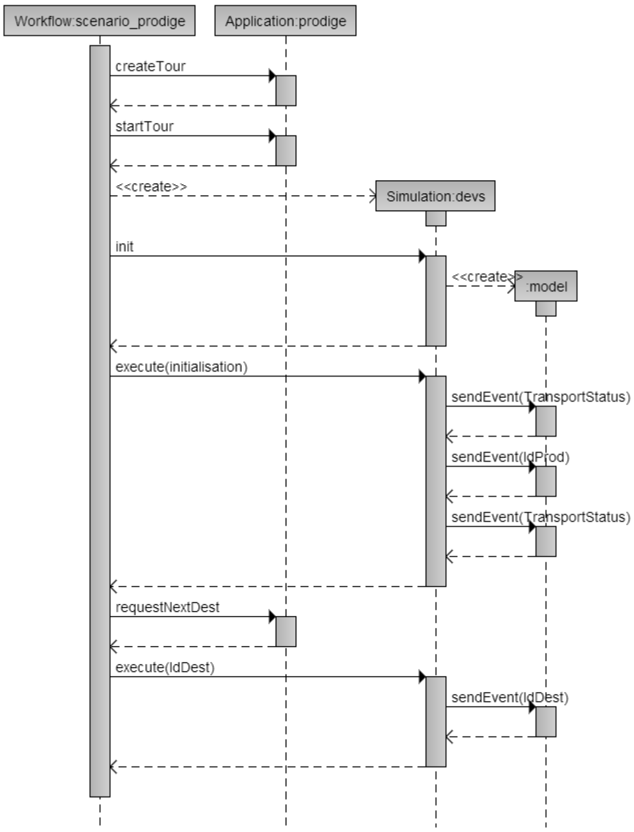


Figure 6 Interoperability sequence diagram.

4. FRAMEWORK

We have implemented the architecture and concept described in the previous section. Figure 7 represents the solution framework.

The top M&S box illustrates the virtual experiment while the bottom Real System box illustrates the real experiment. The virtual experiment is defined using Taverna workflow and DEVS simulation. The Taverna workflow mimics the behaviour of managers, clients and drivers while the DEVS simulation mimics the behaviour of smart devices. Communication between Taverna and DEVS are done through Web services thanks to RISE. The Taverna workflow communicates with the real PRODIGE application and Google Maps in the Cloud through Web service. The real experiment needs real human to manage the PRODIGE application (manager, clients) and drive trucks (divers). Communication between human and PRODIGE are done using a light web application (manager, client) or a mobile application on smart device (driver).

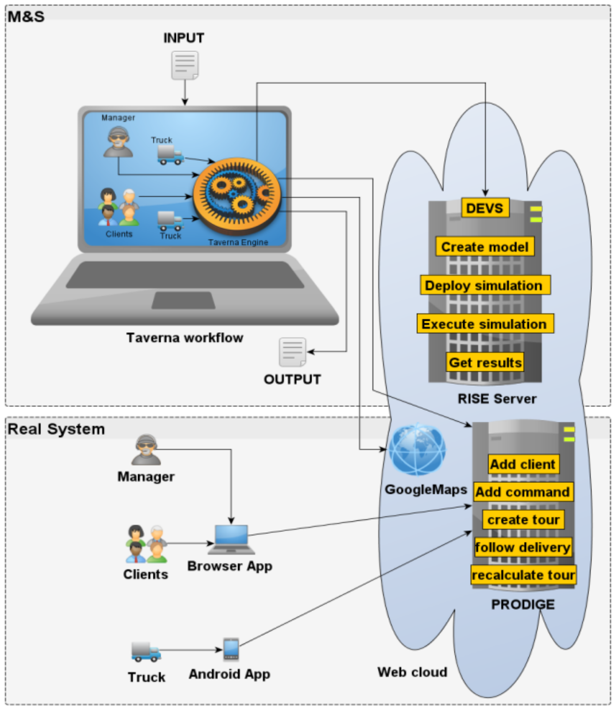


Figure 7 PRODIGE and simulation framework architecture.

4.1. Scenario PRODIGE

We created several data input sets as well as several workflows to simulate different situations and experience the PRODIGE solution before placing it onto market. Packages must be picked up and delivered regarding the two following situations:

- the delivery time windows are wide enough for it to be feasible with a single truck;
- the delivery time windows overlap and several trucks are needed to make the delivery on time.

Those two situations are done using the same generic workflows. We built another workflow to take into account hazards such as traffic jam or truck failure. Indeed, in those cases the workflow must take into account specific decision that could involve building new delivery.

4.2. Execution Example

Main experimentation workflow takes as input a XML configuration file that describes the whole experimentation. The workflow plays the role of all the actors (manager, clients, and drivers) and fills the PRODIGE system. Then, the workflow execute in parallel tours for each driver involved. The workflow retrieves the information needed on Google Maps and using G-DEVS simulation to mimic the behaviour of a real truck. The result of the execution of this workflow is directly visible in the PRODIGE web application on which you can view the current path of a truck making its tour in the region of Bordeaux, France as shown on Figure 8 . We can also imagine an experiment mixing virtual truck and real truck since there is no difference from the PRODIGE platform perspective.

5. CONCLUSION

This work has permitted to introduce a new platform for simulation of logistics and transportation. It recalled existing works that already proposed to use the G-DEVS formalism for the description of the logistic platform components. Then, it introduces the Taverna tool that will be the interoperability link to connect the services and the simulation components. Then it describes the G-DEVS model that has been proposed to serve as the clock ordering component in the system since the TAVERNA and more generally the services do not address the time synchronization consideration. The main demonstration of this paper was to show the interest of interoperability in such simulation. Here the approach was still pragmatic but the future works will propose to make the G-DEVS Clock model more generic to be reused in several service handling tools.

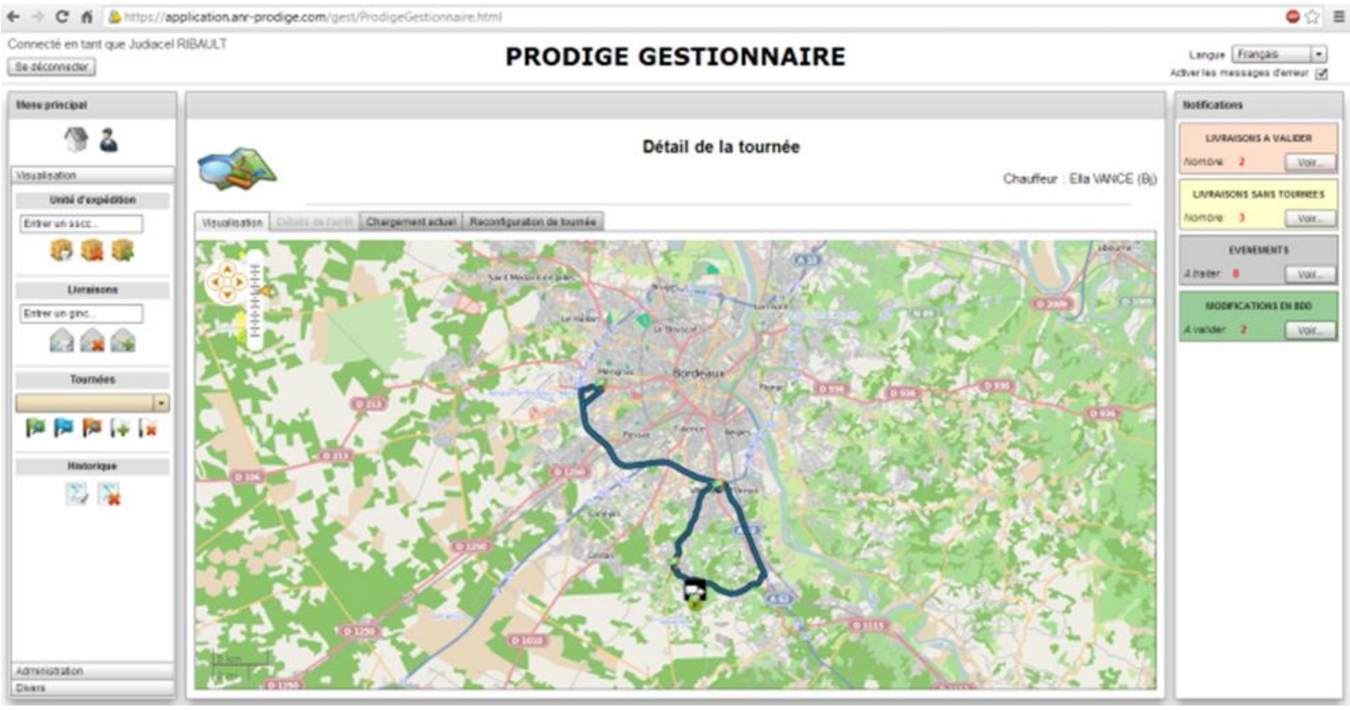


Figure 8 PRODIGE Web application.

ACKNOWLEDGEMENT

This research was partially funded by the PRODIGE project. PRODIGE is a French National Research Agency funded project (ANR-09-VTT-09-01) Nov@log labelled.

REFERENCES

Chandy, K. M., & Misra, J. (1979). *Distributed simulation: A case study in design and verification of distributed programs.* Software Engineering, IEEE, (5), 440-452.

Chen, D., Doumeingts, G. 2003. European Initiatives to develop interoperability of enterprise applications - basic concepts, framework and roadmap, *Journal of Annual reviews in Control*, 27, no.2, 151-160.

Giambiasi N., Escude B., Ghosh S. (2000) *GDEVS A Generalized Discrete Event Specification for Accurate Modeling of Dynamic Systems*. SCS Tr, 17(3) 120-134.

Richardson, L., & Ruby, S. (2007). RESTful web services. O'Reilly Media, Incorporated.

Weske, M. (2012). Business Process Management Archi-tectures. Business Process Management, 333-371.

Ribault, J., & Wainer, G. (2012, March). Using workflows and web services to manage simulation studies (WIP). Proceedings of the 2012 Symposium on TMS/DEVS Integrative M&S Symposium (p. 50).

Ribault, J., & Wainer, G. (2012, May). Simulation Processes in the Cloud for Emergency Planning. In Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012) (pp. 886-891). IEEE Computer Society.

Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., & Oinn, T. (2006). *Taverna: a tool for building and running workflows of services*. Nucleic acids research, 34(suppl 2), W729-W732.

Tan, W., Missier, P., Madduri, R., & Foster, I. (2009). *Building scientific workflow with taverna and bpel: A comparative study in cagrid.* In Service-Oriented Computing–ICSOC 2008 Workshops (pp. 118-129). Springer Berlin/Heidelberg.

Zeigler, B. P., Praehofer, H., & Kim, T. G. (2000). *Theory of modeling and simulation: Integrating discrete event and continuous complex dynamic systems.* Ac. Pr.

Kuhl, F., Dahmann, J., & Weatherly, R. (2000). *Creating computer simulation systems: an introduction to the high level architecture.* Prentice Hall PTR.

PSL, NIST, "R ationale". National Institute of Standards and Technology (NIST). 5/10/2003, last updated 1/15/2007.

Roque M., Vallespir B., Doumeingt G. 2006, "UEML:Coherent Languages and Elementary Constructs Determination", Network-Centric Collaboration and Supporting Frameworks, IFIP International Federation for Information Processing Volume 224, , pp 23-30

Al-Zoubi, K., & Wainer, G. (2010, December). *Managing simulation workflow patterns using dynamic service-oriented compositions.* Proceedings of the 2010 Winter Simulation Conference (WSC), (pp. 765-777). IEEE.

Al-Zoubi, K., & Wainer, G. (2010, December). *Rise: Rest-ing heterogeneous simulations interoperability.* In Simulation Conference (WSC), Proceedings of the 2010 Winter (pp. 2968-2980). IEEE.

Zacharewicz G., Frydman C., Giambiasi N. (2008) G-DEVS/HLA Environment for Distributed Simulations of Workflows. Simulation 84(5): 197-213

Zacharewicz G.; Labarthe O.; Chen D.; Vallespir B. (2011a) "A Multi Agent/HLA Platform for Enterprises Interoperability: Short-Lived Ontology Based. Electronic Supply Network Coordination in Intelligent and Dynamic Environment: Modeling and Implementation, IGI Global, pp. 319-346,

Zacharewicz G., Deschamps J.C., François J., (2011b). Distributed platform for advanced freight transportation systems, Computers in Industry 62, 6 597-612.

THE IMPACT OF EXTERNAL DISTURBANCES ON THE PERFORMANCE OF A CELLULAR MANUFACTURING SYSTEM

Sameh M Saad[(a)], Carlos R. Gómez[(b)] and Nabil Gindy[(b)]
[(a)]Faculty of Arts, Computing, Engineering and Sciences
Sheffield Hallam University
Sheffield, UK
[(b)]Department of Materials, Mechanical and Manufacturing Engineering
Nottingham University
Nottingham, UK

[(a), (b)] S.Saad@shu.ac.uk

ABSTRACT
Cellular manufacturing has been proposed as an approach to cope with the uncertainty characteristic of customer driven markets. However, even cellular manufacturing systems are prone to the effects of varying demand patterns. In this study, the effects of some aspects related to demand variation such as the arrival of material, the variety of products and the variation in product mix are investigated to identify those system characteristics that -within the context of cellular manufacturing systems- represent an advantage in the presence of such disturbances. To do so, discrete event simulation is used to conduct the experimentation by modelling a cellular manufacturing system. Additionally, statistical design of experiment is employed to identify the factors contributing to higher system performance. The results show that, in spite of the demand related disturbance, machines with low set-up duration and highly skilled operators constitute the most important characteristics of an efficient manufacturing cell.

Keywords: cellular manufacturing system, manufacturing disturbances, simulation

1. INTRODUCTION

The current economic environment is characterized by customers having more power than producers in terms of shaping market demand. As a result of customer driven market, low volume and high variety have become an important characteristic in manufacturing. In the presence of these exigencies companies must develop the capability to respond in the shortest time, with the highest levels of quality and with the lowest possible cost. One manufacturing approach to meet the expectations customer driven markets is the cellular manufacturing configuration. This system configuration is characterized by the grouping of different types of machines according to the process combinations occurring within a family of parts, which means that material flows differ for different parts of the same family. However manufacturing cells are not exempt from the influences of disrupting factors or disturbances. According to Deane and Yang (1992), some of those factors affecting cell performance are: machine capacity constraints, complexity of job routing requirements, demand volume, demand pattern and product mix characteristics. In this study particular attention is paid to the last three factors which are originated outside the limits of the system and therefore are considered as external disturbances.

Customer driven markets have an important influence on the arrivals of materials into the manufacturing system. On the one hand, periods of low demand lead to a low utilization of the system's resources; on the other hand, periods of high demand lead to an increase in the arrival rate. Even though an increase in the arrival rate is associated with an increase in throughput and therefore an increase in income, similarly, an increase in the arrival rate inevitably leads to an increase in costly work-in-process (WIP) inventory. The arrival of materials and its influence on system's performance has been approached by a number of authors, among those Tielemans and Kuik (1996) studied the relationship between batching of arrived orders and WIP in order to reduce lead time; they recognised the impact long waiting times could have on system's performance. Chikamura *et al.* (1998) tested the influence of several lot arrival distributions on 7 production dispatching rules. The authors noticed that under most of the arrival scenarios, the best results were observed by a dispatching rule considering variables such as set-ups, waiting times and processing times. Govil *et al.* (1999) focused on the time of new lot arrivals in order to determine ways to predict average queue length at manufacturing resources. Prabhu (2000) claimed that arrival time determines the evolution of events in the manufacturing system, the sequence in which parts are processed and the machine idle time between processing parts. Given the important impact of arrival time on system's performance, this author proposed the arrival time to be selected as a control variable in manufacturing systems. Moreover, Van Ooijen and Bertrand (2003) investigated the effects varying arrival rates have on throughput and WIP for a job shop; they concluded that an acceptable throughput would not necessarily imply a high arrival rate. In addition, they identified a trade-off between the costs associated with controlling the arrival rate and the revenues obtained by throughput.

Another implication of customer driven markets is an increasing demand for more variety. A higher product choice leads to more problems occurring in manufacturing systems; this is due to the level of complexity increasing along with variety. Research on product variety is a divergent topic; some views claim a significant impact of product variety on manufacturing performance, whereas other views suggest that there is actually no impact. MacDuffie *et al.* (1996) identified a trade-off associated to product variety; the authors noticed that whereas there is a higher revenue resulting from a wider variety, there are related higher costs and a loss of economies of scale as well. Although Fisher and Ittner (1999) acknowledged some common negative effects of variety, they also recognised that variety leads to benefits such as increased revenue. Berry and Cooper (1999) claimed that, in order to gain competitive advantage through product variety, it is necessary a proper alignment between marketing and manufacturing strategies in terms of process and infrastructure along with pricing and inventory. Randall and Ulrich (2001) argued that variety does not necessarily mean higher performance; they stated that regardless of variety strategies, the proper alignment between the supply chain and the product variety strategy is what is important. Thonemann and Bradley (2002) investigated the effects of product variety on supply chain performance and found that variety has an important effect on costs, especially when set-ups are significant. Fujimoto *et al.* (2003) stated that more variety causes less efficiency and higher costs; they presented a methodology to manage variety by synthesizing product-based and process-based approaches. Zhang *et al.* (2007) evaluated the impact of response time and product variety strategies on system's performance; they concluded that a higher performance is achieved when both strategies are combined.

The consideration of a wider product variety inevitably leads to another problem for manufacturing systems, which is the variation within the product mix. This is caused by a varying demand for different products, especially when a number of products are at different stages in their life cycle. Variation in product mix also has an important effect on system's performance mostly due to an increased complexity in the system; such complexity is caused by more processing flows and varying production quantities. Among some authors investigating product mix, Deane and Yang (1992) investigated the impact of product mix on the performance of a manufacturing cell; the authors found that in the presence of set-up times, reductions in terms of flow time can be realized by increasing the homogeneity of products. Anderson (1995) analyzed product mix heterogeneity and performance; she confirmed that an increase in manufacturing costs is associated with increases in the number and severity of set-ups and with an increased heterogeneity in process specifications of the product mix. This same author (2001) later found that product variety may have worst consequences for quality than for efficiency. Seifoddini and Djassemi (1996) recognized that changes in product mix lead to performance deterioration and presented a procedure for performance evaluation under product mix variations. The same authors (1997) also compared the performance of two manufacturing system configurations, namely job shop and manufacturing cell for a range of product mix variations; they found the cellular configuration showed the best results only when there were small changes in product mix. Liang *et. al*. (2011) studied the combination of virtual cells idea to construct new manufacturing systems in response to changing market dynamics. Egilmez *et. al* (2012) developed a non-linear mathematical model to the stochastic cellular manufacturing systems design problem to cope with a particular risk level, then simulated the obtained results to validate the proposed model and assess the performance of the designed cellular manufacturing system

The purpose of this study was to identify those components in a cellular manufacturing system contributing to maintain a higher system performance under the influence of external disturbances such as demand variations in both volume and pattern. This has been achieved by using the combined advantages of discrete event simulation and statistical design of experiments. The former was used to model a cellular manufacturing system with its main components; the latter provided the analysis structure for the identifications of those components with a major significance in terms of system performance. The consideration of both tools provided the capacity to adopt a wider perspective in the analysis and, therefore, not only did it facilitate the study of particular system components but also facilitated the study of the interactions occurring within the system.

2. RESEARCH METHOD

2.1. Simulation model

Discrete event simulation has been used to represent a semi-automated cellular manufacturing system consisting of 9 different work centres. Each work centre comprises one input buffer, one machine and one output buffer. All the work centres are connected by an automated material handling system. Each work centre is assisted by one of the six operators within the cell whose job basically consists on loading, controlling and unloading machines. Figure 1 graphically represents the cellular system previously described.

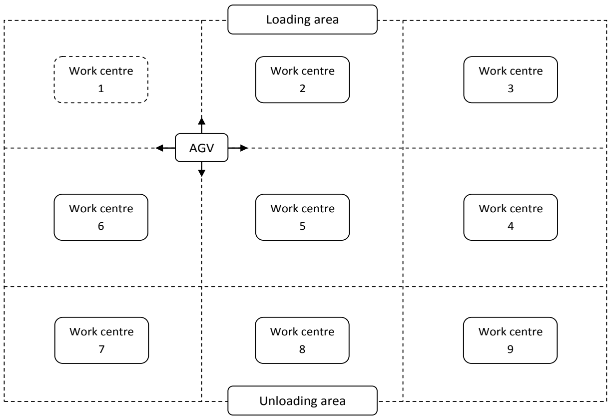


Figure 1: Cellular Manufacturing System Layout

2.2. Simulation model operation assumptions

The following are list of the assumptions considered during the development of the simulation model:

2.2.1. Parts

- Parts arrive in the system one at a time and following an exponential distribution with an average inter-arrival time of 45 minutes
- There are five different products involved; each product with different processing requirements, i.e. different processing times and routes. Process routing is fixed for each of the products.

2.2.2. Machines

- Each machine represents a specific manufacturing process within the system, therefore their different operative features.
- Machines can process only one piece at a time.
- Although all of the machines are assumed to follow a normal distribution in both processing and set up times, the times are different from each other.
- There is a different usage cost per minute associated to each machine.
- Machines do not have any automation level, therefore each machine do require an operator.
- It is assumed that all machines breakdown from time to time, consequently a different efficiency level has been predefined for each machine.
- When machines fail, repairs are assumed to be done by external personnel (not considered for the purposes of this research). Machine repairs are assumed to follow an exponential distribution with different average times for each machine.

2.2.3. Buffers

- Blocking does not occur.
- Buffer capacity is limited; all the buffers have the same capacity. There is a storage cost per item per minute associated to the capacity, i.e. the higher the capacity the higher the storage cost.
- Parts in buffers are prioritized according to FIFO dispatching rule, i.e. parts are dispatched either into a machine or vehicle considering a first come first served rule.

2.2.4. Operators

- Operators have different abilities; in consequence labour cost is associated to the skill level.
- Operators are assumed not to be always available, therefore different availability percentages and absence times have been specified for each operator.
- Travelling times for operators have not been considered.

2.2.5. AGV

- The material handling system is totally independent from human operators.
- The AGV travels at a constant speed along a fixed route connecting all the work centres.
- Material handling costs are omitted and no vehicle breakdowns are assumed.
- The AGV's travelling time is determined based on its speed.

2.2.6. Finished product

- Revenue per finished product regardless of its type has been assumed.

2.3. Model Verification

Model verification can be carried out in three different and complementary ways: Checking the code, performing visual checks, and inspecting output reports (Robinson, 1994). Code checking was facilitated by the capabilities of the simulation software, which made possible to interactively check the coding line by line. Visual checks were performed by keeping track of parts progressing throughout the system, allowing the behaviour of all the components intervening along the process to be monitored. Additionally, the model was run in an *event-by-event* mode in order to complement the verification process. This verification procedure made possible to guarantee that each element within the model would behave as it was originally intended. The last method of model verification consisted in checking the outputs of the main components within the model; to do so 30 replications, each with a run time of 400 simulation-hours, were conducted. After analysing some of the most important system outputs it was possible to confirm that all the model components performed according to what had been defined during the model coding process.

2.4. Model validation

Model validation provides the confidence during the experimentation stage and is basically concerned with the extent to which a certain model is representative of a real system. The level of representation will be judged upon the viability of making decisions based on the information provided by the simulation model. Ideally, a model would be better validated when compared to a real system (Pidd, 1993); however, models do not always represent real systems. Because the latter is the case in the present research, it was not necessary to compare the model with either empirical data or the behaviour of a real system (Maki and Thompson, 2006). Validation techniques are classified in two groups, namely subjective techniques and objective techniques. Objective validation-techniques do require the existence of real systems in order to establish input-output comparisons between systems. Subjective techniques, as their name imply, does not necessarily require the existence of a real system since they are more dependent on the experience and "feelings" of its developers (Banks, 1998). The proposed model has been validated using a sensitivity analysis as a subjective validation method. The sensitivity analysis capability is a built-in feature in Simul8; its function is to test the assumed probability distributions in terms of how sensitive the results are to changes in these inputs. A number of probability distributions particularly related to machine processing times and set-ups have been randomly selected to be tested. The sensitivity analysis confirmed the validity of the assumptions.

2.5. Experimental design

A minimum model warm-up period of 50 hours was calculated by using Welch's graphical method. A minimum run length of 220 hours was also calculated by a graphical approach. Both approaches are described in Robinson's (1994).

To measure the performance of the modelled manufacturing cell three different response variables were selected, namely number of completed parts, total cost and average time in the system. Regarding the design factors, three aspects related to work centres and three aspects related to the material handling sub-system were selected, those are the following:

1. The skill level of operators. It is determined by the number of different machines a single operator is able to control.
2. The capacity of inter-storage buffers. It is related to the maximum number of parts the system is able to hold.
3. The duration of machine set-ups. It is the time it takes for machines to switch from producing one specific type of part to producing a completely different type of part.
4. The number of AGVs. It is related to the total number of material handling vehicles within the system.
5. The speed of AGVs. It is the distance covered by material handling vehicles during a specific period of time.
6. The loading capacity of AGVs. The maximum number of parts a material handling vehicle can transport between work centres.

To investigate the effects of external disturbances on the performance of the modelled manufacturing cell, four different scenarios were considered as shown in the following subsections.

2.5.1. Irregular pattern of raw material arrivals

This scenario simulates the situation in which, due to some external cause such as demand variation or supply delays, the pattern of raw-material-arrivals into the system is disrupted in such a way that there is a higher variation in both the time between arrivals and the arriving number of parts. The variation in the arrival pattern in this scenario with respect to the arrival pattern in the baseline model is contrasted in figures 2 and 3 below.

In order to generate the pattern of arrivals in figure 3 a gamma distribution has been used. The parameters used in the gamma distribution generating the interarrival times in this scenario are a shape parameter (α) of 10 minutes and a scale parameter (β) of 38 minutes. Additionally, a normal distribution with an average of 10 parts and a standard deviation of 5 parts has been also considered to generate the variation in the arriving batches.

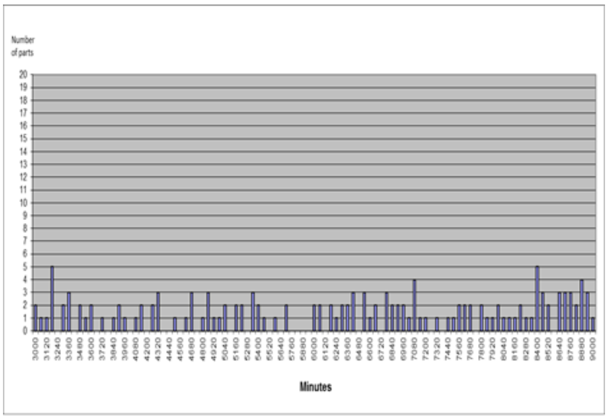


Figure 2: Pattern of Material Arrivals per Time Interval in the Baseline Model

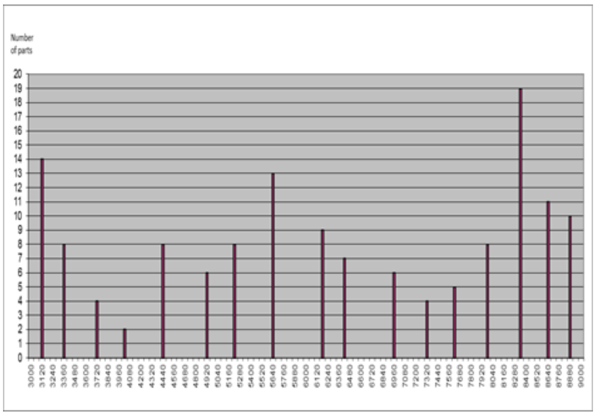


Figure 3: Pattern of Material Arrivals per Time Interval in the Disturbance Scenario

2.5.2. Increased arrivals of raw material scenario

As opposed to the previous scenario, in this scenario the pattern of material arrivals is not modified but amplified in order to simulate a condition where the MS needs to cope with an unexpected increase in production orders. Figure 4 shows the difference between arrivals of raw material in the baseline model and arrivals of material in the disturbance scenario.

To simulate the disturbance condition shown in the figure above, the original interarrival time in the baseline simulation model has been decreased from 45 to 30 minutes with no change in the probability distribution originating the arrivals.

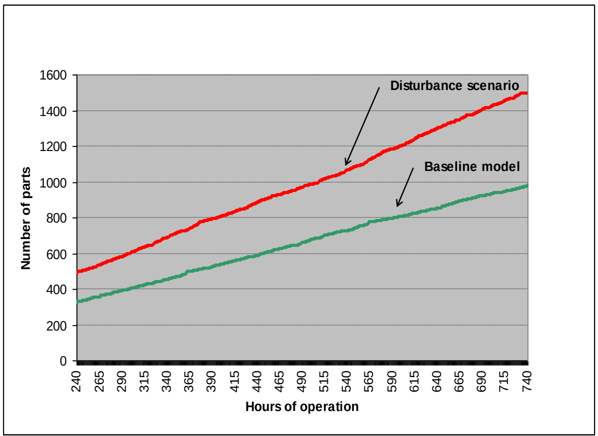


Figure 4: Number of Raw Materials Arrivals

2.5.3. Increased product variety scenario

The range of parts produced by the system has been increased from 5 to 10 parts in this scenario; each part has different processing characteristics. Since the purpose is to investigate product variety and not product mix variation, the product mix range in this scenario with respect to the baseline model has been increased by only 2%. Table 1 compares the increased product variety in this scenario with respect to the product variety in the baseline model.

Table 1: Comparison of Product Varieties

Product	Baseline model	Disturbance scenario
1	23%	10%
2	17%	9%
3	20%	7%
4	18%	8%
5	22%	6%
6	N/A	13%
7	N/A	14%
8	N/A	11%
9	N/A	12%
10	N/A	10%
TOTAL	**100%**	**100%**

2.5.4. High variation in product mix scenario

As opposed to the previous scenario where a wider choice of products is investigated, this scenario explores the demand variation existing among products. To simulate this scenario, the same five original products defined in the baseline model are considered, however, the product mix has been adjusted in order to reflect a bigger difference in the demand for each product in relation to the rest of the products. Table 2 illustrates a comparison between the original product mix and the mix considered in this scenario.

Table 2: Comparison of Product Mix

Product	Baseline	Disturbance
1	23%	18%
2	17%	3%
3	20%	26%
4	18%	8%
5	22%	45%
total	**100%**	**100%**
Range	**6%**	**42%**

It can be noticed from the table 2 above that the range in the product mix for the disturbance scenario is considerably larger than the range in the baseline model.

3. MODEL REPLICATIONS AND DATA ANALYSIS

The number of necessary replications for each simulation scenario was determined by calculating a maximum error estimate out of a series of initial model replications. The maximum error estimate together with a desired error was taken into account to determine the required number of replications for each model. According to such calculation, a minimum of 250 replications per model were enough to guarantee statistical reliability.

Considering that there were 6 design factors involved, each at two levels, a 2^6 full factorial design was employed. Given the high variation in the resulting data related to the responses cost and time, the original data has been normalized using a log transformation. Subsequently an analysis of variance was conducted to identify the significant factors. Main effects plots and interaction plots were used to identify factor levels and factor interactions respectively. Minitab was the statistical software used to analyse the data generated by each simulation scenario.

4. RESULTS

The following sections report on the obtained results including the analysis of variance. Due to space limitations, all the analysis of variance tables and the interactions effects are not included but will be presented at the conference.

4.1. Irregular pattern of raw material arrivals

After conducting an analysis of variance and calculating the factor effect estimates for each of the considered responses, the most important effects were confirmed by the main effects plots in figures. 5, 6 and 7.

In terms of the number of completed parts, figure. 5 and the factor effect estimates in appendix 2 indicated that four influential factors are: high operator skills, high buffer capacity, high number of vehicles and low duration of machine set-ups; all with a combined percent contribution of approximately 61%. Although high vehicle speed also appeared as main effect in figure. 5, its percent contribution was only of 1%.

In figure 6 two key factors to minimize cost were identified, namely low buffer capacity and low duration of machine set-ups; both with a combined percent contribution of 96%. Although low operator skills and low number of vehicles also appeared in figure. 6 as main effects, the calculation of effect estimates in appendix 2 indicated a low percent contribution of 1.5%.

Moreover, high operator skills and low duration of machine set-ups were identified by figure. 7 as the most important factors to achieve minimum time in the system; both factors with a combined percent contribution of approximately 89% according to appendix 2. Low buffer capacity, high number of vehicles and high vehicle speed, although also identified as main effects, showed a combined percent contribution of only 3%.

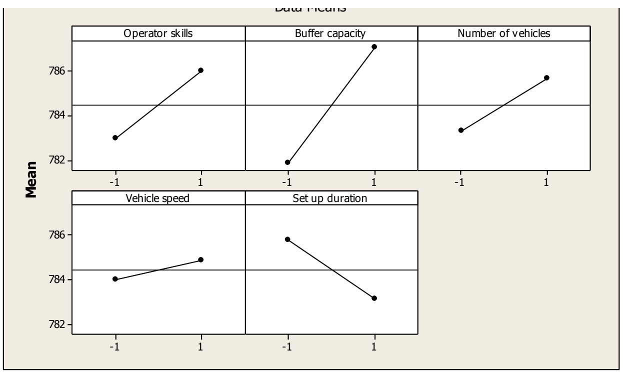


Figure 4: Main Effect Plot for Parts

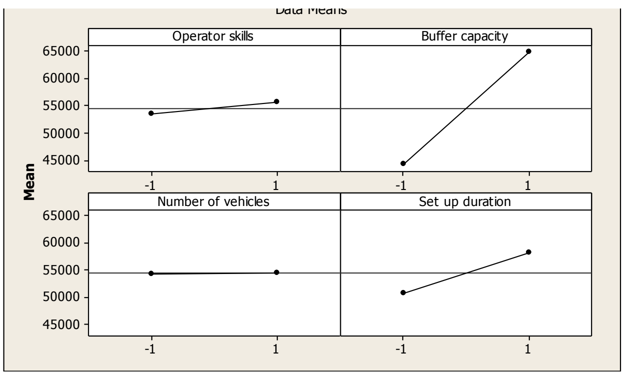


Figure 5: Main Effect Plots for Cost

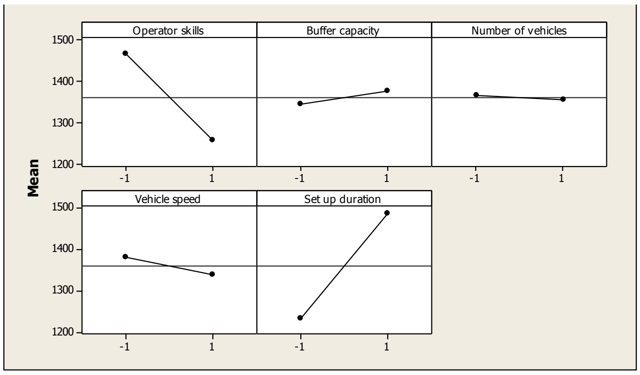


Figure 6: Main Effect Plot for Time

4.2. Increased arrivals of raw material

The main effect plots in figures. 8, 9 and 10 validated the significant factors previously identified by the analysis of variance.

As shown by figure 8, to achieve a maximum number of completed parts, high operator skills, high buffer capacity and low duration machine set-ups were the most significant factors with a combined percent contribution of approximately 67%.

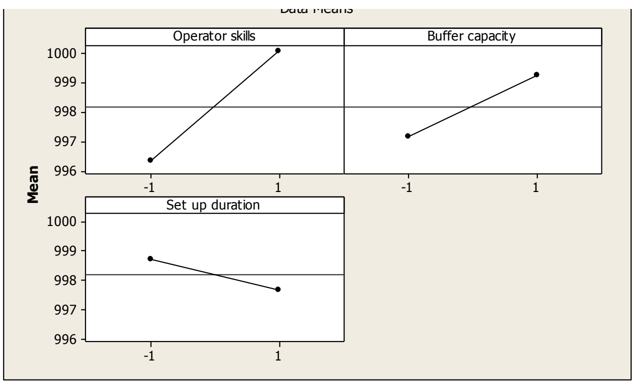


Figure 7: Main Effect Plot for Parts

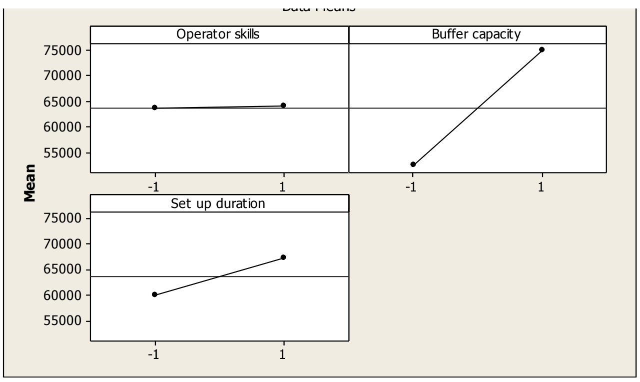


Figure 8: Main Effect Plot for Cost

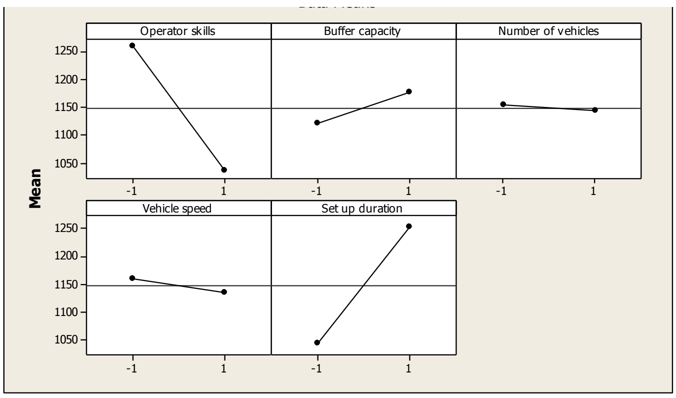


Figure 9: Main Effect Plot for Time

As shown by figure 8, to achieve a maximum number of completed parts, high operator skills, high buffer capacity and low duration machine set-ups were the most significant factors with a combined percent contribution of approximately 67% according to appendix 3.

Figure 9 confirms that low buffer capacity and low duration of machine set-ups, both with a combined percent contribution of approximately 97% according to appendix 3, were the most significant factors to achieve minimum cost.

Concerning a minimum time in the system, figure 10 shows that high operator skills, low duration of machine-set ups and low buffer capacity are the three most influential factors with a combined percent contribution of approximately 88% consistent with appendix 3. A high number of vehicles and high vehicle speed, both with a combined percent contribution of nearly 1%, were not influential enough.

4.3. Increased product variety

After conducting an analysis of variance the most significant factors in a scenario characterised by a wider product variety have been confirmed by figures 11, 12 and 13.

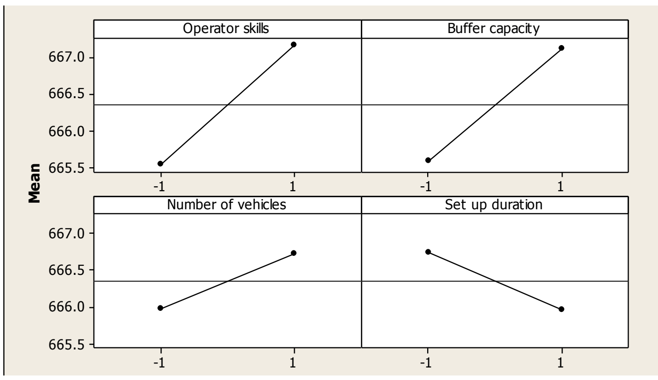


Figure 10: Main Effect Plot for Parts

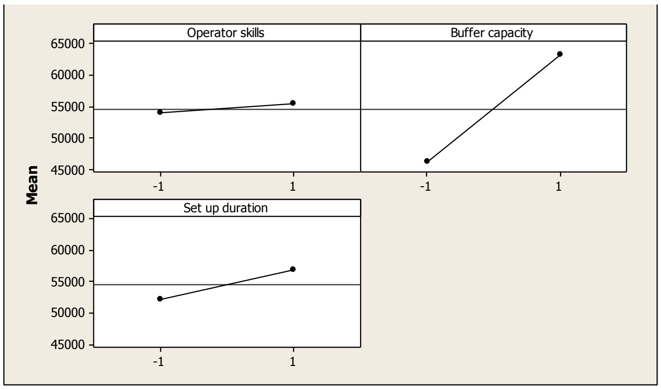


Figure 11: Main Effect Plot for Cost

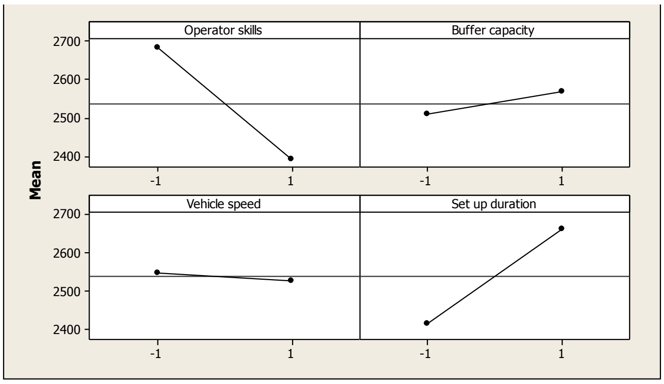


Figure 12: Main Effect Plot for Time

Figure 11 indicates that, to achieve a maximum number of completed parts, the first and most significant factor was high operator skills with a percent contribution of 26%, followed by high buffer capacity with a percent contribution of 23%; the third and fourth important factors were low duration of machine set-ups and high number of vehicles with percent contributions of 6% and 5.6% respectively.

Figure 12 shows that there were only two main factors for achieving minimum cost, those were low buffer capacity and low duration of machine set-ups, both with a combined percent contribution of approx. 97%.

Figure 13 confirmed that, to minimize time in the system, high operator skills and low duration of machine set-ups were key factors with a combined percent contribution of approx. 87% (see appendix 4 for the percent contributions of each factor in terms of the three considered response variables).

4.4 High variation in product mix

Figures 14, 15 and 16 show the most influential factors in achieving higher performance in a scenario characterized by high variation in product mix. See appendix 5 for the analysis of variance from where the initial significant factors were identified.

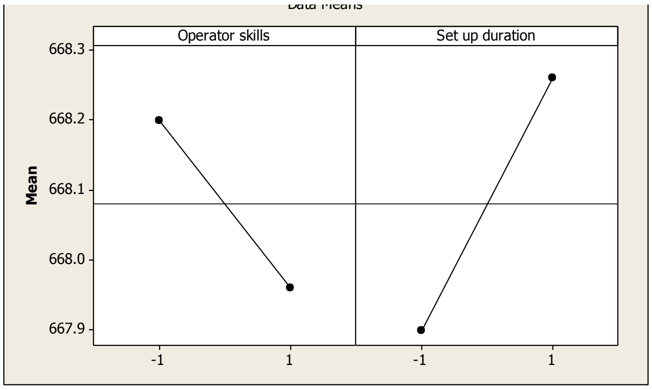


Figure 13: Main Effect Plot for Parts

Figure 14 shows that, in terms of a maximum number of completed parts, the only two significant factors were high duration of machine set-ups and low operator skills, both with a combined percent contribution of approx. 24%. The analysis of variance in appendix 5 shows that an important interacting factor to achieve a higher number of parts was high vehicle speed.

Figure 15 confirms that, in terms of minimum cost, low buffer capacity was the most influential factor with a percent contribution of 92% according to appendix 5. Low operator skills and low duration of machine set-ups were also important factors with a significantly lower percent contribution of 4% each.

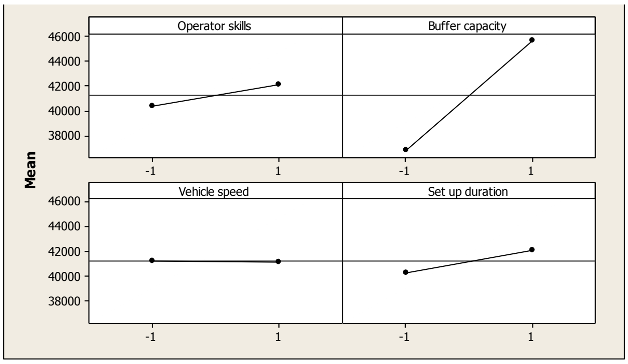


Figure 14: Main Effect Plot for Cost

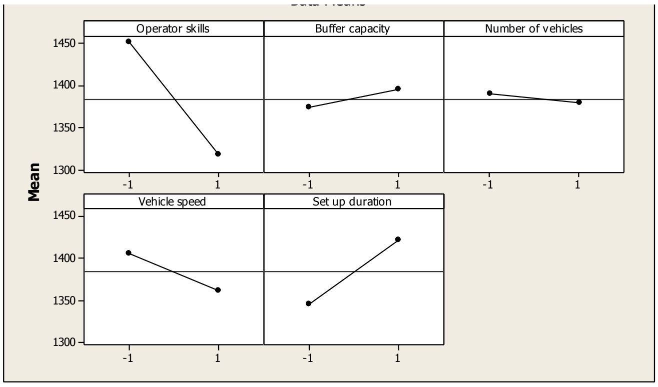


Figure 15: Main Effect Plot for Time

Although figure 16 shows the existence of five significant factors in terms of minimum time in the system, according to appendix 5 only three factors were truly significant; those were high operator skills, low duration of machine set-ups and high vehicle speed; all with a combined percent contribution of approx. 92%.

5. DISCUSSION AND FUTURE WORK

As it was mentioned at the beginning of this study, demand aspects such as volume and pattern are among some of the factors affecting manufacturing cell's performance. As Van Ooijen and Bertrand (2003) claimed, periods of high and low demand lead to unbalanced workload and variation in resource utilization. The same authors identified the trade-off existing between a higher throughput resulting from an increase in the arrival rate and higher costs particularly associated to high levels of WIP inventory. In addition to approaches like arrival rate control policies, other mechanisms to cope with material arrivals associated problems have been identified in this study. To achieve a maximum number of completed parts in scenarios characterized by either irregular or increased arrivals of raw material, the most important factors identified in this study were highly skilled operators and high buffer capacity. These same factors also were the most influential to achieve a maximum number of parts in a scenario characterized by increased product variety. Highly skilled operators are important especially during periods of high demand in order to guarantee better resource utilization. A higher buffer capacity is similarly necessary to store the excess of WIP inventory originated during those periods. In scenarios characterized by a high variation in product mix, dedicated operators are more suitable since old products, going through the last stage of their life cycle, will experience a low demand allowing more resource attention to be paid on products with higher demand.

The trade-off existing between the responses throughput and high costs resulting from high levels of WIP inventory, has been mentioned in the previous paragraph. The experiments conducted in this study have shown that, in all of the considered disturbance scenarios, the two most important factors were low buffer capacity and low duration of machine set-ups. High costs resulting from high WIP inventory levels can be overcome by limiting WIP levels; this can be achieved by setting work centres with low buffer capacity. Scenarios of variation in the demand pattern like product variety and product mix are particularly associated to higher costs. On the one hand, a product variety scenario implies higher costs associated to lower economies of scale, more set-ups and lower labour productivity (Thonemann and Bradley, 2002). On the other hand, a product mix scenario involves increased manufacturing costs resulting from increased heterogeneity in process specifications of a product mix (Anderson, 1995). Therefore low duration of machine set-ups is another characteristic that needs to be considered to reduce costly WIP inventories, particularly those caused by variation in the demand pattern.

To cope with a changing environment, a crucial task to a quicker throughput and to an improved performance is lead time reduction. In order to achieve a reduced job flow time, improvements in delivery speed have been proposed along with improvements in WIP inventory and response to market requirements (Deane and Yang, 1992). In the manufacturing system analysed in the present study, two essential features to achieve the minimum time in the system were highly skilled operators and low duration of machine set-ups. These two characteristics were the most significant ones for all of the considered scenarios. Similarly, high speed of vehicles was other factor that resulted significant exclusively for the scenario involving high variation in product mix.

Looking at the whole picture, the system's features that consistently resulted significant for all the considered scenarios, in terms of the three considered performance measures, were low set-ups duration in the first place followed by highly skilled operators. Cellular manufacturing systems with similar operating characteristics to those specified in this study, and which are constantly facing frequent changes in the volume and pattern of the demand, may find that, counting on versatile machines able to accomplish quick changeovers and skilled human operators able to keep the system operating under different circumstances, are key characteristics to maintain an acceptable performance.

Future work on the topic could adopt a wider perspective on the origin of disturbances affecting manufacturing systems. Both internal and external disturbances could be considered to investigate their effect on performance and to identify aspects providing manufacturing systems the capability to cope with a number of disrupting situations. To complement the present study, it would be interesting to consider a range of variation in the intensity of disturbances and identify how certain system characteristics become significant at varying disturbance intensities. Moreover, other systems layouts could be investigated to confirm the advantages offered by cellular manufacturing against uncertainty. Another important aspect to consider in future research is related to the implications for the manufacturing system to hold a certain degree of flexibility, i.e. the effects, in terms of different performance measures, of system adjusting to a number of situations, together with the trade-offs involved.

REFERENCES

Anderson, S. W., 1995. Measuring the impact of product mix heterogeneity on manufacturing overhead costs. *Accounting Review,* 70**,** 363-387.

Anderson, S. W., 2001. Direct and indirect effects of product mix characteristics on capacity management decisions and operating performance. *The international journal of flexible manufacturing systems,* 13**,** 241-265.

Banks, J., 1998. *Handbook of Simulation: Principles, methodology, advances, applications, and practise*. 1998, New York: John Wiley & Sons.

Berry, W. L. and Cooper, M. C., 1999. Manufacturing flexibility: methods for measuring the impact of product variety on performance in process industries. *Journal of Operations Management,* 17**,** 163-178.

Chikamura, A., Nakamae, K. and Fujioka, H., 1998. Influence of lot arrival distribution on production dispatching rule scheduling and cost in the final test process of LSI manufacturing system. *IEE Proc.- Sci. Meas. Technol.*, 145**,** 26-30.

Deane, R. H. and Yang, J. Q., 1992. Product mix selection and closed manufacturing cell flow time performance. *International Journal of Production Economics,* 28**,** 157-169.

Egilmez, G., Suer, G.A. and Huang, J., 2012. Stochastic cellular manufacturing system design subject to maximum acceptable risk level, *Computers & Industrial Engineering*, 63, 842-854.

Fisher, M. L. and Ittner, C. D., 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science,* 45**,** 771-786.

Fujimoto, H., Ahmed, A., Iida, Y. and Hanai, M., 2003. Assembly process design for managing manufacturing complexities because of product varieties. *International Journal of Flexible Manufacturing Systems,* 15**,** 283-307.

Govil, M., Minis, I. and Proth, J., 1999. Prediction of average queue length in manufacturing resources at the time of new lot arrival. *Applies Stochastic Models in Business and Industry,* 15**,** 111-121.

Liang, F., Fung, R.Y.K. and Jiang, 2011. Modelling approach and behaviour analysis of manufacturing resources in virtual cellular manufacturing systems using resource element concept. *International Journal of Computer Integrated Manufacturing*, 24, 12, 1168-1182.

Macduffie, J. P., Sethuraman, K. and Fisher, M. L., 1996. Product variety and manufacturing performance: Evidence from the International Automotive Assembly Plant Study. *Management Science,* 42**,** 350-369.

Maki, D. and M. Thompson, 2006. *Mathematical Modeling and Computer Simulation,* Belmont, Ca.: Thomson Higher Education.

Pidd, M., 1993. *Computer simulation in management science*. 3rd. ed. 1993, Chichester: John Wiley & Sons.

Prabhu, V., 2000. Performance of real-time distributed arrival time control in heterarchical manufacturing systems. *IEE Transactions* 32**,** 323-331.

Randall, T. and Ulrich, K., 2001. Product variety, supply chain structure, and firm performance: Analysis of the US bicycle industry. *Management Science,* 47**,** 1588-1604.

Robinson, S., 1994. *Successful Simulation: A Practical Approach to Simulation Projects,* London, McGraw-Hill.

Seifoddini, H. and Djassemi, M., 1996. Sensitivity analysis in cellular manufacturing system in the case of product mix variation. *Computers & Industrial Engineering,* 31**,** 163-167.

Seifoddini, H. and Djassemi, M., 1997. Determination of a flexibility range for cellular manufacturing systems under product mix variations. *International Journal of Production Research,* 35**,** 3349-3366.

Thonemanu, U. W. and Bradley, J. R., 2002. The effect of product variety on supply-chain performance. *European Journal of Operational Research,* 143**,** 548-569.

Tielemans, P. and Kuil, R., 1996. An exploration of models that minimize lead time through batching of arrived orders. *European Journal of Operational Research,* 95**,** 374-389.

Van Ooijen, H. and Bertrand, J., 2003. The effects of a simple arrival rate control policy on throughput and work-in-process in production systems with workload dependent processing rates. *International Journal of Production Economics,* 85**,** 61-68.

Zhang, X., Chen, R. Q. and Ma, Y. B., 2007. An empirical examination of response time, product variety and firm performance. *International Journal of Production Research,* 45**,** 3135-3150.

AUTHORS BIOGRAPHY

Professor Sameh M. Saad, BSc (Honours), MSc, PhD, CEng, MIET, MILT, is Professor of Enterprise Modelling and Management, Postgraduate Research Coordinator and MSc/MBA Course Leader, in the Department of Engineering, Faculty of Arts, Computing, Engineering and Sciences, Sheffield Hallam University, UK. His research interests and experience include modelling and simulation, design and analysis of manufacturing systems, production planning and control, reconfigurable manufacturing systems and next generation of manufacturing systems including fractal and biological manufacturing systems. He has published over 130 articles in various national and international academic journals and conferences, including keynote addresses and a book.

Dr Carlos R. Gómez, Graduated with a PhD degree from University of Nottingham.

Professor Nabil Gindy, BSc (Honours), MSc, PhD, CEng a leading academic at The University of Nottingham, who died on 03 May 2013 at the age of 62. Professor Gindy played a key role in the Faculty of Engineering at the University for two decades, and had worked since 2009 at the University of Nottingham Ningbo China (UNNC) where he was Vice-Provost for Research and Dean of the Graduate School. He joined the University as a Professor in Manufacturing Engineering in 1993, bringing with him an international reputation for expertise in his field.

MODELLING FRESH GOODS SUPPLY CHAIN CONTAMINATION

Agostino G. Bruzzone, Marina Massei
DIME University of Genoa
Via all'Opera Pia 15, 16145 Genoa, Italy
Email: {agostino, massei}@itim.unige.it
URL: www.itim.unige.it

Matteo Agresta, Angelo Ferrando
Simulation Team
Via Cadorna 2, 17100 Savona, Italy
Email: {matteo.agresta, angelo.ferrando}@simulationteam.com
URL: www.simulationteam.com

ABSTRACT
This paper proposes models of supply chain devoted to investigate food contamination with special attention to fresh goods; these phenomena are becoming more a more critical and the paper proposes models of both demand evolution, countermeasures for mitigating the impact, operations and logistics; an experimental analysis is provided in order to validate the models and the proposed approach.

keywords**:** Vulnerability, Supply Chain Contamination, Fresh Food Supply chain, Risks, Supply Chain Modeling; Supply Chain Recalling

INTRODUCTION
Supply chains are growing in term of complexity and evolve to be more and more vulnerable against risks due to many factors, in particular the high number of interdependencies introduces vulnerabilities in food and beverage logistics. Indeed the food and beverage industry is characterized by several risks affecting the logistics processes that have different origins; for instance they could be originated from market, from suppliers; sometime risks arise from processes internal to the supply chain itself (i.e. storage, distribution, …) other times they are originated from the external environment (i.e. epidemics, wars, politic problems, …). For what concerns food supply chain, events like contaminations are extremely dangerous in terms of effects, due to the fact that they introduce safety risks affecting health status of a large number of people; this fact obviously have a big impact in term of economic risks for Company stability.
These events are currently becoming more frequent due to several reasons (i.e. intensive food production, use of innovative solutions that could introduce risks of contamination, outsourcing versus not very reliable providers) so it clear the importance of preventing and mitigating these phenomena; Modeling & Simulation (M&S) is a powerful methodology to address this challenging problem, especially considering the high number of variables and interdependencies that are present in the supply chains and that becomes even more critical in food logistics (i.e. traceability, transport/storage conditions, variability of demand, recalling of products in case of contaminations, …).
In this paper is proposed a simulation based model devoted to study fresh fish supply chain; this model was developed based on Discrete Event stochastic simulator and it is implemented in Extendsim®. After the description of the conceptual model the paper proposes a case study related to a fresh fish supply chain specific for the North Italy (Bruzzone et al. 2009); the results are presented as demonstration of potential of this approach as well as example for its operational use; in addition the future research opportunities are summarized.

1. DEFINITION OF SUPPLY CHAIN RESILIENCE

Movement of people, goods and information have always been fundamental components of human society. Contemporary economic processes have been accompanied by a significant increase in mobility and higher level of accessibility, Rodrigue et. al. (2006).
Due to contemporary economic and globalizations of markets, supply chains have become more and more complex, particularly in the last two-three decades. Such complexity has the advantages to increase the business opportunities but makes the supply chains more vulnerable against risks; such events can lead to extremely catastrophic situations both for companies and for final customers.
The historical origins of the "Globalization Process" are the subject of on-going debate: some experts situate that process in the modern era and other attribute to that phenomenon a long history.
By the way, according to many economist experts, market globalization has started in 1980s mainly due to exponential improvement in transportation and telecommunications infrastructures and services.
Such easy way to transfer information and goods, has facilitated the globalization of the supply chain, first for global companies, and later for medium size enterprises, that have increased their flexibility at the "cost" of making each process more complex; the evaluation of the performance by using M&S was addressed by many authors due to the system complexity (Macias et al. 2004; Baruwa & Piera 2008;

Curcio and Longo, 2009; Merkuryev et al. 2008; De Felice et al. 2010; Merkuryeva et al. 2011; Vonoflen et al. 2011). Such complexity is due to the increasing number of interdependencies between all the legs along the supply chain, from the procurement phase up to the final customer. Outsource some process makes the supply chain itself more and more flexible and agile to the market, for example giving the possibility to outsource parts of the logistic or distribution (i.e. 3PL-Third Parties Provider, 4 PL-Fourth Parties Providers,…) or parts of the production process by sourcing globally from suppliers. Anyway, more the supply chains is slim and flexible, more it becomes vulnerable and exposed to risks due to the high number of interrelations. The "resilience" is a way to measure such vulnerability; Barroso et al (2011) define supply chain resilience (SCR) as "the ability to react to the negative effects caused by disturbances that occur at a given moment in order to maintain the supply chain objectives", and Falasca et al (2008) consider SCR as "the ability of a supply chain system to reduce the probabilities of a disruption, to reduce the consequences of those disruptions once they occur, and to reduce the time to recover normal performance". That definition is quite similar to Sceffi (2005), that define SCR as the "ability of an organization to successfully confront the unforeseen".

For a detailed review of the possible definitions of "supply chain resilience" see Stravos et al. (2012).

Furthermore many authors (Merkuryev et al. 2009) have defined the basic elements that affect supply chain resilience as: flexibility, agility, velocity, visibility and redundancy and in Longo & Oren (2008) is presented an overview of many examples of real case studies of supply chain disruption.

The importance of that topic is proved not only in academic research but also "in the business world", both for global firm and for Small and Medium Enterprises (SMEs). For example, recent surveys for Global Firms, such as "Understanding Supply Chain Risks; Mc Kinsey Global Survey", underlines that 65% of the surveyed executives believe that supply chain risk is growing and Jüttner & Ziegenbein (2009) confirm that this topic is important also for SMEs, because they are often exposed to the same risks as their large international firm counterparts but they miss the necessary resources, structures and processes.

This paper focuses on modeling and simulating a particular risk on a supply chain: the risk of contamination.

In particular for certain products, such as fresh food, the contamination of products can cause extremely dangerous situation, both in term of health and in term of money. In fact, when such kind of events happens, the consequences can be really serious: loose of brand image, loose of sales, cost for recalling the products and health problems or diseases.

Food contamination refers, in general, to the presence in food of chemicals or microorganisms that can cause consumer illness.

In this work a fresh fish supply chain is considered and a simulation model is defined in order to create a tool for supporting decision, prevent and mitigate risks by means of simulation techniques.

This work consists of 6 Section; Section 1 is an introduction to the topic of supply chain resilience and Section 2 provides a state of the art on the use of M&S to face supply chain risks. Section 3 is a short description of different source of contamination in fresh fish supply chain and in Section 4 the description of the model is presented. Section 5 provides a case study for the fish supply chain in the North of Italy and finally, Section 6 provide conclusion and possible next steps for future researches.

2. STATE OF THE ART

Supply chain stability should be attacked by many threats; these are usually emerging from potential risks affecting logistics and operations as well as market environment; in "Understanding Supply Chain Risks; a Self Assessment Work Book (2003)" risks are classified in external or internal. The external risk identifies epidemics, wars, politic problems, natural disasters, which can be considered totally external to the supply chain and are hardly predictable and quite impossible to face. Risks coming from the market or from suppliers are considered external to the company but internal to supply chain, and finally, risks related to the process and activities are considered totally internal to the company (Fig.1).

Also Christopher and Peck (2003) categorized risks into 3 similar macro-areas, and identifies five categories:

1. Internal to the firm:
   - i) Process
   - ii) Control
2. External to the firm but internal to the Supply Chain:
   - i) Demand
   - ii) Supply
3. External to the Supply Chain:
   - i) Environmental

Longo & Oren (2008), trying to clarify the complex problem of supply chain resilience, have also identified a possible classification of some areas of research:

1. Supply chain vulnerability, security and resilient management
2. Methods for demand forecasting and supply risk analysis in supply chain
3. Information management and visibility along the supply chains
4. Supply chain Life Cycle Costing
5. Modelling / Simulation devoted to support supply chain resilience

For what concerns the last point, Modeling and Simulation (M&S) is proved to be a useful instrument to consider the complex interrelations between all the constraints and the variables in the supply chain, in particular for fresh food, which needs much more processes instead that traditional foods. (Bruzzone et al. 2008).

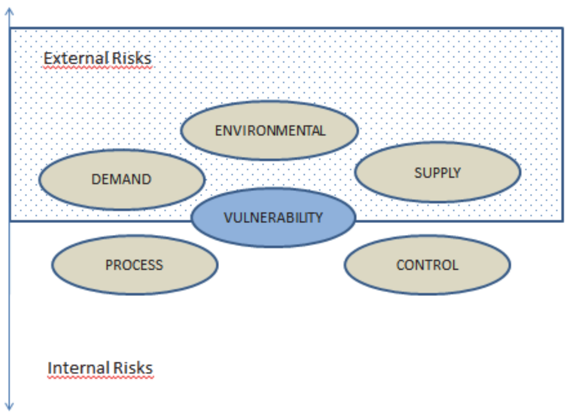


Figure 1: External and Internal Risks for the supply chains

Examples and applications of M&S for fresh food supply chain are reported in Bruzzone & Tremori (2006), that apply M&S in fresh fish and meat supply chains, providing the opportunity to test methodologies and to develop decision support systems based on optimization techniques.
Another work, Busato & Berruto (2009) describes and simulates the recall process of contaminated product by using a discrete event simulation; such kind of simulation is used also in R. Rossi (2012), for investigating the impact of dual sourcing strategies on quality of fresh fruit traded in international supply chains.
Bruzzone & Tremori (2009) focuses on modeling adverse events impacts on retail networks; the model simulates phenomena such as terrorism actions, contaminated or defective product lots, including the influence of media broadcasting news and fear of the market, by making use of System Dynamics Techniques.
For what concerns the economical aspect of catastrophic events on the supply chain Ernst & Young (2011) gives an overview of how companies can quantify recall losses and maximize recovery of those losses from suppliers and insurers, in particular for recall cost. Furthermore, the demand variability in case of these events and the probability distribution of loses is investigated by Sun and Yu (2005).

3. THE FRESH FISH SUPPLY CHAIN RISKS

In order to build a model by which simulate the contamination phenomena, it is necessary to have a good understanding of the supply chain for fresh food products.
Fish can be classified by their origin/production in

- Farmed fish
- Fished fish

or by their storage condition in:

- Fresh fish
- Frozen fish

These differences have a strong influence in the structure of the supply chain, because they determine the range of temperature to be respected in the whole process (i.e. handling, transport, processing,…).
For example distribution is really critical because each time the cold room or the refrigerated truck is opened, there is a change on the temperature, which can cause and accelerate the reproduction of micro-organism, causing contamination.
Also the quality of fodder and water plays a primary role both for the fished product and for the farmed one.
Farmed fish has a high probability to be contaminated by suppliers that provides the animal fodder while fished fish can be contaminated because it is fished in an area where the level of pollution is high or exceptional event are happened discharging in the sea toxic substances.
Furthermore, the transportation in cases (tanks of aluminum, wooden or polystyrene with dry ice), or manufacturing at sea by sea water (cutting of the head, evisceration, storage, freezing, or packaging), or transport (i.e. vessels can remain at sea for days, then there is truck, plane and or train), or finally the storage places (cool box, refrigerated containers, cells, shelves) can be point of risk and possible causes of contamination and alteration of the quality of the product.

4. MODEL DESCRIPTION

In this section the mathematical model is described; the final aim is to reproduce the dynamic behavior of the whole supply chain for fresh fish from the supply side (fishing/farming) up to the final customer in case of a contamination event.
In this work a Discrete Event Simulation (DES) approach is used. DES is a discrete-state, event-driven system: this means that its state evolution depends entirely on the occurrence of asynchronous events over time, Cassandras & Lafortune, (2006) ,Discrete event Simulators are powerful instruments to model the complex dynamics in systems where many variables interact each other.
When contamination occurs, the "stationary" conditions of the goods flow that moves towards the supply chain are perturbed by the contamination of a certain quantity of products. That infected items can be potentially contaminated in each leg of the supply chain with a given probability P and will reach the market or the final consumer in a given time, depending on the structure of the supply chain itself (i.e. number of legs, frequency of orders, quantity of stock in the chain).
Obviously, later the contamination is discovered, more the consequences will be heavy because the contaminated product may be already in the shops or, in the worst case, already consumed.
The main events and main time of occurrence are defined in Tab. 1; the process considered starts from the contamination instant and ends when all the effect of the contamination are totally finished in each leg of the chain.

Table 1: main events in the studied scenario

τ_α	Time when contamination occurs
τ_β	Time when contamination is discovered by the company
τ_γ	first case of disease/illness/death
τ_ν	the media gives the first announcement of the event
τ_o	Time when the media campaign reach the maximum level
τ_ϕ	time when the fear reach the maximum level
τ_ε	time when the product reach the market
τ_ρ	time when the media campaign stops
τ_σ	time when the fears is equal to zero
e_α	contamination event
e_β	company realize the problem
e_γ	first announcement from the media
e_δ	media campaign stops
e_σ	fears is equal to zero

In Fig. 2 the logical time sequence of the events is reported: all the process starts with the first event that is the product contamination e_α and the whole process ends when the problem is completely solved and customer's fear stops in τ_σ.

The simulation starts in a stationary situation where CEDI, wholesaler, and retailers are refurbished of the requested quantity of food $Q_i(\tau)$, in order to satisfy a demand $D_i(\tau)$ of each i –cathegory of fish. At a given time τ_α, a quantity Q of fish is contaminated in a given part of the supply chain (i.e. fishing/farming, CEDI, wholesalers, retailers) and starts to propagate along the chain.

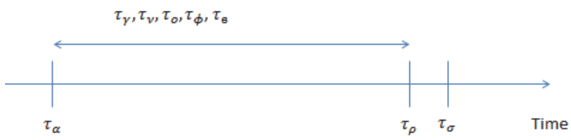


Figure 2: time sequence of the main events

All the other events, don't have a fixed schedule, because they can happen with different combinations and sequences. For example company can realize the problem of contamination before that the product reach the market, or in other cases it can be too late and the problem may already be on the shelves; furthermore diseases and effects on health can happen also before that the company realize of the contamination.

The intensity and frequency of quality control influence the probability of contamination in each leg and the quantity of product to be recalled; furthermore the structure of the supply chain determine the lead time and, consequently the speed in propagations of infected products in the both directions (distribution and recalling).

After contamination, there are two/three possibilities:

I. Company is the first that realize the fact and start recalling process
II. Media is the first to give the notice of the first contamination
III. No one understand the problem and contamination spread to many customers

In case I is the company itself that understand the problem through quality controls and starts the recalling process, before that people is contaminated.

There is a low damage in term of image of the brand and the main damages are in terms of lost sales due to the recalling of the products and the empty shelves. This is the best situation since the company can contain the damages. In case II the media give the notice and the company has many damages in terms of loss of image and compensations for contaminated people. There are also secondary effects: for example the demand of the other products of the same company may decrease because the low level of trustiness in the brand.

Case III is obviously the worst, because the contamination continues to spread postponing I or II and making the situation more and more serious.

As told before, when the process of recalling starts, a quantity of products is retired from the market; this is a cost not only for the process itself, but also because it determine lost sales due to the empty shelves and sales will be affected for a given time period Δ_τ that is needed to refurbish the market with the new and uncontaminated product. Contamination in production can be modeled as (1):

$$\Delta\tau_{Tot} = \Delta\tau_C + \Delta\tau_P + \Delta\tau_L \qquad (1)$$

Where:

$\Delta\tau_{Tot}$ = Total delay in market refurbishing
$\Delta\tau_C$ = Time to check the problem
$\Delta\tau_P$ = Time to change production/supplier
$\Delta\tau_L$ = Lead time

That is because production may stop for a given time in order to find the source of contamination $\Delta\tau_C$, and after that a time period $\Delta\tau_P$ is needed in order to check/change/substitute some production process or supplier and restart with the production, and finally, the product reach the market after the lead time $\Delta\tau_L$.

In this model, the demand risk after a supply chain contamination is supposed to be function of demand in normal condition; in order to considerer the effect of media campaign the Fear Level Function is introduced. That function depends on the amount of information provided by media, which are represented by the Media Campaign Function.

Fig. 3 defines the relation between the shape of the media campaign and the fear level.

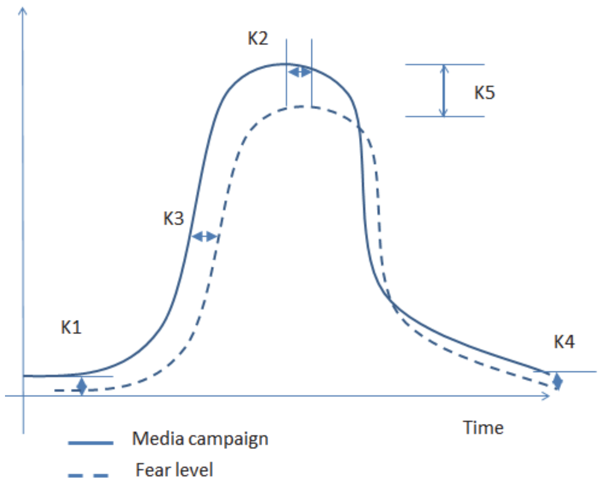


Figure 3: Fear level and Media Campaign in case of contamination event

Parameters k_i allow obtaining the fear function in time, in function of the Media Campaign curve:

$K_1(\geq 0)$: media know the information before market

$K_2(\geq 0)$: maximum peak delay

$K_3(\geq 0)$: market delay to information

$K_4(\geq 0)$: fear stops only when media campaign stop

$K_5(\geq\leq 0)$: maximum peak difference

Fear level is supposed to be a proportional function of the Media Campaign curve; as is possible to note in Fig.3 the market have some delay respect the information provided by the media; some of the factors may be :

1. market don't trust to the emergency
2. market don't receive the notice
3. market have some "inertia" in change the behavior

The implicit assumption is that the market's fear will stop only after that the media campaign will stop ($K_4(\geq 0)$:).

Initial demand $D_i(\tau)$, that is the demand of product *i* before the contamination is supposed to be known, in order to evaluate the $D_i'(\tau)$, that is the demand after the event of contamination.

$D_i'(\tau)$ is the demand function after that the market have the notice of contamination; equation (2) express $D_i'(\tau)$ as a function of $D_i^*(\tau)$ that is the average demand before that event, the fear level, and a decreasing factor depending on time.

$$D_i'(\tau) = D_i^*(\tau) - F_i(\tau)a + \varepsilon(\tau) \quad (2)$$

Where:

$\varepsilon(\tau) =$ decreasing factor

$F_i(\tau) =$ fear function

a, b > 0 that are parameters to be determined

In Fig 4 the qualitative shape of $D_i'(\tau)$ is illustrated.

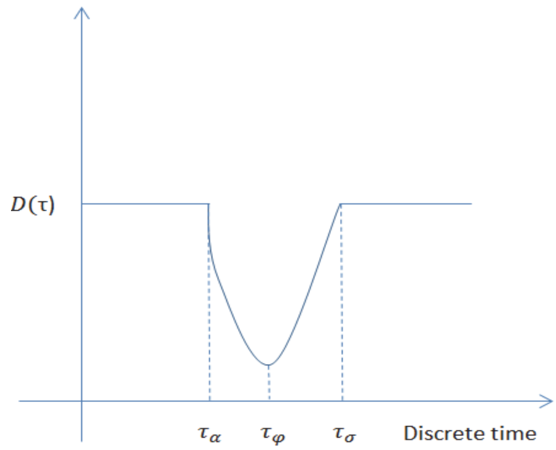


Figure 4: Qualitative shape of demand function in case of supply chain contamination

In general we can say that fear is driven by two opposite forces: is alimented by media and it is reduced by time $\varepsilon(\tau)$.

Other factors that could be considered, in general, when contamination occurs are the following:

- The demand $D_i'(\tau)$ of the contaminated i -class of food falls down af [3]
- The demand for similar products increases [4]
- The total fish demand falls down [5]

When

$$D_i'(\tau) << D_i(\tau) \quad (3)$$

$$D_j' << D_i(\tau) \quad (4)$$

$$\sum_\tau D_i'(\tau) + \sum_\tau D_j'(\tau) < \sum_\tau D_i(\tau) + \sum_\tau D_j(\tau) \quad (5)$$

5. CASE STUDY

In this section a simulation of a real supply chain of the fresh fish distribution in the North of Italy is reported. Four CEDI respectively located in Milan, Genova, Venice, and Bologna are considered and the relative market share considered is reported in Fig. 5.

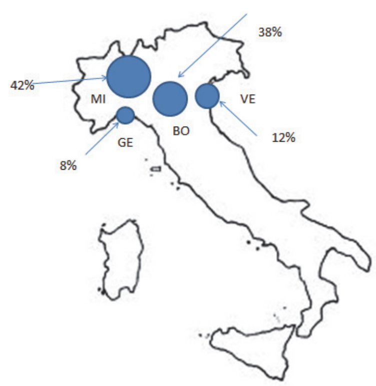


Figure 5: market share of main fish CEDI in the North of Italy

The entire supply chain is simulated with the software ExtendiSim®; in particular both the physical flow and the information flow is considered. Physical flow moves goods from supply to the customers, whilst information flows moves in the opposite direction providing the market demand and the respective orders from Retailers, Wholesalers and CEDI.
In Tab 2 the parameters determining the effectiveness of the recall process in each link of the supply chain are reported:

Table 2: parameters for effectiveness of recall

Parameter	Description	Value
Time when contamination occurs	The tracking system recognize the problem	Time of contamination, identification of the infected leg
Time when the problem is discovered	The recalling process starts	Time and leg of the chain where recall starts
Transit time	Longer is the transit time lower is the probability that contaminated product reach the market	Total delay of the interested legs of the supply chain
Inventory level	Gives the amount of goods stocked in the chain	Quantity and location
Information time	It's the time needed for sharing the information along the supply chain	Delay
Traceability index	Gives the % of product that needs to be recalled, it is a function of the level of traceability and integration of the supply chain	% of product that needs to be recalled
Time for organizing the product recall	Time that company need to start the process of recall	Delay
Time for physical recalling	Lead time for the recalling of all the product from the market	Delay
Time to restart new production/distribution of new product	Time needed by the company to produce/supply new non contaminated products	Delay

In Fig.6 there is a screenshot of the model: both the information and the goods flow is simulated along the supply chain. We make use of hierarchical blocks that allow building the model with a "black box" approach, by refining the details all the simulation also in a second time, without modifying the whole model.

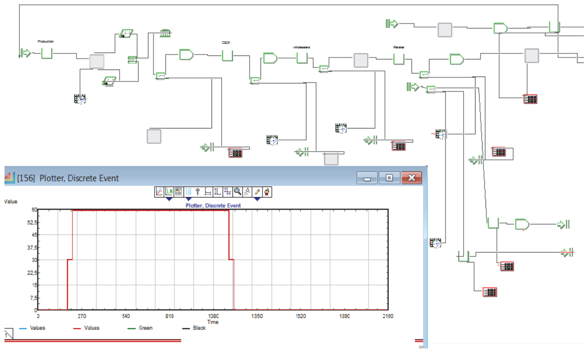


Figure 6: Screenshot of the model

The simulation starts with a demand profile that is known and propagates up to the production that produces the requested quantity to satisfy such demand. For sake of simplicity the demand is considered constant in time; after a certain time all the model reach a stationary condition..
After that the simulation is in stationary condition, in each leg is assigned a certain probability of a contamination event in function of the number of quality controls performed in each section; when contamination occurs, all the other probability are set to zero by the system and a quantity of goods Q starts to propagate in the chain.
When the company understand the fact, the recall process is activated and the goods starts to be retired from the market; in order to consider the effect of traceability systems, we can suppose that the quantities that needs to be recalled in each leg is just a fraction of the total stored quantity in each leg, that decrease in function of the efficiency of the tracking systems of the company considered. At the same time that the recalling process starts, the system simulates the contamination in the market by evaluating the number and gravity of contamination cases with a stochastic approach; ; it is evident the importance in this case study to adopt experimental design for addressing the effects introduced by stochastic factors and to properly estimate the confidence band (Montgomery 2000). Finally the model gives the possibility to introduce the cost for each process and activity performed, for example simulating the transport cost, or inventory cost, as well as the cost of lost sales due to the absence of the product in the shelves.

6. CONCLUSIONS
A model for the simulation of fresh fish supply chain phenomena is described and simulated by using Discrete Event Systems (DES) techniques.,

Media campaign and market fear have been considered and used in the determination of the demand function of a given product after a contamination event in the supply chain. Media induce and aliment the market fear, which determines the reduction of the previous demand.

In future research a more detailed model could be considered, by capturing the behavior of the single customer or groups of customers by a disaggregate simulation of the demand profile. For four different categories and different way to react to such kind of event may be considered:

I. Spike panic (no demand for all the type of fish)
II. Moderate panic (no demand only for the contaminated variety of fish)
III. Slow raising fear (decreasing demand of that particular contaminated fish)
IV. No fear (no change in demand)

Furthermore a more detailed simulation of the supply chain may be useful, in order to capture the influence of technologies like tag or sensor, that increase the traceability, reducing the risk and speeding up the recall phenomena.

7. REFERENCES

Barroso, H.P., Machado, V.H. & Machado, V.C. (2011) Supply Chain Resilience Using the Mapping Approach. Supply Chain Management, Pengzhong Li (Ed.)

Baruwa, O.T., Piera, M.À. (2008) "A derivative Control Mechanism for Supply Chain performance improvement",2008,"Proceedings of 22nd European Conference on Modelling and Simulation, Campora S.G., September

Bossomaier T, A.G. Bruzzone, M. Massei, D. Newth, J. Rosen (2009) Pandemic dynamic object and reactive agents, Int. Workshop on Modelling and Applied Simulation, Puerto de la Cruz, Tenerife

Bruzzone A.G., A. Tremori (2008) Safety & Security in retail: modeling value chain dynamics. Proc. of the 2008 Spring simulation multiconference San Diego, USA

Bruzzone A.G., M. Massei, E. Bocca (2009) Fresh Food Supply Chain Simulation Based Case Studies in Logistic, pp. 127-146

Bruzzone A.G., S. Viazzo, F. Longo, E. Papoff, B.C., (2004) Simulation and Virtual Reality to Modelling Retail and Store Facilities Proc. of SCSC2004,
San Josè, CA.

Busato P. R. Berruto (2009) Use of simulation models to study the dynamic of recall of non-conform perishable produce through the supply chain Int. European Forum on System Dynamics and Innovation in Food Network, Innsbruck-Igls, Austria

Christopher, M. and Peck, H. (2004) Building the resilient supply chain. International Journal of Logistics Management, 15(2), pp. 1-13

Curcio, D., Longo, F., 2009. Inventory and internal logistics management as critical factors affecting the Supply Chain performances. *International Journal of Simulation and Process Modelling* 5(4), 278-288.

De Felice F., G. Di Bona, D. Falcone, A. Silvestri (2010) "New Reliability allocation methodology: the Integrated Factors Method", International Journal of Operations & Quantitative Management Volume 16 Number 1, ISSN: 1082-1910

Ernst & Young 2011 Capturing Recall Costs Measuring and Recovering the Losses, GMA

Falasca M., Zobel, C.W., & Cook, D. (2008) A Decision Support Framework to Assess Supply Chain Resilience, in the proceedings of the 5th International ISCRAM Conference – Washington, DC, USA, May 2008, pp. 596-605.

Jüttner U., A. Ziegenbein (2009) Supply Chain Risk Management for Small and Medium-Sized Businesses Journal of Supply Chain Risk 124 pp. 199-217

Longo F., T. Ören, (2008) Supply chain vulnerability and resilience: a state of the art overview, Proc. of European Modeling & Simulation Symposium, Campora S. Giovanni (CS) Italy

Macías E.J., Mercedes Pérez de la Parte (2004) "Simulation and Optimization of Logistic and Production Systems Using Discrete and Continuous", Petri Nets. Simulation 80(3): 143-152

Massei M. (2006) Logistics and process solution for supply chain of fresh food in retail, Proc. of MAS2006, Bergeggi October

Merkuryev Y., Galina Merkuryeva, Jonas Hatem, Bram Desmet (2008) "Supply Chain Simulation in the ECLIPS Project" Proceedings of the Second Asia International Conference on Modelling and Simulation, Kuala Lumpur, Malaysia, May, pp. 684-690.

Merkuryev Y., Galina Merkuryeva, Miquel Angel Piera, Antoni Guasch (2009) Simulation-based Case Studies in Logistics: Education and Applied Research. Springer-Verlag, London. 232 p.

Merkuryeva G., Yuri Merkuryev, Hendrik Vanmaele (2011) "Simulation-Based Planning and Optimization in Multi-Echelon Supply Chains" Simulation: Transactions of The Society for Modeling and Simulation International. Vol. 87, No.8, August, pp. 698-713.

Montgomery D.C. (2000) "Design and Analysis of Experiments", John Wiley & Sons, New York

Pfohl H.C., H Kohler, D. Thomas (2010) State of the art in supply chain risk management research: empirical and conceptual findings and a roadmap for the implementation in practice Journal of Logistic Research 2(1) pp. 34-44

Rodrigue J.P., C. Comitos, B. Slack (2006) The Geography of Transport Systems, Routledge

Rossi R., W.A. Rijpkema, J.G.A.J.Van der Vorst (2012) The impact of dual sourcing ion food

supply chain networks: the case of Egyptian strawberries, WICaNeM, Nerthelands
Sceffi Y. (2005) Building a Resilient Supply Chain Harvard Business Review 1(8) pp. 1-4
Sceffi Y., J. Rice (2005) A Supply Chain View of the Resilent Enterprise MITSloan Management Review, 47(1) pp. 41-48
Stravos T. P. (2012) Supply chain Resilience; Definition of Concept And Its Formative Elements Journal of Applied Business Research (28)5 pp. 921-930
Sun C., Yu H (2005) Supply Chain Contract under Product Cost Disrumption. Proc of the International Conference on Services Systems and Serivice Management, (1) pp. 708-711
Vonolfen S., M. Affenzeller, A. Beham, E. Lengauer, S. Wagner (2011) "Simulation-based evolution of resupply and routing policies in rich vendor-managed inventory scenarios" Central European Journal of Operations Research, Vol. 19, No. 4, 2011, pp. 23

APPLYING MODELING AND SIMULATION TO EVALUATE STATISTICAL ACCURACY AND INVENTORY CONTROL PERFORMANCE OF LUMPY DEMAND FORECASTING METHODS

Adriano O. Solis[(a)], Francesco Longo[(b)], Letizia Nicoletti[(c)], Aliaksandra Yemialyanava[(d)]

[(a)] School of Administrative Studies, Faculty of Liberal Arts & Professional Studies, York University, Toronto, Canada
[(b) (c) (d)] Department of Mechanical, Energy and Management Engineering, University of Calabria, Rende, Italy

[(a)]asolis@yorku.ca, [(b)]francesco.longo@unical.it, [(c)]letizia.nicoletti@unical.it, [(d)] alexandra.yemialyanava@gmail.com

ABSTRACT
A number of time series methods – 13-month simple moving average (SMA13), single exponential smoothing (SES), Croston's method, and the Syntetos-Boylan approximation (SBA) – are well-referenced methods in the literature on intermittent or lumpy demand forecasting. We apply these four methods to an industrial dataset involving more than 1000 stock-keeping units (SKUs) in the central warehouse of a firm operating in the professional electronics sector. Earlier studies have argued that the negative binomial distribution (NBD) satisfies both theoretical and empirical criteria for modeling intermittent demand. We have found that the NBD often does not provide a good fit. We apply an alternative approach, using a two-stage distribution involving the uniform and negative binomial distributions, in modeling actual demand. We use modeling and simulation to evaluate the four methods in terms of statistical forecast accuracy and, more importantly, inventory system efficiency.

Keywords: lumpy demand forecasting, forecast accuracy, scale-free error statistics, inventory control, modeling and simulation

1. INTRODUCTION

When there are time intervals with no demand occurrences for an item of inventory, demand is said to be *intermittent*. Demand is *erratic* when there are large variations in the sizes of actual demand occurrences. Demand that is both intermittent and erratic is referred to as *lumpy* demand.

Syntetos, Boylan, and Croston (2005) proposed to categorize demand patterns into four classes using cutoff values of CV^2 = 0.49 and ADI = 1.32 (where CV^2 and ADI are, respectively, the squared coefficient of variation of demand and average inter-demand interval) – for the stated purpose of assigning the best forecasting method. The four categories are (i) *smooth*, when ADI < 1.32 and CV^2 < 0.49; (ii) *erratic* (but not very intermittent), when ADI < 1.32 and CV^2 > 0.49; (iii) *intermittent* (but not very erratic) when ADI > 1.32 and CV^2 < 0.49; and (iv) *lumpy*, when ADI > 1.32 and CV^2 > 0.49. These cutoff values and resulting categories have been cited in various other studies involving intermittent or lumpy demand (e.g., Ghobbar and Friend 2002, 2003; Gutierrez, Solis, and Mukhopadhyay 2008; Boylan, Syntetos, and Karakostas 2008; Mukhopadhyay, Solis, and Gutierrez 2012).

A number of studies (e.g., Syntetos and Boylan 2006; Boylan, Syntetos, and Karakostas 2008; Syntetos, Babai, Dallery, and Teunter 2009) have proposed using a negative binomial distribution (NBD) to model the demand distribution of an item exhibiting intermittent demand. The NBD is a discrete probability distribution which may be specified by the density function:

$$f(x;r,p) = \binom{r+x-1}{x} p^r (1-p)^x I_{\{0,1,2,\ldots\}}(x), \tag{1}$$

having two parameters p and r, where the real number p satisfies $0 < p \le 1$ and r is a positive integer. The real number p is a probability of "success" in a Bernoulli trial, while r is a target number of successes (e.g., Feller 1957; Mood, Graybill, and Boes 1974). The random variable X in this case represents, in a succession of the Bernoulli trials, the number of failures preceding the rth success. The NBD has mean

$$\mu = E[X] = \frac{r(1-p)}{p} \tag{2}$$

and variance

$$\sigma^2 = V[X] = \frac{r(1-p)}{p^2}. \tag{3}$$

Since $V[X] = E[X]/p$, it follows that the variance of the NBD is greater than its mean. When r = 1, the NBD reduces to a geometric (or Pascal) distribution with discrete density function

$$f(x;p) = p(1-p)^x I_{\{0,1,2,\ldots\}}(x). \tag{4}$$

Syntetos and Boylan (2006) have argued that the NBD satisfies both theoretical and empirical criteria.

To generate an NBD to approximate the distribution of a random variable with mean μ and variance σ^2, we simultaneously solve (2) and (3) to obtain:

$$\hat{p} = \frac{\mu}{\sigma^2} \tag{5}$$

and

$$\hat{r} = \frac{\mu^2}{\sigma^2 - \mu}, \tag{6}$$

as initial estimates of the NBD parameters. These expressions for $\hat{p}$ and $\hat{r}$, however, represent values in the set $\Re$ of real numbers, while the NBD parameter r is supposed to be integer-valued. In applying (5) and (6), we generally obtain a non-integer value of $\hat{r}$. Thus, in seeking to simulate the actual demand distributions, we have investigated the rounded up and rounded down values of $\hat{r}$ while adjusting the value of $\hat{p}$.

In the intermittent demand forecasting literature, many papers have been published on the relative performance with respect to statistical measures of accuracy of various forecasting methods, most notably simple exponential smoothing (SES), Croston's method (Croston 1972), and an estimator proposed by Syntetos and Boylan (2005). Schultz (1987) suggested that separate smoothing constants, α_i and α_s, be used for updating the inter-demand intervals and the nonzero demand sizes, respectively, in place of Croston's single smoothing constant α. We note, however, that Mukhopadhyay, Solis, and Gutierrez (2012) investigated separate smoothing constants, α_i and α_s, in forecasting lumpy demand and did not observe any substantial improvement in forecast accuracy.

Syntetos and Boylan (2001) pointed out a positive bias in Croston's method arising from an error in his mathematical derivation of expected demand. They proposed (Syntetos and Boylan 2005) a correction factor of $\left(1 - \alpha_i/2\right)$ – where α_i is the smoothing constant used in updating the inter-demand interval estimate – to be applied to Croston's estimator of mean demand. The revised estimator is now often referred to (e.g., Gutierrez, Solis, and Mukhopadhyay 2008; Boylan, Syntetos, and Karakostas 2008; Babai, Syntetos, and Teunter 2010; Mukhopadhyay, Solis, and Gutierrez 2012) in the intermittent demand forecasting literature as the Syntetos-Boylan approximation (SBA).

We also evaluate the 13-month simple moving average (SMA13) method, which is based upon dividing the 52 weeks in a year into 13 four-week "months". SMA13 has been applied in a number of recent intermittent demand forecasting studies (e.g., Syntetos and Boylan 2005, 2006; Boylan, Syntetos, and Karakostas 2008) in view of its being built into some commercially available forecasting software.

This paper is organized as follows. In section 2, we discuss the forecasting methods under evaluation, the statistical measures of forecast accuracy that we use, and the nature of the industrial dataset and how data partitioning is performed. In the next section, we propose a two-stage approach to the modeling of demand distribution. We proceed to report on our empirical investigation of forecasting performance, based upon statistical accuracy measures, on the performance block of the actual data and on the simulated demand distribution. The performance of the forecasting methods in terms of inventory systems efficiency is reported in section 4. We present our conclusions in the final section.

2. FORECASTING METHODS AND DEMAND DATA

2.1. Forecasting Methods and Accuracy Measures

Four methods that are well-referenced in the intermittent demand forecasting literature are evaluated in this paper: SMA13, SES, Croston's, and SBA. For the SES, Croston's, and SBA methods, low values of the exponential smoothing constant α of up to 0.20 have generally been suggested for lumpy demand (e.g., Croston 1972; Johnston and Boylan 1996). We test four α values of 0.05, 0.10, 0.15, and 0.20 as used in a number of recent studies (e.g., Syntetos and Boylan 2005; Gutierrez, Solis, and Mukhopadhyay 2008; Mukhopadhyay, Solis, and Gutierrez 2012).

In this paper, we apply three scale-free error statistics. The first is *mean absolute percentage error* (MAPE), which is the most widely used accuracy measure for ratio-scaled data. The traditional MAPE definition, which involves terms of the form $|E_t|/A_t$ (where A_t and E_t, respectively, represent actual demand and forecast error in period t), fails when demand is intermittent. We applied an alternative specification (e.g., Gilliland 2002) of MAPE as a ratio estimate, which guarantees a nonzero denominator:

$$\text{MAPE} = \left(\sum_{t=1}^{n} |E_t| \Big/ \sum_{t=1}^{n} A_t \right) \times 100 . \tag{7}$$

This specification of MAPE has been used in evaluating lumpy demand forecasting (e.g., Gutierrez, Solis, and Mukhopadhyay 2008; Mukhopadhyay, Solis, and Gutierrez 2012).

A second scale-free error statistic we use is the *mean absolute scaled error* (MASE). It was fairly recently proposed as a forecast accuracy measurement applicable for intermittent demand, without problems seen in other error statistics (Hyndman and Koehler 2006). Using the naïve method, one-period-ahead forecasts are generated. A scaled error is defined as

$$q_t = \frac{e_t}{\left(\sum_{i=2}^{n} |Y_i - Y_{i-1}|\right) / (n-1)}, \qquad (8)$$

which is independent of the scale of the data. A scaled error is less than one if it arises from a better forecast than the average one-step, naïve forecast computed in-sample. It is greater than one if the forecast is worse. MASE is simply the mean value of $|q_t|$.

The third scale-free error statistic we apply in this paper is *percentage best* (PB) – the percentage of time periods in which a particular method performs better than any of the other methods with respect to a specified criterion. PB has been applied in a good number of intermittent demand forecasting studies (e.g., Syntetos and Boylan 2005; Gutierrez, Solis, and Mukhopadhyay 2008; Mukhopadhyay, Solis, and Gutierrez 2012). We used absolute error for assessing performance under the PB approach.

2.2. Industrial Dataset and Partitioning

In this paper, we apply the four methods to an industrial dataset involving more than 1000 SKUs in the central warehouse of a firm operating in the professional electronics sector. The raw data consist of actual stock withdrawals reported in the company's enterprise resource planning system. We first aggregated the transactional data into usage quantities per calendar week, and further aggregated the weekly quantities in terms of 13 four-week "months" in a calendar year. The monthly usage quantities do not constitute actual demand quantities, since the inventory on hand at the time of a stock withdrawal may not meet the required quantity. Since demand is not traditionally tracked as well as actual usage in a transaction-based system, we treat monthly usage quantity as a surrogate measure of monthly demand.

This process yielded 66 months of "demand" data, which we broke down into initialization, calibration, and performance measurement blocks (as in Boylan, Syntetos, and Karakostas 2008) with our blocks consisting of 23, 23, and 20 months, respectively.

For each of the SES, Croston's and SBA methods, we selected α based upon the minimum MAPE attained in the calibration block for use as the smoothing constant in the performance block.

Given that the various SKUs represent end items, sub-assemblies, components, and spare parts that are used for building projects, retail sales, or servicing of professional electronic products, it is understandable that we found many of them to actually exhibit erratic or lumpy demand based on the earlier cited categorization scheme (Syntetos, Boylan, and Croston 2005). We have, however, failed to find a SKU with intermittent demand (ADI > 1.32 and CV^2 < 0.49) according to this scheme.

In this paper, we report findings on a sample of fifteen SKUs, with demand statistics presented in Table 1. The first six SKUs (1–6) are categorized as having erratic demand, while the remaining nine SKUs (7–10) exhibit lumpy demand.

Table 1: Sample of 15 SKUs

SKU #	**1**	**2**	**3**	**4**	**5**
Mean	7.1667	14.1667	11.4091	36.2273	8.9394
Std Dev	8.1930	16.3251	10.3804	32.1050	9.5545
CV^2	1.3069	1.3279	0.8278	0.7854	1.1424
ADI	1.1186	1.1000	1.1379	1.0820	1.2692
z (% of Zero Demand)	10.61%	9.09%	12.12%	7.58%	21.21%
Category	ERRATIC	ERRATIC	ERRATIC	ERRATIC	ERRATIC

SKU #	**6**	**7**	**8**	**9**	**10**
Mean	3.5152	1.0152	8.3333	29.0152	5.3667
Std Dev	4.0959	1.3977	11.8612	57.6463	7.9723
CV^2	1.3577	1.8957	2.0259	3.9472	2.2068
ADI	1.2941	1.7838	1.6500	2.2000	1.6216
z (% of Zero Demand)	22.73%	43.94%	39.39%	54.55%	38.33%
Category	ERRATIC	LUMPY	LUMPY	LUMPY	LUMPY

SKU #	**11**	**12**	**13**	**14**	**15**
Mean	3.1818	2.3333	3.9394	5.2879	2.2273
Std Dev	3.7700	2.9053	5.3805	6.2580	2.2790
CV^2	1.4039	1.5504	1.8655	1.4006	1.0470
ADI	1.6098	1.5000	1.5000	1.3469	1.5000
z (% of Zero Demand)	37.88%	33.33%	33.33%	25.76%	33.33%
Category	LUMPY	LUMPY	LUMPY	LUMPY	LUMPY

2.3. Modeling of Demand Distributions

We sought to simulate demand distributions of the SKUs under study, performing 100 runs each consisting of 100 four-week "months", for a total of 10,000 months in each experiment. We used AnyLogic as our simulation platform. Some code was written in Java in order to address mathematical modeling which could not be readily undertaken within the standard AnyLogic library (the authors have a remarkable experience in developing simulation model for inventory and supply chain problems investigation, Curcio and Longo, 2009; Bruzzone and Longo, 2010). We used the mean $\bar{x}$ and the variance s^2 of the 66-month actual demand time series as values of μ and σ^2, respectively, in (5) and (6) to generate initial estimates $\hat{r}$ and $\hat{p}$ of the NBD parameters. These initial estimates are then adjusted to obtain acceptable NBD parameters r and p.

In the case of SKUs 1 and 7 (out of the 15 SKUs that we report on in the current paper), we found the proposed NBD approximation to yield a simulated distribution which fairly closely follows the actual demand distribution. The simulation results are reported in Table 2. In each case, the adjusted r value is 1 (i.e., the NBD reduces to a geometric distribution) and $p = \Pr(X = 0)$ closely approximates – in fact, slightly exceeds – the actual proportion z of zero demand occurrences.

However, we have found that, for most of the other SKUs currently under study, it is not possible to accordingly obtain adjusted values of $\hat{r}$ and $\hat{p}$ that would lead to an NBD with mean, standard deviation, CV^2, and ADI that are reasonably close to those of the actual demand distribution. To illustrate, in Table 3, we present NBD approximation results for SKUs 2 and 8. In each of these two cases, the NBD reduces to a geometric distribution, with the adjusted r value being 1 However, $p = \Pr(X = 0)$ is less than the actual proportion z of zero demand occurrences, especially in the case of SKU 8 for which demand is lumpy (p = 7.87% versus z = 39.39%).

To more directly address the proportion z of periods with zero demand, we simulate the demand distribution by way of a two-stage probability distribution (earlier applied by Solis, Nicoletti, Mukhopadhyay, Agosteo, Delfino, and Sartiano 2012).

Table 2: Two SKUs with Good NBD Approximations

SKU #	1	7
Mean	7.1667	1.0152
Std Dev	8.1930	1.3977
CV^2	1.3069	1.8957
ADI	1.1186	1.7838
z (% of Zero Demand)	10.61%	43.94%
Category	ERRATIC	LUMPY
r^	0.8566	1.0981
p^	0.1068	0.5196
SIMULATION		
r	1	1
p	0.1177	0.4800
Mean	7.4544	1.0795
Std Dev	7.8854	1.4783
CV^2	1.1310	1.9056
ADI	1.1319	1.9026
z (% of Zero Demand)	11.64%	47.51%
Simulated vs Actual Mean	104.0%	106.3%
Simulated vs Actual Std Dev	96.2%	105.8%
Simulated vs Actual CV^2	86.5%	100.5%
Simulated vs Actual ADI	101.2%	106.7%
Difference in Simulated vs Actual z	1.03%	3.57%

Table 3: Two SKUs with Poor NBD Approximations

SKU #	2	8
Mean	14.1667	8.3333
Std Dev	16.3251	11.8612
CV^2	1.3279	2.0259
ADI	1.1000	1.6500
z (% of Zero Demand)	9.09%	39.39%
Category	ERRATIC	LUMPY
r^	0.7953	0.5247
p^	0.0532	0.0592
SIMULATION		
r	1	1
p	0.0591	0.0787
Mean	15.8662	11.7009
Std Dev	16.3433	11.9177
CV^2	1.0716	1.0484
ADI	1.0641	1.0873
z (% of Zero Demand)	5.99%	8.00%
Simulated vs Actual Mean	82.6%	190.4%
Simulated vs Actual Std Dev	73.0%	137.8%
Simulated vs Actual CV^2	80.7%	51.8%
Simulated vs Actual ADI	96.7%	65.9%
Difference in Simulated vs Actual z	-3.10%	-31.39%

The first stage is modeled with a uniform distribution initially based upon $z_1 = z$, where z is the actual proportion of zero demand periods in the 66-month time series. For the second stage, we determine the mean $\overline{x}_{nz}$ and variance s^2_{nz} of the nonzero demands and use these to calculate first approximations of the parameters $\hat{p}_{nz}$ and $\hat{r}_{nz}$ in line with (5) and (6). The corresponding negative binomial probability $P_0 = \Pr(X = 0) > 0$ is then used to adjust z_1 (as applied in the first stage) downward, as follows:

$$z_1 = \frac{z - P_0}{1 - P_0}, \quad (9)$$

provided $z > P_0$. As a result, the proportion of zero demand periods arising from the two-stage distribution is closer to z. We refine the parameter estimate $\hat{p}_{nz}$ as the values of mean, standard deviation, CV^2, and *ADI* of the actual and simulated demand distributions are compared.

We present in Table 4 the simulation results for the demand distributions of the 15 SKUs using the two-stage approach, except for SKUs 1 and 7 (which were modeled using only the NBD approximation). For the other 13 SKUs, we selected z_1 for stage 1 and the parameters r and p of the NBD in stage 2 based on what appeared to yield the best combination of values of mean, standard deviation, CV^2, and *ADI* of the simulated distribution in comparison with those of the actual distribution.

Table 4: Simulation of Demand Using the Two-Stage Distribution

SKU #	1	2	3	4	5
Mean	7.1667	14.1667	11.4091	36.2273	8.9394
Std Dev	8.1930	16.3251	10.3804	32.1050	9.5545
CV^2	1.3069	1.3279	0.8278	0.7854	1.1424
ADI	1.1186	1.1000	1.1379	1.0820	1.2692
z (% of Zero Demand)	10.61%	9.09%	12.12%	7.58%	21.21%
Category	ERRATIC	ERRATIC	ERRATIC	ERRATIC	ERRATIC
Mean of nonzero demand	8.0169	15.5833	12.9828	39.1967	11.3462
Std Dev of nonzero demand	8.2639	16.4670	10.1038	31.5958	9.4077
r^ nonzero	1.0663	0.9502	1.8916	1.6019	1.6684
p^ nonzero	0.1174	0.0575	0.1272	0.0393	0.1282
SIMULATION					
r	1	1	2	2	2
p	0.1177	0.0589	0.1271	0.0481	0.1491
Pr(*X* = 0)	0.1177	0.0589	0.0162	0.0023	0.0222
Final zero proportion in stage 1	0.00%	3.40%	10.68%	7.36%	19.42%
Mean	7.4544	15.3294	12.2154	37.1695	9.0884
Std Dev	7.8854	16.3021	10.6752	29.6808	8.9894
CV^2	1.1310	1.14615	0.7738	0.6446	0.9853
ADI	1.1319	1.09857	1.1392	1.0807	1.2811
z (% of Zero Demand)	11.64%	9.02%	12.27%	7.46%	21.97%
Simulated vs Actual Mean	104.0%	108.2%	107.1%	102.6%	101.7%
Simulated vs Actual Std Dev	96.2%	99.9%	102.8%	92.4%	94.1%
Simulated vs Actual CV^2	86.5%	86.3%	93.5%	82.1%	86.2%
Simulated vs Actual ADI	101.2%	99.9%	100.1%	99.9%	100.9%
Δ in Simulated vs Actual z	1.03%	-0.07%	0.15%	-0.12%	0.76%

SKU #	6	7	8	9	10
Mean	3.5152	1.0152	8.3333	29.0152	5.3667
Std Dev	4.0959	1.3977	11.8612	57.6463	7.9723
CV^2	1.3577	1.8957	2.0259	3.9472	2.2068
ADI	1.2941	1.7838	1.6500	2.2000	1.6216
z (% of Zero Demand)	22.73%	43.94%	39.39%	54.55%	38.33%
Category	ERRATIC	LUMPY	LUMPY	LUMPY	LUMPY
Mean of nonzero demand	4.5490	1.8108	13.7500	63.8333	8.7027
Std Dev of nonzero demand	4.1246	1.4306	12.5734	71.7573	8.6212
r^ nonzero	1.6603	13.9098	1.3098	0.8013	1.1541
p^ nonzero	0.2674	0.8848	0.0870	0.0124	0.1171
SIMULATION					
r	1	1	1	1	1
p	0.2100	0.4800	0.0730	0.0139	0.1150
Pr(*X* = 0)	0.2100	0.4800	0.0730	0.0139	0.1150
Final zero proportion in stage 1	2.19%	0.00%	34.62%	53.90%	30.32%
Mean	3.7000	1.0795	8.2722	31.6675	5.4287
Std Dev	4.1823	1.4783	11.9503	56.9295	7.7028
CV^2	1.2968	1.9056	2.1254	3.3252	2.0692
ADI	1.2757	1.9026	1.6384	2.2202	1.6033
z (% of Zero Demand)	21.59%	47.51%	39.05%	54.83%	37.57%
Simulated vs Actual Mean	105.3%	106.3%	99.3%	109.1%	101.2%
Simulated vs Actual Std Dev	102.1%	105.8%	100.8%	98.8%	96.6%
Simulated vs Actual CV^2	95.5%	100.5%	104.9%	84.2%	93.8%
Simulated vs Actual ADI	98.6%	106.7%	99.3%	100.9%	98.9%
Δ in Simulated vs Actual z	-1.14%	3.57%	-0.34%	0.28%	-0.76%

SKU #	11	12	13	14	15
Mean	3.1818	2.3333	3.9394	5.2879	2.2273
Std Dev	3.7700	2.9053	5.3805	6.2580	2.2790
CV^2	1.4039	1.5504	1.8655	1.4006	1.0470
ADI	1.6098	1.5000	1.5000	1.3469	1.5000
z (% of Zero Demand)	37.88%	33.33%	33.33%	25.76%	33.33%
Category	LUMPY	LUMPY	LUMPY	LUMPY	LUMPY
Mean of nonzero demand	5.1220	3.5000	5.9091	7.1224	3.3409
Std Dev of nonzero demand	3.5930	2.9294	5.6438	6.3002	2.0109
r^ nonzero	3.3686	2.4108	1.3459	1.5575	15.8781
p^ nonzero	0.3968	0.4079	0.1855	0.1794	0.8262
SIMULATION					
r	3	1	1	1	15
p	0.3800	0.2764	0.1675	0.1380	0.8130
Pr(*X* = 0)	0.0549	0.2764	0.1675	0.1380	0.0448
Final zero proportion in stage 1	34.27%	7.86%	19.92%	13.87%	30.21%
Mean	3.1966	2.3502	4.0198	5.3850	2.3899
Std Dev	3.6903	2.9341	5.3374	6.5807	2.3267
CV^2	1.3534	1.5946	1.7937	1.5183	0.9691
ADI	1.6025	1.5124	1.4889	1.3438	1.5016
z (% of Zero Demand)	37.67%	33.75%	32.86%	25.64%	33.56%
Simulated vs Actual Mean	100.5%	100.7%	102.0%	101.8%	107.3%
Simulated vs Actual Std Dev	97.9%	101.0%	99.2%	105.2%	102.1%
Simulated vs Actual CV^2	96.4%	102.9%	96.2%	108.4%	92.6%
Simulated vs Actual ADI	99.5%	100.8%	99.3%	99.8%	100.1%
Δ in Simulated vs Actual z	-0.21%	0.42%	-0.47%	-0.12%	0.23%

We must quickly point out, however, that even the two-stage approximation we apply fails in the case of SKUs with demand distributions that are lumpier. In Table 5, we provide three SKUs (16–18) out of a number of SKUs whose demand distributions we were unable to model adequately.

Table 5: Some SKUs for which the Two-Stage Distribution Fails

SKU #	16	17	18
Mean	59.3182	3.9848	7.7424
Std Dev	117.7808	10.2379	21.9638
CV^2	3.9425	6.6009	8.0475
ADI	1.5349	1.7368	1.4667
z (% of Zero Demand)	34.85%	42.42%	31.82%
Category	LUMPY	LUMPY	LUMPY
Mean of nonzero demand	91.0465	6.9211	11.3556
Std Dev of nonzero demand	136.0571	12.7775	25.8977
r^ nonzero	0.4500	0.3064	0.1956
p^ nonzero	0.0049	0.0424	0.0169
SIMULATION			
r	1	1	1
p	0.0080	0.0850	0.0400
Pr(*X* = 0)	0.0080	0.0850	0.0400
Final zero proportion in stage 1	34.32%	37.08%	28.98%
Mean	83.9741	6.7120	16.9344
Std Dev	117.9103	10.2027	22.8629
CV^2	2.0030	2.3956	1.8701
ADI	1.5199	1.7186	1.4643
z (% of Zero Demand)	34.20%	41.87%	32.68%
Simulated vs Actual Mean	141.6%	168.4%	218.7%
Simulated vs Actual Std Dev	100.1%	99.7%	104.1%
Simulated vs Actual CV^2	50.8%	36.3%	23.2%
Simulated vs Actual ADI	99.0%	99.0%	99.8%
Δ in Simulated vs Actual z	-0.65%	-0.55%	0.86%

For such SKUs where even our two-stage approximation of the demand distribution failed, we have understandably been unable to proceed with the evaluation of forecast accuracy and inventory control performance.

3. FORECASTING PERFORMANCE

3.1. Forecast Accuracy: Performance Block

The exponential smoothing constant α selected from among the candidate values (0.05, 0.10, 0.15, or 0.20) for each of the SES, Croston's, and SBA methods, based upon the minimum MAPE in the calibration block, are shown in Table 6. The resulting error statistics when applying SMA13, SES, Croston's, and SBA methods to actual demand data in the performance block (the final 20 months) are likewise reported in the table. There does not appear to be a method that exhibits a superior performance overall across the 15 SKUs.

3.2. Forecast Accuracy: Simulated Demand

When applying the methods to the simulated demand distributions, however, we see in Table 7 that overall the SBA method outperforms SMA13, SES, and Croston's methods – particularly for SKUs with lumpy demand – based on MAPE and MASE. While the forecast accuracy performance with respect to PB appears to be inconclusive, the reported results nevertheless suggest the overall superiority of SBA over the long run.

Table 6: Error Statistics when Applying Various Methods to Actual Demand in the Performance Block

SKU #	**1**	**2**	**3**	**4**	**5**
Category	ERRATIC	ERRATIC	ERRATIC	ERRATIC	ERRATIC
Smoothing Constants Selected in Calibration Block					
SES	0.20	0.05	0.10	0.05	0.15
Croston	0.20	0.05	0.05	0.05	0.05
SBA	0.20	0.05	0.10	0.05	0.10
MAPE (in %)					
SMA13	71.92	86.83	76.34	61.59	109.87
SES	63.47	94.56	75.16	61.42	102.88
Croston	64.08	93.25	75.44	62.61	104.22
SBA	63.30	92.00	75.05	63.02	104.66
Best MAPE	SBA	SMA13	SBA	SES	SES
MASE					
SMA13	0.8676	0.9343	0.7440	0.6685	1.1332
SES	0.7657	1.0175	0.7325	0.6668	1.0612
Croston	0.7731	1.0033	0.7352	0.6797	1.0750
SBA	0.7636	0.9899	0.7314	0.6841	1.0795
Best MASE	SBA	SMA13	SBA	SES	SES
PB (in %)					
SMA13	25.00	60.00	25.00	35.00	25.00
SES	30.00	25.00	25.00	25.00	45.00
Croston	15.00	5.00	10.00	0.00	20.00
SBA	30.00	10.00	40.00	40.00	10.00
Best PB	SES/SBA	SMA13	SBA	SBA	SES

SKU #	**6**	**7**	**8**	**9**	**10**
Category	ERRATIC	LUMPY	LUMPY	LUMPY	LUMPY
Smoothing Constants Selected in Calibration Block					
SES	0.05	0.05	0.10	0.20	0.20
Croston	0.15	0.05	0.05	0.05	0.05
SBA	0.20	0.05	0.05	0.05	0.05
MAPE (in %)					
SMA13	117.48	97.01	80.51	90.97	104.90
SES	126.44	98.60	81.56	95.47	115.53
Croston	112.24	90.72	78.10	93.53	96.07
SBA	102.76	89.98	78.24	93.60	96.03
Best MAPE	SBA	SBA	Croston	SMA13	SBA
MASE					
SMA13	0.7921	0.6635	0.7520	0.7533	0.6766
SES	0.8524	0.6744	0.7618	0.7906	0.7453
Croston	0.7567	0.6205	0.7295	0.7746	0.6197
SBA	0.6928	0.6155	0.7308	0.7751	0.6194
Best MASE	SBA	SBA	Croston	SMA13	SBA
PB (in %)					
SMA13	5.00	30.00	25.00	40.00	25.00
SES	20.00	15.00	20.00	20.00	10.00
Croston	15.00	10.00	10.00	5.00	20.00
SBA	60.00	45.00	45.00	35.00	45.00
Best PB	SBA	SBA	SBA	SMA13	SBA

SKU #	**11**	**12**	**13**	**14**	**15**
Category	LUMPY	LUMPY	LUMPY	LUMPY	LUMPY
Smoothing Constants Selected in Calibration Block					
SES	0.20	0.05	0.10	0.05	0.20
Croston	0.20	0.05	0.10	0.05	0.05
SBA	0.20	0.05	0.20	0.05	0.05
MAPE (in %)					
SMA13	56.58	86.50	81.57	67.36	90.21
SES	53.46	84.31	76.97	67.53	90.51
Croston	61.39	82.85	76.50	67.35	89.32
SBA	63.44	82.27	78.45	67.51	89.24
Best MAPE	SES	SBA	Croston	Croston	SBA
MASE					
SMA13	0.8909	0.7546	1.1952	0.7359	0.6983
SES	0.8418	0.7356	1.1278	0.7378	0.7006
Croston	0.9667	0.7228	1.1209	0.7358	0.6914
SBA	0.9989	0.7178	1.1495	0.7376	0.6908
Best MASE	SES	SBA	Croston	Croston	SBA
PB (in %)					
SMA13	25.00	20.00	15.00	35.00	30.00
SES	40.00	25.00	35.00	20.00	35.00
Croston	15.00	5.00	20.00	5.00	15.00
SBA	20.00	50.00	30.00	40.00	20.00
Best PB	SES	SBA	SES	SBA	SES

Table 7: Error Statistics when Applying Various Methods to the Simulated Demand Distributions

SKU #	1	2	3	4	5
Category	ERRATIC	ERRATIC	ERRATIC	ERRATIC	ERRATIC
Smoothing Constants Selected in Calibration Block					
SES	0.20	0.05	0.10	0.05	0.15
Croston	0.20	0.05	0.05	0.05	0.05
SBA	0.20	0.05	0.10	0.05	0.10
MAPE (in %)					
SMA13	80.64	81.26	69.72	64.62	79.89
SES	81.78	79.36	68.84	62.88	80.10
Croston	81.46	79.07	67.56	62.38	77.82
SBA	78.96	78.45	67.95	62.02	78.06
Best MAPE	SBA	SBA	Croston	SBA	Croston
MASE					
SMA13	0.7637	0.7709	0.7554	0.7578	0.7650
SES	0.7737	0.7529	0.7458	0.7373	0.7667
Croston	0.7711	0.7501	0.7321	0.7317	0.7451
SBA	0.7475	0.7442	0.7363	0.7275	0.7474
Best MASE	SBA	SBA	Croston	SBA	Croston
PB (in %)					
SMA13	34.53	41.24	30.05	37.56	26.29
SES	20.75	15.43	16.32	19.96	26.02
Croston	8.64	9.01	34.03	5.97	33.85
SBA	36.08	34.32	19.60	36.51	13.84
Best PB	SBA	SMA13	Croston	SMA13	Croston

SKU #	6	7	8	9	10
Category	ERRATIC	LUMPY	LUMPY	LUMPY	LUMPY
Smoothing Constants Selected in Calibration Block					
SES	0.05	0.05	0.10	0.20	0.20
Croston	0.15	0.05	0.05	0.05	0.05
SBA	0.20	0.05	0.05	0.05	0.05
MAPE (in %)					
SMA13	87.57	107.04	111.30	129.28	106.65
SES	86.65	103.93	109.81	130.74	108.27
Croston	87.96	102.21	105.77	121.36	102.34
SBA	85.61	101.45	105.08	120.55	101.68
Best MAPE	SBA	Croston	SBA	SBA	SBA
MASE					
SMA13	0.7801	0.8241	0.8199	0.8495	0.8031
SES	0.7720	0.8000	0.8088	0.8585	0.8148
Croston	0.7836	0.7869	0.7792	0.7974	0.7709
SBA	0.7628	0.7810	0.7741	0.7922	0.7660
Best MASE	SBA	SBA	SBA	SBA	SBA
PB (in %)					
SMA13	23.43	36.84	31.14	22.46	22.99
SES	30.52	13.46	19.21	32.65	46.90
Croston	9.65	15.71	12.73	12.20	8.43
SBA	36.40	33.99	36.92	32.69	21.65
Best PB	SBA	SMA13	SBA	SBA	SES

SKU #	11	12	13	14	15
Category	LUMPY	LUMPY	LUMPY	LUMPY	LUMPY
Smoothing Constants Selected in Calibration Block					
SES	0.20	0.05	0.10	0.05	0.20
Croston	0.20	0.05	0.10	0.05	0.05
SBA	0.20	0.05	0.20	0.05	0.05
MAPE (in %)					
SMA13	95.95	96.13	101.66	93.39	84.25
SES	97.25	94.61	101.00	90.98	85.34
Croston	95.54	93.67	100.91	90.28	82.39
SBA	93.74	92.95	98.69	89.70	82.25
Best MAPE	SBA	SBA	SBA	SBA	SBA
MASE					
SMA13	0.7939	0.7914	0.7999	0.7805	0.7958
SES	0.8041	0.7788	0.7944	0.7603	0.8056
Croston	0.7907	0.7714	0.7938	0.7545	0.7786
SBA	0.7759	0.7654	0.7766	0.7497	0.7774
Best MASE	SBA	SBA	SBA	SBA	SBA
PB (in %)					
SMA13	24.63	39.72	27.54	39.87	23.95
SES	29.35	15.36	12.52	14.57	32.88
Croston	12.97	12.13	24.18	11.43	16.65
SBA	33.05	32.79	35.76	34.13	26.52
Best PB	SBA	SMA13	SBA	SMA13	SES

4. INVENTORY CONTROL PERFORMANCE

Demand forecasting and inventory control performance have traditionally been considered independently of each other (Tiacci and Saetta 2009). In reality, forecast accuracy may not translate into inventory systems efficiency (Syntetos, Nikolopoulos, and Boylan 2010).

Sani and Kingsman (1997) have recommended a periodic review inventory control system to address intermittent demand. An order-up-to (T,S) periodic review system, where T and S denote the review period and the base stock (or ‘order-up-to’ level), respectively, has been applied in recent intermittent demand forecasting studies (e.g., Eaves and Kingsman 2004; Syntetos and Boylan 2006; Syntetos, Nikolopoulos, Boylan, Fildes, and Goodwin, 2009; Syntetos, Babai, Dallery, and Teunter 2009; Syntetos, Nikolopoulos, and Boylan 2010; Teunter, Syntetos, and Babai 2010) that look into both forecast accuracy and inventory control performance.

In this paper, we assume a (T,S) system with full backordering, with inventory reviewed on a monthly basis (T = 1). The reorder lead time for most SKUs is about one month (L = 1). The literature suggests a safety stock component to compensate for uncertainty in demand during the ‘protection interval’ T+L. For each SKU, we calculated s_{tr}, the standard deviation of monthly demand during the initialization and calibration blocks (the first 46 months of actual usage quantities). We apply a ‘safety factor’ k to yield a safety stock level of $k \cdot s_{tr}$. This approach differs, of course, from that suggested (e.g., Silver, Pyke, and Peterson 1998) under an assumption that daily demand is identically and independently normally distributed during the protection interval. If F_t is the forecast calculated at the end of month t, and I_t and B_t are, respectively, on-hand inventory and backlog, the replenishment quantity at the time of review is

$$Q_t = (T + L) \cdot F_t + k \cdot s_{tr} - I_t + B_t . \tag{10}$$

4.1. Service Levels

Silver, Pyke, and Peterson (1998) identified the two most commonly specified service level criteria in inventory systems. One is a target average *probability of no stockout* (PNS) per review period. The other is a target *fill rate* (FR), or average percentage of demand to be satisfied from on-hand inventory. FR is noted to have considerably more appeal for practitioners. We simulated inventory control performance with respect to two values of the target FR (95% and 98%) and two values of the target PNS (90% and 95%). We performed simulation searches to find the safety factor k that would yield the target FR or PNS.

4.2. Average Inventory on Hand

For a target FR of 98%, resulting averages of inventory on hand are reported in Table 8. We proceeded to index the average inventory on hand, as reported in Table 8, using SBA as base (SBA index = 100). Indices corresponding to a target FR of 98% are reported in Table 9 and depicted graphically in Figure 1. This figure suggests the overall superiority of SBA over SMA13 and SES in terms of resulting average on-hand inventory levels, but does not seem to indicate a similar comparison with respect to Croston's method. In fact, Figure 1 appears to suggest that Croston's method leads to slightly better average on-hand inventory than SBA for erratic demand (SKUs 1–6). When conducting *t* tests, the mean indices for SMA13 and SES exceed 100 at the 1% and 5% levels of significance, respectively. However, testing the mean index for Croston's method does not yield a statistically significant conclusion.

Table 8: Average Inventory on Hand for a 98% Target Fill Rate

SKU #	1	2	3	4	5
SMA13	25.5024	53.2450	28.7932	76.4457	24.6869
SES	25.8652	51.8753	28.1311	72.4345	24.4925
Croston	25.7067	51.7261	27.3639	72.3506	23.7126
SBA	25.5968	51.7647	27.9113	72.3894	23.9568

SKU #	6	7	8	9	10
SMA13	13.9749	5.2988	44.7181	240.5282	28.8782
SES	13.5985	5.0813	44.0254	242.5799	28.8763
Croston	13.9830	5.0115	43.1597	238.3088	28.3105
SBA	14.0235	5.0069	43.2253	238.3386	28.2883

SKU #	11	12	13	14	15
SMA13	10.1452	10.0599	19.8695	22.4635	5.3829
SES	10.2395	9.7714	19.6909	21.6916	5.4407
Croston	10.2056	9.7606	19.9021	21.6139	4.9807
SBA	10.0786	9.8166	19.7692	21.5828	4.9790

Table 9: Indices of Average Inventory on Hand for a 98% Target Fill Rate

SKU #	1	2	3	4	5
SMA13	99.6	102.9	103.2	105.6	103.0
SES	101.0	100.2	100.8	100.1	102.2
Croston	100.4	99.9	98.0	99.9	99.0
SBA	100.0	100.0	100.0	100.0	100.0

SKU #	6	7	8	9	10
SMA13	99.7	105.8	103.5	100.9	102.1
SES	97.0	101.5	101.9	101.8	102.1
Croston	99.7	100.1	99.8	100.0	100.1
SBA	100.0	100.0	100.0	100.0	100.0

SKU #	11	12	13	14	15
SMA13	100.7	102.5	100.5	104.1	108.1
SES	101.6	99.5	99.6	100.5	109.3
Croston	101.3	99.4	100.7	100.1	100.0
SBA	100.0	100.0	100.0	100.0	100.0

Similar results arise for the 95% target FR, as well as for the 90% and 95% PNS target levels – except that the mean index for SES with a target PNS of 95% does not exceed 100 at the 5% significance level. Figure 2 graphically shows indices for the 90% target PNS.

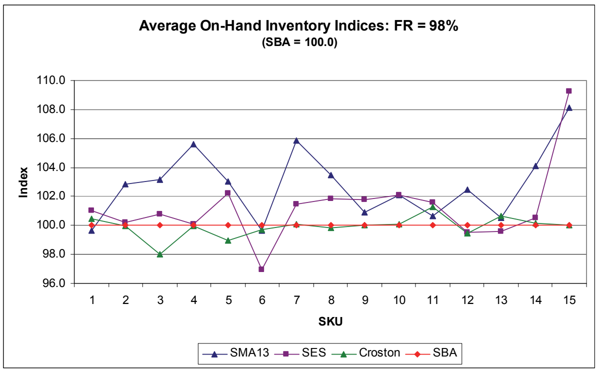


Figure 1: Average On-Hand Inventory Indices for a 98% Target Fill Rate

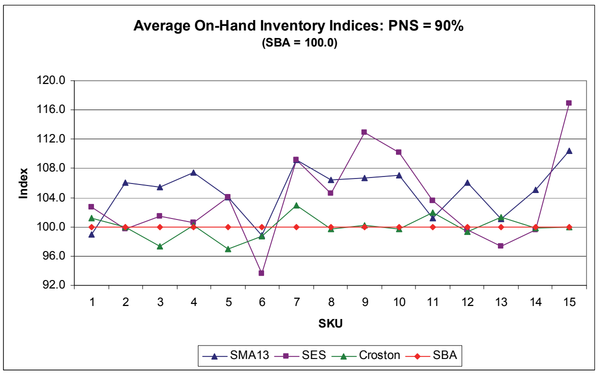


Figure 2: Average On-Hand Inventory Indices for a 90% Target Probability of No Stockout

4.3. Cumulative Backlogs

In Table 10, the SBA row shows the average (across 100 replications) of the cumulative backlogs over 100 months when the target FR is 98%. The rows corresponding to SMA13, SES, and Croston's methods indicate respective differences with the SBA value. In the last three rows, a negative figure (in parentheses) represents a lower average, while a positive figure means a higher average. The differences, in absolute terms, are all less than one, indicating that there is hardly any difference in performance with respect to 100-month cumulative backlogs for the given target FR. The same observation holds for a target FR of 95%.

Table 10: Comparison of Mean 100-Month Backlogs (SBA vs. Other Methods) for a 98% Target Fill Rate

SKU #	1	2	3	4	5
SBA	15.16	30.89	24.28	73.73	18.47
SMA13	0.00	0.04	0.05	(0.22)	(0.18)
SES	(0.03)	(0.01)	0.05	(0.05)	(0.02)
Croston	(0.02)	(0.01)	0.04	(0.03)	0.03

SKU #	6	7	8	9	10
SBA	7.49	2.27	17.28	66.53	11.19
SMA13	0.00	(0.10)	(0.12)	(0.33)	(0.08)
SES	0.02	0.00	(0.10)	(0.42)	(0.08)
Croston	(0.04)	0.00	(0.01)	(0.01)	(0.01)

SKU #	11	12	13	14	15
SBA	6.44	4.73	8.04	11.07	4.66
SMA13	0.05	(0.03)	(0.07)	(0.20)	(0.09)
SES	0.01	0.00	(0.01)	(0.04)	(0.01)
Croston	(0.01)	0.00	(0.01)	(0.01)	0.00

Table 11 provides analogous results for a 90% target PNS. With the exception of SKU 9 in which SES and SMA13 respectively exhibit favorable 18-unit and 5-unit average advantages over SBA, the 100-month cumulative backlogs appear to be more or less similar overall across the four methods. We note that SKU 9 is quite lumpy, characterized by relatively high values of CV^2 (3.9472), *ADI* (2.2), and *z* (54.55%). Similar observations arise under a 95% target PNS.

Table 11: Comparison of Mean 100-Month Backlogs (SBA vs. Other Methods) for a 90% Target PNS

SKU #	1	2	3	4	5
SBA	85.64	171.14	100.18	252.37	81.42
SMA13	(0.17)	0.83	(0.66)	2.56	1.45
SES	(0.40)	1.20	(0.71)	(0.85)	0.90
Croston	0.31	0.50	0.23	(0.51)	1.26

SKU #	6	7	8	9	10
SBA	48.62	19.72	133.36	699.79	89.81
SMA13	0.45	(0.13)	1.28	(5.26)	(1.45)
SES	(0.19)	(2.20)	(1.41)	(18.02)	(1.13)
Croston	(0.09)	(0.86)	0.28	(1.20)	0.01

SKU #	11	12	13	14	15
SBA	35.67	35.78	64.28	74.35	20.03
SMA13	(0.03)	0.20	(0.88)	1.14	0.10
SES	0.17	0.41	0.93	1.03	(1.76)
Croston	(0.26)	0.09	0.01	0.37	0.03

5. CONCLUSION

Earlier studies have argued that the negative binomial distribution (NBD) satisfies both theoretical and empirical criteria for modeling intermittent demand. We tested the NBD on an industrial dataset involving more than 1000 stock-keeping units (SKUs) in the central warehouse of a firm operating in the professional electronics sector. We have established that the NBD often does not provide a good fit for most of the SKUs tested. We used a two-stage approach (applied preliminarily by Solis, Nicoletti, Mukhopadhyay, Agosteo, Delfino, and Sartiano 2012) involving uniform and negative binomial distributions. In the current paper, we report on 15 SKUs, of which six exhibit erratic demand while the remaining nine have lumpy demand. The simulated demand distributions arising from the two-stage modeling approach more closely approximate the actual demand distributions of the SKUs under consideration. The SMA13, SES, Croston's, and SBA methods are well-referenced in the literature on intermittent or lumpy demand forecasting. We investigated the statistical accuracy of these forecasting methods using three scale-free error statistics. In testing statistical accuracy on the performance block (the final 20 months of the 66-month actual distribution), we found none of the methods under consideration to be consistently superior to the others. However, SBA is found to be the best performing method overall when the four methods are tested over the longer term (100 replications of 100 months, or a total of 10,000 months). We subsequently simulated the inventory control performance of each method, applying the demand estimates on the basis of the simulated demand distribution for a given SKU. A (*T*,*S*) periodic review inventory control system with full backordering, with a one-month review period and a one-month replenishment leadtime, was assumed. Using either a target fill rate or a target probability of no stockout, we have found SBA to yield the lowest average levels of inventory on hand in almost all cases. At the same time, the expected cumulative backlogs under SBA are comparable to those using the other forecasting methods. Accordingly, it appears that SBA generally leads to better inventory systems efficiency.

REFERENCES

Babai, M.Z., Syntetos, A.A., and Teunter, R., 2010. On the empirical performance of (*T*,*s*,*S*) heuristics. *European Journal of Operational Research*, 202 (2), 466-472.

Boylan, J.E., Syntetos, A.A., and Karakostas, G.C., 2008. Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, 59 (4), 473-481.

Bruzzone, A.G., Longo, F., 2010. An advanced system for supporting the decision process within large-scale retail stores. *Simulation* 86 (12), 742-762.

Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23 (3), 289-304.

Curcio, D., Longo, F., 2009. Inventory and internal logistics management as critical factors affecting the Supply Chain performances. *International Journal of Simulation and Process Modelling* 5(4), 278-288.

Eaves, A.H.C. and Kingsman, B.G., 2004. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55 (4), 431-437.

Feller, W., 1957. *An Introduction to Probability Theory and Its Applications, Vol. I*, 2nd ed. New York, USA: John Wiley & Sons.

Gilliland, M., 2002. Is forecasting a waste of time? *Supply Chain Management Review*, 6 (4), 16-23.

Ghobbar, A.A. and Friend, C.H., 2002. Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8 (4), 221-231.

Ghobbar, A.A. and Friend, C.H., 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30 (14), 2097-2114.

Gutierrez, R.S., Solis, A.O., and Mukhopadhyay, S., 2008. Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111 (2), 409-420.

Hyndman, R.J. and Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22 (4), 679-688.

Johnston, F.R. and Boylan, J.E., 1996. Forecasting for items with intermittent demand. *Journal of the Operational Research Society*, 47 (1), 113-121.

Mood, A.M., Graybill, F.A., and Boes, D.C., 1974. *Introduction to the Theory of Statistics*, 3rd ed. New York, USA: Mc-Graw-Hill.

Mukhopadhyay, S., Solis, A.O., and Gutierrez, R.S., 2012. The Accuracy of Non-traditional versus Traditional Methods of Forecasting Lumpy Demand. *Journal of Forecasting*, 31 (8), 721-735.

Sani, B. and Kingsman, B.G., 1997. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48 (7), 700-713.

Schultz CR., 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society*, 38 (5), 453-458.

Silver, E.A., Pyke, D.F., and Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*. New York, USA: John Wiley & Sons.

Solis, A.O., Nicoletti, L., Mukhopadhyay, S., Agosteo, L., Delfino, A., and Sartiano, M., 2012. Intermittent demand forecasting and stock control: an empirical study. *Proceedings of the 11th International Conference on Modeling & Applied Simulation*, 367-374, September 19-21, Vienna, Austria.

Syntetos, A.A. and Boylan, J.E., 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics*, 71 (1-3), 457-466.

Syntetos, A.A. and Boylan, J.E., 2005. The accuracy of intermittent demand estimates, *International Journal of Forecasting*, 21 (2), 303-314.

Syntetos, A.A., Boylan, J.E., and Croston, J.D., 2005. On the categorization of demand patterns. *Journal of the Operational Research Society*, 56 (5), 495-503.

Syntetos, A.A. and Boylan, J.E., 2006. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics*, 103 (1), 36-47.

Syntetos, A.A., Babai, M.Z., Dallery, Y., and Teunter, R., 2009. Periodic control of intermittent demand items: theory and empirical analysis. *Journal of the Operational Research Society*, 60 (5), 611-618.

Syntetos, A.A., Babai, M.Z., Dallery, Y., and Teunter, R., 2009. Periodic control of intermittent demand items: theory and empirical analysis. *Journal of the Operational Research Society*, 60 (5), 611-618.

Syntetos, A.A., Nikolopoulos, K., Boylan, J.E., Fildes, R., and Goodwin, P., 2009. The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118 (1), 72-81.

Syntetos, A.A., Nikolopoulos, K., and Boylan, J.E., 2010. Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, 26 (1), 134-143.

Teunter, R.H., Syntetos, A.A., and Babai, M.Z., 2010. Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research*, 203 (3), 619-624.

Tiacci, L. and Saetta, S., 2009. An approach to evaluate the impact of interaction between demand forecasting method and stock control policy on the inventory system performances. *International Journal of Production Economics*, 118 (1), 63-71.

AUTHORS BIOGRAPHY

Adriano O. Solis is Associate Professor of Logistics Management and Area Coordinator for Management Science at York University, Canada. After receiving BS, MS and MBA degrees from the University of the Philippines, he joined the Philippine operations of Philips Electronics where he became a Vice-President and Division Manager. He went on to obtain a PhD degree in Management Science from the University of Alabama. He was previously Associate Professor of Operations and Supply Chain Management at the University of Texas at El Paso. He has been a Visiting Professor in the Department of Mechanical, Energy, and Management Engineering, at the University of Calabria, Italy.

Francesco Longo obtained his degree in Mechanical Engineering, *summa cum laude*, in 2002 and his Ph.D. in Mechanical Engineering in 2005 from the University of Calabria, Italy. He is currently an Assistant Professor in the Department of Mechanical, Energy, and Management Engineering at the University of Calabria, where he also serves as Director of the Modeling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES). He is an Associate Editor of *Simulation: Transactions of the Society for Modeling and Simulation International*. He has been active in the organization and management of a number of international conferences in modeling and simulation, serving as General Chair, Program Chair, Track Chair.

Letizia Nicoletti is currently a PhD student in Management Engineering at the University of Calabria, Italy. Her research interests include modeling and simulation for inventory management, as well as for training in complex systems, specifically marine ports and container terminals. She also actively supports the organization of international conferences in the modeling and simulation area.

Aliaksandra Yemialyanava is a student in the *Laurea Magistrale* (master's degree) program in Management Engineering at the University of Calabria's Department of Mechanical, Energy, and Management Engineering. She obtained her first degree in Management Engineering in 2011. She was recently an exchange graduate student/visiting research scholar at the School of Administrative Studies, York University, Canada.

Author's Index