THE 13TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION

SEPTEMBER 10-12 2014
BORDEAUX, FRANCE

EDITED BY

AGOSTINO G. BRUZZONE
FABIO DE FELICE
MARINA MASSEI
YURI MERKURYEV
ADRIANO SOLIS
GREGORY ZACHAREWICZ

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2014

ISBN 978-88-97999-34-8 (Paperback)
ISBN 978-88-97999-40-9 (PDF)

ISBN 978-88-97999-34-8 (Paperback)
ISBN 978-88-97999-40-9 (PDF)

THE 13TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION
MAS 2014

SEPTEMBER 10-12 2014 - Bordeaux, France

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

LOCAL SPONSORS

UNIVERSITY OF BORDEAUX

IMS LAB

AQUITAINE REGION

I3M 2014 INDUSTRIAL SPONSORS

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2014 Media Partners

Inderscience Publishers – International Journal of Simulation and Process Modeling

Inderscience Publishers – International Journal of Critical Infrastructures

Inderscience Publishers – International Journal of Engineering Systems Modelling and Simulation

Inderscience Publishers – International Journal of Service and Computing Oriented Manufacturing

IGI Global – International Journal of Privacy and Health Information Management

Halldale Media Group: Military Simulation and Training Magazine

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

Euromerci

EDITORS

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

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

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

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

ADRIANO SOLIS
YORK UNIVERSITY, CANADA
asolis@yorku.ca

GREGORY ZACHAREWICZ
IMS UNIVERSITÉ BORDEAUX 1, FRANCE
Gregory.zacharewicz@ims-bordeaux.fr

International Multidisciplinary Modeling & Simulation Multiconference, I3M 2014

General Co-Chairs

Agostino Bruzzone, *MITIM-DIME, University of Genoa, Italy*
Yuri Merkuryev, *Riga Technical University, Latvia*

Program Co-Chairs

Francesco Longo, *DIMEG, University of Calabria, Italy*
Emilio Jiménez, *University of La Rioja, Spain*

The 13th International Conference on Modeling and Applied Simulation

General Co-Chairs

Marina Massei, *Liophant Simulation, Italy*
Adriano Solis, *York University, Canada*

Program Chair

Fabio De Felice, *University of Cassino, Italy*

MAS 2014 International Program Committee

Michael Affenzeller, *Upper Austrian Univ. of AS, Austria*
Thècle Alix, *IMS Université Bordeaux 1, France*
Naamane Aziz, *LSIS, France*
Eleonora Bottani, *University of Parma, Italy*
Agostino Bruzzone, *University of Genoa, Italy*
José M. Cecilia, *Universidad Católica San Antonio, Spain*
Chun-Hung Chen, *George Mason University, USA*
Diego Crespo Pereira, *University of A Coruna, Spain*
Gerson Cunha, *LAMCE COPPE UFRJ, Brazil*
Geneviève Dauphin-Tanguy, *Ecole Centrale de Lille, France*
Fabio De Felice, *University of Cassino, Italy*
David Del Rio Vilas, *University of La Coruna, Spain*
Helder Manuel Ferreira Coelho, *University of Lisbon, Portugal*
Claudia Frydman, *LSIS, France*
Luca Gambardella, *IDSIA, Switzerland*
Andrea Grassi, *University of Modena and Reggio Emilia, Italy*
Yilin Huang, *Delft University of Technology, Netherlands*
Witold Jacak, *Upper Austrian Univ. of AS, Austria*
Janos Sebestyen Janosy, *Centre For Energy Research Hungarian Academy Of Sciences, Hungary*
Sergio Junco, *Universidad Nacional de Rosario, Argentina*
Pasquale Legato, *University of Calabria, Italy*
Francesco Longo, *MSC-LES, University of Calabria, Italy*
Marina Massei, *Liophant Simulation, Italy*
Andrea Matta, *Politechnic of Milan, Italy*
Igor Mayer, *TU Delft, Netherlands*
Rina Mary Mazza, *University of Calabria, Italy*
Roberto Montanari, *University of Parma, Italy*
Somnath Mukhopadhyay, *University of Texas at El Paso, USA*
Letizia Nicoletti, *University of Calabria, Italy*
Mustapha Ouladsine, *LSIS, France*
Eleonora Pantano, *University of Calabria, Italy*
Guillerme Pereira, *Universidade do Minho, Portugal*
Horacio Emilio Pérez Sánchez, *Universidad de Murcia, Spain*
Antonella Petrillo, *University of Cassino, Italy*
Chunming Rong, *University of Stavanger, Norway*
Maria Celia Santos Lopes, *LAMCE/COPPE/UFRJ - Brasil*
Peer Olaf Siebers, *University of Nottingham, UK Georgios Ch.*
Georgios Sirakoulis, *Democritus University of Thrace, Greece*
Jaroslav Sklenar, *University of Malta, Malta*
Alessandra Tesei, *NATO STO CMRE*
Adriano Solis, *York University, Canada*
Alberto Tremori, *University of Genoa, Italy*
Giuseppe Vignali, *University of Parma, Italy*
Hong Wan, *Purdue University, USA*
Wai Peng Wong, *Science Malaysia University, Malaysia*
Gregory Zacharewicz, *IMS Université Bordeaux 1, France*

Tracks and Workshop Chairs

Inventory Management Simulation
Chair: 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, University of Cassino, Italy; Antonella Petrillo, University of Cassino, Italy

Simulation in Energy Generation and Distribution
Chair: Janos Sebestyen Janosy, Centre for Energy Research Hungarian Academy of Sciences, Hungary

Promoting Enterprise Interoperability by Service Modeling & Simulation
Chairs: Thecle Alix, IMS Universite Bordeaux 1, France; Gregory Zacharewicz, IMS Universite 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 Systemes, France

Workshop on virtual and augmented reality
Chairs: Gerson Cunha, LAMCE/COPPE/UFRJ, Brasil; Maria Celia Santos Lopes, LAMCE/COPPE/UFRJ - Brasil

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

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 de Murcia, Spain; José M. Cecilia, Universidad Católica San Antonio, (Murcia) Spain

Workshop on Serious Games applied to Security, Crisis Management and Safety
Chairs: Alessandra Tesei, NATO STO CMRE; Igor Mayer, TU Delft, Netherlands

Tracks on Applied Control and Automation
Chairs: Sergio Junco, Universidad Nacional de Rosario, Argentina; Geneviève Dauphin-Tanguy, Ecole Centrale de Lille, France

Human-centred and Human-focused Modelling and Simulation (common track MAS-EMSS)
Chairs: Agostino Bruzzone, DIME University of Genoa, Italy; Peer Olaf Siebers, University of Nottingham, UK

CHAIRS' MESSAGE

Welcome to the 13th International Conference on Modeling & Applied Simulation!

As first news that we would like to share with MAS authors and attendees, we are glad to announce that MAS proceedings are now indexed by SCOPUS starting from the proceedings of MAS 2007. Being indexed in SCOPUS is for sure a valuable result that will be useful to all the MAS community. Hopefully, very soon during 2014 we will see indexed by SCOPUS also the MAS 2014 proceedings.

Also this year MAS has confirmed its own nature as a conference able to provide its attendees with an extended "simulation" framework made up by different application areas for supporting real networking and cooperation opportunities. Indeed, traditionally this is the main goal of the MAS conference: providing a set of different Modeling & Simulation applications in different domains including simulation based design and decision support systems (also based on serious games), simulation based optimization, automation, Modeling & Simulation of food processes and operations, inventory and production simulation. The possibility of sharing knowledge on Modeling & Simulation applications in different domains is the main MAS stronghold. Indeed, nowadays this is becoming more and more important; cross-disciplinary approaches are common in Industry, in Logistics, in Defense, in Healthcare as well as in many other areas including Government, Sustainable Development and Energy.

As every year MAS is co-located with the International Multidisciplinary Modeling & Simulation Multiconference (I3M) providing the attendees with the additional opportunity to attend the sessions of all the other I3M conferences (including EMSS, HMS, DHSS, IWISH and SESDE).

The venue of MAS 2014 is the wonderful city of Bordeaux, France where wine is bottled poetry!

Adriano Solis
York University
Canada

Marina Massei
Liophant Simulation,
Italy

Fabio De Felice
University of Cassino,
Italy

ACKNOWLEDGEMENTS

The MAS 2014 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 2014 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.

I3M 2014 INTERNAL STAFF

AGOSTINO G. BRUZZONE, *MISS-DIPTEM, UNIVERSITY OF GENOA, ITALY*
MATTEO AGRESTA, *SIMULATION TEAM, ITALY*
CHRISTIAN BARTOLUCCI, *SIMULATION TEAM, ITALY*
ALESSANDRO CHIURCO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*
MARZIA MATTIA, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
LETIZIA NICOLETTI, *CAL-TEK SRL*
ANTONIO PADOVANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
EDOARDO PICCO, *SIMULATION TEAM, ITALY*
GUGLIELMO POGGIO, *SIMULATION TEAM, ITALY*
ALBERTO TREMORI, *SIMULATION TEAM, ITALY*

I3M 2014 LOCAL ORGANIZATION COMMITTEE

GREGORY ZACHAREWICZ, *UNIVERSITY OF BORDEAUX, FRANCE*
THECLE ALIX, *UNIVERSITY OF BORDEAUX, FRANCE*
JEAN CHRISTOPHE DESCHAMPS, *UNIVERSITY OF BORDEAUX, FRANCE*
JULIEN FRANCOIS, *UNIVERSITY OF BORDEAUX, FRANCE*
BRUNO VALLESPIR, *UNIVERSITY OF BORDEAUX, FRANCE*

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

Index

EVALUATION AND MODELING SOME ENGINEERING PROPERTIES OF SAFFLOWER SEEDS

Feizollah Shahbazi
Department of Agricultural Machinery, Faculty of Agriculture, Lorestan University, Iran
Author's e-mail: shahbazi.f@lu.ac.ir

ABSTRACT
Several engineering properties of three safflower cultivars (IL-111, LRV51-51 and Zarghan279) at moisture contents of 10, 15, 20 and 25% were determined and compared. All the linear dimensions, geometric mean diameter and sphericity of safflower seeds increase linearly with increase in seed moisture content. The values of geometric properties were higher for IL-111cultivar than the LRV51-51 and Zarghan279 cultivars. The values of the bulk densities decreased, whereas the thousand grain mass, true density and porosity were increased with increase in seed moisture content. All the gravimetric properties for the three cultivars of safflower were significantly different ($P<0.05$). The values of the terminal velocity for all cultivars were significantly increased as the moisture content increased. The terminal velocity for the three cultivars of safflower were significantly different ($P<0.05$). On the two different surfaces, the coefficient of static friction of the IL-111 cultivar was significantly greater than that of the other cultivars. The static coefficient of friction was greatest against plywood and the least for galvanized steel. The values of the angle of repose were increased with increase in the moisture content. The values of the angle of repose for Zarghan279 cultivar were higher than the IL-111, LRV51-51 cultivars.

Keywords: Safflower, physical properties, sphericity, surface area, porosity, true and bulk density, terminal velocity, static coefficient of friction.

1. INTRODUCTION

Safflower (Carthamus tinctorius L.), which belongs to the Composite family, is cultivated in several parts of the world due to its adaptability to different environmental conditions. It is a rich source of oil (35-40%) and has high linoleic acid content (75-86%). The safflower oil is used for a cultivar of purposes, and especially as biodiesel for the production of fuel for internal combustion engines (Baumler et al., 2006; Sacilik et al., 2007). Safflower production increased recently due to increasing research on alternative energy sources. In Iran, in the last few years the safflower cultivation area has increased and was 15,000 ha in 2005-06. The average seed yield was about 900 kg/ha (Shahbazi et al., 2011).

The physical properties and terminal velocity of safflower seeds, like those of other crops, are essential for selecting the design and operational parameters of equipment relating to handling, harvesting, aeration, drying, storing, dehulling and processing. The properties of the seeds that are important for machine design and quality of processing production line and reducing the waste are: gravimetrical properties (unit mass, bulk density, true density, and porosity), geometrical properties (shape, size, geometric mean diameter and sphericity), and static coefficients of friction and terminal velocity. These properties are affected by numerous factors such as the species cultivar, size, form, superficial characteristics and moisture content of the grain (Sacilik et al., 2007).

Hence, it is necessary to determine these properties for suitable machine design and operational parameters. Many studies have been conducted to determine the physical and engineering properties for different types of seeds, like as soybeans (Deshpande et al., 1993; Paulsen, 1978); oilbean (Oje and Ugbor, 1991); canola and wheat (Bargale et al, 1995); lentil (Carman, 1996); cumin (Singh and Goswami, 1996); sunflower (Gupta and Das, 1997; Gupta and Das, 2000); karingda (Suthar and Das, 1996); black pepper (Murthy and Bhattacharya, 1998); legume (Laskowski and Lysiak, 1999); locust bean (Ogunjimi et al., 2002); pigeon pea (Baryeh and Mangope, 2002); cotton (Ozarslan, 2002), chickpea (Konak et al., 2002); calabash nutmeg (Omobuwajo et al., 2003); Vetch (Yalcin and Ozarslan, 2004); Sweet corn (Coskun et al., 2005); flaxseed (Coskuner and Karababa, 2007).

Limited research has been conducted on the physical properties of Iranian cultivars safflower seeds. Gupta and Prakash (1992) reported some of those properties for safflower JSF-1-type seeds. However, volumetric expansion coefficient, equivalent diameter and fracture characteristics of safflower seed and their variations at various levels of moisture content have not been investigated. Erica et al. (2006) investigate the effect of moisture content on some physical properties of safflower seeds typically cultivated in Argentina. They indicated that volume, weight of seed, the expansion coefficient and porosity increase lineally with the increase in seed moisture content. In addition, they revealed that an increase in moisture content yielded a decrease in bulk density trend and true density varied nonlinearly. Isik and Izil (2007) investigated some moisture-dependent properties of sunflower for only the Turkey sunflower seed cultivar. They showed that the thousand grain mass, true density and porosity increased while the bulk density decreased with an increase in the moisture content range of 10.06–27.06 % (d.b.).

The objective of this study was to investigate the effects of moisture content and cultivar on physical attributes of three major Iranian cultivars of safflower seeds. The parameters measured at different moisture content (10-25% w.b) were size, geometric mean diameter, sphericity, thousand seed mass, bulk density, true density, porosity and terminal velocity for three major Iranian cultivars of safflower seeds (IL-111, LRV51-51 and Zarghan279).

2. MATERIALS AND METHODS

Three safflower seeds of cultivars namely IL-111, LRV51-51 and Zarghan279 were obtained from the farms in the Lorestan province, Iran. A mass of 15 kg from each cultivar was weighted and transported to the lab. The seeds were manually cleaned to get rid of foreign matters, broken and immature seeds. The initial moisture content of the seeds was determined by the vacuum oven method moisture (Official Method 14003, AOAC, 1980). The initial moisture content of the seeds was found 7.95, 6.89 and 8.64% for IL-111, LRV51-51 and Zarghan279, respectively. The seeds with the desired moisture content were obtained by adding calculated amounts of distilled water, thoroughly mixing and sealing them in separate polyethylene bags. The samples were kept at 5^{o}C in a refrigerator for at least a week to allow uniformity of moisture distribution. Before starting a test, the required quantity of the seeds was taken out of the refrigerator and allowed to warm up to room temperature. All the physical properties of the seeds were obtained for four moisture contents in the range 10-25% (w.b.) that is a usual range since harvesting, transportation, storage and processing operations of safflower seed. The tests were carried out with three replications for each moisture content.

To determine the average size of the seed, a sample of 100 seeds was randomly selected. The three linear dimensions of the seeds, namely length (*L*), width (*W*) and thickness (*T*) were measured using a micrometer reading to 0.01 mm. The geometric mean diameter, D_g, of the safflower seed was calculated by using the following relationship (Mohsenin, 1986):

$$D_g = (LWT)^{1/3} \qquad (1)$$

The criterion used to describe the shape of safflower seed was sphericity. The sphericity, φ, of safflower seed was determined using the following formula (Mohsenin, 1986):

$$\varphi = \frac{D_G}{L} \qquad (2)$$

The surface are (*S*) and projected area (A_p) of safflower seed was found by analogy with a sphere of the same geometric mean diameter using the following formulas (Mohsenin, 1986):

$$S = \pi \times D_g^2 \qquad (3)$$

$$A_p = \frac{\pi}{4} L \times W \qquad (4)$$

The 1000 unit mass was determined using precision electronic balance to an accuracy of 0.01g. The seed volume and true density (ρ_t), as a function of moisture content and seed cultivar, were determined by water displacement method (Adejumo et al., 2007). A bunch of 100 seeds of known average weight was dropped into a container filled with water. The bulk seeds were put into a container with known mass and volume (500 ml) from a height of 150 mm at a constant rate bulk density was calculated from the mass of bulk seeds divided by the volume containing mass (Tekin et al., 2006). This was achieved by filling a container of 500ml with grain from the height 0.15m striking the top level and then weighing the contents and the bulk density was determined from the measured mass and volume (Davies, 2010). For each of the moisture content, 10 replications were done and average was taken. The porosity (ε) of the bulk seed was computed from the values of the true density (ρ_t) and bulk density (ρ_b) of the seeds by using the relationship given by Mohsenin (1986):

$$\varepsilon = (1 - \frac{\rho_b}{\rho_t}) \times 100 \qquad (5)$$

To measure the terminal velocity of the samples, a vertical air column was designed, constructed and used. The apparatus is shown in Figure 1. It consists of a vertical transport column made of Plexiglas so that the suspended seeds could be seen from the outside, AC electric motor, fan and electric inverter. For each test, a sample was dropped into the air stream from the top of the air column, up which air was blown to suspend the seeds. Input air was adjusted by changing the velocity of the electric motor through an inverter set until the seeds began to float. The air velocity near the location of the seeds suspension was measured by a hot – wire anemometer having a least count of 0.1 m/s.

The static coefficient of friction for safflower seeds was determined with respect to two selected surfaces (galvanized steel and plywood). An open-ended galvanized iron cylinder, 80 mm diameter and 50 mm height, was filled with sample and placed on an adjustable tilting surface so that the cylinder dose not touch the surface. The surface was raised gradually until cylinder started to slide down. The angle of inclination (α) was read from graduated scale. The coefficient of friction (μ) was calculated from the following relationship (Mohsenin, 1986):

$$\mu = \tan g\alpha \qquad (6)$$

To determine the empting angle of repose, an apparatus consisting of a plywood box of 300 × 300 × 300 mm^3 with a removable front panel was used. The box was filled with the safflower seed samples at the desired moisture content, and the front panel was quickly removed, allowing the samples to flow and assume a natural slope. The empting angle of repose was calculated from the measurements of the vertical depth and radius of spread of the sample (Gupta and Das, 1997; Baryeh, 2002).

The result obtained from the study carried out on the three cultivars of safflower seeds at four different moisture contents at five replications at each moisture content, were subjected to analysis of variance (ANOVA), Duncan multiply range test and as well as Linear regression analysis using Statistical Analysis System.

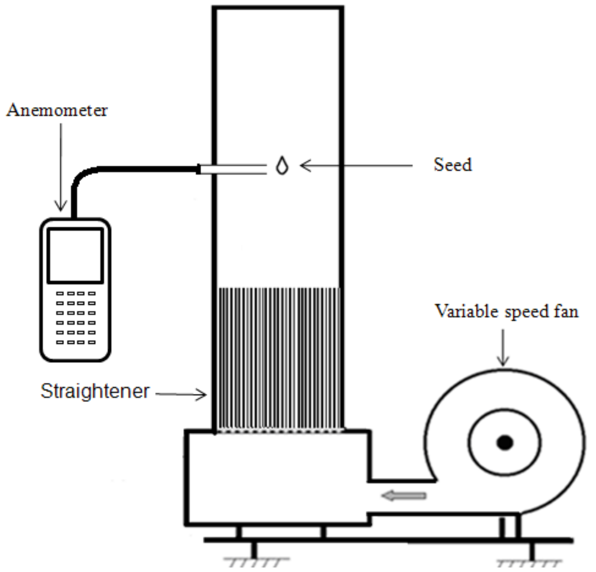


Fig. 1. Schematic diagram of wind tunnel for terminal velocity measurement.

3. RESULTS AND DISCUSSION

3.1. Geometric properties

The mean axial dimensions, geometric mean diameter, sphericity, surface area and projected area of three safflower cultivars at different moisture content, are presented in Table 1. It was observed that the axial dimensions increased with an increase in moisture content. In the sample of IL-111 about 89% had a length in the range of 8-8.5 mm, about 87% had a width of 4.2-4.5 mm while about 79% had a thickness in the range of about 3.2-3.6 mm. LRV51-51 recorded about 84% of its length fall into the range of 8-8.5 mm, about 88% had a width of 4.3-4.6 mm and about 66% had a thickness 3.2-3.6 mm. The experiment also revealed that 82% of Zarhghn279 had its length in the range of 4.4-7.6mm, likewise, 89% of the measured widths were in the range of 3.5-4 mm and 76% had a thickness fall within the range of 2.6-3.1 mm. The analysis of variance ANOVA result indicated that the differences among the moisture content level were significantly different at 5% probability level for the three cultivars. This trend was in agreement with the result reported by Davies and Zibokere (2011). Table 1 showes that as the moisture content increased from 10 to 25%, the geometric mean diameter of the IL-111, LRV51-51 and Zarghan279 cultivars increased from 4.97 to 5.36, 4.76 to 5.32 and 4.41 to 4.64 mm, respectively. The variations among the values were significantly different at 5% probability level.

The sphericity of safflower cultivars seeds increased as the result of increasing the their moisture content (Table 1). Similar trends have been reported by Sahoo and Srivastava (2002) for okra seed, Sacilik et al. (2003) for hemp seed, Coskuner and Karababa (2007) for flaxseed, Altuntas et al. (2005) for fenugreek seed and Davies and El- Okene (2009) for soybean. As the moisture content increased from10 to 25%, the sphericity of the IL-111, LRV51-51 and Zarghan279 cultivars increased from 61.75 to 63.04%, 58.89 to 62.63% and 59.49 to 59.67%, respectively. The variations among the values were significantly different at 5% probability level.

The effect of moisture content on the surface area of safflower seed cultivars is presented in Table1. The obtained result indicated that the surface area of seeds increased linearly with increasing in their moisture content. The surface area of the safflower cultivars of IL-111, LRV51-51 and Zarghan279 increased from 77.61 to 90.25 mm^2, 71.22 to 89.01 mm^2 and 61.15 to 67.49 mm^{2,} respectively, with increasing the moisture content from 10-25% (Table 1) The observed values were significant different ($p<0.05$). Similar trend was reported by Tunde-Akintunde and Akintunde (2008) for beniseed. The projected area of the safflower cultivars of IL-111, LRV51-51 and Zarghan279 increased from 27.36 to 31.42 mm^2, 27.44 to 31.09 mm^2 and 22.17 to 24.70 mm^{2,} respectively, with increasing the moisture content (Table 1).

3.2. Gravimetric properties

The results of Duncan's multiple range tests for comparing the mean values of the ggravimetric properties of safflower seeds at different moisture contents and cultivars are presented in Table 2. It is evident from the data in Table 2 that, as the moisture content of the seeds increased, the one thousand mass increased. Similar results were also reported by Davies and El -Okene (2009) for soybean, Tekin et al (2006) for Bombay bean, Selvi et al. (2006) for linseed, Cetin (2007) for barbunia and Garnayak et al. (2008) for jatropha seed. With increasing moisture content from 10 to 25% the mean values of the one thousand safflower seeds mass increased by 1.26 times. The average values for the one thousand seeds mass were found to be 38.76, 41.91, 44.93 and 47.88 g for moisture contents of 10, 15, 20 and 25%, respectively. The one thousand seeds mass for IL-111 is higher than other cultivars (Table 2). The moisture content and the cultivar had significant effects on the one thousand mass at 1% probability level. Moreover, according to Duncan's multiple range test results, the one thousand mass mean values at different moisture contents and different cultivars were statistically different from each other ($P<0.05$). In Figure. 2 the one thousand seeds mass is plotted against the moisture content, for each safflower cultivar. The figure reveals that, at all the safflower cultivars considered, the one thousand seeds mass increases as the moisture content increases. Its mean values increased from 41.14 to 50.13, 40.58 to 49.56 and from 34.92 to 43.94 g, for IL-111, LRV51-51 and Zarghan279 cultivars, respectively, as the moisture contents increased from 10 to 25%. Regression analysis was used to find and fit the best general models to the data. The results showed that as the moisture content of the seeds increased, the one thousand safflower seeds mass increased linearly. So the dependence of the one

Table 1: Mean and standard error for axial dimension, sphericity surface area and projected area of three varieties of safflower at different moisture contents.

Safflower variety	Moisture content (%)	Seed Dimensions Length (mm)	Width (mm)	Thickness (mm)	geometric mean diameter (mm)	Sphericity (%)	Surface are (mm^2)	Projected area (mm^2)
IL-111	10	8.05 (0.50)	4.33 (0.80)	3.52 (0.11)	4.97 (0.03)	61.75 (0.60)	77.61 (0.94)	27.36 (0.41)
	15	8.22 (1.32)	4.4 (0.26)	3.55 (0.60)	5.04 (0.11)	61.36 (0.41)	79.95 (2.76)	28.40 (2.14)
	20	8.34 (0.11)	4.68 (0.10)	3.76 (0.11)	5.27 (0.37)	63.22 (0.43)	87.36 (1.27)	30.64 (1.34)
	25	8.50 (0.15)	4.70 (0.81)	3.85 (0.08)	5.36 (0.26)	63.04 (1.36)	90.25 (1.79)	31.42 (0.68)
LRV51-51	10	8.08 (0.10)	4.32 (0.08)	3.04 (0.20)	4.76 (0.04)	58.89 (0.28)	71.22 (1.33)	27.44 (0.79)
	15	8.21 (0.12)	4.41(0.81)	3.09 (0.18)	5.05 (0.12)	61.48 (1.27)	80.16 (3.99)	28.46 (0.63)
	20	8.29 (0.15)	4.44 (0.11)	3.55 (0.05)	5.14 (0.04)	62.07 (0.73)	83.14 (1.34)	28.92 (1.09)
	25	8.53 (0.19)	4.66 (0.19)	3.70 (0.08)	5.32 (0.07)	62.63 (1.90)	89.01 (2.71)	31.09 (1.79)
Zarghan279	10	7.41 (0.11)	3.80 (0.05)	2.88 (0.10)	4.41 (0.08)	59.49 (0.49)	61.15 (2.44)	22.17 (0.68)
	15	7.50 (0.17)	3.80 (0.25)	3.04 (0.16)	4.34 (0.12)	57.93 (0.25)	59.41 (3.01)	22.43 (0.48)
	20	7.61 (0.10)	3.92 (0.15)	3.10 (0.90)	4.52 (0.06)	59.47 (0.54)	64.21 (1.84)	23.41 (0.59)
	25	7.77 (0.12)	4.05 (0.11)	3.20 (0.15)	4.64 (0.07)	59.67 (0.69)	67.49 (1.56)	24.70 (0.45)

(*); Standard deviation values are in parentheses.

Table 2. Effects of moisture content and safflower variety on the gravimetric properties of safflower seeds.

Independent variable	Gravimetric properties			
Moisture content (%)	One thousand seed mass (g)	True density (kg/m^3)	Bulk density (kg/m^3)	Porosity (%)
10	38.76 d	1017.05 d	632.18 a	37.65 d
15	41.91 c	1040.55 c	615.04 b	40.88 c
20	44.93 b	1064.62 b	596.91 c	43.92 b
25	47.88 a	1081.53 a	584.89 d	45.91 a
Safflower variety				
IL-111	45.44 a	1055.86 a	606.61 b	42.47 b
LRV51-51	44.72 b	1049.80 b	588.48 c	43.86 a
Zarghan279	39.44 c	1055.86 a	626.67 a	39.95 c

The columns not followed by the same letter are significantly different at the 5% level of significant as judged by Duncan tests.

thousand mass of safflower seeds (M_{1000}, g) on the seed moisture content (M, %) was expressed by the following best-fit equations for each cultivar:

$$M_{1000} = 0.679M + 33.74 \quad R^2=0.998 \quad \text{for: IL-111} \quad (7)$$

$$M_{1000} = 0.694M + 32.22 \quad R^2=0.990 \quad \text{for: LRV51-51} \quad (8)$$

$$M_{1000} = 0.665M + 27.46 \quad R^2=0.989 \quad \text{for: Zarghan279} \quad (9)$$

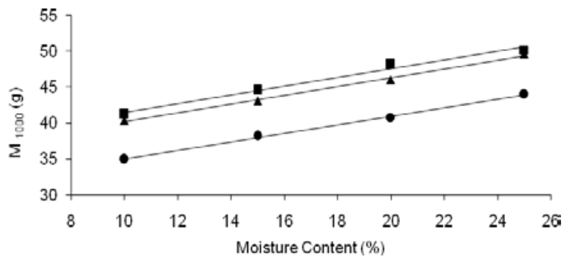


Figure 2. Variation of the one thousand safflower seeds mass with moisture content according to the safflower cultivars: ■ IL-111, ▲ LRV51-5 and ● Zargan279.

The true density of the seeds was evaluated according to the moisture content and safflower cultivar. The true density increased with increasing moisture content (Table 2). An increase in true density with an increase in moisture content was reported for cumin seeds (Singh and Goswami, 1996), pigeon pea (Baryeh and Mangope, 2002), safflower (Erica et al., 2006), flaxseed (Coskuner and Karababa, 2007), sunflower (Isik and Izli, 2007) and jatropha seed (Garnayak et al., 2008). These seeds have a higher weight increase in comparison with their volume expansion on moisture gain. However, Deshpande et al. (1993), Ozarslan (2002) and Konak et al. (2002) have found that the true density of soybeans, cottonseed and chickpea seed respectively decreases as the seed moisture content increases.

The average values of the true density varied from 1014.05 to 1081.53 kg/m^3, between the lowest and the highest moisture content (Table 2). Seifi et al. (2010) reported the true density of safflower seed (Goldasht cultivar) in the range of 1010 to 1070 kg/m^3 in the moisture range from 3.9 to 22%.w.b. The average values for the true density were found to be 1055.86, 1049.8 and 1044.9 kg/m^3 for the for IL-111, LRV51-51 and Zarghan279 cultivars, respectively. The higher value of true density were found at the IL-111 due to the larger seeds of this cultivar (Table 1). Both the moisture content and the cultivar of the safflower significantly affected the true density (P<0.01). According to Duncan's multiple range test results, the true density mean values at different moisture contents and different cultivars were statistically different from each other (P<0.05) (Table 2).

Figure. 3, shows the variation of the safflower seeds true density with the moisture content at each cultivar. As follows from the relations presented in the figure, for all the cultivars considered, the true density of seeds increased with increase in their moisture content. Its mean values increased from 1019.31 to 1089.70, 1012.45 to 1079.7 and from 1010.38 to 1075.17 kg/m^3, for IL-111, LRV51-51 and Zarghan279 cultivars, respectively, as the moisture contents increased from 10 to 25%. Regression analysis showed that the true density increased as a polynomial with increasing with increasing moisture content at all cultivars. The relationship between the true density (ρ_t, kg/m^3) and moisture contents (M, %) at each safflower cultivar can be expressed by the following best-fit regression equations:

$$\rho_t = -0.054M^2 + 6.604M + 958.6 \quad R^2=0.999 \quad \text{for: IL-111} \qquad (10)$$

$$\rho_t = -0.148M^2 + 9.734M + 929.6 \quad R^2=0.998 \text{ for: LRV51-51} \qquad (11)$$

$$\rho_t = -0.084M^2 + 7.333M + 945.2 \quad R^2=0.999 \quad \text{for: Zarghan279} \qquad (12)$$

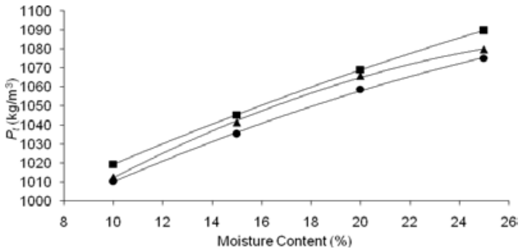


Figure 3. Changes of the true density with moisture content according to the safflower cultivars: ■ IL-111, ▲ LRV51-5 and ● Zargan279.

The bulk density of safflower seeds decreased from 632.13 to 584.89 kg/m^3 with increasing moisture content from 10 to 25% (Table 2). This effect of moisture content was also reported by Deshpande et al. (1993), Carman (1996), Visvanathan et al. (1996), Ogut (1998) and Garnayak et al. (2008) for soybean, lentil seeds, neem, white lupin and jatropha seed, respectively. The values of the bulk density varied from 632.13 to 584.89 kg/m^3 between the lowest and highest moisture contents. Its values varied between 630.13 to 587.25, 616.38 to 566.12 and 650.05 to 601.29 kg/m^3 for the for IL-111, LRV51-51 and Zarghan279 cultivars, respectively, at different moisture contents that were studied. It was observed that Zarghan279 cultivar with smallest dimensions seeds had the greatest bulk density. This could be adduced to the fact that small seeds are likely to be well compacted than the larger size. This is in agreement with findings of Adegbulugbe and Olujimi (2008) for three cultivars of cowpea and Altuntas and Yildiz (2007) for faba bean.

The values of the bulk density were significantly affected by moisture content and safflower cultivar (P<0.01). According to the Duncan multiple range test results, the bulk density mean values at different moisture contents and different cultivars were statistically different from each other (P<0.05) (Table 2). Figure. 4 shows the variation of bulk density with moisture content for all the cultivars. The values of this interaction varied from 587.25 to 650.05 kg/m^3 that occurred in the IL-111 cultivar at the lowest moisture content and in the Zarghan279 cultivar at the highest moisture content, respectively. The models fitted to the data using the regression technique showed that the bulk density decreased linearly with increases in the moisture content for all cultivars. So the following equations were found for the relationship between bulk density (ρ_b, kg/m^3) and moisture content (M, %), for each cultivar:

$$\rho_b = -3.180M + 663.9 \quad R^2=0.978 \quad \text{for: IL-111} \qquad (13)$$

$$\rho_b = -3.803M + 656.9 \quad R^2=0.994 \text{for: LRV51-51} \qquad (14)$$

$$\rho_t = -3.731M + 693.8 \quad R^2=0.996 \text{ for:Zarghan279} \qquad (15)$$

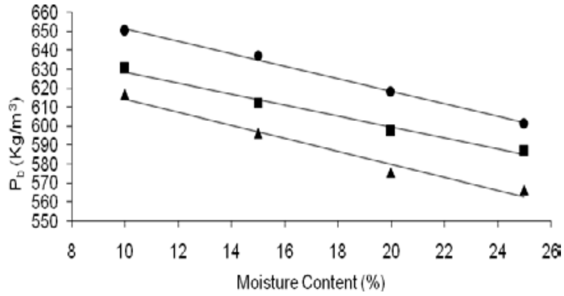


Figure 4. Variation of the bulk density with moisture content according to the safflower cultivars: ■ IL-111, ▲ LRV51-5 and ● Zargan279.

The porosity of safflower seeds was evaluated as a function of moisture content and cultivar. From the data in the Table 2, it can be seen that the porosity of safflower seeds increased from 37.65 to 45.91% as their moisture content increased from 10 to 25%. Similar trends were reported for sunflower seed (Gupta and Das, 1997), lentil seeds (Carman, 1996), pigeon pea (Baryeh and Mangope, 2002) and safflower (Erica et al., 2006), but different to that reported for soybean (Deshpande et al., 1993), safflower (Gupta and Prakash, 1992) and pumpkin seeds (Joshi et al., 1993).

Porosity is the property of grains that depends on its bulk and true density and this dependence is different for every seed. The average values for the porosity were found to be 42.47, 43.89, and 39.95% for the for IL-111, LRV51-51 and Zarghan279 cultivars, respectively. It was observed that LRV51-51 cultivar with larger linear dimensions had the greatest porosity (Table 2). The moisture content and the cultivar had a significant effect on the property (P<0.01). In addition, according to the Duncan multiple range tests, the values for the property were completely different for the moisture contents and cultivars (Table 2). Figure. 5 presents the relationship between the property and moisture content for all the safflower cultivars studied. As moisture content of the seeds increased, the property increased in all the cultivars (Figure 5). The greatest property value was obtained as 47.56% in the LRV51-51 cultivar at the moisture content of 25%, while the lowest value was found to be 35.6% in the Zarghan279 cultivar at a moisture content of 10%. It was found that the property of seeds increased as a linear function of their moisture content for all the cultivars. The following relationships were found between the property (ε, %) and moisture content (M, %), for each safflower cultivar:

$$\varepsilon = 0.588M + 31.88 \quad R^2=0.979 \text{ for: IL-111} \qquad (16)$$

$$\varepsilon = 0.632M + 32.47 \quad R^2=0.984 \quad \text{for: LRV51-51} \qquad (17)$$

$$\varepsilon = 0.635M + 28.49 \quad R^2=0.986 \quad \text{for: Zarghan279} \qquad (18)$$

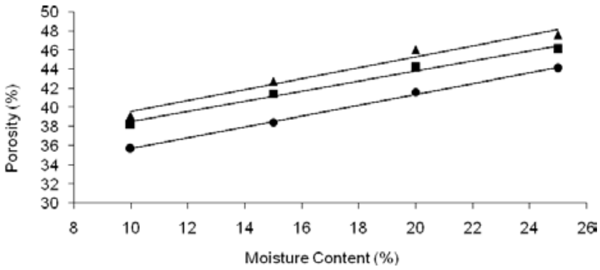


Figure 5. Variation of porosity of safflower seeds with moisture content according to the safflower cultivars: ■ IL-111, ▲ LRV51-5 and ● Zargan279.

3.3. Terminal velocity

The terminal velocity of safflower seeds increased with increasing moisture content (Table 3). This effect of moisture content was also reported by Joshi et al. (1993) for pumpkin seeds, Carman (1996) for lentil seeds, Singh and Goswami (1996) for cumin seeds, Suthar and Das (1996) for karingda seeds, and Aydin and Akar (2005) for gumbo fruit. The values of the terminal velocity varied from 7.45 to 8.37 m/s between the lowest and highest moisture contents. The reason for this difference may be attributed to the increase in mass of the seeds per unit, when their frontal areas were presented to the airflow to suspend the material. The other reason is probably that the drag force is affected by the moisture content of particle. Its values varied between 7.96 to 8.9 m/s, 7.6 to 8.46 m/s, and 6.81 to 7.75 m/s for the IL-111, LRV51-51 and Zarghan279 cultivars, respectively, at the different moisture contents that were studied. Terminal velocities for IL-111 cultivar were observed to be higher than those obtained for LRV51-51 and Zarghan279 cultivars. This result can be explained by the fact that the seeds of IL-111 were bigger than that of LRV51-51 and Zarghan279. 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. 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 values terminal velocity were significantly affected by moisture content and safflower cultivar (P<0.01). According to the Duncan multiple range test results, these values were different from each other for the distinct cultivars. There were no statistically significant differences between moisture contents of 20 and 25% (Table 3).

Figure. 6 shows the variation of the terminal velocity with moisture content for all the safflower cultivars. The values of this interaction varied from 6.8 to 8.9 m/s that occurred in the Zarghan279 cultivar at the lowest moisture content and in the IL-111 cultivar at the highest moisture content, respectively. The models fitted to the data using the regression technique showed that the terminal velocity increased linearly with increases in the moisture content for all the safflower cultivars. So the following equations were found for the

Table 3. Effects of moisture content and safflower variety on the terminal velocity of safflower seeds.

Independent variable	Terminal velocity (m/s)
Moisture content (%)	
10	38.76 d
15	41.91 c
20	44.93 b
25	47.88 a
Safflower variety	
IL-111	45.44 a
LRV51-51	44.72 b
Zarghan279	39.44 c

The columns not followed by the same letter are significantly different at the 5% level of significant as judged by Duncan tests.

relationship between the terminal velocity (V_t, m/s) and moisture content (M, %), for each safflower cultivar:

$V_t = 0.066M + 7.336$ R^2=0.977 for: IL-111 (19)

$V_t = 0.067M + 6.92$ R^2=0.973for: LRV51-51 (20)

$V_t = 0.062M + 6.273$ R^2=0.989 for: Zarghan279 (21)

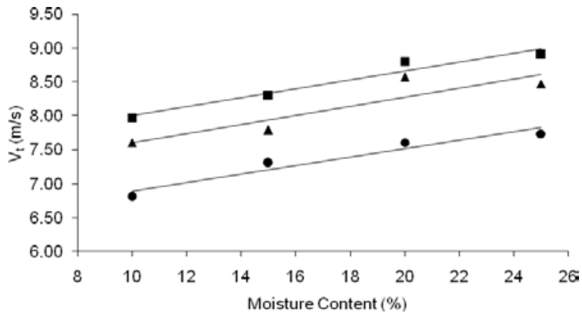


Figure 6. Changes of the terminal velocity of safflower seeds with moisture content according to the safflower cultivars: ■ IL-111, ▲ LRV51-5 and ● Zargan279.

3.4. Frictional properties

The results for the static coefficients of friction for the three safflower cultivars (IL-111, LRV51-51 and Zarghan279) on the two structural surfaces (plywood and galvanized steel) against different levels of moisture contents are presented in Table 4. As it can be found from the data in Table 4, the static coefficient of friction increased (probability <0.05) with increase in moisture content for all the surfaces and for all cultivars. The increase in static coefficients of friction with increased moisture is similar to that obtained by Aviara et al. (1999) for almond nut, Baryeh (2002) for millet, Coskun et al. (2005) for sweet corn, Cetin (2007) for barbunia bean and Razavi et al. (2007) for almond nut. This may be explained by increased cohesive force of wet seeds with the structural surface, since the surface becomes stickier as moisture content increases. The friction coefficient against the galvanized steel surface for the IL-111, LRV51-51 and Zarghan279 cultivars increased significantly from 0.39 to 0.57, 0.37 to 0.51 and from 0.36 to 0.46, respectively, with increase in moisture content from 10 to 25%. Whereas, against the plywood surface, the data were from 0.42 to 0.62, 0.37 to 0.55 and 0.38 to 0.53 for the IL-111, LRV51-51 and Zarghan279 cultivars, respectively

(Table 4). For three cultivars and at all the moisture contents, the least static coefficient of friction was on the galvanized steel surface. This may be owing to smaller cohesive force between safflower seeds and the galvanized steel than the plywood surface. In addition, the static coefficients of friction for IL-111 cultivar were higher than that of the LRV51-51 and Zarghan279 cultivars at similar moisture contents of the seeds and the same surfaces. This shows that, the values of static coefficients of friction of safflower seed increased with an increase in seed size.

The results of empting angle of repose for studied cultivars of safflower seeds at different moisture levels are shown in Table 4. As Seen from the data in Table 4, the emptying angle of repose increased with an increase in moisture content. The increase of the angle of repose for the safflower cultivars is attributed mainly to increase in size of the seeds in higher moisture content, also the higher moisture content cause higher sickness of seeds surfaces and then lowers easiness of rolling seeds on each other as reported by Razavi et al. (2007). The values of the angle of repose for the IL-111, LRV51-51 and Zarghan279 cultivars were increased from 39.78 to 52.63 °, 41.15 to 53.98 ° and from 44.56 to 59.34°, respectively, with increase in moisture content from 10 to 25% (Table 4). The values of the angle of Zarghan279 cultivar with small size seeds were higher than that of the IL-111, LRV51-51cultivars with large size seeds. This can be due to the higher sphericity of the shape for the seeds of large size in comparison with small size, which permits the easiness of sliding of seeds on each other and causes a higher value for radius of spread of the seed. This is similar to that reported by Konak et al. (2002) for chickpea and Paksoy and Aydin (2004) for edible squash.

Table 4. Mean and standard error for frictional properties (coefficient of friction and empting angle of repose) of three varieties of safflower at different moisture contents.

Safflower variety	Moisture content (%)	Coefficient of friction		Empting angle of repose (Degree, °)
		Galvanized iron	Plywood	
IL-111	10	0.391 ±0.011	0.423 ±0.080	39.78 ±3.43
	15	0.447 ±0.032	0.473 ±0.026	44.15 ±2.02
	20	0.489 ±0.011	0.534 ±0.010	47.83 ±4.81
	25	0.576 ±0.015	0.624 ±0.081	52.63 ±3.08
LRV51-51	10	0.375 ±0.018	0.372 ±0.042	41.15 ±4.04
	15	0.401 ±0.012	0.434±0.081	46.24±3.42
	20	0.463 ±0.015	0.476 ±0.011	49.53 ±2.98
	25	0.513 ±0.021	0.554 ±0.019	53.98 ±3.76
Zarghan279	10	0.367 ±0.017	0.382 ±0.051	44.56 ±3.45
	15	0.384 ±0.014	0.412 ±0.025	48.57 ±4.05
	20	0.421 ±0.013	0.491 ±0.016	53.09 ±3.32
	25	0.463 ±0.012	0.534 ±0.011	59.34 ±2.56

4. CONCLUSIONS

Based on the investigation conducted on the some engineering properties of three cultivars of safflower seeds namely IL-111, LRV51-51 and Zarghan279 at moisture contents of 10, 15, 20 and 25% (w.b.). The following conclusions were drawn: the average values of geometric properties of safflower seeds include length, width, thickness, geometric mean diameter, sphericity, surface area and projected area ranged from 7.85 to 8.26 mm, 4.15 to 4.47 mm, 3.22 to 3.60 mm 4.71 to 5.10 mm, 60.04 to 61.77%, 69.99 to 82.25 mm2 and from 25.65 to 29.07 mm2, accordingly, as the moisture content increased from 10 to 25%. The values of geometric properties were higher for IL-111cultivar than the LRV51-51 and Zarghan279 cultivars. The values of the bulk densities of the safflower cultivars decreased from 630 to 587 kg/m3, 616 to 566 kg/m3 and from 650 to 602 kg/m3 for IL-111, LRV51-51 and Zarghan279 seeds, respectively. The thousand grain mass ,true density and porosity were found to increase from 41.14 to 50.13g, 1019 to 1089 kg/m3 and 38.18% to 46.10% for IL-111, from 40.29 to 49.58g, 1012 to 1079 kg/m3 and 39.11% to 47.56% for LRV51-51 and from 34.92 to 43.94g, 1010 to 1075 kg/m3 and 35.60 to 44.07% for Zarghan279. The terminal velocities of the safflower seeds for all cultivars increased linearly with increase in moisture content. The values obtained showed that terminal velocities increased from 7.96 to 8.90 m/s, 7.60 to 8.64 m/s and 7.78 to 6.80 m/s for IL-111, LRV51-51 and Zarghan279 seeds, respectively. The static coefficient of friction was determined for two structural surfaces namely, galvanized steel and plywood. The coefficient of friction increased with increase in moisture content and the plywood surface had higher coefficient of static friction for all the three cultivars. The values of the angle of repose for the IL-111, LRV51-51 and Zarghan279 cultivars increased from 39.78 to 52.63°, 41.15 to 53.98° and from 44.56 to 59.34°, respectively, with increase in moisture content from 10 to 25%.

REFERENCES

Adegbulugbe, T.A and Olujimi, O.A. (2008). Evaluation of Varietal and Moisture Content Effect of Some Cowpea Cultivars (*Vigna Unguiculata (L) Walp*) on their bulk and particle densities and porosity. http//www.ajol.info/journals/nifoj.

Adejumo, O.I., Alfa, A.A. and Mohammed A. (2007). Physical Properties of Kano white Cultivar of Bambara Groundnut. Nigerian Academic Forum, (12)1, 68-77.

Altuntas, E. and Yildiz M. (2007). Effect of moisture content on some physical and mechanical properties of faba bean *(Vicia faba L.)* grain. Journal of Food Engineering, 78:174-183.

Altuntas, E., Ozgoz E and Taser O.F. (2005). Some physical properties of fenugreek (Trigonella foenum-graceum L.) seeds. J. Food Eng. 71: 37–43.

AOAC. (1980). Association of the official analytical chemists, method 14.003. In: W. Howitz, Ph.D. (Ed.), Official Methods of Analysis of the Association of the Official Analytical Chemists (13th ed.). Washington DC, USA.

Aviara, N.A., Geandzang, M.I and Hague M.A. (1999). Physical properties of guna seeds. J. Agric. Eng. Res. 73: 105-111.

Aydin, C and Akar R. (2005). Some physical properties of gumbo fruit cultivars. J. Food Eng. 66: 387–393.

Bargale, P. C., Irudayaraj, J and Marquis, B. (1995). Studies on Rheological behavior of canola and wheat. Journal of Agricultural Engineering Research, 61, 267–274.

Baryeh, E.A and Mangope, B.K. (2002). Some physical properties of QP-38 cultivar pigeon pea. J. Food Eng. 56: 59-65.

Baryeh, E.A. (2002). Some physical properties of millet. J. Food Eng. 51: 39–46.

Baumler, E., Cuniberti, A., Nolasco S. M., and Riccobene I. C. (2006). Moisture dependent physical and compression properties of safflower seed. Journal of Food Engineering, 72(2), 134–140.

Carman, K. (1996). Some physical properties of lentil seeds. J. Agric. Eng. Res. 63: 87–92.

Cetin, M. (2007). Physical properties of barbunia bean (*Phaseolus vulgaris*L. *cv. 'Barbunia'*) seed. J. Food Eng. 80: 353-358.

Coskun, M.B., Yalcin, I and Ozarslan C. (2005). Physical properties of sweet corn seed (Zea mays saccharata Sturt.). J. Food Eng. 74: 523–528.

Coskuner, Y and Karababa, E. (2007). Some physical properties of flaxseed (Linum usitatissimum L.). J. Food Eng. 78: 1067–1073.

Davies, R.M and Zibokere, D S. (2011). Effect of Moisture Content on Some Physical and Mechanical Properties of Three Cultivars of Cowpea (Vigna unguiculata (L) Walp). Agricultural Engineering International: CIGR Journal, Manuscript No.1700. Vol. 13, No.1.

Davies, R.M. 2010. Some physical properties of Arigo seed. Int. Agrophysics, 24:89-92.

Davies, R.M., and El-Okene A.M. (2009). Moisture-dependent physical of Soybeans. Int. Agrophysics, 23:299-303.

Deshpande, S.D., Bal, S and Ojha, T.P. (1993). Physical properties of soybean. J. Food Eng. Res. 56: 89-98.

Erica, B., Cuniberti, A and Susana, M. (2006). Moisture dependent physical and compression properties of safflower seed. J. Food Eng. 2: 134-140.

Garnayak, D.K., Pradhan, R.C., Naik, S.N and Bhatnagar, N. (2008). Moisture-dependent physical properties of Jatropha seed (Jatropha curcas L.). J. Indian Crops Products 27: 123-129.

Gupta, R.K and Das, S.K. (1997). Physical properties of sunflower seeds. J. Food Eng. 66: 1-8.

Gupta, R.K and Das, S.K. (2000). Fracture resistance of sunflower seed and kernel to compressive loading. Journal of Food Engineering, 46, 1–8.

Gupta, R.K., Prakash, S. (1992). The effect of seed moisture content on the physical properties of JSF-1 safflower. Journal of Oilseeds Research, 9(2), 209–216

Isik, E., Izli N. (2007). Physical properties of sunflower seeds. Int. J. Agric. Res. 8: 677-686.

Joshi, D.C., Das S.K and Mukherjee, R.K. (1993). Physical properties of pumpkin seeds. J. Agric. Eng. Res. 54: 219-229.

Kahrs, J. (1994). Aerodynamics properties of weeds seeds. Int. Agrophysics, 8, 259-262

Konak, M., Carman, K and Aydin, C. (2002). Physical properties of chick pea seeds. Biosystems Eng. 82: 73-78.

Laskowski, J., and Lysiak, G. (1999). Use of compression behaviour of legume seeds in view of impact grinding prediction. Powder Technology, 105, 83–88.

Mohsenin, N.N. (1986). Physical Properties of Plant and Animal Materials. 2nd Revised and Updated Edition. Gordon and Breach Science Publishers. New York.

Murthy, C. T., and Bhattacharya, S. (1998). Moisture dependant physical an uniaxial compression properties of black pepper. Journal of Food Engineering, 37, 193–205.

Ogunjimi, L. A. O., Aviara, N. A., and Aregbesola, O. A. (2002). Some engineering properties of locust bean seed. Journal of Food Engineering, 55, 95–99.

Ogut H., 1998. Physical properties of white lupin. J. Agric. Eng. Res. 69: 273-277.

Oje, K., and Ugbor, E. C. (1991). Some physical properties of oil bean seed. Journal of Agricultural Engineering Research, 50, 305–313.

Omobuwajo, T. O., Omobuwajo, O. R., and Sanni, L. A. (2003). Physical properties of calabasch nutmeg (Monodora myristica) seeds. Journal of Food Engineering, 57, 375–381.

Ozarslan, C. (2002). Physical properties of cotton seed. Biosystems Eng. 83: 169–174.

Paksoy, M., and Aydin, C. (2004). Some physical properties of Edible squash seeds. J. Food Eng. 65: 225–231.

Paulsen, M. R. (1978). Fracture resistance of soybeans to compressive loading. Transactions of ASAE, 1210–1216.

Razavi, M.A., Rafe A., Mohammadi moghaddam, T and Amini A. (2007). Physical properties of pistachio nut and its kernel as a function of moisture content and cultivar. J. Food Eng. Res. 56: 89-98.

Sacilik K., Tarimci, C and Colak, A. (2007). Moisture content and bulk density dependence of dielectric properties of safflower seed in the radio frequency range. Journal of Food Engineering, 78, 1111–1116.

Sacilik, K., Ozturk, R and Keskin R. (2003). Some physical properties of hemp seed. Biosystems Eng. 86: 191–198.

Sahoo, P.K., and Srivastava, A.P. (2002). Physical properties of okraseed. *Bio-systems Engineering*, 83, 441–448

Seifi, M.R., Alimardini, R., Akram, A and Asakereh, A. (2010). Moisture –Depend physical properties of safflower (Goldasgt). Advance Journal of Food Science and Technology, 2(6), 340-345.

Selvi K.C, Pinar, Y and Yesiloglu, E. (2006). Some physical properties of linseed. Biosystems Eng. 95: 607–612.

Shahbazi, F., Nazari, G.M., Taheri, G.A and Mohtasebi, S.S. (2001). Physical properties of safflower stalk. Int. Agrophys, 25(3), 281-286

Singh K.K and Goswami, T.K. (1996). Physical properties of cumin seed. J. Agric. Eng. Res. 64: 93-98.

Suthar, S.H and Das, S.K. (1996). Some physical properties of karingda seeds. J. Agric. Eng. Res. 65: 15-22.

Tekin, Y., Isik H., Unal, H and Okursoy, R. (2006). Physical and mechanical properties of Turkish Goynuk Bombay beans (*Phaseolus vulgaris L.*). Pakistan Journal of Biolojical Scince, 9, 2229-2235.

Tunde-Akintunde T.Y. and Akintunde, B.O. (2008). Effect of moisture content and cultivar on selected properties of beniseed. Agric. Eng. Int.: CIGR EJournal, 9, 1-13

Visvanathan, R., Palanisamy, P.T., Gothandapani, L and Sreenarayanan V.V. (1996). Physical properties of neem nut. J. Agric. Eng. Res. 63: 19–25.

Yalcin, I and Ozarslan, C. (2004). Physical properties of vetch seed. Biosys. Eng. 88: 507–512.

USING LEAN PRINCIPLES & SIMULATION FOR PRODUCTIVITY IMPROVEMENT: THE CASE OF A CERAMIC INDUSTRY

Tiago Silva[a], Ana Luísa Ramos[b], José Vasconcelos Ferreira[b]

[a]DEGEI - Department of Economics, Management and Industrial Engineering, University of Aveiro, Campo de Santiago, 3810-193 Aveiro, Portugal
[b] GOVCOPP Research Unit / DEGEI - Department of Economics, Management and Industrial Engineering, University of Aveiro, Campo de Santiago, 3810-193 Aveiro, Portugal

[a]tgs@ua.pt, [b]aramos@ua.pt, josev@ua.pt

ABSTRACT
In this work is illustrated an application of lean principles & simulation in order to increase shop-floor productivity at Vista Alegre Atlantis, SA, a well-known Portuguese ceramic industry. The case study reported in this paper is the outcome of the business internship program sponsored by the Department of Economics, Management and Industrial Engineering of University of Aveiro for the students in the Industrial Management and Engineering master program. A simulation model of the current operation of the finishing section was developed, along with the creation of a value stream mapping and the identification of waste, in order to ascertain its limitations and problems. The relevant operational performance measures such as throughput, work-in-process, and queue statistics, were analyzed to allow the proposal of a set of changes to the existing manufacturing operations. The outcome of the simulation study was taken into account by the decision-makers and the recommendations are being implemented.

Keywords: lean manufacturing, productivity, simulation, ceramic industry

1. INTRODUCTION

Nowadays industry is facing a change in market conditions and customers' requirements. Many organizations are competing in a "red ocean" (Kim and Mauborgne 2005) struggling to reduce production costs and maintain a certain margin of profit.

Lean manufacturing concepts and tools are proving to be a good practice for those organizations who want to become more competitive through waste reduction and value-added creation, despite some criticism regarding aspects such as the human factor and the capacity to deal with variability (Hines et al. 2004) or the critical differences between the application of lean principles in a discrete manufacturing environment and in a continuous process manufacturing environment (Howell 2010). Typically, as frequently referred in the literature, the major benefits of adopting lean manufacturing principles and tools include inventory and lead time reduction, improved product quality and essentially waste (*muda*) elimination, i.e., everything that our customer is not willing to pay for (Melton 2005).

Uncertainty in demand has become the new challenge in ceramic industry (Grahl 2003). In order to fulfil costumer requirements, ceramic industries must be able to adapt their level of productivity and their time of response to market as well as improving their level of quality, being "focused on continuous improvement" (Howell 2011).

The study reported in this paper was carried out at Vista Alegre Atlantis, SA, (Figure 1) a distinguished Portuguese ceramic industry that is facing the problems mentioned above and that needs to increase shop-floor productivity while maintaining high levels of quality and flexibility.

Figure 1: Ceramic art from Vista Alegre (source: http://www.myvistaalegre.com/pt/)

The study focus the elimination of waste and the creation of value to the customer. Using lean manufacturing principles and tools and simulation techniques was possible to analyze different scenarios seeking for the one who meet the new market challenges.

2. SYSTEM IN ANALYSIS: THE FINISHING SECTION

Through the observation of *gemba* it was possible to analyze the current layout of the finishing section, the main processes and the current flow of materials and information. This section finishes the pieces produced in the two high-pressure machines Netschz and Sama. As can be observed in Figure 2, the finishing section has three workstations (P1, P2 and P3) being P1 and P2 responsible for making the finishing operations and P3 for storing the pieces in the transporter (to be transported for the downstream process) and recording the pieces produced. When needed, P3 makes some finishing operations. If the pieces that are being conformed in the high-pressure machines have a long finishing time, then some are processed completely in workstation P1 and others are sent unfinished to P3.

Between P1 and P3 there is a round turntable which is used as a WIP buffer and between P2 and P3 there is a conveyor which also serves as a WIP buffer.

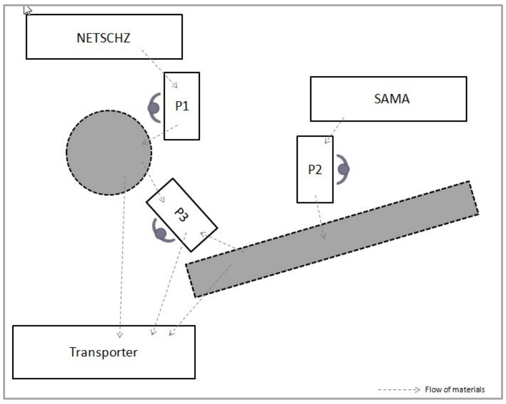


Figure 2: Layout of the finishing section

In this first analysis, it was possible to observe that exists a considerable amount of stock in buffers during the process. With this first information obtained in the field it was made a Value Stream Mapping (VSM) and a process chart (Figure 3) in order to map the activity, providing a better understanding of the process in analysis and in order to find opportunities for improvement.

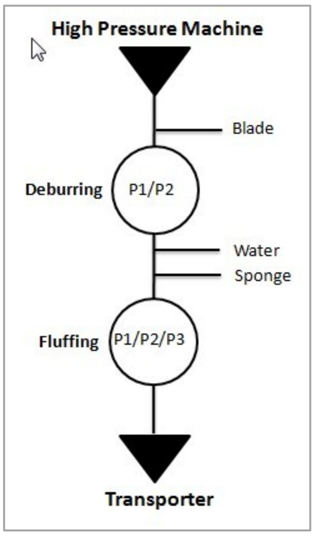


Figure 3: Process chart for the finishing process

This section works in continuous labor with four shifts, each one works eight hours a day. The machines don't have high setup times or other significant constrains.

While looking at the process in *gemba* it was possible to identify several types of *muda* such as waiting, transportation and movement.

It is believed that this section is operating below its capacity due to flow inefficiencies. The time that pieces are waiting to be processed causes high costs associated with non-quality. The actual level of productivity is estimated in 80% and every shift finishes an average of 2000 pieces.

3. DEVELOPING THE SIMULATION STUDY

Simulation is becoming a key strategy in order to describe and analyze different scenarios in industrial plants, because it supplies fundamental data of the new system without implementing it, becoming a cheaper solution (Bruzzone et al. 2013). Simulation can be utilized to explore and document potential opportunities for improvement and it is especially useful in the presentation of results to the direction board (Adams et al. 1991).

Longo (2011) states that modelling and simulation is the best methodology for solving problems in real world complex systems.

Bruzzone et al. (2013) suggest that a simulation model must follow a set of steps in order to achieve the maximum potential of the methodology. In recent years, a lot of research in how to develop a simulation study has been made and it is possible to conclude that the required steps to achieve the best path include problem formulation, conceptual modelling and data collection, operational modelling, verification and validation (V&V), experimentation, and output analysis (Kelton et al. 2010).

The use of simulation is particularly advantageous when the complexity or operational variability of the systems under study renders the application of purely analytical models impossible.

3.1. Formulating the problem

A simulation model of the current operation of the finishing section was developed. The main objective of this simulation study was to document the current state of the section in analysis, identify waste (*muda*) in the process, and improving productivity in 15 % by using lean tools and concepts in the model.

The relevant operational performance measures such as throughput, work-in-process, and queue statistics, were analyzed to allow the proposal of a set of changes to the existing manufacturing operations.

The model was developed using Arena® software from Rockwell Software. This benchmark software is the adopted environment for the simulation courses at University of Aveiro, providing the required features to develop, analyze and animate valid and credible simulation models.

3.2. Conceptual modelling and input data collection

In developing the simulation model particular care was taken to model the finishing process as close to reality as possible. In this stage it was necessary to determine which data would be necessary to use in the model and if this information was available. Talking to a ceramic engineer of the plant it was possible to find that most of the information required was stored in the enterprise information system SAP, but data was not trustable (in some cases).

For example, the data provided by SAP for the processing times of the components were not valuable since were outdated. The solution was to measuring the times in the field. In this stage, a constrain appeared. In this section, thousands of references are processed, and the processing times vary substantially for different references. Thus, an ABC analysis was made in order to identify the most relevant references and simplify this input parameter. It was possible to conclude that four references represent 22% of the total production. These four references were used to determine the processing times used as input in the simulation model.

The time between arrivals for the pieces who would be processed in the model, coming from the two distinct high-pressure machines, has been determined using SAP and was considered as a deterministic input parameter.

The availability of data for the processing times of the tasks involved in the finishing process allowed the fitting of proper distributions to these data. The distributions and its parameters were selected using the Arena's software module Input Analyzer (Figure 4). The distributions obtained were analyzed trough visual inspecting, square error value and p-value, in order to guarantee a "good" fit.

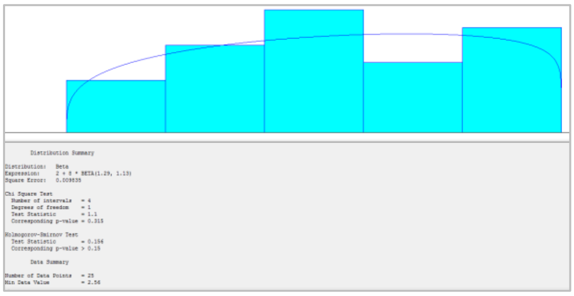


Figure 4: Fitting values to a standard distribution using input analyzer

Regarding material handling operations, a round turntable and a conveyor transport the pieces, and both have buffer functions as well. They were modeled as conveyors and the data necessary was gathered on the field, such as length or velocity. As neither maintenance procedures nor equipment failures influence significantly the regular operation of the system, these were ignored.

One of the *muda* found was movement, that is, human resources have to leave their working stations to get the pieces they need to work and then pass those pieces to the next process. This *muda* was modeled considering the operator in the workstation as the transporter resource. The priority of the process was considered "high" and the priority of the transportation was considered "low". This way the operator finishes the component before getting another one. Since the transportation time varies, input data were gather on the field and then fitted to a standard distribution using input analyzer.

3.3. Operational modelling

As already mentioned, the simulation model of the actual ceramic finishing system was developed using the Arena simulation software. This model was used to: (i) allow for a better understanding of the actual system, (ii) identify critical aspects and opportunities for improvement, (iii) gain the confidence of the decision makers and (iv) try lean solutions in order to improve the productivity of the system-in-analysis.

The run parameters of the model were defined as following:

- Replication length: 1 day of operation;
- Number of replications: 10.

The number of replications was determined through a trial-and-error approach until confidence intervals were reasonable.

The operational model was developed using several modules from Arena templates and it was developed a 2D animation model (Figure 5) illustrating the dynamic behavior of resources, transporters, conveyors, and buffers.

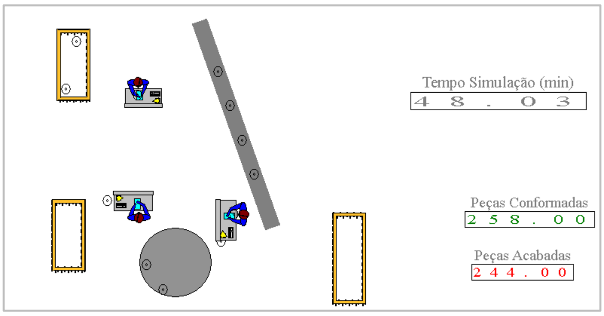


Figure 5: Animation model of the finishing section actual layout

After the operational modelling phase of the study it was conducted the V&V phase and the results were analyzed.

3.4. Verification and Validation (V&V)

The model was verified and validated using different techniques such as animation, internal validity, predictive validation, structured walkthrough, and examination of model traces. The team member who accompanied the project on site was crucial in this process, as he combined the knowledge of the simulation tool being used with the perception gained on the finishing process details.

The animation and the comparison of predicted performance measures with the known behavior of the current system (predictive validation), were the dominant techniques employed, as they allowed the involvement of the decision makers in the validation process.

When the model was experimented, for the first time, the throughput value was 33% lower than the expected value. This occurred due to a fail in the input data. It was considered that the capacity of the human resources was one, when in fact, when transporting components they carry one, two or more depending on the size and weight of the pieces. This parameter was refined and the output values become validated.

The verification and validation process was crucial for gaining the decision-makers' confidence in the outcome of the simulation study.

3.5. Output Analysis

After V&V the model was run and the output data analyzed. A bottleneck analysis was made in order to determine possible causes for the existing low level of productivity.

Analyzing the output given by the simulation model it was possible to extract the following results (statistical estimates of the performance measures based on 95% confidence intervals):

- The total number of pieces created by the high-pressure machines was 2532.
- The average throughput of the finishing section was 2027 pieces.
- Work in progress for "pieces for finishing" was, in average, 26,7 ± 1,15 and "pieces for transporter" was of 145,48 ± 2,42.
- Workstation 2 is the one which retains the pieces more time, this value is in average 55.9 ± 0.6 seconds and workstation 3 is the one which retains the pieces less time (25.8 ± 1.2 seconds).
- The waiting time since the components are created in Netschz until they start to being processed is in average 20,2 ± 1.05 minutes, and in Sama it is 31,3 ± 0,55 minutes, which causes a high number of components waiting to be processed after the high-pressure machines.
- The operator 2 is being in utilization an average of 98,6% ± 0.01. Operator 3 is the one who is less occupied being in utilization an average of 55,6% ± 0.01.

Observing this information, it was possible to conclude that the number of finished components represents approximately 80% of productivity, which is very similar to the data observed on the shop-floor. Operator 2 is the bottleneck , being in use most of the time, while operators 1 and 3 have some idle time.

A large number of pieces is being produced by the high-pressure machines but the actual organization of the finishing section does not have capacity to finish all the production. This bottleneck creates a considerable number of intermediate stock and work in progress. These factors have a negative effect on the quality level of the final product.

3.6. New scenario for the finishing section

Given the attained simulation results, the objective was then to redesign the finishing section in order to achieve a growth of 15% in productivity.

Some changes were made in the layout of the section to provide a better flow of materials and to concentrate the waste in one operator, the "logistics operator". The function of this operator is to do all the operations (e.g., transportation) that do not create value to the final consumer but are necessary in the process. If he has some idle time, he will help the other operators to finish some pieces.

Figure 6 depicts the proposed solution for the new layout of the finishing section. The idea is to have a cellular layout with operators that are concentrated exclusively in creating value.

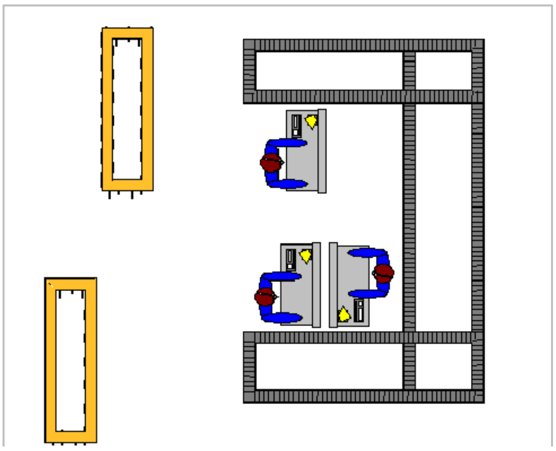


Figure 6: New layout for the finishing section

For this simulation it were considered the same conditions of the previous model for the creation of pieces. This time is deterministic so, maintaining the same replication parameters it is expected to obtain an equal number of created pieces. In the processing time of the pieces by operators 1 and 2 it were excluded movement and transportation times, because these operators are now focused only in creating value.

The results for this new simulation scenario were the following:

- The total number of pieces created by the high-pressure machines was 2532.
- The average throughput of the finishing section was 2403 pieces.
- The work in progress for Netschz pieces was in average 40,9 ± 1,15 pieces and for Sama was 40.9 ± 1,14 pieces;
- Value added time for Netschz pieces was in average 0.64 minutes and for Sama 0.63 ± 0.01 minutes.
- The waiting time since the components are created in Netschz until they start to being

processed is in average 13,9 ± 0.42 minutes, and in Sama it is 13,8 ± 0,41 minutes.
- All resources are being used approximately in average 97% of the time.

4. DISCUSSION

Analyzing the values obtained in the different simulations we can conclude that with the introduction of relatively small lean-based modifications the system-in-analysis was able to finish more 16% components that those which are finished today (the objective was 15% of improvement).

The proposed cellular layout provided a continuous flow, which reduced substantially the work in progress, and the time components were waiting to be processed. All the operators are occupied almost 100% (Figure 7 and Figure 8) of the time, and operators 1 and 2 are dedicated to operations that create value, while operator 3 has concentrated all the *muda* operations that are required but do not create value (supply, per example).

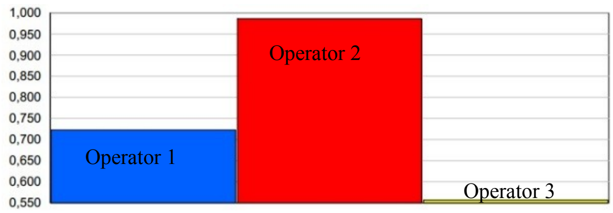


Figure 7: Operators' utilization before modifications

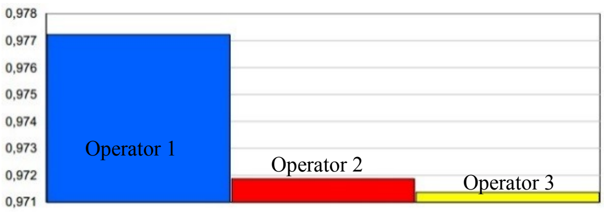


Figure 8: Operators' utilization after modifications

As one can see, the operation of the finishing section for the new scenario is smoother, that is, the workload is now more evenly distributed, the WIP is considerably lower, and the productivity is increased by 16%, as desired.

Other scenarios are being studied by the company such as the option of including one more operator in the section and the opportunity to increase the output rate of the high-pressure machines.

5. CONCLUSION

The development of successful projects involving both the universities and the industry is, generally, difficult to undertake. In the project presented in this paper this difficulty in communication was overcome, due to the fact that one of the university team members worked fulltime within the company throughout the duration of the project. He not only established a privileged communication channel between the university and the company, but also directed management and staff attention to the project.

During the development of the simulation study, formulating the problem and gathering data were the critical steps (the most complex and most time expensive).

Simulation studies can become a powerful tool for analyzing the actual state of a factory or section, and for analyzing possible modifications to the actual state, using DoE and comparing the outputs. This is especially efficient if all intervenient are focused on the goal.

By testing new scenarios in a simulation environment it is possible to save money spend in disrupting systems or implementing "poor" operational solutions. Using simulation techniques involves considerable costs of software and training but, the benefits of using this tool to dynamic evaluate complex systems are unique and can faster outweigh the initial investment.

The company's goals were fully attained and the suggested modifications to its manufacturing operations are being implemented, as a result of the outcome of the simulation study.

This successful case study of university/industry interaction in the simulation field can be used as a showcase to the benefits that SME's can get from the use of simulation.

ACKNOWLEDGMENTS

The authors would like to thank to Vista Alegre Atlantis, SA management and staff, and in particular to Eng. Joana Ribeiro, for allowing us free access to all the information used, for their contribution to the project and for their willingness to allow us to share this experience.

This work is funded by FEDER funds through COMPETE (Operational Programme Factors of Competitiveness) and by National Funds through FCT (Foundation for Science and Technology) in the context of the project PEst-C/CJP/UI4058/2011.

REFERENCES

Adams, M., Componation, P., Czarnecki, H. and Schroer, B., 1999. Simulation as a tool for continuous process improvement, in *Proceedings of the 1999 Winter Simulation Conference*, 766 - 773.

Bruzzone, A., Madeo, F., Dallorto, M., Poggi, D., and Ferrando, A., 2013. Different modeling and simulation approaches applied to industrial process plants, *Simulation Series*, 45(5), 39 - 46.

Grahl, C., 2003. Going Lean. *Ceramic Industry*, 153(6), 6.

Hines, P., Holweg, M. and Rich, N., 2004. Learning to evolve: A review of contemporary lean thinking, *International Journal of Operations & Production Management*, 24(10), 994 – 1011.

Howell, V., 2010. Lean Manufacturing. *Ceramic Industry*, 160(9), 16 - 19.

Howell, V., 2011. Kaizen Events. *Ceramic Industry*, 161(12), 30 - 32.

Kelton, W., Sadowski, R. and Swets, N., 2010. *Simulation with Arena*, 5th ed., McGraw-Hill Higher Education, New York.

Kim, C., and Mauborgne, R., 2005. *Blue Ocean Strategy: How to Create Uncontested Market Space and Make the Competition Irrelevant*. Cambridge, MA: Harvard Business School Press.

Longo, F., 2011. Supply Chain Management Based on Modeling & Simulation: State of the Art and Application Examples in Inventory and Warehouse Management, in *Supply Chain Management*, InTech.

Melton, T., 2005. The benefits of lean manufacturing: What lean thinking has to offer the process industries. *Chemical Engineering Research and Design*, 83(6 A), 662 - 673.

AUTHORS BIOGRAPHY

TIAGO SILVA was born in Esmoriz, Portugal and went to Aveiro University, where he studies Industrial Engineering and Management. Currently he is doing an internship at Vista Alegre Atlantis, SA, in the ceramics sector. His e-mail address is `tgs@ua.pt`.

ANA LUÍSA RAMOS was born in Portugal in 1974. She received the M.Sc. degree in computers engineering from the University of Coimbra, Coimbra, Portugal, in 2002 and the Ph.D. degree in industrial management from the University of Aveiro, Aveiro, Portugal, in 2011. She is an Assistant Professor of Management and Industrial Engineering with the University of Aveiro, where she is member of the Department of Economics, Management and Industrial Engineering Executive Board and Council. Her research work focuses on modelling and simulation, and model-based systems engineering, mainly in the industrial and transportation sectors.

JOSÉ VASCONCELOS FERREIRA was born in Portugal in 1960. He received the M.Sc. degree in operational research and systems engineering from the Technical University of Lisboa, Lisboa, Portugal, in 1989, and the Ph.D. degree in engineering sciences from the University of Porto, Porto, Portugal, in 2005. He is an Assistant Professor of Management and Industrial Engineering (MIE) with the University of Aveiro, since 2005, where he is Director of the MIE doctoral program, and Member of the Department of Economics, Management and Industrial Engineering Council. He has more than twenty years of experience collaborating with the main urban mass transit companies in Portugal. His research work focuses on systems engineering, logistics, multivariate data analysis, and operational planning at mass transit companies.

SUPERVISED GROWTH NETWORK SGWRN MODEL CAN ACHIEVE A BALANCED CONTROL OF TWO-WHEELED ROBOT

Fu-jin LI, Mei-jie HUO, Hong-ge REN ,Dong-mei FEI

College of Electrical Engineering, He bei United University, He Bei Tangshang 063000,China

Email:1042623870@qq.com

ABSTRACT

Based on the unsupervised growth structure of the network GWRN model, GWRN model and radial basis function network architecture combined. Proposed supervised structure of the network can be grown SGWRN models. Model used is the growth of self-organized growth algorithm when needed, by inserting new neurons, changing the connection among competition layers and adjusting the connection weights between layers, achieve the required accuracy of the model output, simulation experiments show that, SGWRN model realizes the two balancing robot control, there are certain anti-interference ability and practicality.

Keywords: SGWRN model; Growth algorithm; Connection weights; Robot

1. INTRODUCTION

Self-organizing neural networks generally use unsupervised learning, each neuron model via inhibition of brain mechanism simulated human side since the distribution of the input mode of organizational learning , but also can learn the topology of the input vector. The development of the network can be grown based on the in self-organizing feature map network , the network according to their different requirements to adjust the size and structure of the network.. However, for a particular cyberspace mapping space, expected to produce a specific response to the unsupervised self-organizing neural network can be grown on some limitations, so need to establish a supervised network structure can be grown. This paper presents a supervised self-organizing network model can be grown, will be based on self-organizing network and RBF[1] network structure of the organic combination of growth, namely SGWRN model[2]. This model uses a network GWRN needed growth of self-organizing algorithm, hidden layer neuron growth and learning process involves the insertion of neurons, changing the connection between competition layer neurons and adjusting the connection weights among layers, model in the process of growth of self-learning, until the model output accuracy. SGWRN model is applied to balance control of two-wheeled balancing robot. The initial number of model neurons is less, hrough continuous learning, competitive layer neuron number is increasing gradually grown to be able to grasp the model of the robot balance control[3].

2. SUPERVISED CAN GROW SRLF-ORGANIZING NETWORK MODEL

Self-organizing neural network is usually unsupervised learning, generally they will enter the mapping data in an orderly manner to the low-dimensional topology, and could often be used to input data into subsets, so that the interior of a subset of data items is similar, and different subsets of data items are not similar. Based on the supervision of neuron growth structure, the unsupervised learning GWRN model extended to supervised learning can growth structure of neural network model, namely the proposed have supervision can grow SGWRN self-organizing network model.

Supervised growing structure network (SGWRN) model combined unsupervised growing network GWRN model with radial basis function (RBF) [4] . For the SGWRN model can learn under the guidance of the supervised signal, so the growth of network structure of GWRN model only two layer to three layer, its structure as shown in figure 1. The basic network structure by the Input Layer (Input Layer, IL), Competitive Layer (Competitive Layer, CL) and Output Layer, the Output Layer, OL) constitute the three Layer. Among them, the input layer and competition constitute GWRN model. Connection weight matrix between the input layer and competitive layer is W and Connection weight matrix between the competitive layer and output layer is V . Which the dimension of the input data and the number of neurons in input layer are the same, the dimensions of the output is the same as the number of neurons in the output layer, According to input data to adjust the number of neurons of the competitive layer, timely and growth, connection weights between the network layers will be adjusted in accordance with guidance of the supervised signal[5].

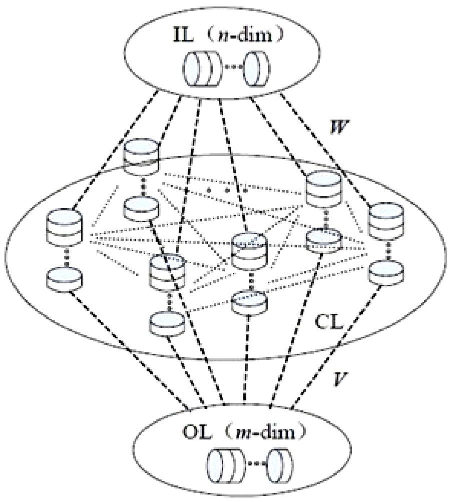


Fig1 SGWRN structure

SGWRN model algorithm consists of two parts, namely the network competitive layer neuron growth algorithm and the learning process of network under the supervised signal.

the growth and learning process of SGWRN model include that insert new neurons in the competitive layer, change the connections between neurons of the competitive layer and adjust the connection weight matrix W and V of the layer and layer, Network model learning in the process of adding neurons, until the output meet the precision requirement of the model[6].

First of all, the initialization supervised growing structure network. The initial competitive layer with 2 neurons, ie $K=\{n_1,n_2\}$ (K represents a collection of neurons mapping space); mapping space neuronal connections set C; using RBF as activation function of the neuron in the competitive layer, the threshold value b_1; number of neurons in the input layer and output layer are respectively n and m.

In best match neurons (BMU) s and the best match between neurons (SBMU) ss insert new neurons process is as follows:

1)Insert a new neurons r， $K=K\cup\{r\}$

2) to initialize the connection weights between new neurons r and the input stimulus ξ

$$\mathrm{w}_\mathrm{r}=(w_r+\xi)/2 \tag{7}$$

3) remove the connection between s and ss

$$C=C/\{(s,ss)\} \tag{8}$$

4) Create a new connection between r and s, r and ss

$$C=C\cup\{(r,s),(r,ss)\} \tag{9}$$

If the sum of squared errors e can not reach the precision requirements, perform the following growing algorithm;

① Select the input / output data pairs（ξ,ξ）, calculate the Euclidean distance $\|\xi-w_i\|$ between competition layer neurons i and input vectors ξ.

② Select the best match of neurons and sub-optimal match to create a connection. Among them s and ss meet:

$$\|\xi-w_s\|=\min_{n\in A}\|\xi-w_n\| \tag{10}$$

$$\|\xi-w_{ss}\|=\min_{n\in A}\|\xi-w_n\| \tag{11}$$

③According to the following formula to adjust the best matching neurons s and the weights of the field neurons i.

$$\Delta w_s=\varepsilon_b\times h_s\times(\xi-w_s) \tag{12}$$

$$\Delta w_i=\varepsilon_n\times h_i\times(\xi-w_n) \tag{13}$$

Among them $0<\varepsilon_n<\varepsilon_b<1$， h_s and h_i meet:

$$h_s(t)=1-\frac{1}{\alpha_b}(1-e^{-\alpha_b t/\tau_b}) \tag{14}$$

$$h_t(t)=1-\frac{1}{\alpha_n}(1-e^{-\alpha_n t/\tau_n}) \tag{15}$$

④If the best matching neuron activation value a less than the activation threshold a_T, and the start count is less than the threshold h_T ,the network is growing(that is inserted new neurons r between s and ss).

Among them, $a=\exp(\|\xi-w_s\|)$；

⑤If there is available training samples, return①.

⑥ Numerical calculation competitive layer activation function of neurons i

$$d_i=\exp(-(\|w_i-\xi\|\cdot b_1)^2/2) \tag{16}$$

⑦ By (d_i,ζ) adjust v, and threshold of output layer neurons b_2

⑧ Calculate the output of the output layer

$$output=D\cdot V+b_2 \tag{17}$$

Calculate the error sum of squares e

⑨ If e arrived in accuracy, exit; Otherwise, insert new neurons, the network continue to grow.

3. SGWRN MODEL IS APPLIED TO TWO BALANCING ROBOT CONTROL

Supervised can grow self-organizing neural network model can be used in supervised learning. For balance control of two-wheeled balancing robot[7] [8], Corresponding to different states of the different time robots, If it can get the corresponding control quantity as monitoring signal, You can have a supervision can be growth dynamic structure model is applied to two wheel robot balance control. Based on supervised can grow structure of the neural network two rounds of balancing robot learning control system is shown in figure 2

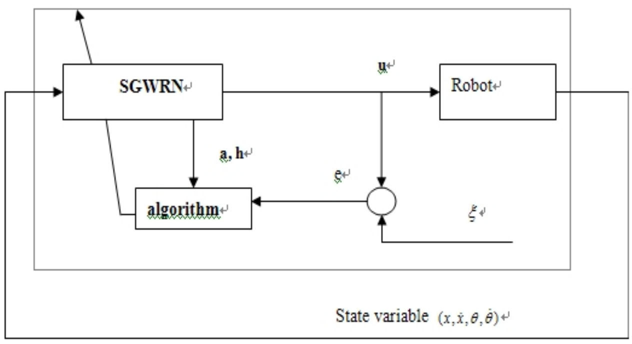


Figure 2 two wheel balance robot control model structure diagram

The dashed box is based on the growth structure of the neural network controller. The dynamic structure model for offline training, after training for two rounds of balancing robot for on-line control. Learning algorithm to obtain the activation values a of neurons from SGWRN and error sum of squares e. Dynamic structure model of the competitive layer consists of two neurons not connected at first, under the guidance of the learning algorithm, Dynamic structure model began to grow, continue to control the robot to learn the rules, until it meets the accuracy requirements. After training SGWRN as control output to control the robot[9] [10].

Here, robotic system for two rounds of balance, the state of the system is 4-dimensional vector, control of the robot's forces or its acceleration. Therefore, network model is established, the number of neurons in the input layer IL is 4, number of neurons in the output layer OL is 1, the initial number of neurons in the competitive layer is 2. Dynamic structure in the process of learning, the number of CL neurons incrementing, gradually grow into a master model robot balance control skills.

SGWRN algorithm flow:

1) Initialization: $A=\{n_1,n_2\}$, $C=\Phi$

2) CL neuron activation using RBF function, Thresholds b_1,

3) M linear output neurons

4) when the error sum of squares does not meet the accuracy requirement, Iteration λ times

① Select from a training set of input/output data (ξ,ς)

② Determine the BMU and SBMU, and establish a new connection C

③ By the formula, adjusting weight matrix W

④ When the network does not match the input stimuli, network growth and by the formula (7) calculated new neurons in the initial weights Calculated by the formula (7) new neurons in the initial weights, then equation (8), (9) to update the connection between neurons, the executive v; otherwise, direct execution

⑤ Calculate activation function value of CL neurons

$$d_i = \exp(-(\| w_i - \xi \| \bullet b_1)^2 / 2) \tag{18}$$

⑥ By (d_i,ς) adjust V, and OL neuron threshold value 2

⑦ Calculate OL output

$$output = D \bullet V + b_2 \tag{19}$$

⑧If e meet the accuracy requirements, exit the loop, otherwise, add a new neurons, the network continues to grow.

4. SIMULATION EXPERIMENT

SGWRN model can be applied to solve the problem of balance control, SGWRN in to a certain degree of offline learning control law, Control can be performed online.

In two rounds of the balancing of the robot control as an example, design a dynamic structure model learning model, and apply the model to control.

Sample vector is expressed as $(x,\dot{x},\theta,\dot{\theta},\varsigma) = (0,0,5pi/180,0)$,randomly selected 50 samples for study. Set parameters of SGWRN were $\varepsilon_b = 0.6$, $\varepsilon_n = 0.25$, $\alpha_b = 0.8$, $\tau_b = 5$, $\alpha_n = 1.3$, $\tau_n = 13$, $h_T = 0.38$, $\alpha_T = 0.6$. The sampling period is $T = 0.01$,requires e less than $1.25e-12$,Training CL neurons number is 60 SGWRN model are obtained.

The initial set up tilt angle is 5 degrees, the initial position is zero, with training of supervision to control the robot dynamic structure model, and comparing with no supervision and control the robot dynamic structure model. In the process of the control ,acting on the robot control, no supervision and dynamic structure model output as shown in figure 3, supervised the dynamic structure model output as shown in figure 4, the horizontal axis shows the sampling points.

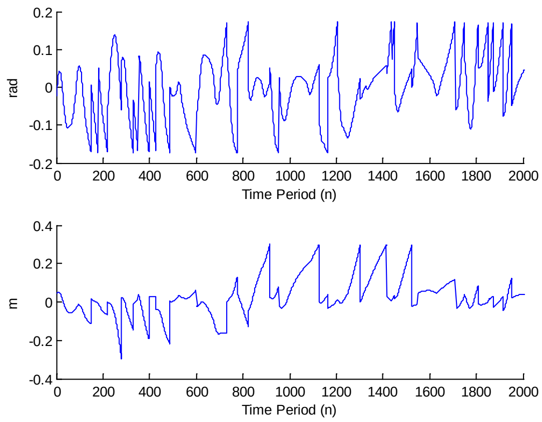


Figure 3 GWRN model control the robot state

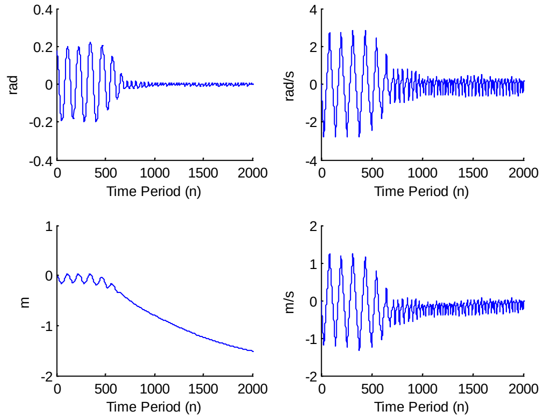


Figure 4 SGWRN model of state control of the robot

Corresponding to each time the state of the robot (the location of the car and tilt Angle) as shown in figure 4, the robot's state variables, fast convergence SGWRN model control effect is good

Through the simulation experiment shows that, Figure 3 unsupervised learning network model can't very good control Angle and displacement of the robot, The figure 4 state variables of the dynamic structure model can in good control. Two figure in comparison, It is concluded that the proposed dynamic structure model has good practicability, shorten the time to learn, verify the validity of the model.

5. CONCLUSION

This paper presents a supervised structure of the network can be grown (SGWRN) model, the model combined no supervision can grow network with RBF function, to achieve balance control two-wheeled robot. Growth dynamic structure in monitoring network signal rapidly under the guidance of learning, increasing competition layer neurons, then the network model in offline training, after training is good for two rounds of robot via online balance control. The network model effectively solves the mapping space for a particular network, expected to produce a specific response. The Simulation results show that, a dynamic structure model with no supervision structure model is compared, Supervised more dynamic structure model can be finished in a short time the balance control of the robot, and when two rounds of the robot parameters change too much , still can well control the balance of the two rounds of the robot. Shows the monitoring dynamic structure can better achieve the equilibrium of two rounds of balance control of the robot, practical.

REFERENCES

[1] Zhou Junwu Sun Chuanyao etc. Radial basis function (RBF) network. *Research and realization of the mining and metallurgy*. 2001, 12 (4) : 71-75

[2] Xie Yanhui. Can grow structure self-organizing network research and its application in the inverted pendulum control [D]. *Beijing: Beijing university of technology electronic information and control engineering college*, 2007

[3] Qiao Junfei,Zhang Ying. Fast pruning algorithm for multilayer feedforward neural network [J]. *Journal of Intelligent Systems*，2008,3(2):173-176.

[4] MOODY J. Prediction risk and architecture selection for neural networks[C]//*Statistics to Neural Networks: Theory and Pattern Recognition Applications,NATO ASI Series F. New York*,1994:178-197.

[5] S.E.Fahlman,C.Lebiere.The cascade-correlation learning architecture[J]. A*dvances in Neural Information Processing Systems*, 1990(2):524-532.

[6] Sevki S Erdogan,G S Ng and K H Chan.Measurement Criteria for Neural Network Pruning[C].*//IEEE Region 10 Annual Conference*,1996: 83-89.

[7] Ehud D Karninl. A simple procedure for pruning back-propagation trained neural network [J]. *IEEE Trans on Neural Networks* 1990,1(2):239-242.

[8] Loh A P, Hang C C, Quek C K, et al. Autotuning of multiloop proportional-integral controllers using relay feedback[J]. *Industrial and Engineering Chemistry Research*, 1993, 32(6): 1102-1107.

[9] Shih H S, Cheng C Y. Use of relay feedback test for automatic tuning of multivariable systems[J]. *AIChE J*,1994, 40(4): 627-646.

[10] Palmor Z J, Halevi Y, Krasney N. Automatic tuning of decentralized PID controllers for TITO processes[J].*Automatica*, 1995, 31(7): 1001-1010

MODELING AND SIMULATION FOR THE PERFORMANCE EVALUATION OF THE ON-BOARD COMMUNICATION SYSTEM OF A METRO TRAIN

Alberto Falcone, Alfredo Garro, Andrea Tundis

Department of Informatics, Modeling, Electronics, and Systems Engineering (DIMES), University of Calabria, via P. Bucci 41C, 87036, Rende (CS), Italy

{alberto.falcone, alfredo.garro, andrea.tundis}@dimes.unical.it

ABSTRACT

The paper presents the evaluation of the dependability performances of a real On-Board Communication System of a Metro train centered on the application of RAMSAS, a recently but promising model-based method for the reliability analysis of systems through Simulation. In particular, after the description of the On-Board Communication System under consideration, of its dependability requirements, and related performance indicators to be evaluated, a SysML-based model of the structure and behavior of the system is presented. Beside the nominal system behavior, specific dysfunctional tasks, able to alter the intended behavior of the system, are introduced in order not only to evaluate through Simulation system dependability performances but also to compare different design choices and parameters settings against the requirements.

Keywords: Model-Based Systems Engineering, Safety, Availability, Reliability, Performance Evaluation, Urban Rail Transport

1. INTRODUCTION

Non-functional requirements analysis and related system performance evaluation are challenging tasks that involve several disciplines ranging from Modeling and Simulation to Systems Engineering. These tasks rely on the modeling of system properties that deals with formally expressing constraints and both functional and non-functional requirements so to enable their verification through real or simulated experiments and/or analytical techniques.

Among non-functional requirements, the *dependability* ones (such as *reliability*, *availability*, *maintainability* and *safety)*, which represent important properties to be satisfied for a wide range of systems (Guillerm, Demmou, and Sadou 2010; Laprie 1992; Stapelberg 2008), become really crucial in *mission-critical* industrial domains, such as nuclear plants, avionics, automotive and satellite (Lahtinen, Johansson, Ranta, Harju, and Nevalainen 2010; Rierson 2013; Navinkumar and Archana 2011; Garro, Groß, Riestenpatt Gen. Richter, and Tundis 2013).

As a consequence, international organizations, research centers, and companies are strongly involved in investigation and standardization activities focused on dependability aspects; for instance (i) IEC-61508, provided by IEC (International Electrotechnical Commission), deals with aspects of Electrical, Electronical and Programmable Electronical Systems (IEC-61508 2010) as well as ISO-26262 which represents the reference standard in automotive domain (ISO-26262 2011); (ii) RTCA - DO 254, by the RTCA Special Committee, provides guidance for design assurance of airborne electronic hardware (RTCA/DO 254 2000); (iii) ECSS-Q80-03 is a standard defined by ESA (European Space Agency) concerning methods and techniques to support the assessment of software dependability and safety (ECSS-Q80-03 2006); (iv) NASA/SP-2007-6105 provides top-level guidelines and best practices as well as crosscutting management processes in systems engineering (NASA 2007). These efforts testify the need of models and methods for representing system requirements and constraints able to support their validation, traceability and verification (Krause, Hintze, Magnus, and Diedrich 2012; Peraldi-Frati and Albinet 2010; Tundis, Rogovchenko-Buffoni, Fritzson, and Garro 2013; Yu, Xu, and Du 2009).

Recently, great attention is devoted towards the railway domain and, particularly, on its safety and reliability. Indeed, human errors, as well as deliberate sabotage, pose a considerable danger to passengers travelling on the modern railways and have disastrous consequences. To protect civilians against both intentional and unintentional threats, rail transportation has become increasingly automated and performance studies are fundamental to increase the lifetime of railway systems (Flammini 2012). One of the main goals of this analysis is to verify whether system working conditions are reliable and safe to reduce dangerous situations or even losses of human lives. This task not only takes into account the analysis of the whole traction chain, but also requires ensuring that the railway infrastructure is properly working. As a consequence, several tests for detecting any dysfunctional behavior need to be carried out (Scott, Dadashi, Wilson, and Mills 2013; Reliability and Safety in Railway 2012).

Unfortunately, even though the modeling and simulation of functional requirements are well supported by several tools and techniques, a lack of methods, models and practices specifically conceived to deal with non-functional requirements and, in particular, with the dependable ones, are still missed; as a consequence, the evaluation of system performances is often delayed to the late stages of the development process with the high risk of having to revise already implemented design choices, and, consequently, to miss project deadlines and budget.

To contribute to fill this lack, the paper exemplifies a comprehensive approach for supporting the evaluation of dependability performances centered on Simulation by taking as the reference system the On-Board Communication System (OBCS) supplied by SELEX (SELEX ES) installed on the Line 5 of the Milan Metro train. Specifically, the experimentation is performed through RAMSAS (Garro and Tundis 2014), a recently proposed model-based method for the reliability analysis of systems through Simulation.

In particular, Section 2 briefly introduces the RAMSAS method, then the description of the On Board Communication System under consideration is provided in Section 3; in Section 4 the dependability requirements and related performance indicators to be evaluated are described. Then the structure and behavior (both nominal and dysfunctional) of the OBCS is modeled in Section 5; whereas, in Section 6, the evaluation of its dependability performances by adopting simulation techniques is presented. Finally, conclusions are drawn and future work delineated.

2. RAMSAS: A MODEL-BASED METHOD FOR DEPENDABILITY ANALYSIS THROUGH SIMULATION

The evaluation of performances of the On-Board Communication System under consideration is performed through RAMSAS (Garro and Tundis 2014), a model-based and simulation-driven method which consists of four main phases: *Reliability Requirements Analysis*, *System Modeling, System Simulation,* and *Results Assessment*. Specifically, in the Reliability Requirements Analysis phase the objectives of the reliability analysis are specified and the reliability functions and indicators to evaluate during the simulation are defined. In the System Modeling phase, the structure and behavior of the system are modeled in SysML (System Modeling Language), the OMG Systems Modeling Language, by using zooming in-out mechanisms (Molesini, Omicini, Ricci, and Denti 2005); moreover, beside the intended system behaviors, specific dysfunctional behaviors and related tasks, which model the onset, propagation and management of faults and failures, are introduced. In the System Simulation phase, the previously obtained models of the system are represented in terms of the constructs offered by the target simulation platform, then simulations are executed so to evaluate the reliability performance of the system also on the basis of different operating conditions, failure modes and design choices. Finally, simulation results are analyzed with respect to the objectives of the reliability analysis; if necessary, new partial or complete process iterations are executed. With reference to a typical V-Model process, RAMSAS can be used: (i) in the testing phases to support the evaluation of unit and system reliability performances; (ii) in the design phases to support the validation and evaluation through simulation of configuration scenarios and setting of system parameters so to evaluate, compare and suggest different design choices and improve the descriptive and predictive capabilities of the reliability system model. RAMSAS has been already experimented in the satellite domain for the reliability analysis of an Attitude Determination and Control System (ADCS) (Garro, Groß, Riestenpatt Gen. Richter, and Tundis 2013), in the avionics domain for the reliability analysis both of a Landing Gear System (Garro, Tundis, and Chirillo 2011) and of a Flight Management System (Garro and Tundis 2012b); and in the automotive domain for the reliability analysis of an Electronic Stability Control (ESC) system (Garro and Tundis 2012a). It combines in a unified framework OMG modeling languages (System Modeling Language) and the popular Mathworks simulation and analysis environments (MATLAB-Simulink).

3. THE ON-BOARD COMMUNICATION SYSTEM

The considered On-Board Communication System (OBCS), supplied by SELEX (SELEX ES) and installed on the Line 5 of the Milan Metro train, is composed by a set of devices required to perform safety tasks and functions as well as the dissemination of information to passengers such as: bidirectional audio communication between the Central Station Operator and Passengers when a situation of emergency occurs, communication between two Central Station Operators, data exchanging for diagnosis among equipment, on board video monitoring, and sending of live/recorded messages to passengers. The general architecture of the OBCS is centered on a Control Unit (CU) subsystem able to manage all system devices and to select the necessary equipment to perform the required functionalities and tasks, as well as to handle, if necessary, voice and data communication by combining Tetra Radio (Terrestrial Trunked Radio) and Wi-Fi components. Figure 1 shows the connections between the CU and the other main communication devices.

Specifically, the components of the considered OBCS are: (i) two Control Unit, (ii) a keyboard for each CU employed in emergency situations by the Central Station Operator on the train, (iii) a Tetra Radio component for each CU for data and audio communication among the Operation Center and the Central Station Operator/passengers on board, and (iv) other components for supporting communication (e.g. Emergency Call Point/ECP, Wi-Fi, Ethernet Switch, Emergency Buttons, Microphones, Speakers, and Amplifiers for Environmental Audio Diffusion).

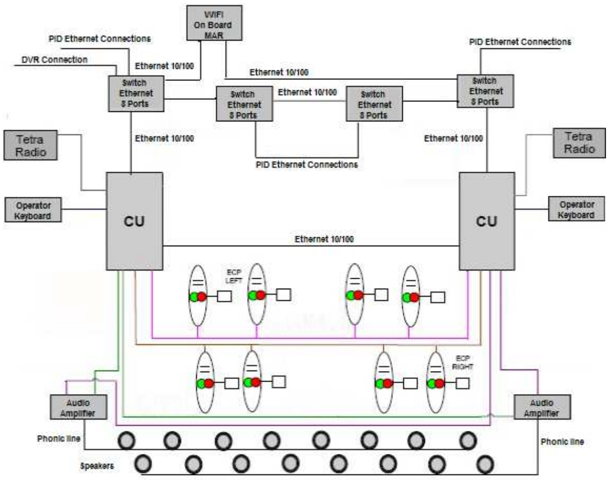


Figure 1: The architecture of the reference On-Board Communication System

4. DEPENDABILITY REQUIREMENTS ANALYSIS

The On-Board Communication System of a Metro train has to fulfill functional and non-functional requirements such as the dependable ones in order to provide a safe mode of transport so as to avoid the occurrence of hazards and to prevent accidents. In general, a system or a component is reliable if it has the ability to perform its required functions under stated conditions for a specified period of time. A more formal definition is based on the concept of MTBF (Mean Time Between Failure) which is defined as the average time that elapses between two successive failures. In this case, the analysis of performance of the system is conducted on the basis of failure events and their effect on the operation of the system. In particular, the following specific fault situations and related performances have been considered:

- Total Block (TB): a block of the system operation for a time more than 3 minutes;
- Partial Block (PB): a block of the overall system operation from 3 minutes up to 10 minutes;
- Delay: a delay of the overall system operation for a time less than 3 minutes.

As it is shown in Table 1, in order to study the behavior of the system subject to the above mentioned type of failures, the following indices are related values have been considered.

Concerning the failures of type TB, the key requirement to be considered is represented by the *Unavailability* calculated as 1-*Availability* which is defined as the ability of an item (e.g. system, subsystem or a component) to perform a required function at a given instant of time or at any required instant of time within a given time interval. In particular, *Availability* is determined according to the equation (1), where the *MPS* represents the minutes of performed service and the *MSS* represents the minutes of scheduled service.

$$Availability = \frac{MSP}{MSS} \tag{1}$$

To study the behavior of the system subject to failures of type PB and Delay, the key requirements to be considered are instead represented by (i) the *Failure Rate* that represents the frequency with which an item (e.g. system, subsystem or a component) fails, expressed, for example, in failures per hour, (ii) the *DownTime_evMAX* that represents the actual duration of the outage resulting from a failure.

Table 1: On-Board Radio System Dependability Requirements

Adverse Event	*Unavailability*	*Failure Rate* (λ) *[hour^{-1}]*	*DownTime_evMAX [hour]*
Total Block	1-0,9999 = 0,0001	-	-
Partial Block	-	2,31E-06	0,16 (~10 min.)
Delay	-	4,33E-06	0,05 (~3 min.)

5. SYSTEM MODELING

This Section describes both the structure and behavior of the OBCS under consideration by exploiting OMG SysML (Systems Modeling Language) as well as zooming in-out mechanisms; moreover, beside the intended system behaviors, specific dysfunctional behaviors and related tasks, which model the onset, propagation and management of faults are introduced in order to fully enable the analysis of the dependability performances through Simulation.

5.1. System Structure Modeling

In this activity a complete (or partial) representation of the structure of the system (or of its parts), that is under analysis, has to be provided. This representation allows for a layered view of the system that is useful to figure out the components involved at some specific layer and the relationships among them. The system structure is modeled by using SysML *Blocks* following a *top-down* approach so as to obtain a hierarchical decomposition of the system (e.g. system, subsystems, equipment, and components). Specifically, each system entity is represented by a SysML *Block* and modeled by both *Block Definition Diagrams* (BDDs) and *Internal Block Diagrams* (IBDs).

As an example, Figure 2 and Figure 3 show respectively the Block Definition Diagram (BDD) and Internal Block Diagram (IBD) relating to the On-Board Communication System, whose components have been already introduced in Section 3 (see Figure 1).

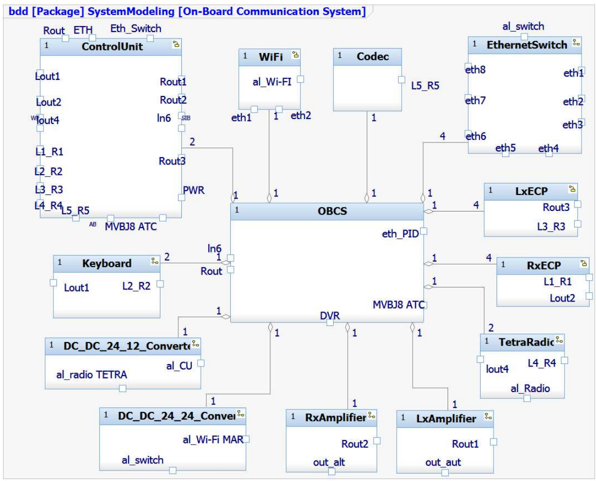


Figure 2: Block Definition Diagram of the On-Board Communication System

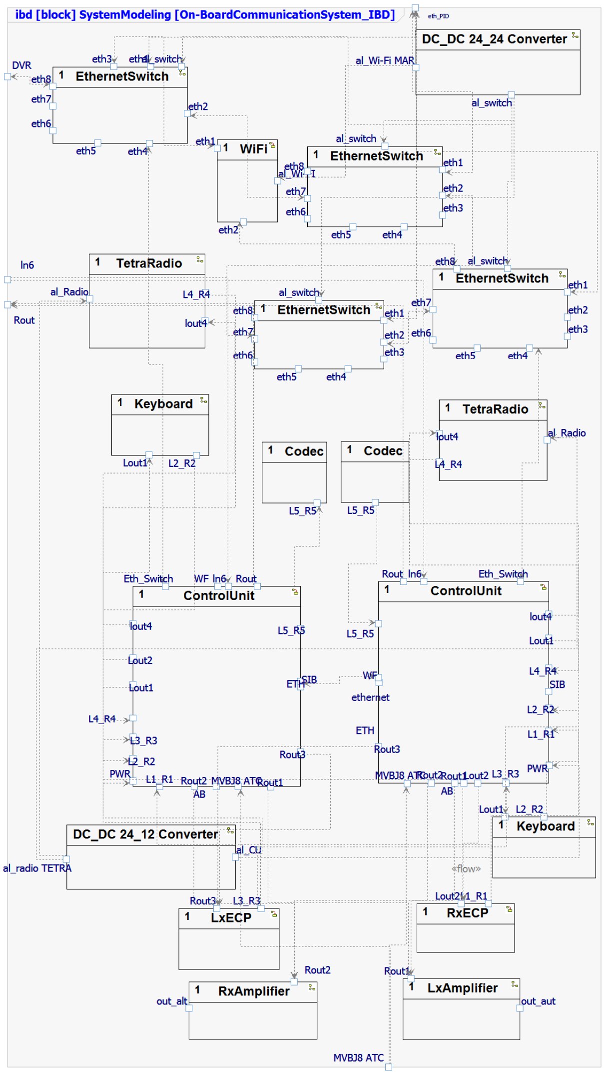


Figure 3: Internal Block Diagram of the On-Board Communication System

By applying *zooming-in* mechanisms (e.g. by breaking down the system) further components can be identified so as to reach a deeper level of decomposition. In the following, the diagrams related to the *ControlUnit* (CU) subsystem are reported; in particular, in Figure 4 and in Figure 5 its BDD and IBD are represented respectively.

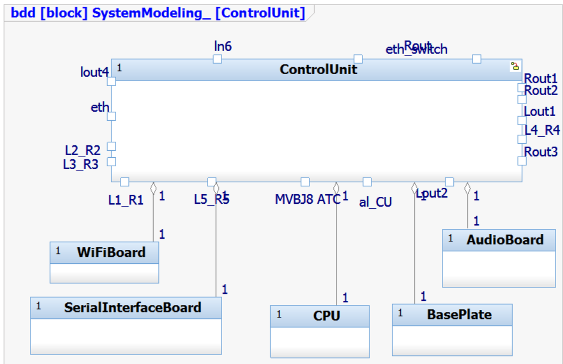


Figure 4: Block Definition Diagram of a Control Unit

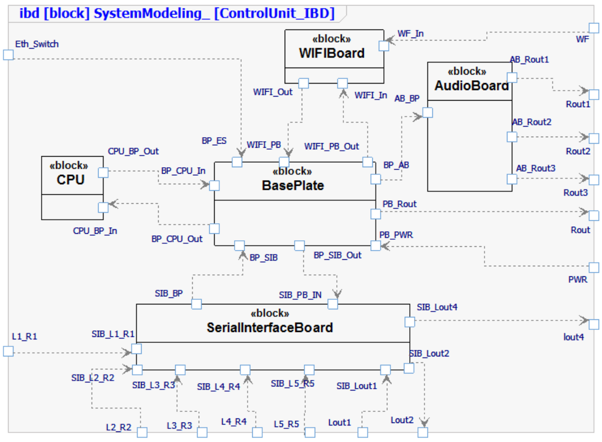


Figure 5: Internal Block Diagram of a Control Unit

5.2. Intended Behavior Modeling

The modeling of the *Intended Behavior* takes into account the hierarchical structure of the system and specifies the intended behavior by following a *bottom-up* approach. Specifically, the behavior of the system entities at the lowest level in the hierarchy, called *leaf level* (e.g. component level), are first specified; then, the behavior of the entities at higher levels of abstraction, called *non-leaf levels* (e.g. equipment and subsystem level), are modeled by specifying how the enclosed entities participate and determine the behavior of each considered enclosing entity.

Different kind of SysML diagrams can be exploited to model the behavior of a given entity: *Activity*, *Sequence*, *Parametric*, and *Statechart Diagrams* according to the characteristics of the behavior and the abstraction level to represent.

In particular, Figure 6 shows the Intended Behavior of a *ControlUnit*, by using a *Statechart,* which is able to handle simultaneously several input signals. Indeed, it can receive and then manage different kinds of signals such as *Passenger Call*, *ForwardMessage, CentralStationOperator-Passenger Call, OnBoardOperator Call, CentralStationOperator-OnBoardOperator Call.* Specifically, when the *CU* is

Active, all its parallel sub-states are in *InWaiting* (see Figure 6).

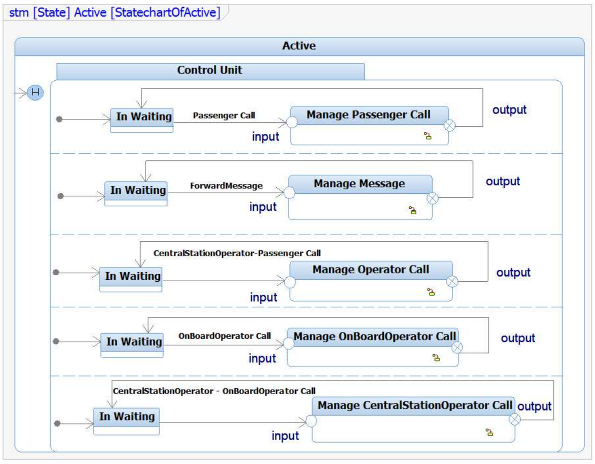


Figure 6: Intended Behavior of a Control Unit

As an example, when a *PassengerCall* is incoming then the behavior of the CU changes its state from *InWaiting* to the *MenagePassengerCall* state that takes in charge such input signal and handles it opportunely by producing a proper output signal.

Figure 7 shows the internal statechart of the *ManagePassengerCall* which aims to initiate, maintain and terminate the call from the passenger to the control center operator. This state is composed by three sub-states: *CallManagement*, which takes care of assigning priority to the call (information/emergency) and to retrieve the identifier of the train where the passenger is located; *ControlCenterConnection* that performs the connection to the control center; finally, the *InCalling* state where resources are allocated to allow communication between the involved parties.

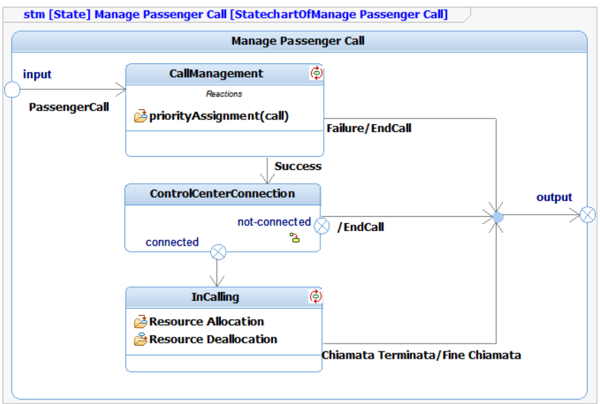


Figure 7: Internal state of the *ManagePassengerCall* of a Control Unit subsystem behavior

A specific path scenario of the above described behavior is shown in Figure 8 by exploiting a SysML Sequence diagram.

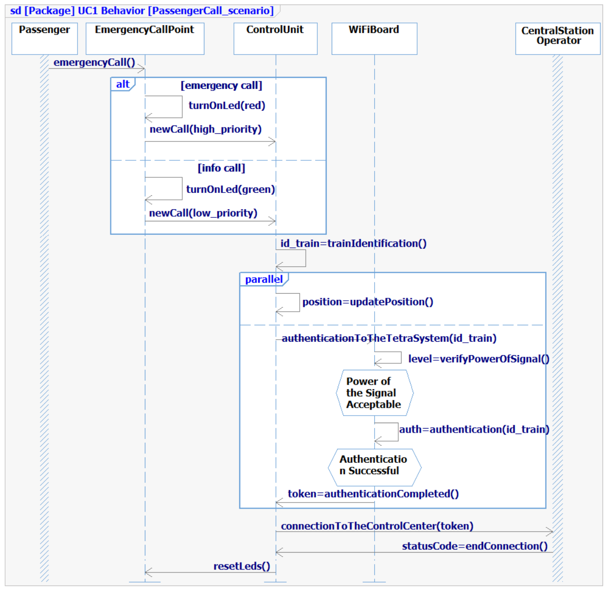


Figure 8: Intended Behavior of a Control Unit when a *Passenger Call* occurs

The modeling of both the *System Structure* and the *Intended Behavior* can be straightforward if during the system design similar structural and behavioral models have been obtained by using a UML/SysML modeling notation.

5.3. Dysfunctional Behavior Modeling

After modeling the intended behavior of systems, dysfunctional behaviors and related tasks, to represent fault and failure events and conditions, are introduced in order to analyze dependability performances through Simulation.

In particular, both the generation, management and the possible propagation of failures are modeled by considering the specific characteristics of components and then realized by combining different probability models based on popular distribution functions such as Weibull and Normal. Figures 9 and 10 show, by using SysML Activity diagrams, the tasks that represent the processes of *FaultManagement* and *FailurePropagation*.

Specifically, the *FaultManagement* task, represented in Figure 9, is able to take in input four types of fault signal: *BasePlateFault, CPUCardFault, WiFiCardFault* e *SoundCardFault*. Then, a specific activity, called *CUFaultManagement*, is in charge of managing opportunely the incoming type of fault. At the end of the management process of faults, two results can be reached: (i) the fault is handled and no other harmful consequences persist/affect in the system, (ii) the fault is not handled, so a *CUFault* signal is sent externally.

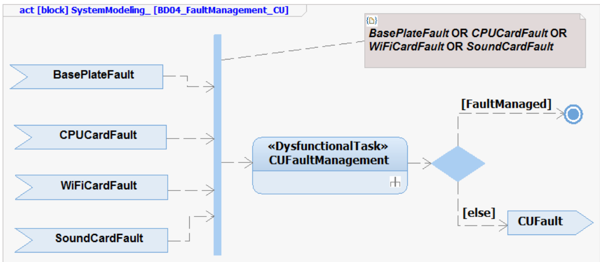


Figure 9: *FaultManagement* task of a Control Unit subsystem

The *FailurePropagation* task, represented in Figure 10, is able to take in input two types of fault signal (i.e. *Failure_24_12_DC_DC_Converter*, *CUFault*) and, after having combined and transformed them, to propagate externally a signal of failure (i.e. *CUFailure*).

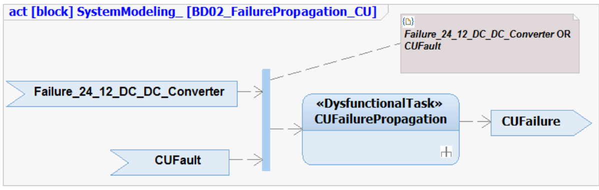


Figure 10: *FailurePropagation* task of a Control Unit subsystem

5.4. Behavior Integration

In the *Behavior Integration* activity, both the intended behaviors and the dysfunctional behaviors modeled in the previous modeling activities are integrated to obtain an overall behavioral model of the system and its component entities. As an example, in order to integrate both the *FaultManagement* and *FailurePropagation* task in the intended behavior of the *ControlUnit* subsystem, two new states have been introduced (see Figure 11) which implement the dysfunctional behavior represented in Figure 9 and Figure 10.

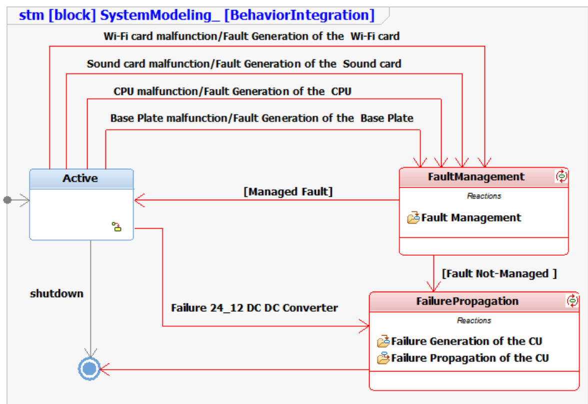


Figure 11: *BehaviorIntegration* of the Control Unit subsystem behavior

In particular, the new state machine that represents the overall behavior of the CU is now modelled by three states:

- *Active*: that performs the intended behavior of the *ControlUnit* as expected under normal operative working conditions;
- *FaultManagement:* that implements the behavior specified by the *FaultManagement* task (see Figure 9), which is responsible for managing fault signals. If the fault is handled, then the state of the *ControlUnit* comes back into the *Active* state, otherwise it changes into the *FailurePropagation* state, as described in the following;
- *FailurePropagation:* that implements both the *FailureGeneration* task for the generation of signals of faults/failures as well as the dysfunctional behavior specified by the *FailurePropagation* task of Figure 10 for the propagation of failures.

In Figure 12 the Sequence diagram shows a failure situation/scenario during the authentication process to the TETRA net when a *PassengerCall* is incoming.

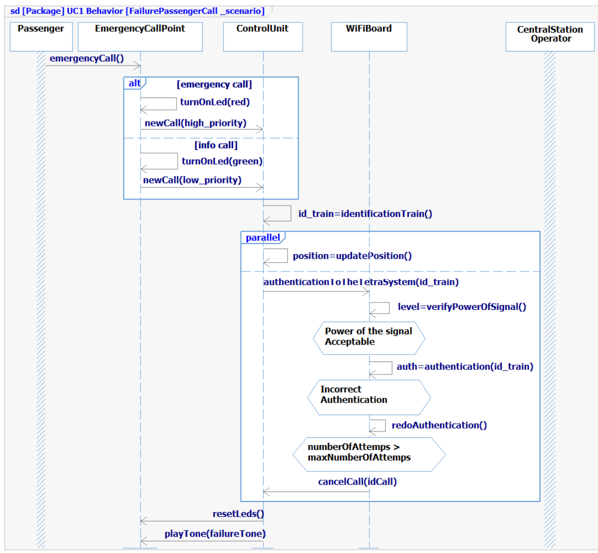


Figure 12: Failure situation of a *PassengerCall*

This activity closes the *System Modeling* phase by delivering the *System Model for Reliability Analysis* (*SMRA*) work-product.

6. SYSTEM SIMULATION

The objective of the *System Simulation* phase is to evaluate through simulation the reliability performances of the system and, possibly, compare different design alternatives and parameters setting.

The following three main activities are performed: *Model Transformation*, *Parameters Setting*, and *Simulation Execution*. Each of these activities is described in the following subsections.

6.1. Model Transformation

In this activity, the models of the system, obtained in the previous phase, are represented in terms of the constructs offered by the target simulation platform.

Indeed, a skeleton of an *Executable System Model (ESM)* is derived from the *System Models for Reliability Analysis* (*SMRA*) obtained in the *System Modeling* phase. In particular, in the current version of the RAMSAS method the *ESM* is generated for the Mathworks Simulink platform which represents a de facto standard for the simulation of multi-domain dynamic and embedded systems. This model transformation is based on a mapping between the basic SysML and Simulink constructs; in particular: (i) a (simple) SysML Block is transformed into a Simulink Block; (ii) a (composite) SysML Block, consisting of other blocks (its parts), is transformed into a Simulink Subsystem Block; (iii) SysML FlowPorts are transformed into Input and Output Simulink Blocks; (iv) SysML Flow Specifications, used to type FlowPorts, are transformed into Simulink Bus Objects. Moreover, the SysML behavioral diagrams which model the intended and the dysfunctional system behavior are transformed in Simulink functions and/or Stateflows, according to specific transformation rules.

As an example, Figure 13 sketches a Simulink model which has been derived from the On-Board Communication System represented, through a SysML notation, in Figure 2 and Figure 3.

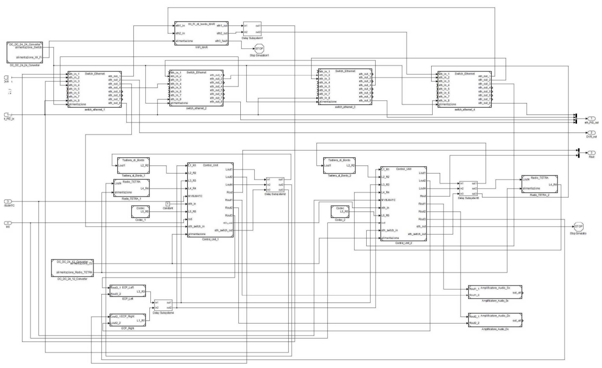


Figure 13: The Simulink model of the On-Board Communication System of the Metro train

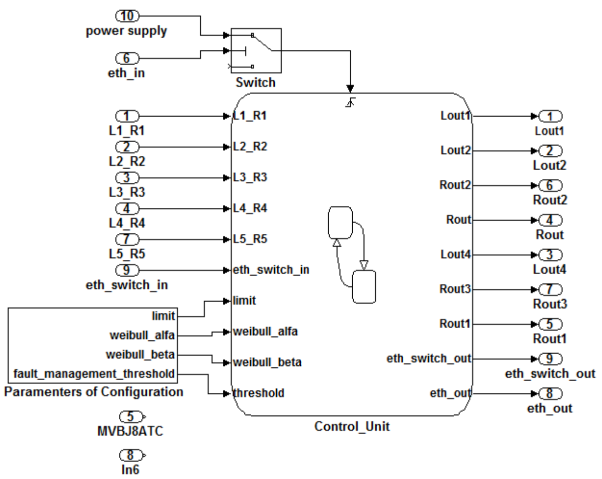


Figure 14: A High Level Simulink model of a Control Unit

As described in Section 3, the system is equipped with two *ControlUnit* (CU) that handle all the functions provided by the system. If both CU get damage, the total blockage of the system occurs. The high level representation of a CU subsystem, along with its signals, is represented in Simulink in Figure 14, which is in charge of carrying on the actions to be taken in response to an external event.

More specifically, Figure 15 shows the internal decomposition, through sub-states, of the CU behavior which in turn has been derived from the SysML diagram of Figure 11 that models the integration of the CU *Intended behavior* with its *Dysfunctional behavior*; whereas, in Figure 16 the internal behavior of the *Active* state and, in particular, of the *ManagePassengerCall*, is shown.

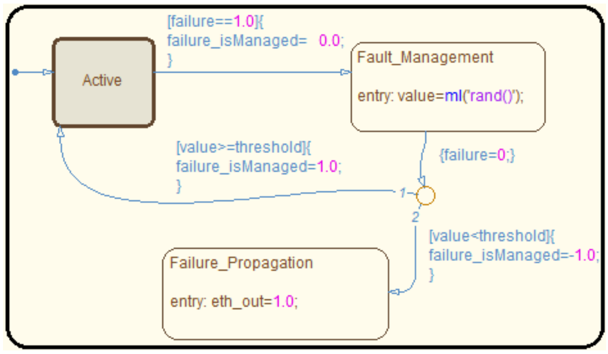


Figure 15: Internal state of a Control Unit subsystem

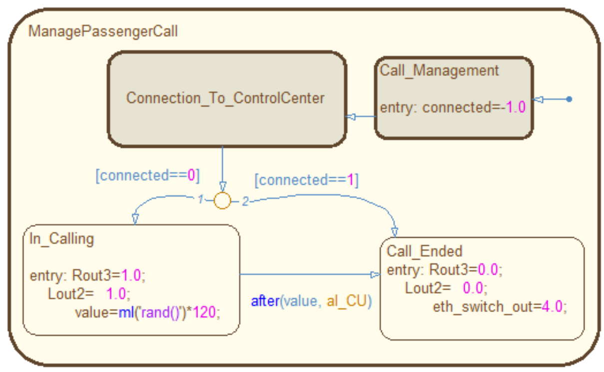


Figure 16: Stateflow of the *ManagePassengerCall* state represented in Simulink

6.2. Parameters Setting

Before starting the simulation, several system and configuration parameters can be set to evaluate system reliability performance in different simulation scenarios. In the *Parameters Setting* activity, the *ESM* is refined so to allow the flexible setting of system configuration and simulation parameters which can be tuned according to both the characteristics of the operative scenario to simulate and the failure modes to analyze (by acting on the settings of the faults and failures generation, propagation and management tasks).

In particular, each component of the system is provided of a module called "configuration parameters" in which is possible to set the parameters of the simulation; as an example, some of them are listed in Table 2.

Table 2: Configuration Parameters with possible values

Parameter	*Description*	*Range*	*Value*
Scale Factor	It represents the statistical dispersion of the Weibull distribution	[0-∞]	1.5
Shape Factor	It defines the shape of the Weibull distribution and the position of its maximum.	[0-∞]	1
Failure Management threshold	It represents the threshold value for the management of failure	[0-∞]	0.1
Failure Generation threshold	It represents the threshold value for the generation of failure	[0-∞]	3.5
…	…	…	…

Then the model is executed according to a synchronous reactive model of computation: at each step, Simulink computes, for each block, the set of outputs as a function of the current inputs and the block state, then it updates the block state. During the simulation *faults* and *failures* are injected and/or properly caused (by *TimedEvent* or *TriggeringEvent*) in order to stress and analyze the behavior of the OBCS.

6.3. Simulation Execution and Results Assessment

The reliability indices of the Metro train system (see Section 3) have been evaluated by considering two alternatives design solutions that differ in the redundancy of the Control Unit. Indeed, the main objective was to evaluate if a second Control Unit is actually necessary to ensure the required system reliability. Indeed, a correct evaluation allows to satisfy the reliability system requirements without unnecessarily complicating the system architecture and increasing production and management costs.

In this context, in the first alternative design solution of the OBCS, the behavior of the system has been analyzed by considering a single Control Unit on board and the value of availability of the overall system has been determined according to the equation (1) reported in Section 4. In the second alternative design solution the availability of the system has been evaluated by introducing redundant elements in the overall architecture of the OBCS and, in particular, by exploiting two Control Units.

Table 3 summarizes the values of *Unavailability*, *DownTime_evMAX* and *Failure Rate (λ)* performed by the On-Board Communication System of the Metro train, respectively by employing one *CU* and two *CUs* in its architectural design.

Table 3: Performances reached respectively with One CU and Two CUs in the OBCS

	Target values	*Design with One CU*	*Design with Two CUs*
TB	Unavailability = 1-Availability = 1-0,9999 = 0,0001	Unavailability = 0,17	Unavailability = 0,0001
PB	λ= 2,31E-06 *[h^{-1}]* *DownTime_evMAX* = 0,16 *[h]*	λ = 4,02E-07 *[h^{-1}]* *DownTime_evMAX* < 0,16 *[h]*	λ = 9,01E-09 *[h^{-1}]* *DownTime_evMAX* < 0,16 *[h]*
Delay	λ= 4,33E-06 *[h^{-1}]* *DownTime_evMAX* = 0,05 *[h]*	λ=5,72E-06 *[h^{-1}]* *DownTime_evMAX* < 0,05 *[h]*	λ=6,33E-06 *[h^{-1}]* *DownTime_evMAX* < 0,05 *[h]*

Finally, the simulation results, obtained from the *Simulation Execution* phase, are analyzed in the *Results Assessment* phase with reference to the objectives of the reliability analysis identified in the initial phase of the process that are reported in Section 4.

In this case, the evaluation through simulation of the dependability performances of the system under consideration has allowed to point out some weaknesses of the OBCS design and thus to produce suggestions for its improvement before its actual realization and deployment/release. In particular, the analysis has revealed that, even though the *Failure Rate (λ)* and the *DownTime_evMAX* are quite similar when one or two Control Units are employed, the value of system reliability improves considerably with two CUs with a consequent increase in its availability. Specifically, the availability of 0,83 obtained by using one Control Unit does not satisfy the value of *availability* required (0,9999) that is easily reached by using two CUs.

7. CONCLUSIONS

Performances evaluation in railway domain and, in particular, dependability requirements analysis is very important to considerably reduce dangers to passengers travelling on the modern railways and to avoid disastrous consequences. To protect civilians against both intentional and unintentional threats, rail transportation has become increasingly automated and performance studies are fundamental to increase the lifetime of railway systems.

In this context, the paper has shown both the modeling of a real On-Board Communication System (OBCS), supplied by SELEX and installed on the Line 5 of the Milan Metro train, and the evaluation of its dependability performance by exploiting simulation techniques.

The concrete experimentation has been supported by RAMSAS, an innovative model-based method for the dependability analysis of systems. According to the RAMSAS method, the definition of a SysML-based model, both for the intended and dysfunctional system behavior, along with the subsequent derivation of a Simulink-based simulation model have been shown.

The experimentation has led to interesting insights about the system design of the OBCS. In particular, the simulations results of the system under analysis have allowed the comparison of different design choices so as to improve the reliability and overall performances of the OBCS. Specifically, the experimentation has highlighted that the OBCS is not able to guarantee its availability if only one Control Unit is exploited, whereas two Control Units are necessary to meet the dependability requirements.

This experience has provided a further demonstration of the effectiveness of the RAMSAS method and its increasing maturity, giving useful insights for guiding its further improvements.

ACKNOWLEDGMENTS
This work has been conducted in the context of TETRis (TETRA Innovative Open Source Services), a research project funded by the Italian Ministry of Education, University and Research (MIUR) within the PON Project - Research and Competitiveness 2007/2013.

REFERENCES

ECSS-Q80-03, 2006. *Space product assurance: Methods and techniques to support the assessment of software dependability and safety*. ESA Publications Division.

European Space Agency (ESA). Available from: http://www.esa.int/.

Flammini F., 2012. Railway Safety, Reliability, and Security: Technologies and Systems Engineering. *IEEE Computer Society*, Italy.

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. *Journal of Computational Science*, 5(3):439-449, Elsevier B.V., Amsterdam (The Netherlands).

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. 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., 2014. On the Reliability Analysis of Systems and SoS: the RAMSAS method and related extensions. *IEEE Systems Journal*. IN PRESS, IEEE Systems Council.

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).

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

International Electrotechnical Commission (IEC). Available from: http://www.iec.ch/.

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

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.

MATLAB-Simulink. Available from: http://www.mathworks.com/.

Molesini A., Omicini A., Ricci A. Denti E., 2005. Zooming multi-agent systems. *Proceedings of the 6th International Workshop on Agent-Oriented Software Engineering*. Utrecht (The Netherlands).

NASA, 2007. *Systems Engineering Handbook*, Revision 1, NASA/SP-2007-6105, NASA.

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.

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).

Reliability and Safety in Railway, 2012. Edited by Xavier Perpinya, pp. 418.

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

RTCA/DO 254, 2000. *Design Assurance Guidance For Airbone Electronic Hardware*.

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

Scott A., Dadashi N., Wilson J. R., Mills A., 2013. *Rail Human Factors: Supporting reliability, safety and cost reduction*. Taylor & Francis.
SELEX ES – A Finmeccanica Company. Available from: http://www.selex-es.com/.
Systems Modeling Language (SysML). Available from: http://www.omgsysml.org/.
TETRA Innovative Open Source Services (TETRis). Available from: http://www.ponrec.it/open-data/progetti/scheda-progetto?ProgettoID=5013.
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 2013)*. April 19, Nottingham (UK).
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).

AUTHOR BIOGRAPHIES

Alberto Falcone

Alberto Falcone received the Laurea Degree in Computer Engineering from the University of Calabria (Italy) in 2011 and a Master title in Industrial Research from the same institution in 2013. He is a PhD student in Information and Communication Engineering for Pervasive Intelligent Environments at Department of Informatics, Modeling, Electronics and Systems Engineering (DIMES) at University of Calabria. His main research interests include Modeling and Simulation of Systems. He has been involved as a Team Leader in the Simulation Exploration Experience (SEE) 2014, a project lead by NASA, in which he has contributed to the design and development of a module able to provide surface-to-surface communication services among the entities that populate a Moon base.

Alfredo Garro

Alfredo Garro is an Associate Professor of Computing Systems at the Department of Informatics, Modeling, Electronics and Systems Engineering (DIMES) of the University of Calabria (Italy). From 1999 to 2001, he has been a researcher at CSELT, the Telecom Italia Group R&D Lab. From 2001 to 2003, he collaborates with the Institute of High Performance Computing and Networking of the Italian National Research Council (CNR). On February 2005 he received the PhD Degree in Systems and Computer Engineering from the University of Calabria. From January 2005 to December 2011, he has been an Assistant Professor of Computing Systems at the DIMES Department (formerly DEIS) of the University of Calabria. His main research interests include: systems and software engineering, reliability engineering, modeling and simulation. His list of publications contains about 80 papers published in international journals, books and proceedings of international and national conferences. He is a member of the IEEE and IEEE Computer Society from 2005. He is a member of the IEEE Reliability Society and IEEE Aerospace and Electronic Systems Society. Currently, he operates as a member of the SPACE Forum Planning and Review Panel (PRP) of the Simulation Interoperability Standards Organization (SISO). He is member of the Executive Committee of the MODRIO (Model Driven Physical Systems Operation) ITEA2 Project and the Technical Contact for his Institution in the Open Source Modelica Consortium (OSMC).

Andrea Tundis

Andrea Tundis received the Laurea Degree in Computer Engineering from the University of Calabria (Italy) in 2009, a Master title in Industrial Research from the same Institution in 2010, and, on February 2014, a PhD Degree in Systems and Computer Engineering from the University of Calabria where he is currently a research fellow. His main research interests include the definition of model-based methods for the reliability and safety analysis of systems as well as models for the formalization and traceability of non-functional requirements. In the last year, he worked at the Programming Environment Laboratory (PELAB) at Linköping University (Sweden) on the extension of the Modelica language for the modeling of system properties.

APPLICATION OF EVOLUTIONARY ALGORITHMS FOR BAKERY PRODUCTION SCHEDULING

Florian T. Hecker[(a)], Bernd Hitzmann[(b)]

[(a), (b)]University of Hohenheim, Department of Process Analytics and Cereal Technology, Garbenstraße 23, 70599 Stuttgart, Germany

[(a)]Florian.Hecker@uni-hohenheim.de, [(b)]Bernd.Hitzmann@uni-hohenheim.de

ABSTRACT

The production in bakeries can be modelled as a no-wait permutation flow-shop based on the constraints and frame conditions given by the real production processes. A modified genetic algorithm, ant colony optimization and particle swarm optimization were used to analyse and optimize the production planning of a bakery production line that processes 40 products on 26 production stages. This setup leads to 8.2 x 10^{47} different possible schedules in a permutation flow-shop model and is thus not solvable in reasonable time with exact methods. The makespan of the production, an objective function of high economic interest, was analysed. In combination with the created model, the applied algorithms proved capable to provide optimized results for the scheduling operation within a restricted computational time of 15 min, reducing the makespan by up to 8.6 %.

Keywords: Bakery Production Planning; Modified Genetic Algorithm; Ant Colony Optimization; Particle Swarm Optimization; Flow-Shop Scheduling

1. INTRODUCTION

It is an economical and often also vital necessity throughout the different industry branches from producing to service companies to work on an optimal level concerning their efficiency or at least as close to this ideal as possible. Even more since the increased awareness of the finite nature of most of the today commonly used resources came up.

To reach a state of efficient production and sustainable or at least efficient use of resources is, besides the degree of efficiency of the utilized machines, for the most part a matter of optimal or at least efficient scheduling.

From its beginnings in the 1950's (Johnson, 1954; Jackson, 1955; Smith, 1956) flow-shop scheduling became an increasingly important tool for decision-making, mainly in the last 20 to 30 years, because of its potential to optimize manufacturing processes and by that creating significant savings for companies of all kinds of industries.

The main elements of which all flow-shop models are build are a set of *j* jobs, which have to be processed on a set of *m* machines. In most manufacturing facilities a job, e. g. a product, has to undergo a series of operations or processing steps related to specific machines. In many cases the jobs have to follow the same route through the production stages and the required machines are assumed to be set up in series. Such a manufacturing environment is then referred to as a flow-shop (FS). Some minor modifications or additional constraints of the aforementioned set up lead to special kinds of FSs. If the order in which the jobs are processed is the same for all stages of the FS, the FS is referred to as permutation flow-shop (PFS). In a more general machine set up that consists of *k* stages with a certain number of parallel machines *m* in each stage and *j* jobs that have to be processed on only one of these machines in each stage, the machine environment is referred to as flexible FS, compound FS or hybrid flow-shop (HFS) (Pinedo 2008).

Any kind of FS scheduling problem that includes a no-wait constraint for the jobs can be labelled as a no-wait flow-shop (NWFS) scheduling problem. This type of scheduling represents a very important application since it can be used for all kinds of time dependent industries like food or pharmaceutical production (Hall and Sriskandarayah 1996, Raaymakers and Hoogeveen 2000), as well as chemical, concrete or steel processing (Fondrevelle, Oulamara, Portmann, and Allahverdi 2009, Grabowski and Pempera 2000, Rajendran 1994).

A commonly faced problem in flow-shop scheduling is that it belongs to the class of NP-hard problems (Garey, Johnson, and Sethy 1976, Lenstra and Rinnooy Kan 1978), which means that the computational requirements for finding an optimal solution increase exponentially with the problem size. Thus large problems cannot be solved using exact methods, but a huge variety of heuristic and metaheuristic solution methods has been developed for flow-shop scheduling.

Due to the increasing importance of tackling complex multi-stage FS scheduling problems to model and solve real world scheduling problems to optimize manufacturing in numerous industrial branches, the range of approaches and solution methods employed is as various as the respective real world archetypes. Nevertheless there are many reviews available that summarize those approaches and methodologies, e. g. the works of Vignier et al. (1999) or Linn and Zhang (1999), Ruiz and Vázquez-Rodriguez (2010), Hejazi and Saghafian (2005), Gupta and Stafford (2006), Minella, Ruiz, and Ciavotta (2008) or Hall and Sriskandarajah (1996).

Three frequently used optimization algorithms for scheduling problems are genetic algorithms (GA), ant colony optimization (ACO) and particle swarm optimization (PSO). Due to the many examples of successful implementation of these methods it was decided to also apply them in this work.

Inspired by the Darwinian principles of evolution, the first GA was introduced by Holland (1975). Genetic and evolutionary algorithms use the same basic operations as in the reproduction and evolution of higher species, like inheritance of genes, mutation, selection and recombination. GAs and modified GAs are widely used for solving complex optimization problems (Fereidoonian and Mirzazadeh 2012, Gómez-Gasquet, Andrés, and Lario 2012, Phanden, Jain, and Verma 2012, Ventura and Yoon 2013, Zhang, Zhou, and Liu 2012, Ziaeifar, Tavakkoli-Moghaddam, and Pichka 2012).

Initially proposed by Dorigo in his Ph.D. thesis (Dorigo 1992) ACO adapts the mechanisms that help ants to find the shortest and thus optimal way between a food source and their formicary. Proven its capability to resolve the TSP (Dorigo and Gambardella 1997), one of the classical NP-hard problems, ACO is another nature inspired and frequently used algorithm to solve combinatorial optimization tasks. There is by now an almost inconceivable variety of applications using ACO, covering all kinds of scheduling, routing and optimization problems. The works of Dorigo, Di Caro and Gambardella (1999), Mullen et al. (2009), Chandra Mohan and Baskaran (2012) and of Tavares Neto and Godinho Filho (2013) provide an impressive overview of how versatile and successful ACO can be employed.

Kennedy and Eberhart (1995) invented PSO as an adaption of the movement and behaviour of bird flocks or fish schools on their search for a food source. Tasgetiren et al. (2007) were the first to tackle makespan and total flowtime minimization of a PFS using PSO. Minimizing C_{max} was also the objective function in the work of Lian, Gu and Jiao (2008) who used a novel PSO adapted to the discrete space of the PFS problem. Pan, Tasgetiren and Liang (2008) created a discrete PSO for a NWFS with makespan and total flowtime criterion and a new position update method as well as several speed-up methods were presented in their work. A *k*-stage NWFS was modelled with makespan as objective function for the scheduling of a polypropylene process by Liu, Gao and Pan (2011) and a hybrid PSO combined with SA was introduced to solve the scheduling problem.

The huge variety of applications of PSO is also presented in the works of Eslami et al. (2012) and Poli, Kennedy and Blackwell (2007).

Although there are these many examples where numerical modelling and optimization have been successfully applied in different industry branches, the baking industry in Germany yet provides no such efforts or applications. Even though this industry branch with its high diversity of products and time dependent production processes is as if predestined for the application of state-of-the-art scheduling methods. In the German baking industry, the production planning is almost completely based on the practical experience of the responsible employee(s) instead of the usage of mathematical methods. Regarding the high diversity of the product range in a common German bakery that includes around 50-100 different products and the high complexity of the scheduling task induced therein, the performance of bakeries is often sub-optimal.

The baking industry in Germany consists of approximately 14,000 producing companies, employs over 290,000 employees and reaches a business volume of almost 13.4 billion Euros per year (Zentralverband des Deutschen Bäckerhandwerks e. V. 2012). The increase of companies' efficiency in respect of energy consumption or staff allocation and man working hours therefore comprises high potential to decrease production costs.

2. MATERIAL AND METHODS

The model implementation, simulation and optimization were performed on a "lenovo ThinkPad R500" with an "Intel Core 2 Duo" 2.26 GHz processor, 2 GB RAM and Microsoft XP 2002 as operating software.

The modelling of a bakery production line with 40 products and 26 production stages, as well as the implementation of MGA, ACO and PSO and the optimization were programmed and performed with MATLAB R2012b (The MathWorks, Inc).

2.1. Modelling procedure

From the scheduling point of view the production in a bakery can be described as a hybrid flow-shop according to commonly used definitions (Pinedo, 2008; Ruiz & Vázquez-Rodriguez, 2010). The machine environment is called a hybrid flow-shop, if jobs have to be processed on only one machine *m* in a stage *k* or can completely bypass it, being the case in a bakery, as long as they are processed on at least one stage.

By considering the scheduling task in a bakery as a permutation flow-shop instead of a 'normal' hybrid flow-shop by adding the constraint, that the order in which the jobs *j* pass through the production is fixed and does not change between production stages (Pinedo, 2008), the number of possible product sequences can be reduced significantly. Although the real process in a bakery does not fulfil these requirements entirely, this model can be used and modified to match with the real production processes, where products can bypass earlier started products and the sequence of products on the first production stage determines all subsequent process tasks, due to the further down specified time dependence in bakery production. Doing so, the number of possible combinations is reduced from $(j!)^m$ to $j!$ and each possible schedule is a permutation of j (Pinedo 2008). Each of those permutations can be used to represent a sequence of products on the first production stage of the bakery which is crucial for sequencing the work flow. Still, high numbers of j, and thus $j!$ different schedules may lead to optimization problems unsolvable with exact methods in reasonable computational time.

Additionally the production in bakeries is commonly subject to a no-wait constraint due to inherent fermentative processes. The production of most baking goods requires proofing and the most commonly used proofing agent is yeast (*Saccharomyces cerevisiae*). The main characteristic of this process is the fermentative decomposition of glucose to CO_2, among other components. The retention of CO_2 produced by the yeast cells is given by the dough matrix surrounding the gas bubbles and leads to a desired volume increase. Up to a certain degree the dough matrix can withstand the structural stress induced by the, over time, increasing gas pressure, but after exceeding the maximum gas retention ability the dough matrix collapses. Due to these processes the production of such goods is not highly but strictly time sensitive from the point on, where the microorganisms get in contact with water and substrates under preferable conditions of temperature and humidity, as it happens in the dough production process. Cooling can be used to regulate or slow down the fermentation speed of yeast but is costly and sometimes accompanied with negative influence on the product quality.

2.1.1. Production modelling

Modelling the production site and ensuring the compliance of the no-wait constraint are done prior to the scheduling optimization.

All products in a bakery follow more or less the same way on consecutive stages through the production, meaning that a product does not return to an already passed stage. The common progression of these production stages is shown in Figure 1.

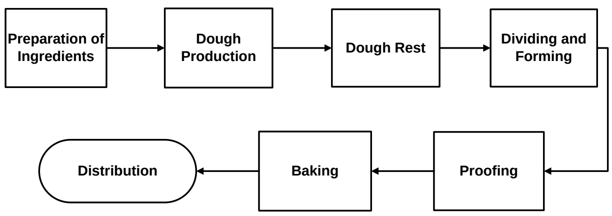


Figure 1: Basic Bakery Production Flow Model

Besides the definition of each products individual work flow through the production and the capacities of the employed machines, the processing times (*PTs*) of the products on each stage represent the most important information for modelling. *PT* also include process steps, where the product is not literally "processed" by means of being influenced by a worker or a machine, like the dough rest since this is a defined and desired waiting time. Table 1 shows an example containing the required processing information of products. The numbers in Table 1 represent the elements of a matrix ***A***. The rows and columns of matrix ***A*** represent the products and the production stages respectively, such that, for example $\underline{\boldsymbol{a}}_{2,3}$ would return the processing time of product 2 on stage 3, meaning in case of Table 1, that product "B" requires a forming time of 10 min. The recipe and the desired characteristics of the finished product determine the respective processing times.

Table 1: Example of bakery production data

Stage / Product	Dough production [min]	Dough rest [min]	Forming [min]	Proofing [min]	Baking [min]
A	5	0	5	35	25
B	4	20	10	35	55
C	8	30	15	50	30
D	6	25	25	40	60
E	9	0	8	55	40
F	10	10	12	35	35

Taking into account that some products do not have to be processed on all present stages (e. g. if a product needs no dough rest) and therefore might skip certain processing steps, a zero entry in matrix ***A*** indicates that the specific product skips the respective stage and is not processed there.

Matrix ***A*** is the basis to form a new matrix ***B***. To do so the starting times (*STs*) of all products on all stages are calculated for the investigated sequence of products (which represents an individual in GA or a path of an ant in ACO or a particle in PSO for optimization). *ST* for the first product on stage 1 is "0" as this represents the start of the production shift and the *STs* on the successive stages are just a summation of the respective previous *PTs*.

Each investigated product sequence is scheduled following the procedure shown in Figure 2. The scheduling conditions include a reconciliation of the calculated *STs* of the currently processed product with the capacities and busy times of the involved machines, to make sure that the no-wait constraint is not violated and the sequence is valid for the following optimization procedure.

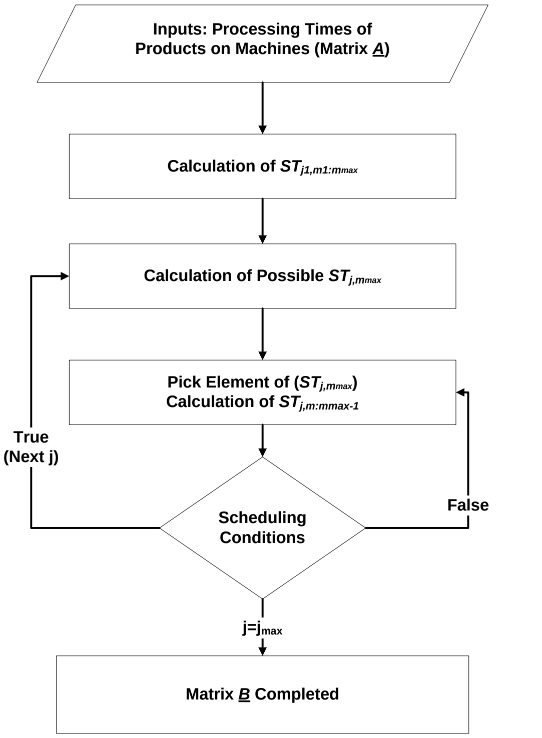


Figure 2: Flow chart of modelling algorithm without optimization procedure

Using the example data of Table 1 and defining that the processes “Dough rest” and “Proofing” are stages with unlimited capacity, matrices ***A*** and ***B*** will be obtained as shown in Figure 3, where ***A*** contains the processing times and ***B*** the starting times of products (row-wise) on the respective stages (column-wise), such that e. g. $\underline{\boldsymbol{b}}_{4,4}$ means that product “D” starts its proofing process 119 min after the shift start.

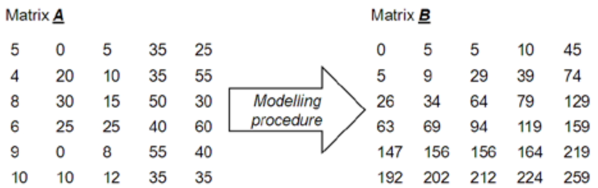


Figure 3: Example matrices A extracted from Table 1 and B with the same product order

To comply with the no-wait constraint, the processing start of some products has to be delayed, as can be seen in the third row of matrix ***B*** shown in Figure 3. The first machine would be available for product “C” at nine min after shift start, but starting the processing at this time would mean that the product has to be baked at 112 min after shift start. At this time the oven is blocked by product “B”, which is baked from 74-129 min after shift start and product “C” would have to wait, thus violating the no-wait constraint.

A real bakery production in Germany including 40 different products and 26 different production stages was investigated in this study. Figure 4 shows the corresponding production line model. Modelling this production line as a permutation flow-shop leads to a total of 8.2 x 10^{47} (= 40!) different possible schedules.

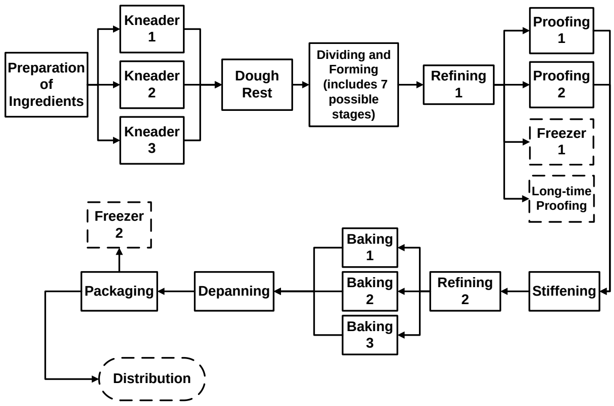


Figure 4: Model of the investigated production

The common way of a product would be to exit the production line as a baked good via the “Distribution” process. Nevertheless some products can exit the production e. g. without passing the baking process due to long-time proofing or being distributed as frozen dough pieces and thus do not influence the production after their specific final process step. The model therefore contains different exit points, shown as shapes with dashed lines in Figure 4. Any product following a long-time proofing procedure will be distributed or baked after the current work shift has finished and therefore exits at the “Long-time Proofing”. And finally products that are distributed frozen have to be stored in a freezer and thus exit the model at the “Freezer 1” or “Freezer 2” stage.

The capacities of the employed machines or stages have a crucial importance for the scheduling process. The production stages can roughly be grouped into stages with limited capacity and stages with (practically regarded) unlimited capacity. The latter can handle any amount of different products without blocking under common production conditions. In the modelled bakery shown in Figure 4 “Dough Rest”, “Proofing 1”, “Proofing 2”, “Long-time Proofing”, “Stiffening”, “Freezer 1”, “Freezer 2” and “Depanning” belong to the group of stages with unlimited capacity. “Baking 1” has a limited capacity of two and “Baking 2” of six, meaning that these stages can process two or six products (or product batches) simultaneously. All other stages of the model have a limited capacity of one and can process only one job at a time.

2.2. Optimization procedures

As a quality measure of the analysed product sequence at least one cost or objective function to be optimized must be defined prior to the actual optimization process. The minimization of the makespan of the production was chosen as objective function in this work.

The makespan (C_{max}) represents the required time to complete a defined production goal and equals the highest end time (ET) of the products and thus can be easily retrieved from matrix $\underline{\boldsymbol{B}}$ and matrix $\underline{\boldsymbol{A}}$ by summation of the start of the last production step of the last product (e. g. $\underline{\boldsymbol{b}}_{6,5}$ in Figure 3) and the corresponding processing time (e. g. $\underline{\boldsymbol{a}}_{6,5}$ in Figure 3). For the example shown in Figure 3 C_{max} would thus be 294 min (259 min + 35 min).

The computational time was restricted to a maximum of 15 minutes to make the optimization procedure feasible for usage prior to a work shift in a real production environment. To match this constraint it was decided in the MGA to limit the number of individuals to 100 and the number of generations to ten. Analogous in the ACO and PSO the number of ants/particles was limited to ten and the number of iterations to 100, respectively.

2.2.1. Modified Genetic Algorithm

The classical GA is primarily not considered to be applied to combinatorial scheduling problems. Therefore the GA used in this work was modified to make it suitable to solve the combinatorial sequencing problem. The conducted modifications follow the ideas of the New Genetic Algorithm (NGA) introduced by Ventura and Yoon (2013) and of partially matched crossover (PMX) by Goldberg and Lingle (1985).

The first step consists of the calculation of the objective function values of all individuals of the initial population and their ordering according to the respective results. Afterwards the best 50 % of the population are selected for the mating and reproduction process. The mating in this MGA simply consists of coupling the selected individuals pair wise, such that the individual with the best objective function value is mated with the second best, the third best is mated with the fourth best and so on. As next step two offspring are created by each mated pair.

The creation of these offspring is performed by a PMX operation. In contrast to the NGA (Ventura and Yoon 2013) four crossover regions were defined such that the jobs on positions 1-5, 11-15, 21-25 and 31-35 in the parent sequences were exchanged as well as the respective jobs included in these sections, to ensure that the offspring sequences represent permutations without job duplications.

After the production of the offspring by PMX, mutation can occur with mutation probability of Mp = 0.1. If mutation occurs, two jobs in the sequence are chosen at random with a uniform distribution and change their respective places. The progress of offspring creation in the employed MGA is illustrated in Figure 5, showing the recombination of the two parent sequences P1 and P2 to create one example offspring.

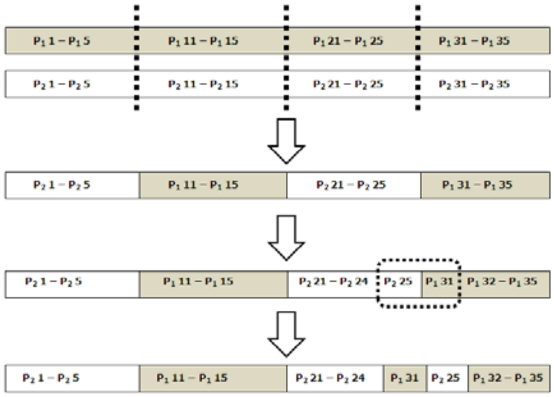


Figure 5: Creation of Offspring in MGA

The four crossover regions are separated by the three vertical dashed lines. After creating an offspring with recombined ‘genes’ mutation occurs (indicated by the dotted frame), causing two products to change their respective positions and creating a mutated offspring.

2.2.2. Ant Colony Optimization

Like GA, ACO is a nature inspired iterative optimization algorithm. In this case the activities and mechanisms observable in sedentary ant colonies are adapted, not the processes that happen during gene transfer in the reproduction of higher organisms. It is well known in biology that such ants often find the shortest and thus optimal way between a food source and their formicary over a certain time span.

The mechanisms behind this effect are also known and based on the special way of communication between ants via pheromones, a special group of evaporative biochemical molecules. If a foraging and randomly wandering ant finds a food source it heads back to its formicary, laying down a pheromone trail. Other ants crossing such a trail will stop wandering randomly and follow this trail to the food and start bringing it to the nest while also laying down pheromones and thus reinforcing the scent. With time the shortest and most frequently used way will thus hold the highest pheromone concentration and attracts the highest number of ants. Pheromones are not persistent and once a food source is exhausted, the previously marked and reinforced pheromone trails are less and less used and the scent dissipates over time.

In ACO an artificial ant “moves” by applying a local stochastic decision and while moving builds a possible solution to the given optimization problem, e. g. a product sequence as in the presented work. Figure 6 shows the process of creating such a possible solution.

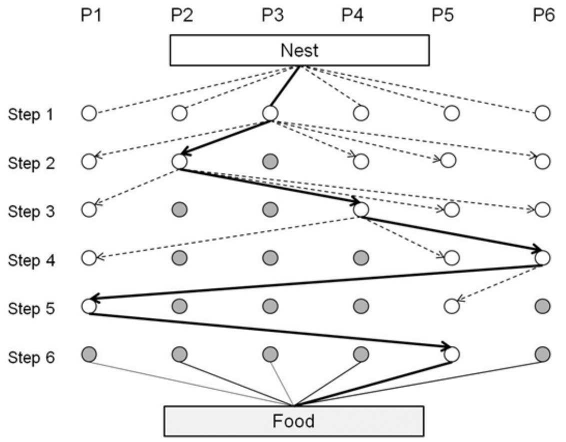


Figure 6: Modified ACO for the Production Planning Optimization

The bold arrows indicate the chosen path of an ant and every position already visited is depicted grey and cannot be chosen any more. In every step the remaining possible paths to the next position are shown as arrows with dashed lines.

This ant's "way" and resulting cost function value are then used as pheromone information to direct the moves of "following" ants by influencing their stochastic decision making process.

The ACO algorithm applied to solve the production planning task by providing an optimized product sequence works with the following steps:

1. Input parameters for ACO: cost function to be optimized (C_{max}), number of ants (n), number of iterations (i), number of products (p), initial pheromone matrix (P_i).
2. Initialize an ant ($a_{n,i}$) with random starting position (pos_p). This position represents the first product in the sequence.
3. Move $a_{n,i}$ to the next position according to a probability factor (pf_p) (where pf_p for an already "visited" $pos_p = 0$) and the pheromone trail in P_i.
4. Repeat 3. until all pos_p are "visited" and thus a possible solution / product sequence is created for each $a_{n,i}$.
5. Evaluate each ant's fitness (cost function value).
6. The sequence of $a_{n,i}$ with the best value (s_{best}) represents the best way so far and is used to lay pheromone upon in its representation in P_i. Thus $a_{n,i}$ in following iterations are more attracted to this way / sequence.
7. Evaporate pheromone according to an evaporation factor (ef) with $0 < ef < 1$ to update P_i into P_{i+1} following equation (1):

   $$P_{i+1} = P_i * ef \quad (1)$$

   The value of ef influences how randomly $a_{n,i}$ are searching for an optimal product sequence. A low value will lead to a more random creation of new product sequences, a high value will lead to rather preserving the best sequence found so far and smaller changes thereof.
8. Repeat 2. – 7. until an exit criterion (maximum i, time limit, result quality, etc.) is met.

2.2.3. Particle Swarm Optimization

There are many examples in nature where animals are forming swarms for higher individual safety or better chances to locate food sources. In such a swarm each individual member searches for a food source by itself while staying within a certain range to its neighbours and thus maintain the swarm structure. Once a swarm member finds food it will move towards it, attracting other swarm members to move in the same direction. The more individuals move to a food source and thus influencing the motion of their swarm neighbours, the whole swarm will move to a certain location over time depending on the ratio of moving individuals compared to the swarm size.

The "swarm" in PSO consists of particles, possible solutions (e. g. product sequences as in the presented work) of a given optimization problem, that represent the individual swarm members in nature. During the iterations of the algorithm the particles are "flying" through the search space and due to a frequent update and comparison of the swarm's best sequence so far and each particle's current value of the cost function, move over time to the optimal solution of the given optimization problem.

The PSO algorithm adapted for the bakery scheduling problem works as follows:

1. Input parameters for PSO: cost function to be optimized (C_{max}), number of particles (or swarm size), number of iterations (i).
2. Initialize a swarm of particles (x_i) with random positions (p_i) and velocities (v_i).
3. Each x_i represents a possible product sequence and the predefined initial v_i defines how each of these sequences is changed from the initial to the next iteration.
4. Evaluate each particle's fitness (cost function value).
5. Compare particle's fitness with its personal best value (*pbest*), update *pbest* if current fitness value is better and set *pbest* position of the particle to the current position.
6. Compare particle's fitness to swarm's best value (*gbest*), update *gbest* if current fitness value is better and set *gbest* parameters to the according particle's parameters.
7. Change particle's velocity and position according to equations (2) and (3) respectively:

   $$v_i = v_i + C1 * (pbest_i - x_i) + C2 * (gbest - x_i) \quad (2)$$

$$x_i = x_i + v_i \quad (3)$$

C1 and C2 are two positive constants in the original PSO and in this application equal “1” since the velocity is used to create a new product sequence.

8. Repeat 3. – 6. until an exit criterion (maximum i, time limit, result quality, etc.) is met.

3. RESULTS

MGA, ACO and PSO were used separately to perform 21 separate optimization runs each. Figures 7 and 8 show the results as C_{max} and the reduction of C_{max} compared to the initial value, respectively. The objective function value of the initial product sequence, which is the representation of the real production schedule used in the modelled bakery, for C_{max} was 1,380 min.

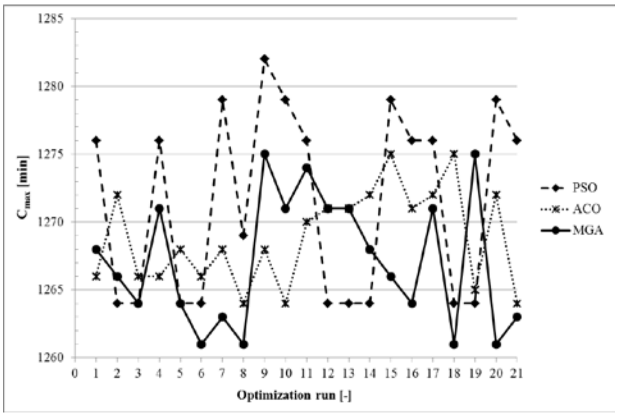


Figure 7: Optimization Results for C_{max} using Different Algorithms

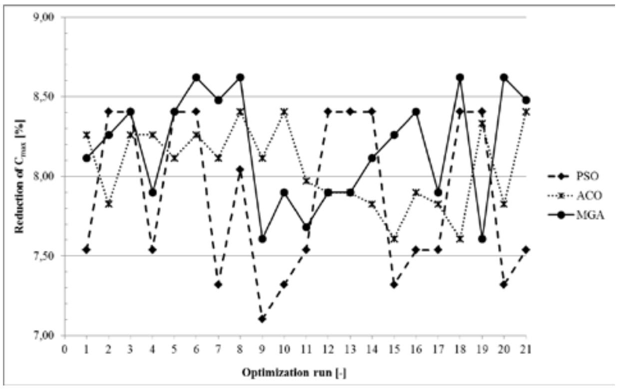


Figure 8: Reduction of C_{max} in % Compared to Initial C_{max} Value

The results show that each algorithm’s optimization run found an optimized solution for C_{max}. With mean values for C_{max} of 1,267 ± 5 min (reduction of 8.18 ± 0.35 %) for MGA, 1,269 ± 4 min (reduction of 8.05 ± 0.26 %) for ACO and 1,271 ± 7 min, all three algorithms performed quite similar. To further analyse if significant differences in the algorithm performances can be obtained, a one-way ANOVA (**AN**alysis **O**f **VA**riance) with significance level $\alpha = 0.05$ was performed. A significant difference in the results is given if the resulting p-values of the ANOVA are smaller than the significance level α and the smaller the p-values the higher the significance of the differences.

The resulting p-values show that the results for optimizing C_{max} of MGA and ACO have no difference ($p = 0.18$), as well as the results of ACO and PSO ($p = 0.15$). The comparison of the results of MGA and PSO instead seems to show a difference with MGA performing better ($p = 0.03$).

The MGA provides the overall best solution of all optimization runs for C_{max} with a reduction of 8.62 % (or 119 min) in four out of 21 runs, PSO provides the overall worst solution with a reduction of 7.10 % (98 min). Still all algorithms even in their worst case performances provide significantly better results for C_{max} than the initial sequence. This means that the production goal of the analysed shift could be reached considerably faster, saving between 98 and 119 min of shift length depending on the optimization result, and thus saving a significant amount of man hours. This becomes especially clear if one has in mind that the reduction of shift length has to be multiplied with the number of involved staff members to calculate this reduction of man hours.

With a mean computational time of 749 ± 28 s, MGA tends to run slightly faster than ACO with a mean computational time of 775 ± 6 s and PSO with 783 ± 17 s. ANOVA results confirmed this, computational times of MGA and ACO ($p = 2.0\text{x}10^{-4}$) and MGA and PSO ($p = 2.7\text{x}10^{-5}$) show a highly significant difference, respectively. The differences in the computational times of ACO and PSO could not be determined clearly, the ANOVA result provided a p-value of $p = 0.0455$ and is thus just slightly below the significance level of $\alpha = 0.05$.

For a better evaluation of the obtained results, additional optimization runs without time restrictions where performed using 2,000 individuals and 500 generations in MGA and 2,000 ants/particles and 500 iterations in ACO/PSO. All three algorithms provided 1,261 min as best result for C_{max}.

4. DISCUSSION

The obtained results are distributed in a certain range, indicating that the sequences obtained in most of the optimization runs have to be local optima. The overall best result (1,261 min) obtained by using MGA seems to represent the global best solution for the problem, as the optimization runs without time restriction also provided this result for all employed optimization methods. This objective function value, that occurred seven times (in four MGA runs and the three optimization runs without time restriction), was provided by four different product sequences.

As with every modelling and optimization task, the quality of the obtained results is directly dependent on the quality and completeness of the data it is based upon. Therefore the optimization results presented have to be regarded critical due to the limited amount of real production data available. The optimization of C_{max} is based only on the data for the bread and bread roll production lines of the case study bakery, possible interactions with other production departments at the baking stages would probably affect the presented results, like the decrease of C_{max} by 7.10-8.62 % (or 98-119 min). Here more data has to be collected from the bakery involved in this work. Also the participation of additional companies providing their production data would further enhance the progress of this project.

Nevertheless even the worst case results for the optimization of C_{max} show significant benefits compared to the initial product sequence's objective function value, regardless of the applied methods.

Since the initial product sequence represents the real production schedule, the calculated results for the respective starting and end times of the products using the model presented in section 2.1.1. were used for verification. These results were checked and approved to be correct in respect to the real production progress by the shift leaders of the bakery modelled in this work. Thus the model used seems to be valid and provides reliable simulation results.

5. CONCLUSION AND OUTLOOK

The application of numerical modelling and optimization algorithms to develop a production planning procedure capable of solving the scheduling task in a bakery proved to be successful. A model of the production processes was designed in MATLAB, that schedules the workflow of a given product sequence according to defined decision parameters. An MGA, an ACO and a PSO have been used separately to solve the optimization task with respect to the optimal makespan of the production.

The applied methods proved capable of solving the given optimization problems in a computational time restricted to a maximum of 15 min, thus providing a scheduling tool that can be employed in a limited time frame prior to a production shift start.

After appropriate customization of the production model it is now possible for decision-makers in baking companies to approach their scheduling task in a fast and promising way capable for usage in praxis, based on the developed modelling and optimization procedure. Since even the worst case results of the applied methods yielded significant benefits compared to the results given by the initial product sequence, using such a mathematical procedure would most probably lead to a considerable increase of baking companies' efficiency.

Running the developed procedure on a PC with higher performance, to calculate more iterations or bigger ant/particle swarms or population sizes in the same computational time restriction, or simply allowing more time for the calculation, would be a promising way to further enhance the obtained results.

Another succeeding step will be to investigate the whole production of the modelled bakery, covering all existing different production departments to provide a companywide optimization approach.

Simultaneously we are also trying to convince additional bakeries to become new project partners and to provide their production information for new modelling and simulation studies.

ACKNOWLEDGEMENTS

This research has been funded by CSM Deutschland GmbH. Mr.

Hermann-Josef Michaelis (DialogMarketing GmbH) supported the collection and availability of the production data. The utilized production data were provided by a baking company in Frankfurt/Main, Germany. This research was not further influenced besides the financial and data collection support.

M. Sc. Marc Stanke, Dipl.-Chem. Olivier Paquet-Durand and M. Sc. Walid B. Hussein contributed with their support in the programming of the optimization algorithms.

REFERENCES

Chandra Mohan, B., Baskaran, R., 2012. A survey: Ant Colony Optimization based recent research and implementation on several engineering domain. *Expert Systems with Applications*, 39, 4618-4627.

Dorigo, M. Gambardella, L.M., 1997. Ant colonies for the travelling salesman problem. *BioSystems*, 43, 73-81.

Dorigo, M., 1992. *Optimization, learning and natural algorithms*. Thesis (PhD). Politecnico di Milano, Italy.

Dorigo, M., Di Caro, G., Gambardella, L.M., 1999. Ant Algorithms for Discrete Optimization. *Artificial Life*, 5, 137-172.

Eslami, M., Shareef, H., Khajehzadeh, M., Mohamed, A., 2012. A Survey of the State of the Art in Particle Swarm Optimization. *Research journal of Applied Sciences, Engineering and Technology*, 4 (9), 1181-1197.

Fereidoonian, F., Mirzazadeh, A., 2012. A genetic algorithm for the integrated scheduling model of a container-handling system in a maritime terminal. *In Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment*, 226 (1), 62-77.

Fondrevelle, J., Oulamara, A., Portmann, M.-C., Allahverdi, A., 2009. Permutation flow shop with exact time lags to minimise maximum lateness. *International Journal of Production Research*, 47 (23), 6759-6775.

Garey, M., Johnson, D., Sethy, R., 1976. The complexity of flow shop and job shop scheduling. *Mathematics of Operations Research*, 1 (2), 117-129.

Goldberg, D.E., Lingle, R., 1985. *Alleles, loci, and the traveling salesman problem. Proceedings of the 1st international conference on genetic algorithms and their applications*, pp. 154-159. July 24-26, Pittsburgh (Pennsylvania, USA).

Gómez-Gasquet, P., Andrés, C., Lario, F.-C., 2012. An agent-based genetic algorithm for hybrid flowshops with sequence dependent setup times to minimise makespan. *Expert Systems with Applications*, 39, 8095-8107.

Grabowski, J., Pempera, J., 2000. Sequencing jobs in some production system. *European Journal of Operational Research*, 125, 535-550.

Gupta, J.N.D., Stafford, E.F., 2006. Flowshop scheduling research after five decades. *European Journal of Operational Research*, 169, 699-711.

Hall, N.G., Sriskandarayah, C., 1996. A survey of machine scheduling problems with blocking and no-wait in process. *Operations Research*, 44, 510-525.

Hejazi, S.R., Saghafian, S., 2005. Flowshop-scheduling problems with makespan criterion: a review. *International Journal of Production Research*, 43 (14), 2895-2929.

Holland, J.H., 1975. *Adaption in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Oxford: University of Michigan Press.

Jackson, J.R., 1955. Scheduling a production line to minimize maximum tardiness. *Research Report 43*, Management Science Research Project, University of California, Los Angeles.

Johnson, S.M., 1954. Optimal two- and three-stage production schedules with set up times included. *Naval Research Logistics Quarterly*, 1, 61-68.

Kennedy, J., Eberhart, R.C., 1995. Particle swarm optimization. *Conference Proceedings Volume 4 1995 IEEE International Conference on Neural Networks*, pp. 1942-1948. Nov. 27-Dec. 1, Perth (Australia).

Lenstra, J.K., Rinnooy Kan, A.H.G., 1978. Complexity of scheduling under precedence constraints. *Operations Research*, 26, 22-35.

Lian, Z., Gu, X., Jiao B., 2008. A novel particle swarm optimization algorithm for permutation flow-shop scheduling to minimize makespan. *Chaos, Solitons and Fractals*, 35, 851-861.

Linn, R., Zhang, W., 1999. Hybrid flow shop scheduling: a survey. *Computers & Industrial Engineering*, 37 (1-2), 57-61.

Liu, H., Gao, L., Pan, Q., 2011. A hybrid particle swarm optimization with estimation of distribution algorithm for solving permutation flowshop scheduling problem. *Expert Systems with Applications*, 38, 4348-4360.

Minella, G., Ruiz, R., Ciavotta, M., 2008. A review and evaluation of multiobjective algorithms for the flowshop scheduling problem. *INFORMS Journal on Computing*, 20 (3), 451-471.

Mullen, R.J., Monekosso, D., Barman, S., Remagnino, P., 2009. A review of ant algorithms. *Expert Systems with Applications*, 36, 9608-9617.

Pan, Q.-K., Tasgetiren, M.F., Liang, Y.-C., 2008. A discrete particle swarm optimization algorithm for the no-wait flowshop scheduling problem. *Computers & Operations Research*, 35, 2807-2839.

Phanden, R.K., Jain, A., Verma, R., 2012. A Genetic Algorithm-based approach for flexible job shop scheduling. *Applied Mechanics and Materials*, 110-116, 3930-3937.

Pinedo, M.L, 2008. *Scheduling - theory, algorithms, and systems*. 3rd ed. New York: Springer.

Poli, R., Kennedy, J., Blackwell, T., 2007: Particle swarm optimization: An overview. *Swarm Intelligence I*, 33-57.

Raaymakers, W., Hoogeveen, J., 2000. Scheduling multipurpose batch process industries with no-wait restrictions by simulated annealing. *European Journal of Operational Research*, 126, 131-151.

Rajendran, C., 1994. A no-wait flowshop scheduling heuristic to minimize makespan. *Journal of the Operational Research Society*, 45, 472-478.

Ruiz, R., Vázquez-Rodriguez, J.A., 2010. The hybrid flow shop scheduling problem. *European Journal of Operational Research*, 205, 1-18.

Smith, W.E., 1956. Various optimizers for single stage production. *Naval Research Logistics Quarterly*, 3, 59–66.

Tasgetiren, M.F., Liang, Y., Sevkli, M., Gencyilmaz, G., 2007. A particle swarm optimization algorithm for makespan and total flowtime minimization in the permutation flowshop sequencing problem. *European Journal of Operational Research*, 177, 1930-1947.

Tavares Neto, R.F., Godinho Filho, M., 2013. Literature review regarding Ant Colony Optimization applied to scheduling problems: Guidelines for implementation and directions for future research. *Engineering Applications of Artificial Intelligence*, 26, 150-161.

Ventura, J.A., Yoon, S.-H., 2013. A new genetic algorithm for lot-streaming flow shop scheduling with limited capacity buffers. *Journal of Intelligent Manufacturing*, 24 (6), 1185-1196.

Vignier, A., Billaut, J.C., Proust, C., 1999. Hybrid flowshop scheduling problems: state of the art. *Rairo-Recherche Operationelle-Operations Research*, 33 (2), 117-183.

Zentralverband des Deutschen Bäckerhandwerks e. V., 2012. Available from: http://www.baeckerhandwerk.de/baeckerhandwerk/zahlen-fakten [accessed 05.11.2013].

Zhang, Y., Zhou, Z., Liu, J., 2012. The production scheduling problem in a multi page invoice printing system. *Computers & Operation Research*, 37, 1814-1821.

Ziaeifar, A., Tavakkoli-Moghaddam, R., Pichka, K., 2012. Solving a new mathematical model for a hybrid flow shop scheduling problem with a processor assignment by a genetic algorithm. *International Journal of Advanced Manufacturing Technology*, 61 (1-4), 339-349.

AUTHORS BIOGRAPHY

Florian T. Hecker was born 1979, finished his apprenticeship as a baker in 2002 and afterwards studied Food Technology at the University of Hohenheim, Stuttgart, Germany, where he received his Engineer‘s Degree in 2008. In May 2008 he started working as a researcher and PhD student in the Department of Process Analytics and Cereal Technology, Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, Germany. His research interests are in production optimization, scheduling theory and image processing based quality evaluation systems.

Bernd Hitzmann studied Physics at the University of Hannover, Germany, where he got also his PhD from the Faculty of Chemistry. Thereafter he worked as Post-doc at the California Institute of Technology, Pasadena, USA, as well as the Research Center of Asea Brown Boveri, Heidelberg, Germany. He got is habilitation in Technical Chemistry at the University of Hannover, Germany. Now he is head of the Department of Process Analytics and Cereal Technology, Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, Germany. His research areas are: Design of experiments, spectroscopy for process monitoring, mathematical modelling, data evaluation, control, automation, and optimization of process in biotechnology and especially in cereal technology.

NEURONAL CONTROL STRATEGIES FOR AC-DC CONVERTER WITH POWER FACTOR CORRECTION

Rija RAJAONAH[(a)], RAJONIRINA Solofanja[(b)], RAZAFINJALA Nirnarison Jean[(c)]

[(a)]Laboratoire d'Automatique, Ecole Superieure Polytechnique d'Antsiranana, Madagascar
[(b)]Laboratoire d'Electronique et Informatique Industrielles,Ecole Supérieure Polytechnique, Madagascar
[(c)]Laboratoire d'Automatique, Ecole Supérieure Polytechnique, d'Antsiranana, Madagascar

[(a)]rajaonah.rija@gmail.com , [(b)]rajonirinas@gmail.com , [(c)]razafinjaka@yahoo.fr

ABSTRACT

This paper deals with the study of two kinds of neuronal controllers applied to a AC-DC converter with power factor correction (Boost PFC): one with an average current mode whose realization is analogical and the other with classic structure whose neuronal controller is numerical. The total harmonic distortion (THD) is taken as criterion of evaluation as well as the regulation speed: the systems must respect standard IEC 61000-3-2. In both cases, the practical project of realization is taken into account. The simulation results showed the effectiveness of the two suggested methods.

Keywords: Nonlinear control, neuronal networks, boost PFC, total harmonic distortion, average current mode, LabView

1. INTRODUCTION

Although of lower cost, the traditional converter with capacitor and diodes generate generate 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).
To bring solution of this problem, various strategies are proposed whose principal objectives are summarized as follows (Benaïssa 2006), ,(Tédjini 2008), (Singh 2003), (Keraï 2003), (Razafinjaka 2013):

- 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 Vs

There are several methods to obtain these objectives according the adopted model. The system is primarily non-linear because of the presence of the static inverter. However, in several cases, modeling the loop voltage as a system of first order is sufficient to obtain good results. In this case, it is assumed that the current loop is perfect. The basic scheme is given in figure 1.

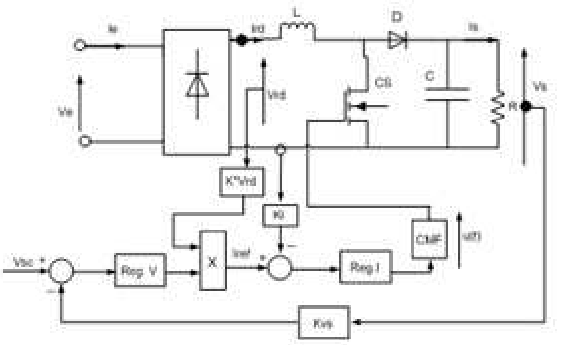


Figure 1: Structure of the boost PFC.

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 the 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, two kinds of boost PFC are proposed: the ACCM boost PFC and the classic boost PFC. For the study, the loop voltage is calculated assuming that the current loop is perfect.

Neuronal controllers are applied. The training uses the descent of gradient method with Levenberg-Marquardt algorithm. Obtaining simple solutions but effective in both cases is held in account. In all that follows, the structure given by Figure 1 is the reference. The current i_L which is equal to ired in the inductor L must follow in all time the current reference iref and it must not falls in zero as showed by Figure 2.

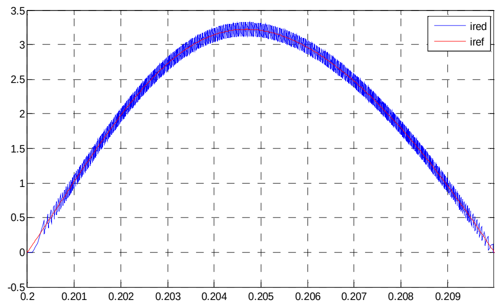


Figure 2: The current in continuous conduction mode

2. BOOST PFC WITH AVERAGE CONTINUOUS CURRENT MODE

Usually, the ACCM Boost PFC is used or high power (Abdel-Rahman 2013), (Dixon 1999) . Here, classic controller as PI is applied for the loop voltage and a neuronal controller is proposed for the current loop. To build the command d(t), several variables are taken in account : iref, Vrd and i_L.

The reference current iref is obtained by multiplying the output controller for the loop voltage with a part of the dressed voltage Vrd. The goal is to have a linear expression for the command u(t):

$$d(t) = a_3 i_L(t) + a_2 iref(t) + a_1 Vrd(t) + a_0 \qquad (5)$$

2.1. Neuronal controller

Here, the classic controller itself it is not to be identified in which case a dynamic network is necessary but the proposition is to identify directly the output d(t). To drive the converter static CS, this output is compared with a signal saw tooth to obtain u(t): it is the role of the shaping circuit CMF (figure 1).

The basic idea starts from a artificial neurons network which the structure is given according to figure 3.

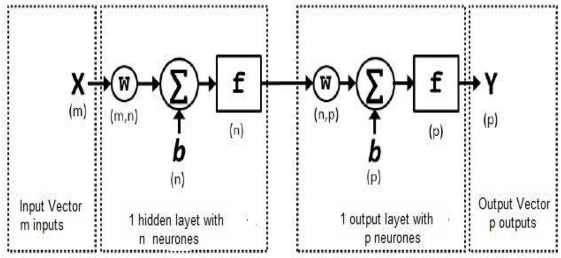


Figure 3: Neuronal network p outputs- m inputs

In our case, there are three inputs and one output. For the study, the followed choice is taken:

- Feed forward network
- Linear function for the activation function
- Type batch training using descent of gradient algorithm as Levenberg-Marquardt

The network used for the identification is so showed according figure 4.

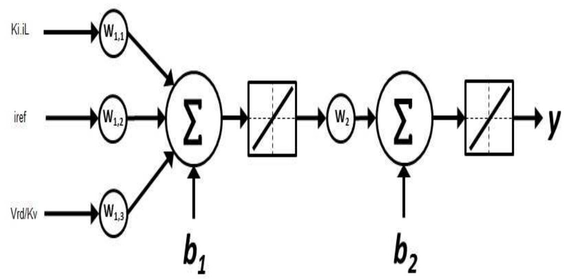


Figure 4: Network used for the identification

After training, a linear neurons network is found which output d(t) is as follow:

$$d(t) = w_2\left(w_{1,1} Ki.i_L(t) + w_{1,2}.i_{ref}(t) + w_{1,3}\frac{V_{rd}(t)}{Kv} + b_1 \right) + b_2 \quad (6)$$

- $w_{i,j}$: synaptic weight of the j^{th} input of the i^{th} neuron
- b_1 : first neuron bias
- w_2 : weight connection between the two neurons
- b_2 : second neuron bias

The relation (6) may be re arranged as:

$$d(t) = w_{iL}.i_L(t) + w_{iref}.i_{ref}(t) + w_{red}\frac{V_{rd}(t)}{Kv} + B \qquad (7)$$

When these different coefficients (weights) are identified, the relation (7) can be translated into analogical diagram using resistances and operational amplifiers as showed by Figure 5.

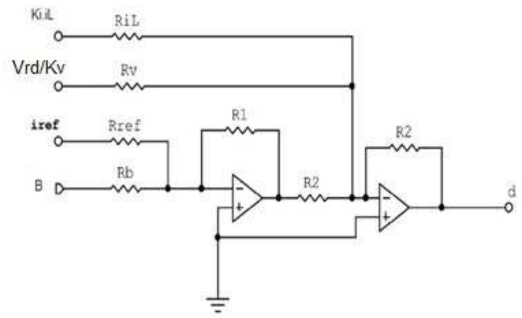


Figure 5: General scheme for a realization

2.1.1 Results

Simulation was made to test the system behaviour following a set value variation and application disturbance materialized by the variation of the load resistance. The total harmonic distortion (THD %) is also used as criterion of appreciation. The results are summarized in figures 6 and 7.

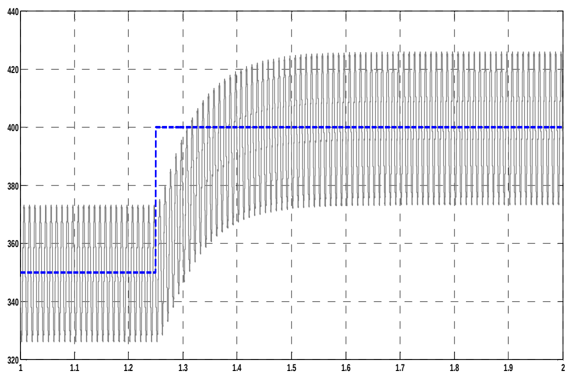


Figure 6: Output Vs response with set point variation

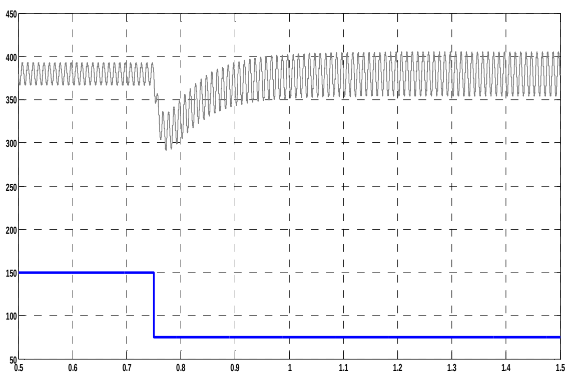


Figure 7: Behaviour by applying load disturbance

In figure 6, the set value is changed from 350[V] to 400[V] and in figure 7, the resistance load R is reduced for its half value (here at t = 0,75[s], R = 30 [Ω] ↓ R=15[Ω]). After Δt = 0,15 [s], the set value is reached and the steady state established.

The input current is sinusoidal and in phase with the input voltage as shown in figure 8. The THD is 5 %.

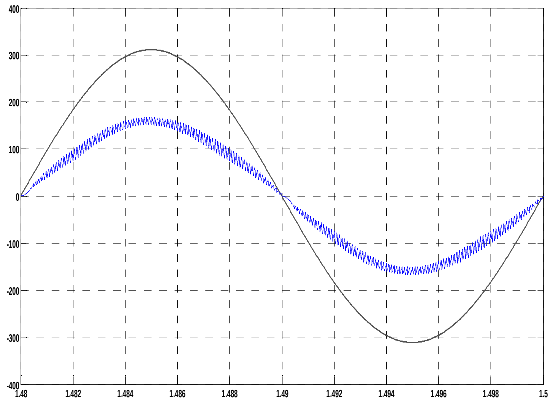


Figure 8: Input current and voltage curves.

2.1.2 Discussion

- The performance of the neural network increases with the number of examples used.
- The ratio of the chopping frequency and the line frequency has effect of the THD. The ratio

  $k = \frac{F_h}{F_L} < 1000$ is adopted ($F_L = 50$ [Hz])
- Changing the weight w_{red} has no effect on the THD. It has effect on the speed regulation.
- More the bias B moves away from the peak of the saw tooth signal, more the THD increases.
- When the absolute value of w_{iref} increases more the THD [%] decreases however, the CS commutation must be taken into account.

This phenomenon is shown in figure 9.

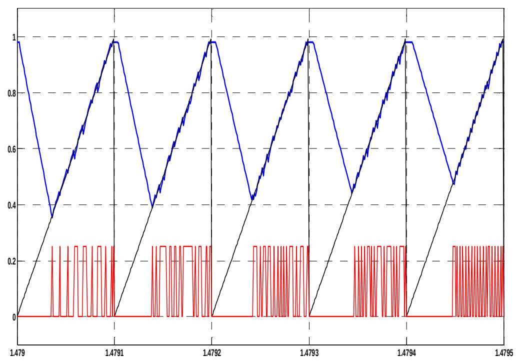


Figure 9: Different signal curves with d(t): blue saw tooth : black u(t): red

In all simulations made, the expression of d(t) according the relation 8:

$$d(t) = -3.K_i.i_L(t) + 5.i_{ref}(t) + \frac{4}{3}.\frac{V_{red}(t)}{Kv} + 12 \qquad (8)$$

Ki and Kv depend of the system.

3. CLASSICAL BOOST PFC

The basic scheme is always given by figure 1. In this case, a hysteresis command is applied for the current loop and an adaptive neuronal controller working as a discrete PI controller for the loop voltage. First, a classic PI controller is applied for the loop voltage. The method consists to design this adaptive neuronal controller by using the measures obtained when classic PI controller is used. It is assumed that the current loop is perfect when this hysteresis command is applied.

3.1. Current loop

The command using sinusoidal band hysteresis is chosen. The current i_L = ired crossing the inductor L is forced to stay in band ±ΔI around the reference as shown in figure 10.

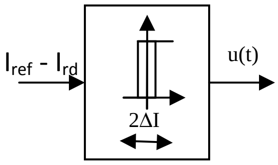


Figure 10: Hysteresis command

Chopping frequency F_H calculation is largely presented in literatures (Razafinjaka 2013), (Feld 2003), (Multon 2003). Its expression is:

$$F_H = \frac{V_M (V_s - V_M)}{2.L.\Delta i.V_s} \qquad (9)$$

L, the inductor, V_M, RMS value of input voltage, Vs the output voltage and Δi , the current bandwidth

The inductor L must be designed to respect the chopping frequency F_H according the values of V_M, Vs and Δi. Figure 11 shows curves F_H vs L (V_M= 220V and Vs =400V)

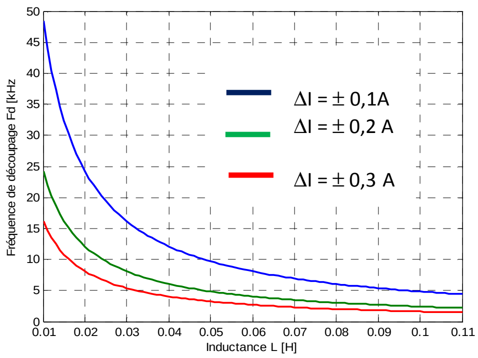


Figure 11: Curves F_H vs L

3.2. Voltage loop

It is assumed that the current loop is perfect (iref = ired). . It is there possible to adopt the following approximation obtained by modelling by assessment of power (Feld 2009), :

$$\frac{V_s(p)}{I_{red}(p)} \simeq \frac{V_s(p)}{I_{ref}(p)} \qquad (10)$$

Hence,

$$\frac{V_s(p)}{I_{ref}(p)} = \frac{V_M}{4.V_s} . \frac{R}{1 + p\frac{RC}{2}} \qquad (11)$$

The opened loop is defined by a first order transfer function.

$$G(p) = \frac{K}{1 + pT} \qquad (12)$$

With $K = \frac{V_M . R}{4 . V_S}$ and $T = \frac{R\ C}{2}$

A PI regulator is sufficient to control such system. Its function transfer may be expressed like followed:

$$G_R(p) = \frac{1 + pA.T_i}{pT_i} \qquad (13)$$

The gain A and the constant time Ti can be determined by imposing a frequency Fc for the closed loop. The transfer function for opened loop Go(p) is:

$$G_0(p) = G_R(p).G(p) = \frac{K}{pT_i} \qquad (14)$$

By using method compensation:

$$A.Ti = T \qquad (15)$$

The transfer function for closed loop is:

$$H(p) = \frac{Go(p)}{1 + Go(p)} = \frac{1}{1 + p\frac{T_i}{K}} \qquad (16)$$

Imposing frequency Fc according the relation (16) gives Ti and then the gain A by relation (15).

Figures 12, 13 and 14 show the simulation results when frequency at closed loop are Fc= 5 [Hz] and Fc=15[Hz]

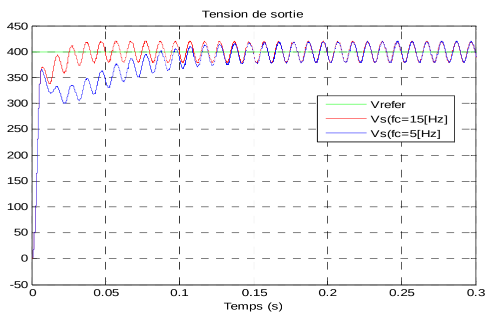


Figure 12: Curves giving Vs at Fc=5[Hz]- Fc=15[Hz]

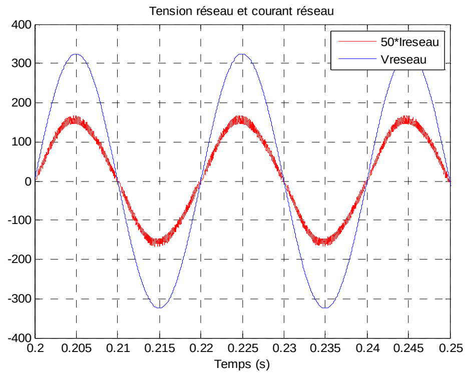


Figure 13: Current and voltage input at Fc = 5[Hz]

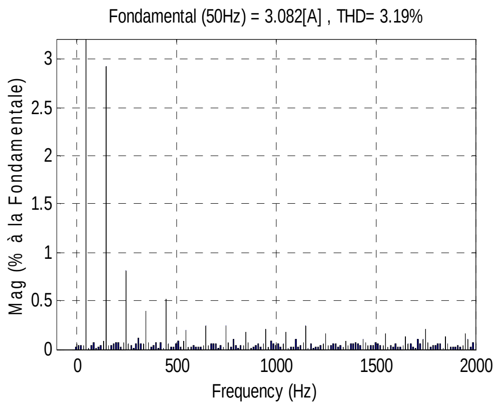


Figure 14: Spectrum current analysis at Fc = 5[Hz] THD= 3,19%

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 %

In all that follows, the results obtained with Fc =5 [Hz] are chosen.

3.3. Neuronal controller for the loop voltage

As said below, an adaptive neuronal controller working as a PI controller is designed using the results obtained by the classic PI controller. In this context, an ADALINE network is chosen for the identification of the parameters using retro propagation algorithm with direct action. A structure of such network is given in figure 15.

The objective consists to identify the different values of the synaptic weights and bias. These values are stored as vectors to determine the two parameters K and Ki for the discrete PI controller (Nampoina 2010).

$$G_R(z) = K + \frac{Ki}{z-1} \qquad (17)$$

Figure 16 shows the structure of the adaptive neuronal controller built around the values of the weights and bias.

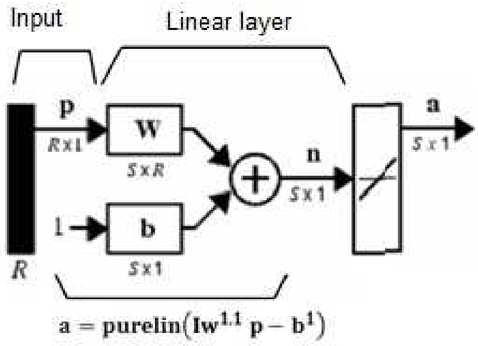


Figure 15: Structure of an ADALINE network

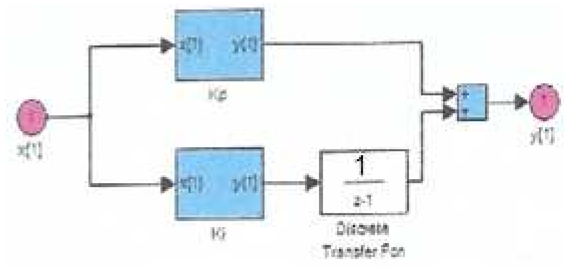


Figure 16: Adaptive neuronal controller implantation

3.3.1 Results

The values of different weights and bias are stored and used in all simulations. The results are presented in Figures 17, 18, 19 and 20.

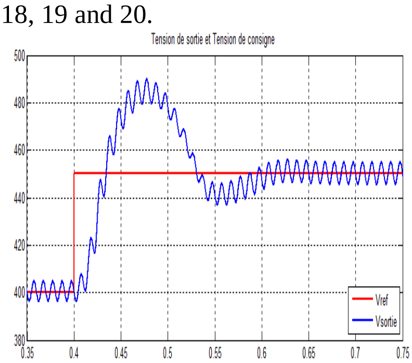


Figure 17: Response Vs with set point variation

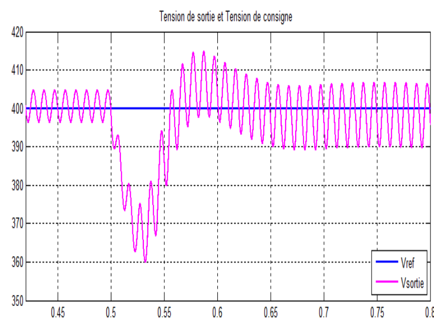


Figure 18: Behavior with load disturbance

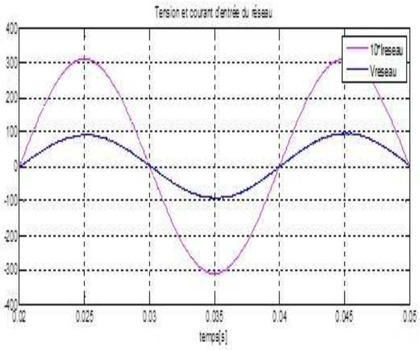


Figure 19: Current and voltage curves

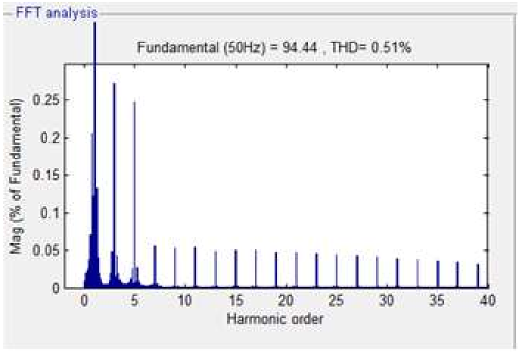


Figure 20: Spectrum current analysis

It is here highlighted that an overshoot D1= 8,9% appears with set point variation (Vsc = 400V →450V). At t = 0,5 [s], a load disturbance is applied (R = 328 [Ω] → R = 164 [Ω]), the steady state is reached after Δt = 0,2 [s].

The regulation performances regulation is less than the classic one but the THD is largely improved. By the same method of calculation, here the total harmonic distortion is THD = 0, 51%.

3.4. Discussion

At first time, the training has generated N = 2, number of hidden layers. Several simulations are then adopted to find an optimal value of this number N by taking THD % as criterion. Figure 21 gives the variation of the THD vs N.

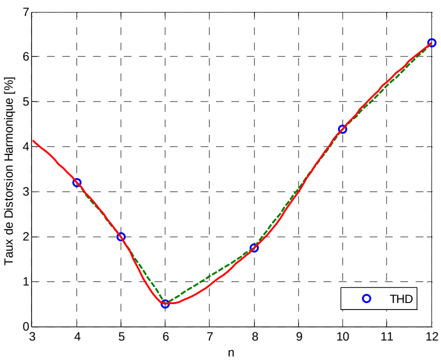


Figure 21: THD (%) vs. the number N

An optimal value $\hat{N} = 6$ is found and applied in the identification of the parameters.

CONCLUSION

In this paper, two kinds of neuronal controller are proposed. For the ACCM Boost PFC, the command d(t) is directly built for the current loop. For the classical Boost PFC, the hysteresis command is applied for the current loop and an adaptive neuronal controller is chosen for the voltage loop. Simplicity of realization are taken into account in both cases. The simulations show the feasibility of the methods and the efficiency of the neuronal controllers.

REFERENCES

Abdel-Rahman, N.A., 2013. CCM PFC Boost Design, Design Note DN 2013-01, *Infineon Technologies North America Corp.*

Benaïssa A., 2006. Commande par mode de glissement d'un convertisseur AC-DC avec correction de facteur de puissance. *Acta Electrotehnica, volume 47.Number 2.* pp 67-72.

Dixon L. 1990. Average current mode control of switching power supplies, *Unitrode Switching Regulated Power Supply Design Seminar.* USA

Feld Gilles. 2009. Etude et simulation d'un convertisseur AC/DC à absorption sinusoïdale de courant. *Publication ENS Cachan.*

Keraï S. 2003. Calcul du convertisseur AC-Dc avec correction du facteur de puissance. *Publication LMER.* Département d'Electronique. Université Abdou Bakr Belkaïd. Alger

Multon et al. 2003. Le redresseur MLI en absorption sinusoïdale de courant. *Revue ENS Cachan.* Antenne de Bretagne.

Nampoina R. 2010. Neuronal command for boost pfc-Implementation on FPGA. *Travail de Mémoire pour l'obtention du diploma d'ingéniorat de l'Ecole Supérieure Polytechnique d'Antsiranana*, Université d'Antsiranana, Madagascar.

Rajaonah R. 2014. Commande neuronale en courant moyen d'un boost PFC. *Mémoire de DEA. Ecole Supérieure Polytechnique d'Antsiranana.* Université d'Antsiranana, Madagascar.

Razafinjaka et al 2013. Non Linear Control for AC-DC Converter with Power Factor Correction, *Paper accepted in . 3rd IFAC Symposium on Telematic Automation.* Yonsei University. Seoul Korea

Singh B. 2003. A review of Single-Phase Unity Power Qualiy AC-DC Converters. *IEEE Trans. Ind. Electronics Vol-50.*

Tédjini H. et al 2008. Contrôle non linéaire avancé du redresseur MLI triphasé en absorption sinusoïdale de courant. *ACTA Electrotehnica. Vol 49-N° 3 pp 231-300.*

AN INTEGRATED SIMULATION MODEL FOR SITE LAYOUT PLANNING OF TUNNELLING PROJECTS

SeyedReza RazaviAlavi[(a)], Simaan AbouRizk[(b)]

[(a)]Ph.D. Candidate, Dept. of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G7
[(b)]Professor, Dept. of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G7

[(a)]reza.razavi@ualberta.ca, [(b)]abourizk@ualberta.ca

ABSTRACT
Overlooking site layout in the planning phase of construction projects leads to loss of productivity and incurs extra costs. In tunneling projects, site layout has a significant impact on material flow and tunneling operations, particularly on congested sites. In addition, construction planning decisions can influence the efficiency of the layout. This paper proposes simulation as a decision making tool to model tunnel construction operations and site layout, and capture their mutual influences. To facilitate building the simulation model, even for users with limited simulation knowledge, a special purpose simulation (SPS) tool was customized and developed. This simulation tool provides an integrated environment to model the parameters of different disciplines including site layout, material procurement, tunnel operations and logistics. The developed tool is of great assistance for the planners to make decisions simultaneously on site layout and other construction planning parameters, and find the most cost-efficient plan.

Keywords: special purpose simulation, tunnel construction, site layout planning, decision making tool

1. INTRODUCTION

Site layout planning, the process of identifying the required type of temporary facilities and determining their size and location, has been studied in the past due to its significant impacts on project productivity, time and cost. Most of these studies, e.g. Zhang and Wang (2008), attempted to improve the location of the facilities by optimizing the sum of weighted distance function ($\sum w \times d$), which strives to minimize transportation costs between facilities. Some studies, e.g. Elbeltagi et al. (2004), used the same function to subjectively optimize the location of facilities by defining qualitative rates assigned to the interaction and closeness constraints between the facilities.

However, this function does not realistically model the material, workers and equipment flow, and the interaction between facilities. Overlooking these important factors leads to inefficiency of the site layout in practice. Simulation can address this drawback by modeling complex construction processes and interactions between facilities. Alanjari, et al. (2014) showed the advantage of simulation over the sum of weighted distance function (SWDF) to reduce the transportation time in material layout planning. They demonstrated that resource interaction, an important factor, is ignored in SWDF, but simulation can consider it in modeling the material handling process to plan more efficient layouts. Tommelein (1999) developed one of the first simulation-based models for planning location and the number of tool rooms in construction projects. Azadivar and Wang (2000) integrated simulation with genetic algorithm (GA) for facility layout planning in the manufacturing industry to minimize transportation time. For stock yard layout planning of precast concrete products, Marasini et al. (2001) also used simulation integrated with GA. Simulation was also utilized for sizing temporary facilities in construction site layout planning (RazaviAlavi and AbouRizk, 2014).

Despite the proven advantages of simulation in site layout planning, its full potential has not been employed in this domain. Aleisa and Lin (2005) believe that two schools of thought, “layout then simulation,” and “simulation then layout,” have been followed for using simulation in site layout planning. The first approach is time efficient and used when the production strategies are predetermined, the stochastic behaviors of the system are insignificant at the early stage of layout planning, and/or the objective is to minimize the travel distance (Aleisa and Lin 2005). The latter approach results in more realistic and efficient layouts to improve throughput levels, and it is more applicable when stochastic demands or complex interactions in the system are significant, operational parameters should be justified prior to layout planning, and/or the objective is minimizing flow congestions (Aleisa and Lin 2005). Both explained approaches isolate decision making on construction planning parameters from site layout parameters while those parameters have mutual impacts. For instance, when the site is congested and limited space exists for storing materials, material delivery decisions should be made to prevent space shortage on the site. On the other hand, decisions on the number of employed crews can increase the production rate, and consequently the consumption rate of the material, which reduces the need for material storage space (size). These dependencies and mutual impacts bring about a new approach that can integrate construction planning and site layout planning, and simultaneously make decisions on those influencing parameters.

Integrating these parameters is critical in tunneling projects, particularly on congested sites. In tunneling projects, the location of some facilities, e.g. material storage areas, affects material transportation time, which is one of the main drivers of project productivity. In addition, the production rate impacts the size of material storage areas. The limited space for these facilities on tunneling sites can influence construction operation decisions, material procurement and logistic plans. These interdependencies highlight the need to consider all influencing parameters in a unified model. As discussed earlier, simulation can provide this integrated environment for modeling purposes. In this study, a special purpose simulation (SPS) tool is developed to model the tunnel site layout and construction operations, along with the pertinent parameters from different disciplines, such as material procurement and logistics. This tool facilitates modeling and is able to examine various scenarios and provide users with comprehensive results to make decisions.

In this paper, first, the application of simulation in modeling tunneling projects is described. The significance of tunneling site layout is then analyzed in detail. The last sections outline the developed SPS and its implementation in a case study, followed by a summary and conclusion.

2. SIMULATING TUNNEL CONSTRUCTION PROCESSES

Due to the complexity, uncertainties and randomness inherent in construction projects, simulation was found to be an effective tool to model, analyze and improve the performance of construction operations (Mohamed and AbouRizk 2006) and has been used in different sectors of construction projects. CYCLONE (Halpin 1973) was one of the earliest tools developed for simulating construction projects. STROBOSCOPE (Martinez and Ioannou 1994) and Simphony (Hajjar and AbouRizk 1996) are programmable and more flexible simulation tools, primarily used in the last two decades.

Due to the repetitive nature of tunnel construction activities and the inherent uncertainties such as the soil type and equipment reliabilities, simulation has been widely used to model, plan, and estimate the time and cost of tunneling projects. Studies by Touran and Asai (1987), Tanaka (1993) and AbouRizk et al. (1997) were among the first notable attempts to simulate the tunneling process. Different aspects of tunnel projects were incorporated in the simulation model in recent years. Ruwanpura and AbouRizk (2001) tried to predict soil transition in tunneling. Ebrahimy et al. (2011a) modeled supply chain management in tunneling using simulation. They substantiated that size of the concrete segment storage can affect the project time. Optimizing the closeness constraints using GA, Zhou et al. (2009) tried to find the optimum layout in tunneling projects. They used simulation to examine the efficiency of the enhanced layout from the optimization. Despite the contribution of this research, it did not consider the influence of material storage size on the project time, proven by Ebrahimy et al. (2011a).

Developing simulation models is not a trivial task due to the requirement for knowledge of the technical domain of the real system, simulation modeling techniques and computer programming (Mohamed and AbouRizk 2006). To overcome these challenges, SPS has been developed to facilitate building simulation models and promote the application of simulation in the industry. SPS was customized for different types of construction projects such as earth moving (Hajjar and AbouRizk 1996, Siadat and Ruwanpura 2013), aggregate production plants (Hajjar and AbouRizk 1998), construction site dewatering (Hajjar et al. 1998), supply chain (Petrovic 2001, Ebrahimy et al. 2011b), industrial fabrication (Sadeghi and Robinson Fayek 2008), construction noise prediction (Gannoruwa and Ruwanpura 2007), and bridge construction (Marzouk et al. 2008).

For simulating the tunneling process, an SPS tool was developed by AbouRizk et al. (1999) using the Simphony platform. The current version of this tool has been developed in Simphony.NET 4.0 with some modifications, and designed for modeling projects executed by tunnel boring machines (TBM). This tool can model three main activities: working shaft and retrieval shaft construction, tail tunnel and undercut construction, and tunnel construction. The working shaft is for equipment, crew and segment access and removing the dirt from the tunnel, while the removal shaft is for recovery of the TBM at the end of the tunnel. The shaft can be either circular or rectangular. Excavation and lining are the main activities in shaft construction. Undercut and tail tunnel are located adjacent to the working shaft and retrieval shaft, respectively, for providing more room for moving or setting up equipment. See Zhou et al. (2008) for more information on shaft, tail tunnel and undercut construction. In tunnel construction, the TBM excavates the soil and fills the muck cars with dirt. The cars transport the dirt to the working shaft, and generally a crane hoists the cars to empty them in the spoil pile. Then, the crane loads the cars with the concrete segments to be transported to the TBM for the next cycle. Meanwhile, lining the tunnel, resetting the TBM, surveying, and rail track extensions, when needed, are performed in the tunnel. See Ruwanpura et al. (2001) for further details on simulating tunnel construction.

Figure 1 depicts the overview of the current version of the tunneling SPS tool and its different elements. Each element has its own properties, which are the user inputs for specifying the characteristics of the tunnel. Table 1 shows the main inputs of the tool elements. For more flexibility of the tool to model different types of tunnels and activities, some simple elements exist inside of some elements, such as the shaft element that can model the user-defined activities, shown in Figure 1. The graphical interface of this tool is user-friendly and intuitive and a user with limited knowledge of simulation can easily build the model. In the next section, the significance of the site layout plan in a tunneling project is described.

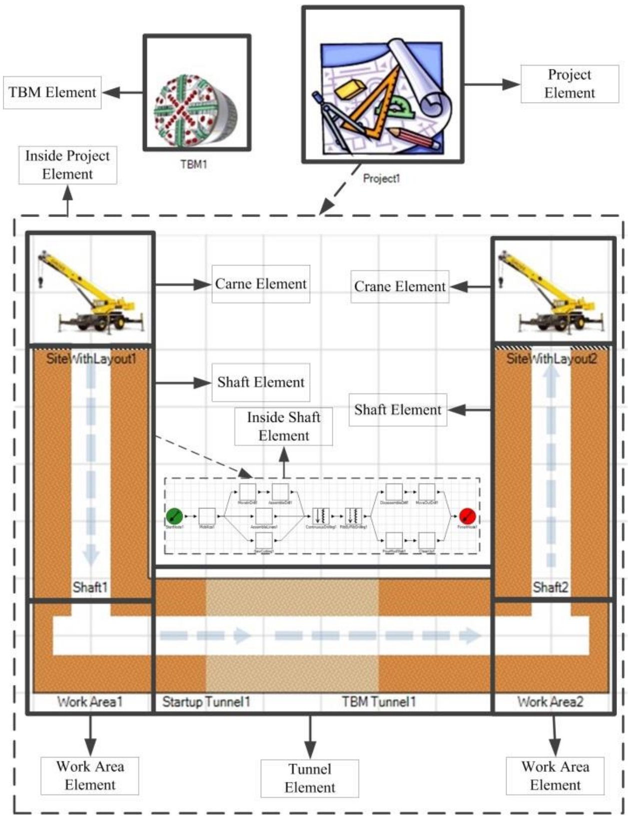


Figure 1: Overview of the Tunneling SPS Tool

Table 1: Tunneling SPS Inputs

Element	Inputs
TBM	Dimensions, resetting duration, and reliability
Crane	Reliability
Shaft	Dimensions and shape, soil spec., and flexible activities for excavation
Work Area	Geometry and dimensions, soil spec., and flexible activities for excavation, train and car spec.
Tunnel	Tunnel length, soil spec., and activity durations and plans

3. THE SIGNIFICANCE OF SITE LAYOUT IN TUNNELING PROJECTS

As discussed earlier, in site layout planning, three attributes of the temporary facilities: type, size and location, are determined. In tunneling projects, the type of facilities include, but are not limited to, the shaft, hoisting equipment (e.g. crane), spoil pile, the segment storage area, the crew trailer (office), and the electrical facilities for supporting the TBM. Among these facilities, the size of the shaft, hoisting equipment, crew trailer and electrical facilities is fixed and predetermined, while the size of the spoil pile and segment storage area is variable and should be determined based on the flow of the dirt and segments, respectively, in the project. To show the flows of these materials and identify their influencing factors, as well as the effect of these facility sizes on construction processes, a causal loop diagram (Sterman 2000) is used. In this diagram, arrows link independent variables to dependent variables and polarities of the arrows (positive or negative) demonstrate how independent variable changes affect the dependent variable (Sterman 2000).

In the dirt flow diagram exhibited in Figure 2, the dirt volume in the spoil pile is the main variable. Since the dirt comes from the TBM excavation, the production rate of the TBM influences the dirt inflow. The dirt is generally removed from the site by trucks. The capacity and the number of the trucks influence the dirt outflow. Since a loader is employed to load the truck, the availability of the loader is another driver of the dirt outflow. The size (capacity) of the spoil pile determines how much dirt can be stored in it. If the available dirt reaches the capacity of the spoil pile, the dirt can no longer be offloaded into the spoil pile. Consequently, lack of space in the spoil pile halts the excavation until the dirt is removed and enough space is available in the spoil pile.

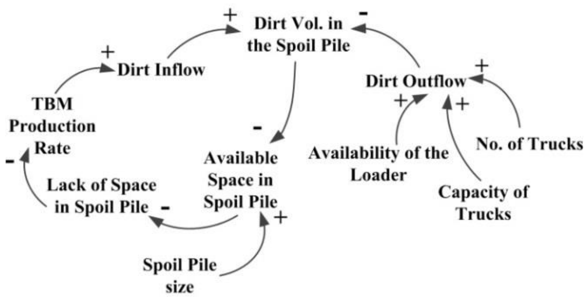


Figure 2: Dirt Flow and its Influencing Factors

For the concrete segment flow shown in Figure 3, the available number of segments in the storage is the main variable. The segments are delivered to the site from a supplier. The size of the incoming segment batches and their inter-arrival time influence the segment inflow. The segments are consumed in lining the tunnel, which depends on the TBM production rate (TBM production rate influences the segment outflow). On the other hand, segment stock-out halts the project because the TBM cannot progress without lining. The size of the segment storage should be considered in making decisions on the size and frequency of the incoming segment batches. If the capacity of the segment storage is full, no more segments can be delivered to the site. It incurs extra costs to the project to resolve space shortage, for example, by providing an off-site storage or delaying the incoming segment batches.

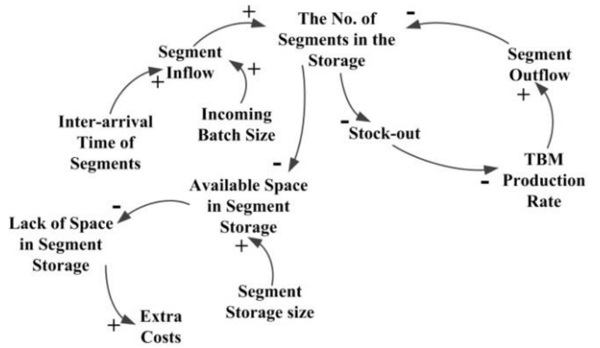


Figure 3: Concrete Segment Flow and its Influencing Factors

Integrating Figure 2 and 3, the complexity and interdependency of the influencing factors in tunneling material flow is observed in Figure 4. It is also shown in Figure 4 that these factors are pertinent to different planning disciplines including site layout, tunneling operations, logistics and material procurement. All these factors and their complex interdependency are sophisticatedly modeled in an integrated simulation environment and their impacts on project cost are estimated.

The location of four facilities: shaft, crane, spoil pile and segment storage, can impact the project time. The closeness of these facilities reduces the transportation time of the dirt and segments. Generally speaking, these durations are more critical in determining the total time of long tunnel construction projects. Thus, it is important to optimally determine where to position these facilities, while the position of the shaft is mostly predetermined on the site. Simulation can measure the effects of these facility positions on the project time and cost. The position of other facilities does not directly affect the project time. Those facilities occupy space on the site, and their positions depend on closeness constraints or user preferences. For instance, planners often prefer to position the crew trailer close to the gate, or the closeness constraints specify that the electrical facilities should be close to the shaft. A general constraint for all facilities is that they should be located inside the site boundaries and should not have any overlaps.

It should be emphasized that size and location of some facilities also have mutual influences. The location of the four above-mentioned facilities influences the production rate of the project, which is the main driver of the size of the spoil pile and the segment storage. In addition, in positioning facilities, their sizes should be considered to avoid overlapping of facilities. In particular, on congested sites, the size of the facilities may be adjusted to be fitted for positioning in a certain location.

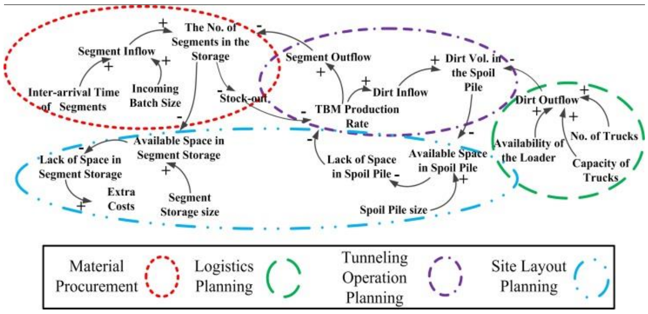


Figure 4: Integration of Dirt Flow and Concrete Segment Flow

4. SPS FOR TUNNEL SITE LAYOUT PLANNING AND CASE STUDY

The SPS for planning the site layout is developed in Simphony and nested in the current version of the tunneling tool to keep the integrity of the site layout tool with Simphony's existing tools. The site layout tool includes a site element, for which size should be determined, and the facility elements, which are dragged and dropped to the site, and are movable. As discussed earlier, the positions of four facilities (i.e. shaft, crane, spoil pile and segment storage) and the size of the spoil pile and segment storage affect the simulation of projects. That is, these facilities have predefined elements in the tool with specific functionalities. Other facilities that do not have simulation roles (e.g. a crew trailer and electrical facilities), use a unique element: "miscellaneous facility" element. For these facilities, the user should determine only their size and position. Table 2 shows the main properties of these elements. To examine the effect of the designed spoil pile and segment storage size on the project time and cost, the user is given an option to select the capacity of these facilities as unlimited and compare the results with the limited capacity. Ultimately, the integrity check of the model is performed once the user wants to execute it, and is done manually, or when the user wants to run the model, and it is done automatically. The main items checked through the integrity check process are as follows:

- Existence of shaft, crane, spoil pile and segment storage on the site.
- Non-overlapping constraints of facilities.
- Being inside the site boundary constraints.

This tool provides the user with comprehensive result reports including tables and charts that intuitively give the user perceptions on the main parameters measured in simulation. These reports help the user make decisions on site layout and other parameters. The major decision-making factor for site layout is the project cost, which is also estimated by simulation. This template is capable of analyzing stochastic input data with diverse types of distributions and running Monte Carlo simulation. The results are accessible for multiple runs in the form of statistical results as well as results for each individual run. Table 3 presents a summary of the simulation tool outputs. An overview of the tool and samples of these reports are demonstrated in a case study.

Table 2: Main Properties of Site Layout Elements

Element	**Properties**
Site	Dimensions and scale
Shaft	Size, shape and location
Crane	Size, location, loading, unloading and hosting durations, and swing speeds
Spoil pile	Size, location, capacity, initial vol. of dirt, and truck capacities, loading travel durations, and truck and loader costs
Segment storage	Size, location, capacity, initial vol. of Segments, size and inter-arrival of segment delivery, and extra storage costs
Misc. facilities	Size and location

Table 3: Site Layout tool Outputs

Output data	**Data format**
Equipment, labor, and rental cost report	Table
Project delays caused by lack of space in spoil pile	Chart and Table
Project delays caused by segment stock-out	Chart and Table
Fullness of spoil pile and Segment storage	Chart and Table
Crane utilization	Chart and Table
Loader utilization	Chart and Table
Truck idle time caused by unavailability of the loader	Chart and Table
Truck idle time caused by unavailability of the dirt	Chart and Table

5. CASE STUDY

In a tunneling project with a length of 1030 meters, two different layouts: Layout (A) and Layout (B) were designed based on the geometry of the site as illustrated in Figure 5 and 6, respectively. Figure 5 and 6 are the snapshots of the tool user interface depicting an overview of the tool. In Layout (A), the spoil pile size is smaller and its distance to the shaft is slightly more than those of Layout (B). In turn, the size of the segment storage and its distance to the shaft in Layout (B) is more than those of Layout (A). In addition to selection of a suitable layout, decisions should also be made on the size of the trucks deployed for removing the dirt, and the frequency of segment deliveries and quantity of the segments in each delivery. The planner of this project can opt between two types of trucks with 10 m^3 and 12 m^3 capacities. Choosing the larger truck incurs more hourly costs, and reduces the risk of the spoil pile over filling. The planner also has two options for supplying segments: 9 segment batches per day or 16 segment batches over two days (each batch includes 4 segments), which have identical costs. However, if the segment storage does not have enough capacity for storing the incoming segments, they are stored off-site, which incurs fixed cost for transportation and daily cost for maintaining that segment batch. As discussed earlier, all these variables are interdependent and can influence one another. This case study aims to determine the most cost efficient plan from the possible scenarios briefly presented in Table 4.

Based on the characteristics of the project, the tunneling process was simulated using the developed tool. The duration of most of the activities, such as excavation rate, rail tack extension, and surveying, was modeled stochastically to account for project uncertainties. Cost data was also incorporated to evaluate the efficiency of the scenarios. Having run the model multiple times, it was revealed that scenario #5 had the minimum total cost on average, as shown in Figure 7. Figure 7 also shows the cost distribution between tunneling costs including equipment and labor costs for tunnel construction, truck and loader costs for removing the dirt, and extra storage costs for the off-site segment storage, if any.

The most cost efficient scenario (#5) has a large spoil pile size, deploys the large truck and orders segments more frequently. Investigating the results indicates that deploying a small truck (10 m^3) is not efficient, and the scenarios with the small truck (#3, #4, #7 and #8) have the highest costs. Among these scenarios, the costs of the scenarios with smaller spoil pile size are higher, because a full the spoil pile halts the tunnel construction process, which entails more tunneling costs due to idleness of the resources. Hence, the extra costs of the large truck are compensated by completing the project earlier. These results confirm that the spoil pile size and decisions on the logistics (i.e. truck size) are dependent and have a significant influence on the tunneling project cost. Modeling the construction process along with site layout and logistics to capture their influences is crucial.

By deploying the large truck, the tunneling process is executed with higher rates, and the demand of the segments becomes higher. As a result, more frequent segment deliveries are desirable in this project. Although this decision incurs extra storage cost, it reduces the risk of segment stock-out, which would lead to delays in the project. To highlight the importance of this decision, scenario #5 and #6 are compared. All the specifications of these two scenarios are identical, except for the segment delivery plan. The results of the model show that the total delay time caused by lack of segments for scenario #5 and #6 are 57 and 289 hours, respectively, which leads to saving $136,507 in scenario #5 as compared to scenario #6.

Similar analysis and comparisons between different aspects of the project performance can be carried out for each scenario using the comprehensive reports of the tool.

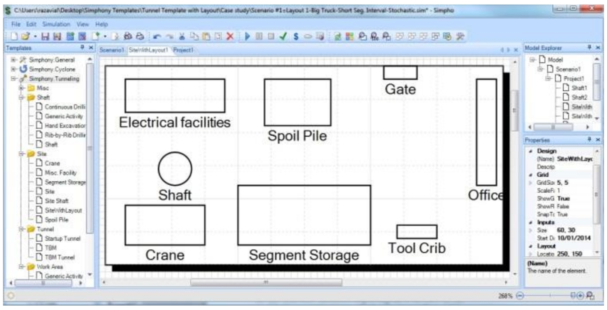


Figure 5: Layout (A)

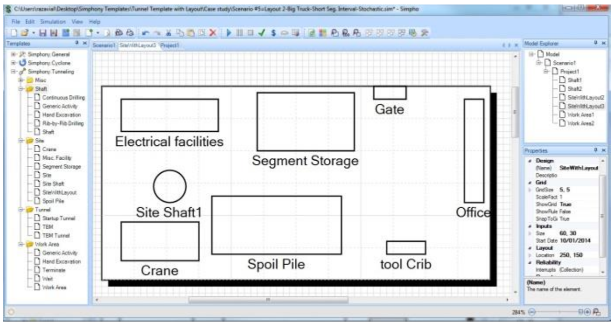


Figure 6: Layout (B)

Table 4: Specifications of the examined scenarios

Scenario	Layout type	Segment Delivery	Truck Capacity
#1	A	9/day	12 m3
#2	A	16/2 days	12 m3
#3	A	16/2 days	10 m3
#4	A	9/day	10 m3
#5	B	9/day	12 m3
#6	B	16/2 days	12 m3
#7	B	16/2 days	10 m3
#8	B	9/day	10 m3

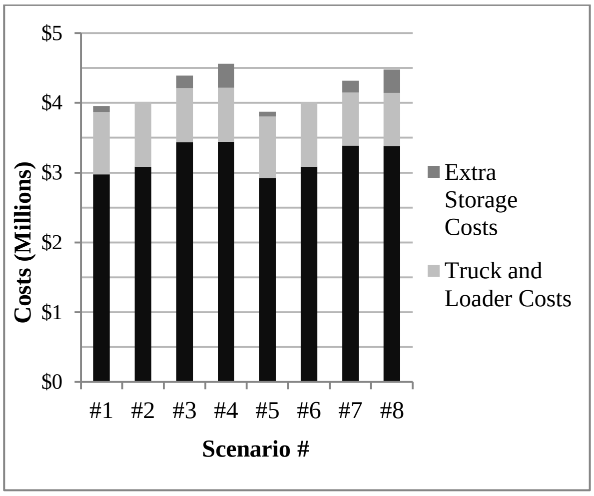


Figure 7: Cost results of the examined scenarios

6. SUMMARY AND CONCLUSION

This paper demonstrated the significance of the site layout plan in tunneling projects. The mutual impacts of site layout parameters, i.e. facility size and location, and construction planning parameters from different disciplines were analyzed and modeled through an integrated simulation environment. To promote the practicality of the simulation tool in the industry, a user-friendly SPS tool for tunneling site layout planning was developed. This tool complements the existing tunneling simulation tool, which models only tunnel construction operations.

The result of this research shows that decisions on construction plan, material procurement, logistics and site layout are dependent in tunneling. Ignoring this dependency leads to loss of productivity and inefficiency of the site layout, which further substantiates the merit of the research. The main contribution of this research is to integrate interdependent parameters from different disciplines implementing simulation to obtain the most cost-efficient construction plan for tunneling projects. The comprehensive and intuitive reports of the simulation model on the project cost and project delays along with other aspects of the project performance are of great assistance for planners to make complicated decisions. This approach could also be adopted for site layout planning of other types of construction projects, and similar tools could be produced.

REFERENCES

AbouRizk, S., Manavazhi, M. and Dozzi, P., 1997. Analysis of productivity on a tunnel operation using computer simulation. *Proceedings of Construction Congress V: Managing Engineered Construction in Expanding Global Markets*, pp. 382-388. Minneapolis, Minnesota.

AbouRizk, S. M., Ruwanpura, J. Y., Er, K. C. and Fernando, S., 1999. Special purpose simulation template for utility tunnel construction. *Proceedings of the 1999 Winter Simulation Conference*, pp. 948-955. Phoenix, Arizona.

Alanjari, P., Razavialavi, S. and AbouRizk, S., 2014. A simulation-based approach for material yard laydown planning. *Automation in Construction*, Volume 40, pp. 1-8.

Aleisa, E. E. and Lin, L., 2005. For effective facilities planning: layout optimization then simulation, or vice versa?. *Proceedings of the 2005 Winter Simulation Conference*. Orlando, pp.1381-1385 FL, USA

Azadivar, F. and Wang, J., 2000. Facility layout optimization using simulation and genetic algorithms. *International Journal of Production Research*, pp. 4369-4383.

Ebrahimy, Y., AbouRizk, S. M., Fernando, S. and Mohamed, Y., 2011a. Simulation modeling and sensitivity analysis of a tunneling construction project's supply chain. Engineering, *Construction and Architectural Management*, 18(5), pp. 462-480.

Ebrahimy, Y., AbouRizk, S. M., Fernando, S. and Mohamed, Y., 2011b. Simphony Supply Chain Simulator: a simulation toolkit to model the supply chain of construction projects. *Simulation*, 87(8), pp. 657-667.

Elbeltagi, E., Hegazy, T. and Eldosouky, A., 2004. Dynamic layout of construction temporary facilities considering safety. *Journal of Construction Engineering and Management*, 130(4), pp. 534-541.

Gannoruwa, A. and Ruwanpura, J. Y., 2007. Construction noise prediction and barrier optimization using special purpose simulation. *Proceedings of the 2007 Winter Simulation Conference*, pp. 2073-2081. Washington, DC.

Hajjar, D. and AbouRizk, S. M., 1996. Building a special purposes simulation tool for earth moving operations. *Proceedings of the 1996 Winter Simulation Conference*, pp. 1313 – 1320. Coronado, CA.

Hajjar, D. and AbouRizk, S. M., 1998. Modeling and analysis of aggregate production operations. *Journal of Construction Engineering and Management*, 124(5), pp. 390-401.

Hajjar, D., AbouRizk, S. and Xu, J., 1998. Construction site dewatering analysis using a special purpose simulation-based framework. *Canadian Journal of Civil Engineering*, 25(5), pp. 819-828.

Halpin, D. W., 1973. *An investigation of the use of simulation networks for modeling construction operations*, Urbana, Illinois: Ph.D. dissertation, University of Illinois.

Marasini, R., Dawood, N. N. and Hobbs, B., 2001. Stockyard layout planning in precast concrete product industry: a case study and proposed framework. *Construction Management and Economics*, Volume 19, pp. 365-377.

Martinez, J. C. and Ioannou, P. G., 1994. General purpose simulation with stroboscope. *Proceedings of the 1994 Winter Simulation Conference*, pp. 1159-1166. San Diego, CA.

Marzouk, M., Said, H. and El-Said, M., 2008. Special-Purpose Simulation Model for Balanced Cantilever Bridges. *Journal of Bridge Engineering*, 13(2), pp. 122-131.

Mohamed, Y. and AbouRizk, S. M., 2006. A hybrid approach for developing special purpose simulation tools. *Canadian Journal of Civil Engineering* , 33(12), pp. 1505-1515.

Petrovic, D., 2001. Simulation of supply chain behaviour and performance in an uncertain environment. *International Journal of Production Economics*, 71(1), pp. 429-438.

RazaviAlavi, S., AbouRizk, S. and Alanjari, P., 2014. Estimating the Size of Temporary Facilities in Construction Site Layout Planning Using Simulation. *Proceedings of Construction Research Congress 2014*, pp. 70-79. Atlanta, Georgia.

Ruwanpura, J. Y. and AbouRizk, S. M., 2001. Design, development and application of soil transition algorithms for tunneling using special purpose simulation. *Proceedings of the 2001 Winter Simulation Conference*, pp. 1512-1520. Arlington, Virginia.

Ruwanpura, J. Y., AbouRizk, S. M., Er, K. C. and Fernando, S., 2001. Special purpose simulation templates for tunnel construction operations. *Canadian Journal of Civil Engineering*, 28(2), pp. 222-237.

Sadeghi, N. and Robinson Fayek, A., 2008. A framework for simulating industrial construction processes. *Proceedings of the 2008 Winter Simulation Conference*, pp. 2396-2401. Miami, FL.

Siadat, J. and Ruwanpura, J. Y., 2013. Effective simulation of earth moving projects. *Proceedings of the 2013 Winter Simulation Conference*, pp. 3282 - 3293. Washington, DC.

Sterman, J., 2000. *Business dynamics: Systems Thinking and Modeling for a complex world*. New York: McGraw-Hill.

Tanaka, Y., 1993. Cycletime simulation of shield-tunneling operation. *Proceedings of the Fifth International Conference on Computing in Civil and Building Engineering* (1993), pp. 1386-1389. Anaheim, Calif.

Tommelein, I. D., 1999. Travel-time simulation to locate and staff temporary facilities under changing construction demand. *Proceedings of the 1999 Winter Simulation Conference*, pp. 978-984. Phoenix, Arizona.

Touran, A. and Asai, T., 1987. Simulation of tunneling operations. *Journal of Construction Engineering and Management*, 113(4), pp. 554-568.

Zhang, H. and Wang, J. Y., 2008. Particle swarm optimization for construction site unequal-area layout. Journal of Construction Engineering and Management, 134(9), pp. 739-748.

Zhou, F., AbouRizk, S. M. and AL-Battaineh, H., 2009. Optimisation of construction site layout using a hybrid simulation-based system. *Simulation Modelling Practice and Theory*, Volume 17, pp. 348-363.

Zhou, F., AbouRizk, S. M. and Fernando, S., 2008. A simulation template for modeling tunnel shaft construction. *Proceedings of the 2008 Winter Simulation Conference*, pp. 2455-2461. Miami, FL.

AUTHOR BIOGRAPHIES

SeyedReza RazaviAlavi is a PhD candidate in Construction Engineering and Management at the University of Alberta. He received his BSc and MSc in Civil Engineering in 2007 and 2010, respectively. His research interests are mainly focused on simulation modeling including discrete event, continuous, and hybrid simulation models, construction site layout planning, and material management.

Simaan AbouRizk holds an NSERC Senior Industrial Research Chair in Construction Engineering and Management at the Department of Civil and

Environmental Engineering, University of Alberta, where he is a Professor in the Hole School of Construction Engineering. He received the ASCE Peurifoy Construction Research Award in 2008. He was elected fellow of the Royal Society of Canada in 2013.

ANALYSING AND OPTIMISING A DUAL RESOURCE CONSTRAINT PRODUCTION SYSTEM USING SIMULATION

K.. Ramaekers, H. Pollaris, L. Claes

Hasselt University, research group Logistics
Universiteit Hasselt, campus Diepenbeek
Agoralaan – gebouw D, 3590 Diepenbeek

katrien.ramaekers@uhasselt.be, hanne.pollaris@uhasselt.be

ABSTRACT
In current markets with rapidly changing customer expectations and technology and severe competition, it is highly important for a company to effectively respond to these changes and the associated uncertainty. This paper focuses on workforce flexibility to deal with the current markets, and more specific on dual resource constrained production systems. A dual resource constrained (DRC) production system is one in which all equipment in the shop is not fully staffed and, furthermore, workers can be transferred from one piece of equipment to another as needed. The main dimensions of a DRC production system are worker flexibility, worker assignment, transfer costs, job release mechanism, job dispatching and specific methodologies for DRC scheduling. In this paper, a serial manufacturing line with flexible workers is studied. The melting line of a zinc-producing company in Belgium is simulated and the influence of DRC-concepts on the efficiency of the production line is tested.

Keywords: simulation, DRC production system, worker flexibility.

1. INTRODUCTION

In current markets with rapidly changing customer expectations and technology and severe competition, it is highly important for a company to effectively respond to these changes and the associated uncertainty. In general, researchers and manufacturing managers contend that the most important concept to cope with this uncertainty is manufacturing flexibility, which is the ability of a firm to manage production resources and uncertainty to meet the customer requests (Zhang et al., 2003). Although the manufacturing flexibility literature tends to emphasize equipment flexibility and neglect the potential impact of workers, the workforce plays a vital role in most production processes (Yildiz and Tunali, 2008).

A dual resource constrained (DRC) production system is one in which all equipment in the shop is not fully staffed and, furthermore, workers can be transferred from one piece of equipment to another as needed (Treleven, 1989).

Human operators, unlike machines, are capable of learning and acquiring new skills. It is often useful to take advantage of a flexible workforce where workers are trained in several skills or departments such that they can be assigned to a variety of jobs as the need arises. (Xu et al., 2011)

DRC production systems are more complicated than their single resource counterparts and present a number of additional technical challenges which must be considered during resource scheduling. (Xu et al., 2011)

The literature on flexible workers and DRC production systems can be classified in two groups: studies employing flexible workers in job shops (e.g. Fryer, 1974; Treleven, 1985, Weeks and Fryer, 1976) and studies employing flexible workers in serial manufacturing lines (e.g. Bartholdi and Eisenstein, 1996; Bischak, 1996; Zavadlav et al., 1996). DRC job shops have received much more attention in literature than serial manufacturing lines. In this paper, a serial manufacturing line with flexible workers is studied. The melting line of a zinc-producing company (Nyrstar) in Belgium is simulated and the influence of DRC-concepts on the efficiency of the production line is tested. In literature, simulation remains the most popular method in DRC research.

The paper is organized as follows. In section 2, a literature review on the main dimensions that exist in DRC production systems is given. In section 3, the simulation study is described and the two scenarios that are developed are explained. Furthermore, results of the simulation study are presented in this section. In section 4, conclusions on the use of simulation for analyzing and optimizing dual resource constraint production systems are formulated.

2. LITERATURE REVIEW

Based on literature, the additional challenges that must be considered in DRC production systems are divided into six dimensions based on Xu et al. (2011): worker flexibility, worker assignment, transfer costs, job release mechanism, job dispatching and specific methodologies for DRC scheduling.

2.1. Worker flexibility

Worker flexibility in general refers to the responsiveness of a system to variations in the supply and demand of workers. Workers are no longer treated like machines and their unique human characteristics are being acknowledged as an important influence on system performance. (Xu et al., 2011)

Worker flexibility can be viewed from a variety of perspectives. Yue et al. (2008) define three basic aspects of worker flexibility: the level of multi-functionality, the pattern of skill overlaps and the distribution of skills.

The *level of multi-functionality* is defined as the number of different skills mastered by a worker. The *pattern of skill overlaps* can be expressed by two parameters: the number of skill overlaps and the number of workers involved in a chain of skill overlaps. For example, if the average number of skill overlaps equals 2, each skill can be mastered by two workers. The number of workers involved in a chain of skill overlaps indicates the number of opportunities to transfer work. The *distribution of skills* describes to what extent workers are trained for the same number of skills. If workers receive the same degree of cross-training, it is referred to as an equal distribution of skills. If workers are trained to work in a different number of departments, it is termed an unequal distribution of skills.

The benefits of worker flexibility are well documented (e.g. Treleven 1989, Malhotra et al. 1993, Felan and Fry 2001). A flexible workforce may enhance the responsiveness of a system to unexpected and unbalanced workloads. As a result, it reduces manufacturing lead time, decreases the level of work-in-process, and improves customer service. Despite these benefits, worker flexibility may incur some costs which can counteract its benefits, such as learning and forgetting costs.

2.2. Worker assignment

There are two aspects of worker assignment to consider when scheduling human operators: when a worker is available to be transferred and where that worker is to be assigned when he is eligible for transfer.

When-rules dictate when a worker is eligible for transfer. The most widely used when-rules in the literature are the centralized and decentralized rules. The centralized rule provides the maximum worker flexibility because this rule implies that the worker is available for transfer whenever a job is completed. The decentralized rule implies that the workers are eligible for transfer when there is no job waiting to be processed at their current workstations. When transfer costs are significant, the decentralized rule is used more. (Xu et al., 2011)

Where-rules dictate which workstation a worker is to be assigned when that worker becomes eligible for transfer. Some widely used where-rules in DRC literature are first in system first serve (FISFS), largest number in queue (LNQ) and earliest due date (EDD). (Xu et al., 2011)

2.3. Worker transfer costs

The transfer costs are the costs and/or delays caused by worker transfer across different work stations. When transfer costs are taken into account, there is a negative impact on overall system performance if workers are transferred too often. Aside from the simple transfer delay times caused when a worker changes workstations and jobs, there is also a loss of productivity due to human learning and forgetting effects. (Xu et al, 2011)

2.4. Job release mechanisms

The job release mechanism is responsible for the pattern of work release onto the work floor. This mechanism has an influence on the distribution of arriving jobs. A good planning system will positively impact the performance of a DRC system by producing lower and more predictable costs. (Xu et al., 2011)

2.5. Job dispatching

Job dispatching determines the order of jobs to be processed. The selection of the most appropriate job dispatching rule to use is largely dependent on the desired performance measure to optimize. Dynamic selection of a dispatching rule based on current conditions is a promising approach. (Xu et al., 2011)

2.6. Specific methodologies for DRC scheduling

As more factors such as various human and job characteristics are considered in a DRC production system, the scheduling problem becomes increasingly larger and more complicated to solve. There is a move towards using less exact solving methods to find good solutions as opposed to one optimal solution. Yildiz and Tunali (2008) proposed an optimal worker flexibility policy for a constant WIP controlled DRC system using simulation optimization (response surface methodology). Their goal was to more efficiently assign workers to workstations.

3. SIMULATION STUDY

A simulation model representing the melting line at Nyrstar is built in Arena. The simulation model is based on data that was obtained through observations of the melting line and interviews with employees working at the production line and working at a strategic level.

The melting line produces both pure zinc and zinc alloys. The pure zinc is made on the SHG-line. This line consists of two parts: the loading part and the casting part. In the loading part, the zinc is loaded from a smelting furnace. The casting line consists of three separate lines. The zinc alloys are made on the ZAMAK-line. This line consist of three parts: the loading part, the mixing part and the casting part. In the mixing part, the alloys are composed in two mixing bowls, depending on the type of alloy. The casting line has four separate lines. In the basic scenario, there are operators working on the loading part, the smelting furnace, the mixing part and on the casting part. As an example, part of the simulation model implemented in Arena in shown in Appendix A.

To validate the simulation model, first, the output of the simulation model (in number of zinc cubes) is compared to the expected theoretical output of the casting line. In Table 1, the expected theoretical output is compared to the results of the simulation model. Second, the utilization rate on the different casting lines is compared with the mean utilization rate of the last 11 months. In Table 2, the theoretical utilization rate is compared to the simulated utilization rate. Both comparisons indicate that the simulation model is a good representation of the real casting line.

Table 1: Validation: Output of casting lines

Casting line	Theoretical output	Simulated output	LCL	UCL
Z1	149305	149227	127984	160512
Z2	107184	113635	112784	119776
Z3	61654	59353	48032	70528
Z4	18500	19545	19200	19584
SHG1	64397	57067	53496	59040
SHG2	77285	69748	65384	72160
SHG3	49068	55897	51200	68000

Table 2: Validation: Utilization rate

Casting line	Theoretical utilization	Simulated utilization
Z1	92.35%	92.72%
Z2	89.32%	80.67%
Z3	85.63%	/
Z4	94.87%	89.02%
SHG1	89.44%	/
SHG2	89.45%	87.94%
SHG3	70.99%	86.66%

A simulation of 7 days of 24 hours is run. The warm-up period is determined using the graphical method (see Figure in Appendix B) and is set to eight hours. 25 replications are made for each scenario of the simulation model.

Two alternative scenarios are developed to introduce DRC-concepts at the production line at Nyrstar.

Nyrstar works with a continuous production system and uses a rotating shift system. Nyrstar is considering changing this rotating shift system from 4 shift to 5 shifts. This would give employees more free evenings and weekends and would have a beneficial effect on the stress factor and the absenteeism of employees. In both alternative scenarios, the shift from 4 to 5 shifts is assumed.

The second adaptation of the simulation model in the alternative scenarios concerns the allocation of operators. The number of operators at the moulding line, a part of the casting line, is reduced with 1. The remaining 2 operators are more flexible and allocated where necessary. This reduction of operators at the moulding line is assumed in both alternative scenarios.

In the second scenario, additionally, the operator at the scratch line, a part of the casting line, is dropped and his tasks are taken over by the operator of the smelting furnace.

The results of both alternative scenarios are calculated and compared with the results of the basic scenario. A paired t-test is used to make the comparisons. The results for the first scenario are shown in Table 3. The results for the second scenario are shown in Table 4.

Table 3: Results Scenario 1

	LCL	UCL
Total costs	20061	20641
Total output	-4426	14422
Utilization operator 1	-0.00856	0.0208
Utilization operator 2	-0.512	-0.436

The difference in total costs between scenario 1 and the basic scenario is significantly different from zero. The total costs for scenario 1 are on average 20351 Euro per week lower than for the basic scenario. The difference in total output between the two systems is not significant. The utilization of the first operator at the moulding line does not change significantly. The utilization of the second operator at the moulding line increases with 40 to 50 percent points. This high difference in utilization is explained by the way the tasks are divided among the two operators. If the scenario is implemented, the tasks should be redivided so that both operators have a comparable utilization rate.

Table 4: Results Scenario 2

	LCL	UCL
Total costs	23547	24278
Total output	-3257	11581
Utilization operator 1	-0.00756	0.0206
Utilization operator 2	-0.512	-0.43
Utilization furnace operator	-0.0643	-0.051

When comparing scenario 2 with the basic scenario, the same conclusions as for the first scenario can be drawn: the total costs decrease significantly while the total output does not differ significantly. In the second scenario, the scratch operator is removed and his tasks are taken over by the furnace operator. The utilization of the furnace operator increases hereby by 5 to 6 percent points.

4. CONCLUSIONS

In general, researchers and manufacturing managers contend that the most important concept to cope with the uncertainty in current markets is manufacturing flexibility, which is the ability of a firm to manage production resources and uncertainty to meet the customer requests. Although the manufacturing flexibility literature tends to emphasize equipment flexibility and neglect the potential impact of workers, the workforce plays a vital role in most production processes. Dual resource constrained (DRC) production systems are systems in which all equipment in the shop is not fully staffed and, furthermore, workers can be transferred from one piece of equipment to another as needed.

In this paper, a serial manufacturing line with flexible workers is studied. The melting line of a zinc-producing company (Nyrstar) in Belgium is simulated and the influence of DRC-concepts on the efficiency of the production line is tested.

The results indicate that the introduction of DRC-concepts has a positive effect on the performance of the melting line of Nyrstar: while the total output does not differ significantly, the total costs decrease significantly. The utilization of some specific operators increases but remains at an acceptable level.

In the meantime, Nyrstar has successfully introduced some of these DRC-concepts in its existing production line. Thanks to this simulation study, the advantages of introducing these concepts was clear to the management. However, it was of utmost importance that the operators at the melting line were properly informed before introducing DRC.

REFERENCES

Bartholdi III, J.J., Eisenstein, D.D. (1996). A production line that balances itself. *Operations Research,* 44(1), pp. 21-34.

Bischak, D.P. (1996). Performance of a manufacturing module with moving workers. *IIE Transactions,* 28(9), pp. 723-733.

Felan, J.T., Fry, T.D. (2001). Multi-level heterogeneous worker flexibility in a dual resource constrained (DRC) job-shop. *International Journal of Production Research,* 39(14), pp. 3041-3059.

Fryer, J.S. (1974). Labor flexibility in multiechelon dual constrained job shops, *Management Science,* 20(7), pp. 1073-1080.

Malhotra, M.K., Fry, T.D., Kher, H.V., Donohue, J.M. (1993). The impact of learning and labour attrition on worker flexibility in dual resource constrained job shops. *Decision Sciences,* 24(3), pp. 641-663.

Treleven, M.D., Elvers, D.A. (1985). An investigation of labor assignment rules in a dual-constrained job shop. *Journal of Operations Management,* 6(1), pp. 51-67.

Treleven, M.D. (1989). A review of the dual resource constrained system research. *IIE Transactions,* 21(3), pp. 279-287.

Weeks, J.K., Fryer, J.S. (1976). A simulation study of operating policies in a hypothetical dual constrained job shop. *Management Science,* 22(12), pp. 1362-1371.

Xu, J., Xu, X., Xie, S.Q. (2011). Recent developments in Dual Resource Constrained (DRC) system research. *European Journal of Operational Research,* 215, pp. 309-318.

Yildiz, G., Tunali, S. (2008). Response surface methodology based simulation optimization of a CONWIP controlled dual resource constrained system. *International Journal of Advanced Manufacturing Technololgy*, 36, pp. 1051-1060.

Yue, H., Slomp, J., Molleman, E., Van der Zee, D.J. (2008). Worker flexibility in a parallel dual resource constrained job shop. *International Journal of Production Research,* 46(2), pp. 451-467.

Zavadlav, E., McClain, J.O., Thomas, L.J. (1996). Self buffering, self balancing, self flushing production lines, *Management Science,* 42(8), pp. 1151-1164.

Zhang, Q., Vonderembse, M.A., Lim, J.S. (2003). Manufacturing flexibility: defining and analyzing relationships among competence, capability, and customer satisfaction. *Journal of Operations Management,* 21(2), pp. 173-191.

APPENDIX A

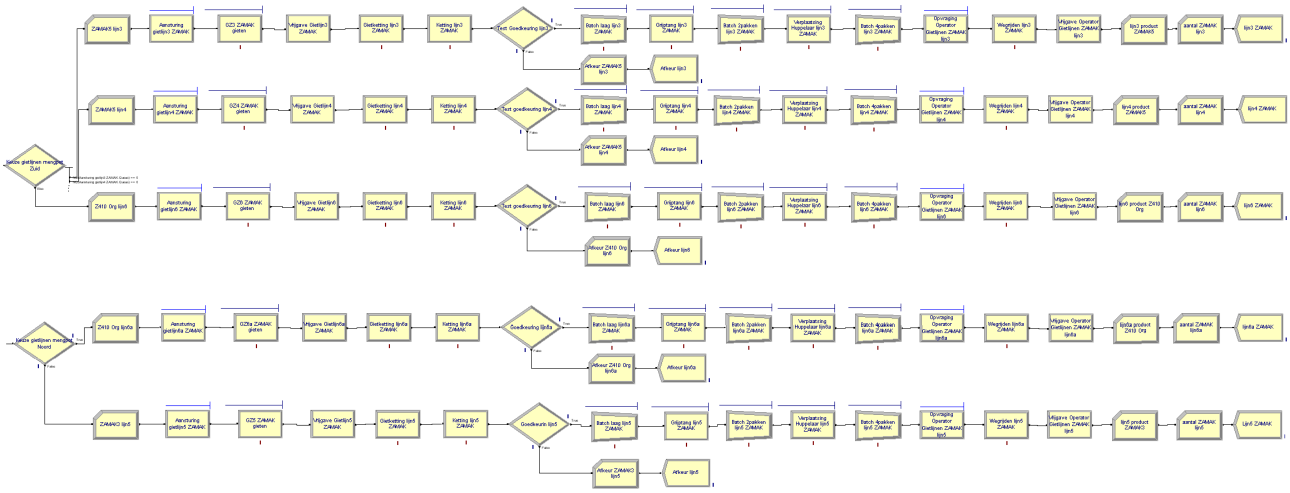


Figure : Casting line ZAMAK

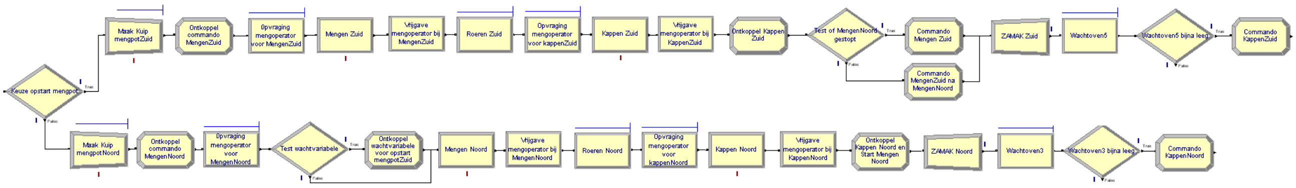


Figure : Mixing part ZAMAK

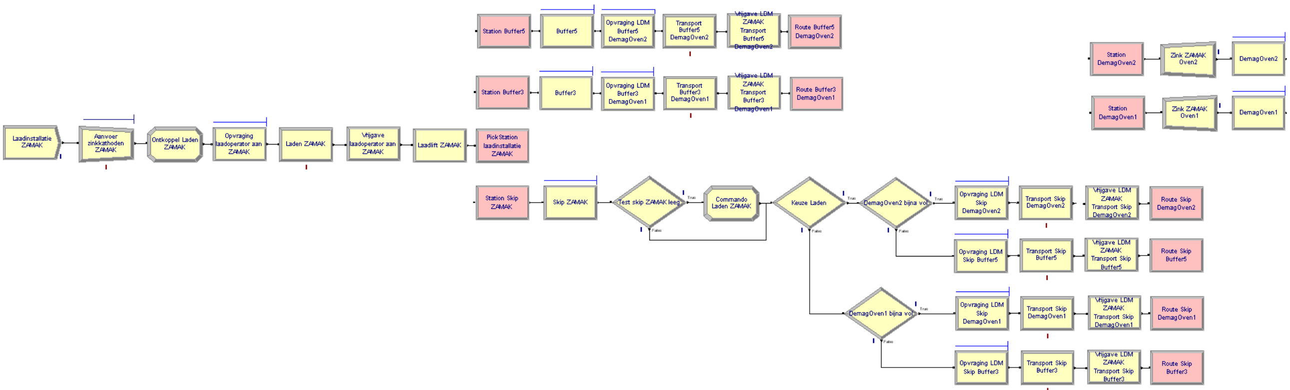


Figure : Loading part ZAMAK

APPENDIX B

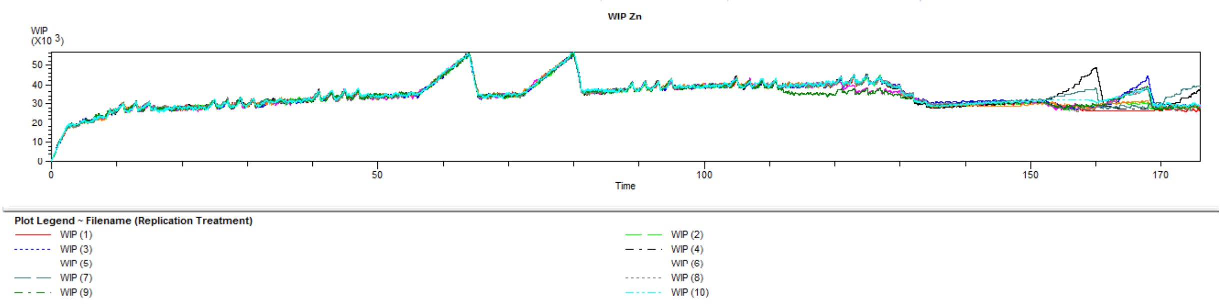


Figure: Warm-up period

AUTHORS BIOGRAPHY

Katrien Ramaekers graduated as Master of Business Economics at the Limburg University Centre in 2002. In 2007, she obtained her Ph.D. in Applied Economic Sciences at Hasselt University, Belgium. Her Ph.D. research is on the integration of simulation and optimisation, especially as a support for complex logistics decision- making. Currently, she is a post-doctoral researcher at Hasselt University and is working on the integration of simulation and optimization.

Hanne Pollaris graduated in 2012 as Master of Applied Economic Sciences: Business Engineer at Hasselt University, Belgium. In October 2012, she started her Ph.D. research on which she will be working until 2016. This research mainly considers practical vehicle routing problems with loading constraints.

Lien Claes graduated as Master of Business Economics at Hasselt University in 2013. The topic of her masterthesis was the simulation of a DRC production system.

EVALUATING THE INFLUENCE OF VIRTUAL REALITY-BASED TRAINING ON WORKERS' COMPETENCIES IN THE MINING INDUSTRY

Shiva Pedram[(a)], Pascal Perez[(a)], Stephen Palmisano[(b)]

[(a)] SMART Infrastructure Facility, Faculty of Engineering & Information Sciences, University of Wollongong, NSW
[(b)] School of Psychology, Faculty of Social Sciences, University of Wollongong, NSW

[(a)]spedram@uow.edu.au, [(b)]pascal@uow.edu.au, [(c)]stephenp@uow.edu.au

ABSTRACT

The Australian mining industry has achieved remarkable performance and safety results through continuous improvement of its training standards. Interactive virtual reality-based training is the most recent technology used to enhance workers' competencies in a safe and controlled environment that allows for replicable testing of extreme event scenarios. Like any other training method, VR-based training needs to be assessed in order to evaluate the advantages and limitations of this innovative technology, compared with more traditional approaches. Our research aims to design and implement an evaluation framework that can be used to assess VR-based training programs developed, amongst other training methods, by Coal Services Pty Ltd, a pioneering training provider for the coal mining industry in Australia. Our research focuses on specific training programs developed for mining rescue brigades. These teams are made up of highly specialized miner volunteers who are in charge of primary response in case of major incident. The evaluation framework aims at identifying adequacy between competency needs, technological capabilities and the implementation of interactive simulation. Research outcomes are meant to provide evidence-based information on the advantages and limitations of VR-based training for mining rescue brigades. The evaluation framework is flexible enough to be applied to other types of training for the mining industry or even to be adapted for use in other industries.

Keywords: Virtual Reality, Interactive Simulation, Mining Industry, Training, Safety

1. INTRODUCTION

Computer simulation as a learning environment has progressively embraced technological innovations ranging from chart-based interfaces to fully immersive environments (Bell, Taseen et al. 1990, Jou and Wang 2012). Virtual Reality (VR) provides both immersive and interactive features, allowing users to 'feel' that they are actually in the training environment (Raskind, Smedley et al. 2005).

Best practice in the mining industry includes extensive initial and professional training for staff involved in field operations. Simulator-based training is now frequently used to both establish and maintain this training. A VR environment, which is an interactive 3-D representation of the mine, has a high potential to enhance miners' safety through improved techniques for training, retraining and up-skilling. Mine sites are complex underground environments, and a 2-D representation of the environment would not be effective for most training purposes since some relationships might not be present and might be neglected unintentionally (Stothard 2007).

In case of emergency, rescue brigades are the first teams to access a mining incident. Their members are highly skilled volunteers, selected by mine managers at each production pit. Rescue brigades attend frequent training sessions in order to perform effectively in an emergency situation. A VR-based training program for rescue brigades provides a safe environment to perform collective drills for various emergency scenarios. During these sessions, trainees can improve their technical and non- technical skills.

Previous research has showed that flight simulators were very successful at bringing learning and theory into practice in a supervised, safe but highly realistic environment (Deaton, Barba et al. 2005). Despite the rapid development of VR-based training in the mining industry there is no formal evaluation of its impact on miner's skills and competences. Furthermore, due to the specificity of underground mining, it would be dangerously misleading to extrapolate training transfer results from other industries (aeronautics, automotive).

Hence, we describe in this paper an experimental design aiming at evaluating VR-based training programs developed by Coal Services Ltd for underground rescue brigades in Australia. Then, we present preliminary results from this on-going research project.

2. BACKGROUND

Historically, the mining industry is one of the most hazardous industrial sectors. Although the industry has achieved significant success in reducing the number of accidents and limiting their consequences, it remains a risky business.

According to NSW Trade and Investment, the average fatal injury frequency rate (FIFR) decreased by 65% between 2007 and 2012. The overall lost time injury frequency rate (LTIFR) also decreased by 58% over the same period while the serious bodily injury frequency rate (SBIFR) decreased by 56% (NSW Trade and Investment 2013). These records suggest that the Australian mining industry has achieved remarkable performance through continuous improvement of its safety procedures. As working environment-related incidents tend to decrease, the bulk of remaining accidents is due to human errors as showed by Williamson (1990) in Australia and Rushworth and colleagues (1999) in the US. Sources of human errors are diverse and need to be integrated into relevant training programs.

2.1. Human Mistakes and Errors

The ‘human factor analysis and classification system’ (HFACS) is a systematic and evidence-based framework aimed to design, assess and enhance the interaction between individuals, technologies (including equipment) and the organisation (Wiegmann and Shappell 2001). HFACS describes human error at each of four levels of failure: 1) unsafe acts of operators, 2) preconditions for unsafe acts, 3) unsafe supervision, and 4) organizational influences and outside factors (Patterson and Shappell 2010).

HFACS has been implemented in various hazardous industries such as civil aviation (Wiegmann and Shappell 2001, Wiegmann, Faaborg et al. 2005, Shappell, Detwiler et al. 2007), air traffic control (Broach and Dollar 2002), logistics (Reinach and Viale 2006, Baysari, McIntosh et al. 2008, Celik and Cebi 2009) and medicine (El Bardissi, Wiegmann et al. 2007).

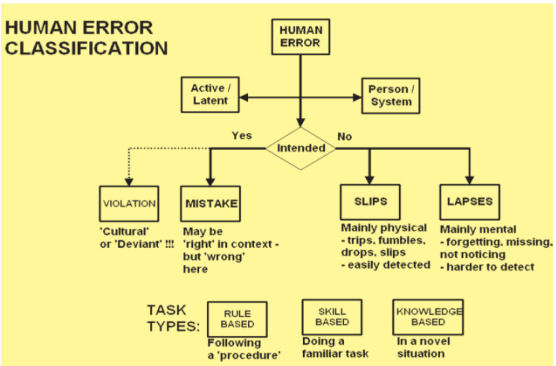


Figure 1: Human Error Classification (Trade & Investment, 2013)

As will be discussed in our evaluation framework, a list of possible human errors will be compiled from expert’s judgments (mine managers) and compared to the content of VR-based training scenarios for rescue brigades. Their degree of alignment and the success rates at performing requested tasks will inform potential improvements to training programs.

2.2. VR-based Training for Rescue Brigades

Misanchuk (1987) lists the three main factors to evaluate the quality of a training session or program: (1) employee’s ability to accomplish the assigned task, (2) relevance of the training materials to what trainees are expected to do and (3) their motivation to undertake training. Figure 3 shows the framework developed by Coal Services Ltd for rescue brigades. Training requirements and guidelines are based on Australian Coal Mine Health & Safety Acts and regulations.

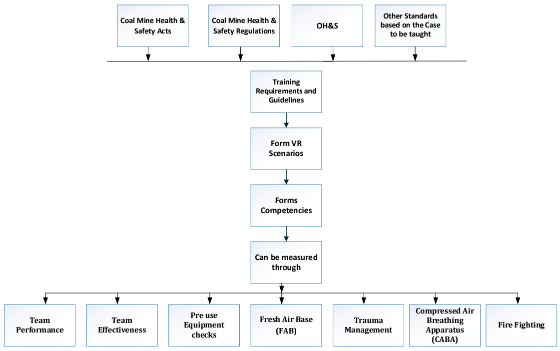


Figure 2: Training Scenario Development and Evaluation (Coal Services Ltd).

Hence, training transfer can be evaluated against: team performance, team effectiveness, pre-use equipment checks, fresh-air base, trauma management, use of the compressed air breathing apparatus and fire-fighting skills.

3. EVALUATION FRAMEWORK

Our evaluation framework (figure 4) includes four nested layers of analysis. Gaps and mismatches at the interface between two layers will help identifying training deficiencies and possible improvements to the current programs set up by Coal Services Ltd.

The outermost layer of the framework corresponds to actual training needs. Interviews with mine managers will constitute the main source of information alongside reviews of the literature produced by the mining industry. The second layer focuses on constraints associated with real-world training (aka traditional training). Coal Services Ltd offers a range of training programs for rescue brigades using various environments: mine sites (real environment), artificial galleries (controlled environment) and 3D simulators (virtual environments). Hence, it will easy to observe responses and performance of rescue brigades across the different training environments. The third layer of focuses on capabilities associated with VR technology. In-depth interviews with VR designers and trainers will help us to better understand potential and actual use of this technology. Finally, the innermost layer corresponds to the learning process experienced by trainees. Over a two-year period, several rescue brigades will be followed through their training programs, focusing on VR-based training sessions and regular safety competitions they have to participate to. Performance during these competitions will be used as a proxy for display of group and individual competences.

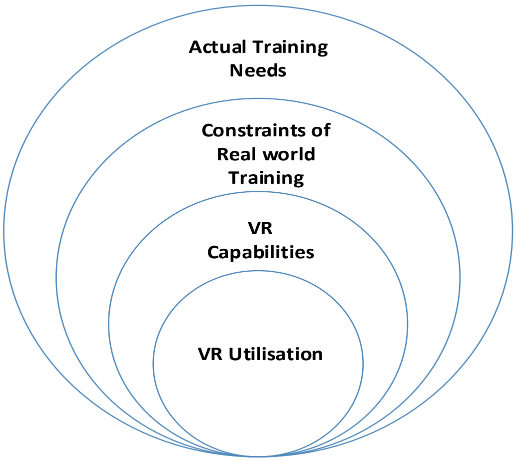


Figure 3: Evaluation Framework

3.1. Actual Training Needs

Subject matter experts (SMEs), such as team supervisors and mine managers, will be interviewed to identify potential training needs. Then, based on their responses, a Likert-scale type questionnaire will be developed. The questionnaire will be distributed to trainees in order to determine: (i) current competency level of trainees in an identified area, (ii) the most relevant training areas to trainee's actual role and (iii) trainee's level of motivation and desire to attend the VR training session. Those training areas which score below 3 for competency and above 3 for relevance will be identified as training needs (McKillip 1987).

3.2. Real-World's Constraints

Focus groups, made up of subject matter experts (SMEs) will be tasked with identifying (i) types of usual human mistakes in mining environments, (ii) constraints associated with real-world training and (iii) potential for VR-based training to overcome these limitations.

3.3. VR-based Training Capabilities

Based on the above, an initial set of desirable VR-based training features will be identified. Interviews with VR designers and trainers will identify: (i) the current capabilities of VRs, (ii) the limitations of VR and potential for upgrade, and (iii) the relevance of VR technology features for training purposes. Another questionnaire will also be developed to identify: (1) the role of simulation features and resources in overcoming the identified training challenges, and (2) the challenges of using each simulation feature. The VR-based training environment used by coal Services Ltd corresponds to a state-of-the-art 360^{o} interactive theatre with 3-D immersive visualization.

3.4. VR-based Training Utilisation

Over a two-year period, several rescue brigades will be followed through their training programs across four training facilities (Woonona, Mangerton, Singleton and Lithgow in NSW). Each trainee will have to fill a short questionnaire before and after a training session in order to record previous experiences, expectations, responses to VR environments and self-assessment of individual performance. Equivalent questionnaires will be provided to trainers as well. Finally, all aspects elicited in each analytical layer will be fed into a competency model (figure 4).

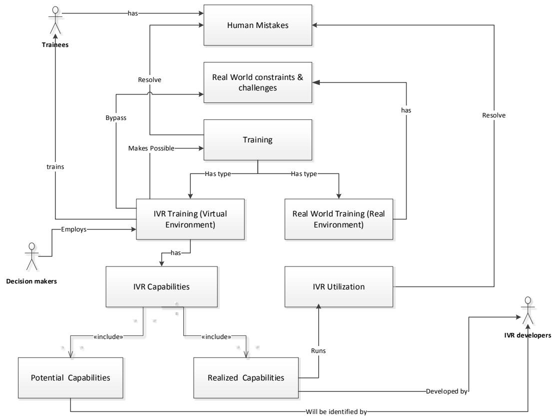


Figure 4: Entity Diagram between different factors affecting competency

Self-assessed performance during training will be first compared with evaluation from trainers. Then, these results will be compared with actual demonstrations of competence during rescue competition. As many rescue brigades will have been through different training sequences (real-world, controlled and virtual environments) and display contrasted individual characteristics (mining experience, team bonding experience, motion sickness…), the quality of training transfer will be assessed using a Hidden Markov Model (HMM).

4. PRELIMINARY RESULTS

This is an on-going research project and results provided below correspond to a preliminary analysis of 25 VR-based training sessions only (93 trainees and 25 trainers in total). Interviews with SMEs and VR technology designers have not been completed yet. Likewise, results from rescue competitions will be collected in coming months.

4.1. Trainee's Perception

Our preliminary analysis concentrates on three subjective evaluation criteria: (i) perceived success, (ii) perceived usefulness and (iii) perceived realism. The first one aims at eliciting trainee's perception of his performance during the VR-based training session. The second one aims at evaluating how much trainees learnt during the session. Finally, the third one focuses on eliciting trainee's opinion on the degree of 'realism' of the VR-based training session.

Nearly 87% of the trainees felt that they had been 'relatively successful' to 'very successful' at completing requested tasks during the training session. Likewise, 90% of the trainees thought that the VR-based training session had been 'useful' or 'very useful' for the development of their skills. Only 22% of trainees thought that the VR environment was 'realistic' or 'very realistic' while 34% felt that it was 'unrealistic' or 'very unrealistic' (44% of trainees neither agreed nor disagreed). Reasons provided to justify the lack of realism ranged from the inability to 'walk' through the scenario (trainees stand in the middle of the theatre and the virtual environment revolves around them) to specific (and unrealistic) details that spoilt the experience.

A cross-tabulation between perceived 'usefulness' and 'realism' criteria shows that 32% of the trainees who found the training session 'useful' or 'very useful' considered the VR environment as 'unrealistic' or 'very unrealistic' while another 45% had a neutral perception of its degree of realism (figure 5). Although these are only preliminary results drawn from a relatively small sample, the apparent dichotomy between both criteria is worth investigating further when more data will be available.

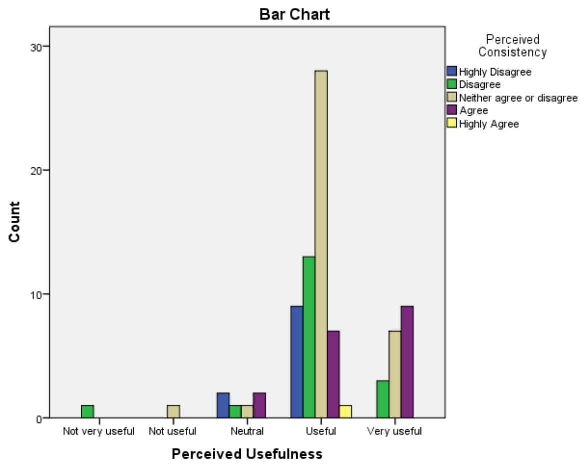


Figure 5: relationship between perceived usefulness and consistency (realism) of the VR-based training session (85 valid trainee's responses).

4.2. Trainer's perception

Although our small sample (25 trainers) does not allow us to draw any statistically significant conclusion, figures show that trainers are slightly more positive about the usefulness of training sessions and performance of rescue brigades during training. 36% of trainers (9 cases) consider that their session was 'very useful' and 'highly successful' while 52% of trainers (13 cases) considered that their session was 'useful' and 'successful'.

A cross-examination of responses from trainees and trainers was conducted using a Mann Whitney U test as most variables don't display Normal distributions. Comparison between perceived usefulness from trainee's and trainer's viewpoints show that there isn't any significant difference between both groups ($Z= -1.563$; $p=.118>.05$). Likewise, comparison between perceived success from trainee's and trainer's viewpoints show that there isn't any significant difference between both groups ($Z= -0.967$; $p=.333>.05$).

These preliminary results start to inform the innermost part of our evaluation framework. Competition outcomes will allow us to evaluate the quality of training transfer according to various training sequences over a two-year period. Later on, interviews with SMEs and VR designers will allow us to inform outer layers of our evaluation framework.

5. CONCLUSION

This research aims to evaluate the effects of VR-based training on the competency of rescue brigades in the mining industry. Hence, a research framework has been proposed to qualitatively evaluate and assess the role of VR-based training in achieving these competencies. The proposed framework can be used as a guide to evaluating the adequacy of implementing interactive simulation in order to improve competency and prevent or deal successfully with accidents in different industries.

This framework aims to collect a broad range of data to help with a qualitative analysis of the impact of mining simulators on improving workforce competency and safety. The evaluation framework is flexible enough to be applied not only to other types of training for the mining industry but also to be adapted to other industries. Our evaluation framework will be applied to four VR-based training facilities operated by Coal Services Ltd in NSW, Australia.

6. LIST OF REFERENCES

Baysari, M. T., et al. (2008). "Understanding the human factors contribution to railway accidents and incidents in Australia." *Accident Analysis & Prevention* 40(5): 1750-1757.

Bell, P. C., et al. (1990). "Visual interactive simulation modeling in a decision support role." *Computers & Operations Research* 17(5): 447-456.

Broach, D. M. and C. S. Dollar (2002). Relationship of employee attitudes and supervisor-controller ratio to en route operational error rates, DTIC Document.

Celik, M. and S. Cebi (2009). "Analytical HFACS for investigating human errors in shipping accidents." *Accident Analysis & Prevention* 41(1): 66-75.

Deaton, J. E., et al. (2005). "Virtual environment cultural training for operational readiness (VECTOR)." *Virtual Reality* 8(3): 156-167.

ElBardissi, A. W., et al. (2007). "Application of the human factors analysis and classification system methodology to the cardiovascular surgery operating room." *The Annals of Thoracic Surgery* 83(4): 1412-1419.

Fox, J., et al. (2009). "Virtual reality." *Journal of Media Psychology: Theories, Methods, and Applications* 21(3): 95-113.

Jou, M. and J. Wang (2012). "Investigation of effects of virtual reality environments on learning performance of technical skills." *Computers in Human Behavior*.

McKillip, J. (1987). *Need analysis: Tools for the human services and education*, Sage Newbury Park, CA.

Patterson, J. M. and S. A. Shappell (2010). "Operator error and system deficiencies: analysis of 508 mining incidents and accidents from Queensland, Australia using HFACS." *Accident Analysis & Prevention* 42(4): 1379-1385.

Raskind, M., et al. (2005). "Virtual Technology Bringing the World Into the Special Education Classroom." *Intervention in School and Clinic* 41(2): 114-119.

Reinach, S. and A. Viale (2006). "Application of a human error framework to conduct train accident/incident investigations." *Accident Analysis & Prevention* 38(2): 396-406.

Rushworth, A., et al. (1999). "Investigate the causes of transport and tramming accidents on coal mines."

Shappell, S., et al. (2007). "Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system." Human Factors: *The Journal of the Human Factors and Ergonomics Society* 49(2): 227-242.

Stothard, P. (2007). Developing Interactive Training Simulations for the Coal Mining Industry. In: *Proceedings of SimTect 2007*. Simulation Australia, Adelaide, Australia.

Trade and Investment Resources and Energy (2013). *Mine Safety Performance Measure*. Online access: http://www.resources.nsw.gov.au/safety/mine-safety-results/safety-performance-measures.

Wiegmann, D., et al. (2005). Human error and general aviation accidents: A comprehensive, fine-grained analysis using HFACS, DTIC Document.

Wiegmann, D. A. and S. A. Shappell (2001). A Human Error Analysis of Commerical Aviation Accidents Using the Human Factors Analysis and Classification System (HFACS), DTIC Document.

Willamson, A. and A.-M. Feyer (1990). "Behavioural epidemiology as a tool for accident research." *Journal of Occupational Accidents* 12(1): 207-222.

AUTHORS BIOGRAPHY

Shiva Pedram

Shiva is currently undertaking a PhD research project at the SMART Infrastructure Facility, University of Wollongong (NSW, Australia). Shiva has a background in civil engineering and her current research focuses on evaluating VR-based training programs for the mining industry in Australia.

Pascal Perez

Pascal is the inaugural Research Director of the SMART Infrastructure Facility, University of Wollongong (NSW, Australia). He is also a Professor of infrastructure modelling within the Faculty of Engineering and Information Science.

Stephen Palmisano

Stephen is an Associate Professor at the School of Psychology, University of Wollongong (NSW, Australia). He is a specialist of evaluating effects of VR-based technology on individuals.

DEVELOPMENT OF A "NUMERICAL HAM" MODEL THAT COUPLES WATER AND SALT TRANSFERS TO PROTEOLYSIS DURING THE DRY-CURED HAM PROCESS

Harkouss R.[(a)], Chevarin C.[(a)], Daudin J.D.[(a)] , Mirade P.S.[(a)]

[(a)]INRA, UR370 Qualité des produits Animaux, F-63122 Saint-Genès-Champanelle, France

[(a)]Pierre-Sylvain.Mirade@clermont.inra.fr

ABSTRACT
Since salting is essential during the dry-cured ham process, reducing salt content could affect the final product quality. The aim of this study was to build a 3D multi-physical numerical model which estimates the biochemical evolution (proteolysis) as well as the distribution of salt and water contents during all the stages of dry-cured ham process. The model built showed a very good prediction of the distribution of salt and water content in the ham. Its robustness was evaluated by comparing the predicted values of proteolysis index (PI) at the end of post-salting stage with the experimental PI values measured in samples extracted from industrial Bayonne hams. The 3D "numerical ham" built can be considered as an original software tool that could be helpful for professionals to reduce the final salt content in dry-cured hams.

Keywords: dry-cured ham, multi-physical numerical model, proteolysis, water and salt transfers

1. INTRODUCTION

In the dry-cured ham elaboration process which includes salting (duration of about 15 days), post-salting (8 weeks), pre-drying (1 week) and drying (5 to more than 18 months) stages, salt is a multifunctional element that influences the final product in terms of quality and safety. Salt contributes to water holding capacity, texture and flavour development and limits the growth of pathogens and spoilage microorganisms (Taormina 2010). On the other hand, the final dry-cured ham quality is also affected by the time-course of proteolysis which depends on various factors, such as temperature, pH, salt and water contents (Toldrá and Flores 2000; Arnau, Guerrero, and Gou 1997).

Although salt is essential during the dry-cured ham manufacturing, many efforts have been made to reduce the final salt content, without altering the final dry-cured ham quality. To achieve this, NaCl reduction was mainly performed through two different approaches: (a) direct reduction of NaCl content or (b) partial substitution of NaCl by other salts. Results of the first approach showed that longer post-salting time was needed for lower-sodium hams to reach the same a_w values as hams normally salted with 100% NaCl. No more than 25% salt reductions can be obtained using this first approach. However, particular caution has to be paid when using this approach to avoid any microbiological stability problem and final defective texture as a result of abnormal proteolysis (Desmond 2006). Given these concerns, NaCl substitution by KCl, $CaCl_2$, $MgCl_2$ or K-lactate constitutes the second approach that can be used.

In food industry, salting and drying are mass transfer processes leading to time-variable water and salt content profiles. It is of high interest to assess the time-course evolution and to quantify these two parameters, e.g. to better understand their effects on the biochemical evolution and textural properties that occur during the dry-cured ham manufacture. To achieve this, numerical modelling and simulation could be a good candidate since it is non-invasive, non-destructive, less expensive, and rapid. However, very few studies aimed at modelling salt and water transfers during the different stages of dry-curing hams. On the other hand, there has not been a single study that predicted the time-course of proteolysis during the dry-cured ham production process. Very recently, Harkouss, Safa, Gatellier, Lebert, and Mirade (2014) established phenomenological models allowing proteolysis to be quantified, for five various pork muscles, as a function of temperature, water and salt content. However, quantitative modelling of salt diffusion, water migration, heat transfer and proteolysis evolution using finite element method has, to our knowledge, never been attempted in dry-cured ham. On account of this, this study aims at presenting a multi-physical 3D model developed under the Comsol® Multiphysics software that combines a determination of proteolysis through the phenomenological models developed by Harkouss et al. (2014) with salt penetration, water migration and heat transfer modelling. For practical reason, this paper is focused on the modelling of what happens in terms of water and salt transfers and proteolysis during the first stages of the dry-curing ham process, those carried out at low temperature: the salting and post-salting stages.

2. MATERIALS AND METHODS

Before integrating the proteolysis model and the heat and mass transfer models into the multi-physical model built using Comsol® multiphysics software, complete 3D ham geometry was built.

2.1. 3D Ham Geometry and Meshing

The process of constructing an accurate 3D ham representation required a series of 2D slices of a fresh ham that can be typically provided by Computed Tomography (CT), which is a rapid and non destructive X-ray imaging technique. Automated tools for noise reduction, smoothing and contrast sharpening were used as well as other data manipulation tools such as re-sampling and rescaling. The process was automated by thresholding, meaning that the different parts of the ham, i.e. rind, muscles, bone and thin film of fat between the SM and BF muscles, were identified by specifying minimum and maximum values for the signal. In total, the segmentation of 181 X-ray CT images of green ham was made using specific software called Mimics® (Materialise, Leuven, Belgium). The segmented objects were then simultaneously connected and meshed based on an orthotropic grid intersected by interfaces defining the boundaries.

The obtained 3D ham geometry was then smoothed and meshed many times in order to obtain a high quality surface mesh. After that, a volumetric tetrahedral mesh with extra smooth boundary was obtained. Finally, the 3D ham model consisting in 202,000 tetrahedral meshes and containing 5 different groups of muscle was imported into Comsol® Multiphysics software (Figure 1).

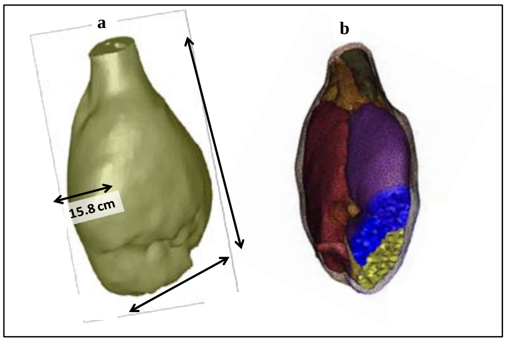


Figure 1: Views of (a) the 3D Ham Geometry Built and (b) the Volumetric Tetrahedral Mesh with the Different Groups of Muscle inside the Ham.

2.2. Model of Proteolysis

Several phenomenological models allowing proteolysis rate to be quantified as a function of several factors have been built and recently published in Harkouss et al. (2014). In the present study, modelling of proteolysis was performed from solving equation (1) and assuming that the proteolysis rate could be calculated at each time step from a global phenomenological model corresponding to Equation (2):

$$\frac{d\,PI}{d\,t} = R_i \tag{1}$$

With

$$R_i = 8.286.10^{-4} - 1.024.10^{-2}.T + 1.147.10^{-4}.S + 1.466.10^{-4}.W - 2.62.10^{-4}.T.S + 3.254.10^{-4}.T.W - 1.746.10^{-6}.S.W \tag{2}$$

where PI is the proteolysis index (%), T is the temperature (°C), S is the salt content (kg salt.kg dried matter^{-1} or % DM) and W is the water content (kg water.kg total matter^{-1} or % TM).

These two equations were solved in all domains of the numerical ham, except the bone which was logically excluded from the PI calculation. Besides, an initial condition was applied to PI. Indeed, a PI value of 2.5% was considered at time zero, which represents the mean average of a huge number of experimental PI measurements carried out on fresh small pork meat samples prepared from different ham muscles.

2.3. Heat Transfer Modelling

Fourier law was introduced to numerically simulate heat transfer inside the ham and predict how ham temperature changes in response to the modification of air temperature during the pre-drying or drying/ripening stages. The simplified equation that was solved for calculating the distribution of the ham temperature T as a function of space and time is:

$$\rho\, c_p\, \frac{\partial T}{\partial t} = \nabla.(k\ \nabla T) \tag{3}$$

where t is the time (s), ρ is the ham density (1072 kg.m^{-3}), c_p is the ham specific heat capacity (3200 J.kg^{-1}.K^{-1}) and k is the ham thermal conductivity (0.45 W.m^{-1}.K^{-1}).

This equation was applied and solved in all domains of the numerical ham, except the bone which was considered as thermally insulated. At time 0, i.e. at the beginning of the salting stage, we assumed that a green ham having an initial temperature of 3°C was placed in a cold room with air at 4°C.

At the air-ham interface, the following thermal boundary condition was imposed in the model, thus allowing the heat flow rate q to be calculated taking into account the thermal convection and the energy exchanged during water evaporation:

$$q = h\,(T_{air} - T_{surface}) + \varphi_{water}\,.L_v \tag{4}$$

In equation (4), h represents the convective heat transfer coefficient. A value of h equal to 7 W.m^{-2}.K^{-1} was considered in the present model as a result of low air velocity (Kondjoyan and Daudin 1997). $T_{surface}$ represents either the temperature of the muscle surface where salt is added or the temperature of the ham rind surface where no salt is added since we modelled here a "limited salt input" salting procedure consisting in salting only the muscle part of the ham without salting the rind. At last, in equation (4), L_v represents the latent heat due to the water evaporation (equal to 2450 kJ.kg^{-1}) and φ_{water} is the water flow rate that evaporates from the muscle or the rind surface (please see Equation (5)).

2.4. Mass Transfer Modelling

We assumed that the mass transfer phenomena that occur during the dry-cured ham elaboration process, i.e.

salt diffusion and water migration, can be modelled by the following Fick equation:

$$\frac{\partial c_i}{\partial t} + \nabla.(-D_i \nabla c_i) = 0 \quad (5)$$

where c_i is the salt or water concentration. In equation (5), D_i represents the diffusion coefficient of salt or water inside the ham. For these coefficients, in the lean meat of ham, we assumed a constant D_{salt} equal to 5.10^{-10} $m^2.s^{-1}$, but a D_{water} varying as a function of the water content expressed on dry basis (X_{water}) according to the following relation (Ruiz-Cabrera et al. 2004):

$$D_{water} = 4.10^{(0.625\,.\,X_{water} - 12)} \quad (6)$$

The Fick equation was solved in all domains, except in the bone which was logically excluded from the calculation. An initial salt content of 0% TM and an initial water content of 75% TM, i.e. values corresponding to fresh meat, were imposed in the model.

As previously indicated, the "numerical ham" model was used first to assess what happens in terms of water and salt transfers and proteolysis during the first 11 weeks of process performed at low air temperature. This period of time included a first stage of salting corresponding to a "limited salt input" salting procedure applied for 15 days. From a numerical point of view, salting on the muscle side was modelled through the application of a brine solution composed of a variable mass of salt added to a constant volume of water. Indeed, at each time step of the salting stage, the mass of salt that penetrated inside the ham was calculated as well as the mass of salt that stayed at the ham surface and thus the new salt concentration of the brine solution in contact with the muscle side of the ham. The salting stage was then followed by 62 days of post-salting during which salt diffused inside the ham. Simultaneously to salt diffusion, water migration occurred from inside the ham to the outer zones where evaporation took place. For that purpose, specific boundary conditions had to be imposed at the ham-air interface in terms of mass flow rates. The water flow rate can be calculated from the following equation:

$$\varphi_{water} = k\,.\left(a_{w,surface}\,.\,P_{sat,T_{surface}} - \frac{RH_{air}}{100}\,.\,P_{sat,T_{air}}\right) \quad (7)$$

with an air relative humidity value RH_{air} of 70% and in the range of temperature 0-40°C:

$$P_{sat,T} = exp\left(-\frac{6764}{T} - 4.915.\log T + 58.75\right) \quad (8)$$

In equation (7), k, which is the water transfer coefficient ($kg.m^{-2}.Pa^{-1}.s^{-1}$), can be calculated from the convective heat transfer coefficient h using the relation:

$$k = \frac{h\,.\,M_{water}}{Cp_{air}\,.\,M_{air}\,.\,P_{atm}\,.\,Le^{2/3}} \quad (9)$$

with M_{water} = 18 $g.mol^{-1}$; Cp_{air} = 1004 $J.kg^{-1}.K^{-1}$; M_{air} = 29 $g.mol^{-1}$; P_{atm} = 1 atm; and Le is the Lewis number which is equal to 0.777, for air. On the rind side of the ham, we arbitrary assumed that the water transfer coefficient k was divided by a factor 20 to indirectly account for the barrier effect against water transfer played by the fat layer located just under the rind.

In equation (7), the term 'surface' is associated either to the muscle side or to the rind side, depending on the boundary condition considered. On the rind side of the ham, we arbitrary considered that a_w was equal to 1 due to the presence of a few mm-thick fat layer. On the other hand, on the muscle side of the ham, owing to the presence of salt, a_w at the muscle surface obeyed to the following relations (Rougier et al. 2007):

$$a_{w_{salted\ meat}} = a_{w_{meat\ with\ no\ salt}}\,.\,a_{w_{salted\ water\ in\ meat}} \quad (10)$$

with

$$a_{w_{meat\ with\ no\ salt}} = 0.993\,.\,exp^{(-\,0.0204\,.\,X_{water}^{-1.96})} \quad (11)$$

$$a_{w_{salted\ water}} = -0.4553\,.\,X_{salt}^{water^2} - 0.5242\,.\,X_{salt}^{water} + 0.999 \quad (12)$$

where X_{salt}^{water} represents the salt content expressed on water basis (kg salt.kg $water^{-1}$).

In terms of salting, the boundary condition applied to the muscle side of the ham consists in writing the equality of water activities between the brine solution and the first layer of meat in contact with this brine solution. This assumption leads to:

$$a_{w_{brine\ solution}} = a_{w_{salted\ meat}} \quad (13)$$

The first term of this equation was calculated using equation (12) and from the salt balance performed at each time step of the salting stage on the salting surface allowing the mass of salt to be calculated and thus, for the brine solution, the term X_{salt}^{water}. The right-hand-side term of equation (13) corresponds to equation (10). Solving equation (13) allows, for the salted meat, the term X_{salt}^{water} to be assessed, and then the salt concentration in the first layer of meat in equilibrium with the brine solution.

2.5. Solving of Equations

Once implemented in the "numerical ham" model, solving all these equations lasted about 3.5 h on a 3 GHz Xeon 8-processors PC with 48 Go of RAM to model what happened during the salting and post-salting stages in terms of time course of proteolysis and water and salt transfers, with a time step of 0.1 day.

3. RESULTS AND DISCUSSION

Here only the results corresponding to the first two stages performed at low temperature, i.e. salting and post-salting, will be shown.

3.1. Distribution of Spatial and Time Values of Salt Content, Water Content and PI

Figure 2 shows the distribution of PI, water and salt contents predicted by the numerical model during the salting stage, after one week or two weeks of process.

In this figure, each view presents the distribution of water content, in the range 50%-75% TM, on the section located in the lower part of the ham geometry, the distribution of salt content, in the range 0%-10% TM, on the section located in the middle part of the ham geometry, and that of PI, in the range 2.5%-12.5%, on the section located in the upper part of the ham geometry. After one week of process, the salt penetration is only visible in the few first centimetres from the salting surface, leading to a salt concentration reaching 10% TM in this region, whereas it is lower in the middle part of the ham (2%-3% TM) and negligible near to the rind side of the ham opposite the salting surface. Concerning the water migration and distribution, Figure 2 shows that just the superficial area of the "muscle side" of the ham has lost water during the first week of process.

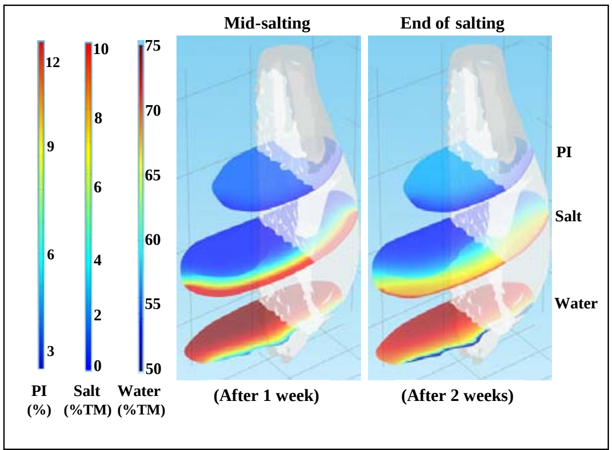


Figure 2: Distribution of PI, Salt and Water Contents Predicted during the Salting Stage, at Mid-Salting and at the End of Salting

After two weeks of process, i.e. at the end of salting, Figure 2 shows that salt has diffused more inside in the ham, increasing obviously the salt concentration, in particular in the middle part of the ham where concentration values reach 5% TM; however, the salt concentration is still relatively low in the deeper zones of the ham. Due to water evaporation, the water content of the ham continued to decrease until the end of salting, but this is still visible close to the salting surface, in particular in the few first millimetres of this surface where the water content decreased to 55% TM. As a result of the short duration of process, PI values did not increase clearly during these two weeks, even if, in the deeper zones of the ham, predicted PI was equal to approximately 4-5% compared to 2.5%-3% close to the salting surface. Actually, owing to the behaviour highlighted by the phenomenological models of proteolysis built, the predicted PI values were logically lower in areas highly dried and highly salted, as close to the salting surface. Inside the ham, the proteolysis logically increased due to high water content values combined to low salt content values, thus leading to a 2-3% increase in PI values during the two first weeks of process (Figure 2).

During the post-salting stage, numerical simulation showed that, some weeks after the end of salting, homogeneity started to take place in the different domains of the ham (Figure 3). This homogeneity is more and more visible, especially when visualizing the distribution of salt content at the end of the post-salting stage. During this stage, salt showed a great ability to penetrate deeply towards the inner zones of the ham. After six weeks, the salt content gently increased in the middle of the ham (4%-5% TM) and slightly in the deepest zones of the ham where values of 2%-3% TM were predicted. Meanwhile, the water content of the ham gradually decreased in the inner zones (68%-70% TM) and the very dried zone still exists close to the salting surface. Due to the evolution of the salt and water content, PI values logically increased inside the ham where indices ranging from 7% to 9% were predicted, leading to the apparition of a gradient of PI between the inner zone of the ham and the zone located in close proximity to the salting surface where salting and drying highly inhibited the proteolysis evolution.

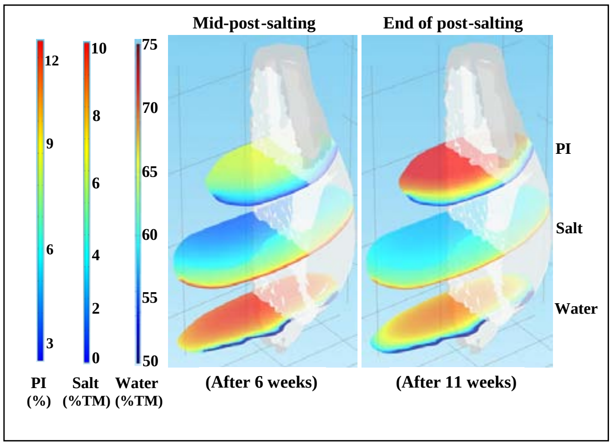


Figure 3: Distribution of PI, Salt and Water Contents Predicted during the Post-Salting Stage, at Mid-Period and at the End of Post-Salting

At the end of the post-salting stage, i.e. after 11 weeks of process, due to diffusion, the distribution of salt appeared as more and more homogeneous inside almost all the ham volume with values of salt content of about 4-5% TM. The salt content even decreased in the zone close to the salting surface where high salt content values were now predicted only on the few first millimetres (Figure 3). During this period, the water content obviously decreased inside the ham (65%-67% TM) and the thickness of the very dried zone close to the salting surface slightly increased (Figure 3). Eleven weeks were finally necessary to raise the predicted PI values in a noticeable manner, especially in the inner domains of the ham where PI exceeded 10%. Figure 3 also shows that a gradient of PI is created between the surface where salt was deposited and the inner zones of the ham.

3.2. Predicted Mean Values of Salt Content, Water Content, PI, and a_w

The "numerical ham" model also allows calculating mean values of the predicted parameters, such as salt content, water content, PI and a_w in the different groups of muscles identified in the ham geometry during its construction using Mimics® software.

Figure 4 shows the time course of mean values of salt content predicted by the "numerical ham" model during the salting and post-salting stages in three different groups of muscles: (1) in the SM muscle, called "grosse noix" in a "French butcher" language, (2) in the 'BF + ST' muscles, called "sous-noix" in a "French butcher" language, and (3) in all the ham volume. Due to salting, the salt content in these three different groups of muscles increased rapidly, especially since the group of muscles is located close to the salting surface. That is why the salt content in the SM muscle increased more rapidly and exceeded 6% TM, in average, at the end of salting compared to the one of the "sous-noix" whose mean value reached about 1% TM over the same period. During the post-salting stage, the salt content of the group of muscles close to the salting surface decreased in favour of the one located inside the ham, due to diffusion (Figure 4).

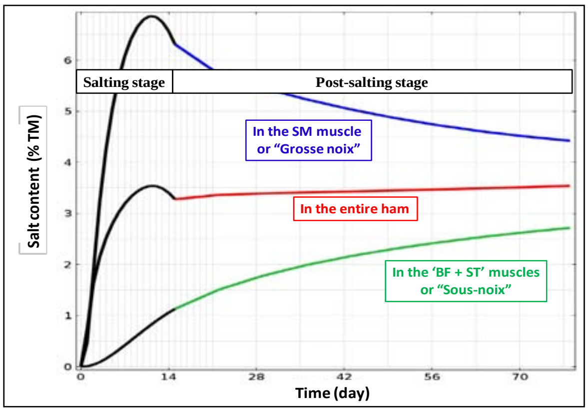


Figure 4: Time Course of Mean Values of Salt Content Predicted during the Salting and Post-Salting Stages and Calculated in the SM Muscle, in the 'BF + ST' Muscle and in the Entire Ham

Regarding the entire ham, the behaviour in terms of time course of salt content was logically intermediate compared to the two other groups of muscles, with a mean value equal to 3.3% TM at the end of the salting stage. This value of salt content then increased slightly during the post-salting stage due to the water evaporation which led to a concentration of salt. Salt content in the entire ham thus reached 3.5% TM, after 77 days of process (Figure 4).

Figure 5 shows the time course of mean values of water content predicted by the model during the salting and post-salting stages in the three same groups of muscles than for salt diffusion. Due to water evaporation, the predicted water content in these three groups of muscles logically decreased, but more or less rapidly according to their distance from the ham surface. Since the "muscle side" surface dried more rapidly than the "rind side" surface of the ham, the water content in the "grosse noix" decreased more rapidly than in the "sous-noix". The difference in water content between these two groups of muscles reached 6% TM, in average, at the end of the post-salting stage. As previously, the behaviour of the entire ham in terms of decrease in water content was intermediate when comparing to the two other groups of muscles, thus leading to a final mean value of water content in the entire ham equal to 66% TM, after 77 days of process.

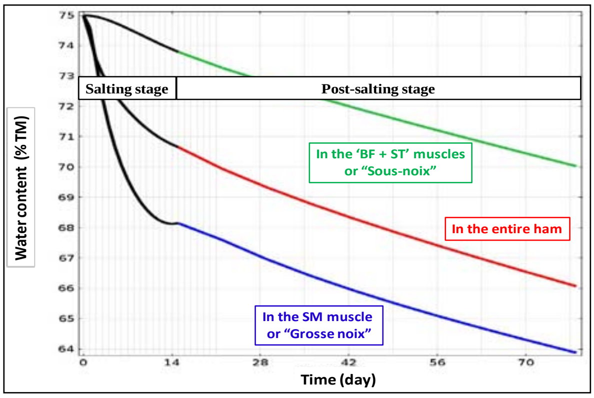


Figure 5: Time Course of Mean Values of Water Content Predicted during the Salting and Post-Salting Stages and Calculated in the SM Muscle, in the 'BF + ST' Muscle and in the Entire Ham

Figure 6 shows the evolution of mean values of PI predicted over the 77 days period in the "grosse noix", the "sous-noix" and in the entire ham, from the phenomenological models implemented in the model. Since the temperature has not changed during the two stages modelled here, the predicted mean values of PI were directly linked to the time course of salt content and water content previously described. So the lower mean values of PI were predicted in the most dried and salted group of muscles, i.e. the "grosse noix". The difference in PI values exceeded 3% between the "sous-noix" (12.5%) and the "grosse noix" (9%); the mean PI value of the entire ham was intermediate (10.7%). In fact, these calculated PI values globally were in agreement with those measured on samples extracted from two types of muscle (BF and SM) of industrial Bayonne dry-cured hams, at the end of the post-salting stage. The experimental PI values were equal to 12% ± 1% and 9% ± 1% for BF and SM, respectively, i.e. values which are relatively close to the calculated PI, thus demonstrating the accuracy of prediction of the 3D "numerical ham" model built. Accurate analysis of results reveals that the difference in terms of time course of mean values of PI between the 3 groups of muscles appeared early in the second or third day of salting.

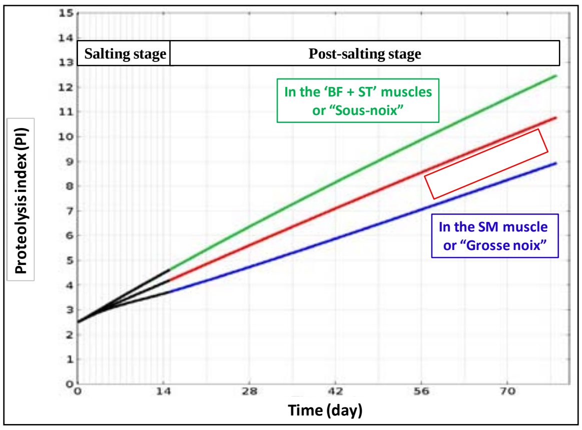


Figure 6: Time Course of Mean Values of PI Predicted during the Salting and Post-Salting Stages and Calculated in the SM Muscle, in the 'BF + ST' Muscle and in the Entire Ham

Figure 7 shows the time course of mean values of a_w predicted during the salting and the post-salting stages in the "grosse noix", in the "sous-noix" and also in the entire ham. The increase of salt content in the muscles combined to the decrease in the water content due to drying globally led to a decrease in mean a_w values in all the groups of muscles with the objective of reaching the microbial stability, which is the aim of the salting and drying operations performed during the dry-cured ham process. It can be noted, in Figure 7, the mean a_w value of the "grosse noix" increased during the post-salting stage as a result of an increase in a_w due to the salt diffusion from the "grosse noix" to the "sous-noix" as shown in Figure 4 not counterbalanced by the decrease in a_w due to the drying that occurred at the same period. However, the mean value of a_w in the entire ham decreased progressively during the post-salting stage, from 0.96, after 15 days, to 0.955, at the end (Figure 7).

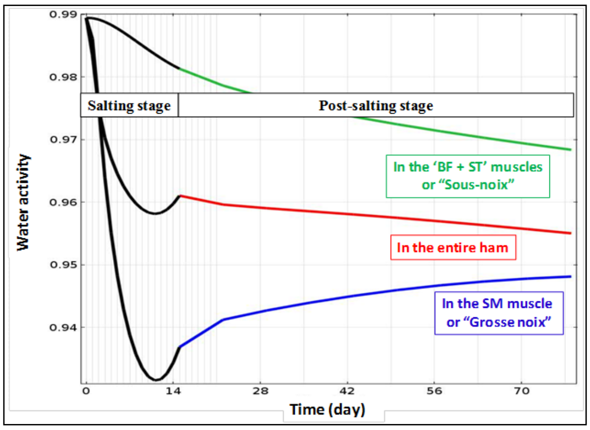


Figure 7: Time Course of Mean Values of a_w Predicted during the Salting and Post-Salting Stages and Calculated in the SM Muscle, in the 'BF + ST' Muscle and in the Entire Ham

4. CONCLUSION

This study presented the development of a numerical model that simulates water and salt transfers in a real ham, and couples the time course prediction of the salt and water content to PI evolution through the implementation of a specific phenomenological model established in a previous study. The present "numerical ham" model is able to predict the time course of salt and water content, PI, and a_w, and profiles can be extracted anywhere in the ham to make the interpretation of the phenomena easier to understand. However, this model needs to be more validated by performing an accurate comparison of the predicted values of salt and water content, and of a_w with experimental values.

To the authors' knowledge, a complete 3D multiphysical numerical model of dry-cured ham was built for the first time. This "3D numerical ham" model constitutes a new numerical tool that would be greatly useful for helping industrials to investigate various process scenarios in order to reduce the final salt content of dry-cured hams, without modifying their final quality in terms of proteolysis, texture and flavour.

ACKNOWLEDGMENTS

This work was funded by the Na- integrated programme (ANR-09-ALIA-013-01) financed by the French National Research Agency. We thank IFIP for providing CT images.

REFERENCES

Arnau, J., Guerrero, L., Gou, P., 1997. Effects of temperature during the last month of ageing and of salting time on dry-cured ham aged for six months. *Journal of the Science of Food and Agriculture*, 74, 193-198.

Harkouss, R., Safa, H., Gatellier, P., Lebert, A., Mirade, P.S., 2014. Building phenomenological models that relate proteolysis in pork muscles to temperature, water and salt content. *Food Chemistry*, 151, 7-14.

Kondjoyan, A., Daudin, J.D., 1997. Heat and mass transfer coefficients at the surface of a pork hindquarter. *Journal of Food Engineering*, 32, 225-240.

Rougier, T., Bonazzi, C., Daudin, J.D., 2007. Modelling incidence of lipid and sodium chloride contents on sorption curves of gelatine in the high humidity range. *LWT- Food Science and Technology*, 40, 1798-1807.

Ruiz-Cabrera, M., Gou, P., Foucat, L., Renou, J.P., Daudin, J.D., 2004. Water transfer analysis in pork meat supported by NMR imaging. *Meat Science*, 67, 169-178.

Taormina, P.J., 2010. Implications of salt and sodium reduction on microbial food safety. *Critical Reviews in Food Science and Nutrition*, 50, 209-227.

Toldrá, F., Flores, M., 2000. The use of muscle enzymes as predictor of pork meat quality. *Food Chemistry*, 69, 387-395.

MODELING AND SIMULATION FOR CASCADING FAILURE CONSIDERING AN AC FLOW POWER GRID

Woojin Kim[(a)], Chulhan Kim[(b)], Hyounglak Jun[(c)], Tae-Eog Lee[(d)]

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

[(a)] woojin.kim@kaist.ac.kr, [(b)]chulhan.kim@kaist.ac.kr, [(c)] hyounglak.jun@kaist.ac.kr, [(d)]telee@kaist.ac.kr

ABSTRACT
These days, there has been a growing interest in the importance of electricity in the human daily life. Importance of providing steady electricity from a power grid infrastructure has been continuously increased. However, the power grid infrastructure can fail by several reasons. The breakdown of a small part of the network can affect the whole. The electrical failure can spread in sequence while causing blackouts, and this process is called Cascading failure. Although extensive research has been carried out on this with DC approximated model, little attention has been paid to AC models. In this paper, we propose an AC flow model to conduct a simulation for cascading failure of an electrical power grid. South Korea's peak amount of electricity demand data is used for simulation. In conclusion, the flow of a power grid is calculated by the AC model which gives more realistic results than the DC model.

Keywords: cascading failure, blackout, AC flow, power grid

1. INTRODUCTION

Nowadays, there has been an increasing interest in the importance of electricity in human daily life. Electricity is used for almost all activities in life and inevitably, there is constant growth of the electricity demand. Hence, importance of providing steady electricity from a power grid infrastructure has been continuously increased. A power grid infrastructure is strongly connected with other infrastructures. For example, communication and railroad networks, and medical systems, are operated by electricity and can have negatively effects if the Power Grid Infrastructure encounters a problem.

The power grid infrastructure can break down due to several reasons. For example, natural disasters can physically damage a power plant or an electrical cord. Moreover, excessive power demand can overload power transmission, and this might cause failure in the overall infrastructure. An initial failure at one infrastructure can result in chain reaction of other failures of dependent infrastructures. The electrical failure can spread sequentially causing blackouts, and this process is called cascading failure. A failure might increase the amount of loaded electricity onto particular power transmission which can result in a situation of cascading failure.

There are several blackout examples caused by cascading failure. In 2012, the power grid of South Korea had been damaged because of typhoon Bolaven and cascading failure on power grid made blackouts on 2 million households. In August 2003, approximately 100 power plants and 50 million people experienced major blackouts due to one transmission line problem near the east coast of North America. This paper seeks to remedy cascading failure problems by forecasting the component that can influence the whole power grid when failure happened.

Forecasting the size of the cascading failure has been widely studied so far. First of all, Bienstock and Verma (2010) proposed a model that can calculate the survival time of the power grid when k components were deleted out of N grid components. Although, this model is not directly related to the cascading failure problem, however, many other researchers have developed cascading model based on it. Carreras and Dobson (2004) proposed a model that can calculate the amount of failures caused by an initial failure. An arithmetical model was developed by Kim, Bae, Bae, and Lee (2013) that can calculate a new power flow when a failure happens.

According to these studies, failures of a few components of a power grid cause serious cascading failure. To avoid the huge cascading failure, demand shedding policies were suggested by Kim, Bae, Bae, and Lee (2013). In addition, other researchers (Bienstock 2011; Pinar, Meza, Donde, Lesieutre 2010) have considered methods that shed load.

There is a large volume of published studies considering cascading failure with DC approximated model (Kim, Bae, Bae, Lee 2013; Pfitzner, Turitsyn, Chertkov 2011; Bernstein, Bienstock, Hay, Uzunoglu, Zussman 2011; Bienstock 2011; Dobson, Carreras, Lynch, Newman 2007; Dobson, Carreras, Newman 2005; Chen, Thorp, Dobson 2005; Carreras, Lynch, Dobson, Newman 2004; Bernstein, Bienstock, Hay, Uzunoglu, Zussman 2011). The DC model is able to provide solution, which contain flows of transmission lines and phases of power plants by using the linear programming or other methods. However, one of the critical constraints of the DC model is that the phase difference between the two-linked-plants that should converge to zero. The model cannot provide any solution when the difference has a value which is not close to zero. One of the major issues in DC model is that it does not reflect the flow that varies based on time series.

There are also several papers assuming deterministic demand quantity (Kim, Bae, Bae, Lee 2013; Bienstock 2011; Dobson, Carreras, Lynch, Newman 2007; Nedic, Dobson, Kirschen, Carreras, Lynch 2006; Dobson, Carreras, Newman 2005; Carreras, Lynch, Dobson, Newman 2004). However, the demand is not deterministic in reality, and it varies by time. Moreover, assuming the demand as a deterministic value might cause failures, and these failures affect cascading procedure. Thus, it is possible that the cascading procedure is misestimated with the assumption.

In the real case, a power plant operates an Alternating Current (AC) motor, and it has its own angular frequency. The frequency might cause a big difference among phases. In addition, the amount of flow changes by time during the rotation of the motor. Although, extensive research has been carried out on DC models, little attention has been paid to AC flow models.

Therefore, to overcome these limitations, a new power grid model is needed to provide solutions regardless of the phase difference. Thus, we suggest an AC flow model for simulating cascading procedure of a power grid.

2. POWER GRID CASCADING MODEL

2.1. AC Flow Calculation

AC flow model has more complicated conditions than a DC Approximation model, because it requires more complex data sets to explain current, voltage, and impedance. Impedance consists of reactance and resistance. We introduce the concept of complex number to deal with these two values independently. Accordingly, impedance is expressed with resistance as a real number and reactance as an imaginary number. This type of impedance influences not only the voltage, but also the flow by Ohm's law, meaning that the voltage and the flow should be expressed by the complex number. Then, the voltage, flow, and impedance have two dimensional values, and the values interact with each other.

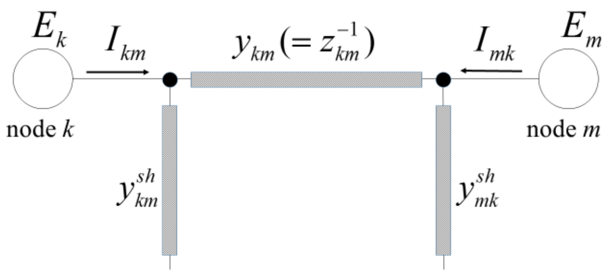


Figure 1: A diagram of node *k*, node *m*, and a transmission lines between these nodes.

The AC power grid model can be constructed by network modeling as shown in Figure 1. Transmission lines of a power gird are expressed by arcs, and other components, such as power plants and transforming stations, are expressed by nodes. The following terminologies can be defined.

E_k *: complex voltage of node k*

V_k *:voltage of node k*

U_k *: maximum voltage of node k*

I_{km} *: complex flow from node k to node m*

z_{km} *: impedance of an arc, connecting node k and m*

r_{km} *: resistance of an arc, connecting node k and m*

x_{km} *: reactance of an arc, connecting node k and m*

y_{km} *: admittance of an arc, connecting node k and m*

y_{km}^{sh} *: admittance of a shunt between node k and m*

g_{km} *: conductance of an arc, connecting node k and m*

b_{km} *: susceptance of an arc, connecting node k and m*

Admittance is inverse number of impedance. Then, the following equations can be found.

$$z_{km} = r_{km} + x_{km} j \tag{1}$$

$$y_{km} = z_{km}^{-1} = (r_{km} + x_{km} j)^{-1} = g_{km} + b_{km} j \tag{2}$$

$$g_{km} = \frac{r_{km}}{r_{km}^2 + x_{km}^2} \tag{3}$$

$$b_{km} = \frac{-x_{km}}{r_{km}^2 + x_{km}^2} \tag{4}$$

However, Andersson (2008) suggested an equation to calculate AC flow. In this equation, voltage, flow, and impedance are expressed with complex numbers.

$$I_{km} = y_{km}(E_k - E_m) + y_{km}^{sh} E_k \tag{5}$$

$$I_{mk} = y_{mk}(E_m - E_k) + y_{mk}^{sh} E_m \tag{6}$$

$$E_k = U_k e^{j\theta_k} = U_k(\cos\theta_k + \sin\theta_k j) \tag{7}$$

$$E_m = U_m e^{j\theta_m} = U_m(\cos\theta_m + \sin\theta_m j) \tag{8}$$

These equations show relationships between the phase of power plants and the flow of transmission lines. And *j* is the symbol of imaginary number. We take the advantage of the suggested equations (5), (6), (7), and (8) to calculate the AC flow.

In reality, an admittance of a shunt is a very small value, compared to an admittance of a transmission line. Moreover, there might be no shunt between some nodes. Therefore, a complex flow can be calculated without considering admittance of shunt as below.

$$I_{km} = y_{km}(E_k - E_m) \tag{9}$$

In the equation (7), if a condition $x_k^2 + y_k^2 = U_k^2$ is satisfied, then x_k and y_k can be introduced, instead of $U_k \cos\theta_k$ and $U_k \sin\theta_k$.

$$x_k = U_k \cos\theta_k \tag{10}$$
$$y_k = U_k \sin\theta_k \tag{11}$$
$$E_k = x_k + y_k j \tag{12}$$

Then, the complex flow is presented with x_k and y_k by using (7) and (9) as shown above.

$$I_{km} = y_{km}\{(x_k + y_k j) - (x_m + y_m j)\} \tag{13}$$
$$= (g_{km} + b_{km} j)\{(x_k - x_m) + (y_k - y_m) j\} \tag{14}$$
$$= \{g_{km}(x_k - x_m) - b_{km}(y_k - y_m)\} + \{g_{km}(y_k - y_m) + b_{km}(x_k - x_m)\} j \tag{15}$$

Therefore, an instantaneous value of AC flow i is the real number of the complex flow. A part of imaginary number indicates the phase of flow. In general, the real part is called flow, and the imaginary part is only used to explain the phase of a flow. In this study, we focus on finding out the flow, so that the phase is determined by the flow automatically.

$$i_{km} = g_{km}(x_k - x_k) - b_{km}(y_k - y_m) \tag{16}$$

Equation (16) is real part of I_{km}.Thus, it is the flow between node k and m, and a node mass can be calculated. The mass of node k is expressed with P_k, D_k, and zero. A node is a supply node if it has a positive node mass, which means the amount of supply generated by the node. On the other hand, node k is a demand node when it has a negative value of node mass. In this case, the amount of demand is indicated by D_k, and the node mass is -D_k. Otherwise, a node is an intermediate node when its node mass is zero.

$$\sum_{m=1}^{n} i_{km} = \begin{cases} P_k \\ -D_k \\ 0 \end{cases} \tag{17}$$

The number of nodes is n, and the node mass is the sum of flows, which are connected to the node. As a result, an AC flow model is introduced by using (16) and (17).

Maximize $\sum_{k=1}^{n} P_k$

Subject to

$$\sum_{m=1}^{n} i_{km} = \begin{cases} P_k \\ -D_k \\ 0 \end{cases} \tag{18}$$
$$g_{km}(x_k - x_k) - b_{km}(y_k - y_m) = i_{km} \tag{19}$$
$$x_k^2 + y_k^2 \le U_k^2 \tag{20}$$

Objective value of this AC flow model is total amount of power generation in order to conserve a given power demand. Constraints (18) and (19) are equivalent to (17) and (16). However, constraint (20) is slightly different from $x_k^2 + y_k^2 = U_k^2$, because U_k means a maximum voltage generated by node k. This constraints indicates that the node can generate less power than U_k. Therefore, constraint (20) can have the inequality.

The flow finding problem contains both linear and quadratic constraints. The quadratic constraints are used not only due to the flows are expressed with complex numbers, but also due to the constraints related with size of the flow. Moreover, the size of the complex flow is constructed by the combination of the square of real and imaginary parts, which should be less than a flow capacity. Due to the constraint, we cannot solve the problem by linear programming which is used in DC models. Therefore quadratic constraints programming is used to find the flows in the power grid.

2.2. Cascading Procedure

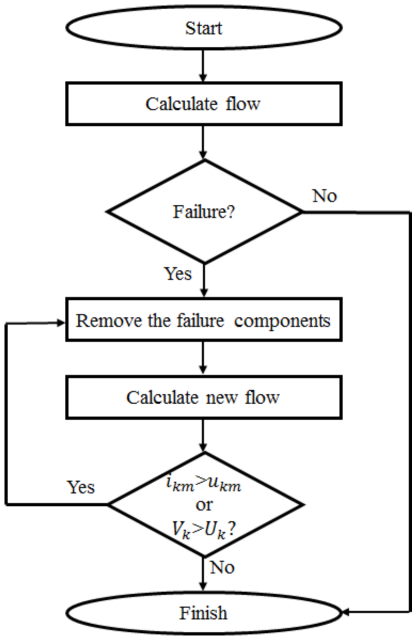


Figure 2: A flow chart for cascading procedure

In order to satisfy the power demand, power plants generate power and transmission lines transmit their flows. However, every component of a power grid has its own capacity. A power plant has a capacity of power generation, and a transmission line can only transmit electric current under its transmission capacity. Therefore, these capacities cause failure when some of the components have values over the capacity, and a failure component cannot work until it is repaired. In this study, the repair process is not considered. Thus, the failure components are removed from the power grid, and the remaining power grid without the failure components is obtained. However, it is difficult to satisfy the demand with the remaining power grid. It is because the demand could not be satisfied with the original power grid even though it has more components than the remaining power grid. Hence, the power grid without failure components is likely to contain additional failure. This procedure continues until finding a power grid which satisfies the power demand. This procedure is called a cascading procedure.

As shown Figure 2, we introduce a flow chart for cascading failure of an electrical power grid. This procedure starts with information about capacities and demands. By using (16), we can get the flow of transmission lines, and the amount of generation of power plants. Failure of a component is determined by checking its capacity. A constraint $i_{km} < u_{km}$ or $V_k < E_k$ means that a component, which is related with this condition, does not violate its capacity. If the conditions are violated by some components of the power grid, then the failure components should be deleted, and new flow and generation will be calculated without those components. This power grid repeats the same procedure until there is no failure.

A number of components is expected to be failed from this procedure, even though the number of initial failure components is very small. As a result, cascading failure can cause huge damage on the power grid. Therefore, analyzing this cascading failure is important to satisfy the power demand stably.

3. DEMAND AND SUPPLY

To calculate the flow, information about the power demand should be known. Most of other studies have been considered the demand as constant and fixed value. This value is given from demand forecasting and it indicates that the forecasted demand is not a real amount of demand. Moreover, the real amount of demand is changed by the time. Thus, calculating flow with the forecasted demand might lead to uncertainty. Furthermore, using the fixed demand can make flow of the power grid infeasible.

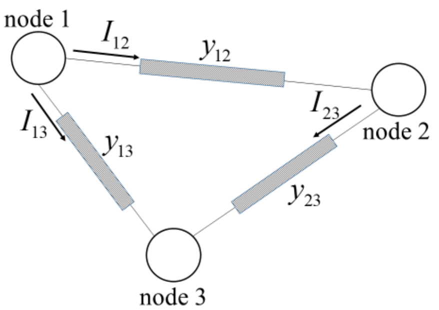


Figure 3: An example of a small power grid

Figure 3 shows a small power grid. Suppose that this power grid requires flow I_{12}=90, I_{13}=100, I_{23}=0, and all of the admittance are same. In this example, the amount of power generation can be calculated by (9).

$$I_{12} = y_{12}(E_1 - E_2) = 90 \tag{21}$$

$$I_{13} = y_{13}(E_1 - E_3) = 100 \tag{22}$$

$$I_{23} = y_{23}(E_2 - E_3) = 0 \tag{23}$$

Equation (23) indicates $E_2 = E_3$, and this relation leads to $y_{12}(E_1 - E_2) = y_{13}(E_1 - E_3)$ because y_{12}=y_{13} is given. Moreover, this relation means I_{12}=I_{13}, which is contradiction to (21) and (22). Therefore, this power grid has no feasible solution under this condition.

However, if a condition $I_{12} = 90$ changes to $I_{12} = 100$, then this power grid has a feasible solution with the increased amount of the total demand, because it can satisfy $y_{12}(E_1 - E_2) = y_{13}(E_1 - E_3)$.

This example of small power grid shows that a power grid might have failure components even though it can afford larger amount of the power demand without any failure. Furthermore, there is no reason to stick to the fixed value of the demand, which is not exactly same with the real demand.

Therefore, we introduce a model that the power demand and supply are in some ranges. In this study, the ranges are calculated by using the fixed demand.

$$I_{km} \in [0.9 \times I_{km}^{Fixed}, 1.1 \times I_{km}^{Fixed}] \tag{24}$$

$$I_{km} \in [1.1 \times I_{km}^{Fixed}, 0.9 \times I_{km}^{Fixed}] \tag{25}$$

Equation (24) works when I_{km} is a non-negative number. Otherwise, I_{km} satisfies (25).

4. DEMAND SHEDDING IN CASCADING PROCEDURE

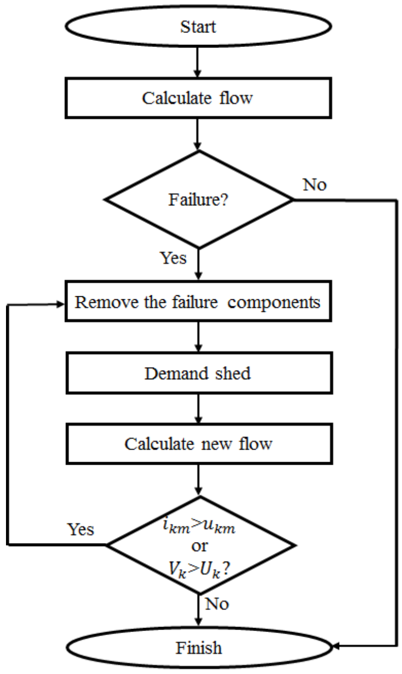


Figure 4: A flow chart of cascading procedure considering demand shedding

As treated in section 2.2., a small size of initial failure can cause huge damage on the entire power grid infrastructure. The huge damage occurs because a power grid cannot find any feasible solution in cascading procedure with a given power demand. Therefore, if the demand is reduced when failure occurs, then the power grid might find feasible solutions. This demand reduction is called demand shedding, and it relaxes the entire problem by reducing the power demand when the flow cannot be obtained to satisfy the power demand after the failure occurred. Therefore, demand shedding process is needed after the failure in cascading procedure. Figure 4 shows an updated flow chart of cascading procedure.

There are many strategies for demand shedding. Demand shedding strategies are suggested by Kim, Bae, Bae, and Lee (2013), and a strategy with the highest demand conservation was founded. The demand conservation indicates that how much demand is left after demand shedding, compare to initial demand. As a result, a proportional shedding strategy is proposed as the best strategy for demand conservation. Thus, in this study, proportional demand shedding strategy is adopted.

Proportional shedding does not give any priority to components of a power grid. When a power grid cannot afford a given total demand, proportional shedding strategy reduces demand of all nodes in same proportion. For example, there are two demand nodes, called A and B. Demand of these nodes are100 and 200, respectively. However, 30 of total demand should be reduced by proportional shedding. Then, the demand of A becomes 90, and the demand of B becomes 180. This is because proportion of reduced demand should be same on every node.

5. SIMULATION

In this study, a cascading simulator, which was developed by Kim, Bae, Bae, and Lee (2013), is used. It is extended with the AC flow model. Demand conservation and cascading procedure are results of the simulator. In addition, this simulation is replicated 30 times, with South Korea’s peak amount of electricity demand data on year 2013. It contains information of 837 nodes and 1148 arcs.

5.1. Failure model

A failure model should be defined before starting the simulation. Failure model determines whether a component of power grid fails or not. Exponential smoothing method is one of the ways, suggested by Kim, Bae, Bae, and Lee (2013), to calculate *effective flow* of an arc, and compare it with its *effective capacity*. If effective flow is bigger than effective capacity, then the arc fails.

$$\hat{f}^{r}_{km} = \alpha f^{r}_{km} + (1-\alpha) f^{r-1}_{km} \tag{26}$$

The effective flow $\hat{f}^{r}_{km}$ indicates that a temporary overloading on a transmission line does not make a failure. The transmission line will fail when it is overloaded continuously.

5.2. Simulation Result

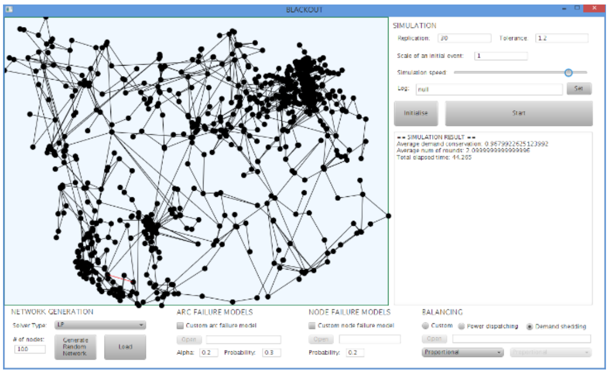


Figure 5: Simulation Result from cascading simulator with AC flow

The simulator shows a cascading procedure, and it helps to find which components are critically influenced by other components when failure is occurred. Result of the simulator contains electricity demand conservation ratio after the initial failure occurrence, and average number of cascading failure phase right before stabilizing of the power grid infrastructure.

As shown in Figure 5, the total elapsed time is 44.265 seconds for 30 times replication. It indicates that it takes about 1.5 second per one cascading procedure to calculate. Moreover, the demand conservation is about 96.80, and it shows that there are only 3.2% loss of demand during cascading procedure.

6. CONCLUSION

This study set out to determine an AC flow model using network modeling to simulate and analyze the cascading procedure when a power grid infrastructure breaks down. Moreover, the AC flow model can deal with the demand as bounded-range data, and provides solutions in the certain range of interval. This approach indicates that the model is flexible to solve a problem and reflects real cases properly.

Furthermore, we proposed a way to calculate a complex AC flow. This calculation may be able to applicable on other AC flow models, in order to derive the AC flow and phase of the components.

For further study, we need to modify the AC flow model. It has quadratic constraints, and it should be solved by quadratic constraints programming. The quadratic constraints programming takes more time than linear programming. Therefore, if the quadratic constraints are expressed in terms of linear constraints, then the total elapsed time might be diminished greatly. In addition, developing failure models and demand shedding strategies might provide a more realistic simulation results.

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 (2011-0029881).

REFERENCES

Chulhan Kim, Kiwook Bae, Soojeong Bae, and Tae-Eog Lee, 2013. Simulation analysis for cascading failure of an electrical power grid. *Proceedings of the 12th International Conference on Modeling and Applied Simulation*. September 25-27, Athens (Greece).

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.

Andersson, G., 2008. Modelling and analysis of electric power systems. *Power Flow Analysis-Fault Analysis Power Systems Dynamics and Stability*. Available from:http://educationdisha.com/upload/modelling_hs08_script_02.pdf[accessed 13 May 2014].

Dobson, I., Carreras, B., Lynch, V., 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.

Kirschen, D.S., Jayaweera, D., Nedic, D.P., Allan, R.N. 2004. A Probabilistic Indicator of System Stress. *IEEE Transaction on power system*, 19 (3) 1650-1657

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.

Bernstein, A., Bienstock, D., Hay, D., Uzunoglu, M., and Zussman, G., 2011. *Power Grid Vulnerability ot Geographically Correlated Filures – Numerical Evaluation*. Available from: http://www.ee.columbia.edu/~zussman/pub_files/CUEE-2011-05-06-power-cascades.pdf[accessed 13 May 2014].

Bienstock, D., 2011. Optimal Control of Cascading Power Grid Filures. Preceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, December 12-15, Orlando (USA).

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

Gurobi Optimization, 2013. *Gurobi optimizer reference manual*. Gurobi Optimization. Available from: http://www.gurobi.com/documentation/5.1/reference-manual[accessed 13 May 2014].

Nedic, D.P., Dobson, I., Kirschen, D.S., Carreras, B.A., and Lynch, V.E., 2006. 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.

AUTHORS BIOGRAPHIES

WOOJIN KIM received the B.S. degree in Hanyang University in 2013. He is in the Ph. D. program in KAIST. His main research interests focus on network modeling of infrastructures, and minimizing damages from disasters on networks especially on electrical power grids.

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 minimizing damages from disasters on networks especially on electrical power grids.

HYOUNGLAK JUN received the B.S. degree in Georgia Institute of Technology in 2013. He is in the M.S. program in KAIST. He mainly studies regression analysis and exponential smoothing method to develop an electricity demand forecasting model

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.

SIMULATION APPROACH FOR AIRCRAFT SPARE ENGINES & ENGINE PARTS PLANNING

José A. Ramírez-Hernández[(a)], Steven Oakley [(b)], Mei Zhang[(c)], Alejandro Scalise[(d)]

[(a), (b), (c), (d)] Operations Research & Advanced Analytics, American Airlines, Inc.

[(a)]Jose.Ramirez-Hernandez@aa.com, [(b)]Steve.Oakley@aa.com, [(c)]Mei.Zhang@aa.com, [(d)]Alejandro.Scalise@aa.com

ABSTRACT

We present the application of a simulation approach utilized for planning the required ownership levels of aircraft spare engines and parts in American Airlines. Such planning is essential to efficiently support both flying and engine maintenance operations. As such, this problem is very important from a financial and operational perspective given the high cost and criticality of these assets. Our models can be utilized in single and multi-location settings and are utilized to estimate the minimum ownership requirements to satisfy a given service level. In addition, our models can provide other important operational information such as out of service related metrics. To illustrate the utilization and versatility of the models, three case studies with actual industry data are presented. Results from these case studies demonstrate the value of the information generated by our models which facilitates the ownership planning and can also be used to support other related decision processes.

Keywords: supply chain, inventory planning, simulation, spare engine and parts planning

1. INTRODUCTION

Airlines own or lease aircraft spare engines to cover the operation while engines are overhauled or repaired. Due to the high cost of these assets, accurate planning is necessary not only to support the airline's flying operations, but also to avoid unnecessary excess in engine ownership that can be expensive for the company. Moreover, accurate planning of the required levels of engine spare parts is required to support the engine overhaul process, which is essential to provide on-time engine repair completions and avoid a negative impact in the spare engines availability.

In this paper, we present the application of a simulation approach to determine the required ownership of both engine spares and engine spare parts. Due to the particular complexity of the maintenance programs conducted and the uncertainty of engine removal and repair processes, a simulation approach is utilized to determine the required spare engine and the engine spare parts inventory levels. Our team has also derived closed-form formulations that can be utilized for fast ownership calculations under specific operational conditions. The simulation approach is preferred when a more detailed modeling of the process is required.

The problem addressed by this application belongs to the field of repairable inventory systems (Guide and Srivastava 1997; Tysseland and Halskau 2007). This problem differs from the classical problem of consumable inventories in the sense that the items in inventory can be repaired. The main reason for the repair operation is the high cost of the items in the system. Usually, such items are cheaper to repair than to replace for new ones. Thus, these systems are characterized by a flow of items being consumed but that later are removed and sent back to repair shops for reconditioning before returning to the system. As mentioned in (Guide and Srivastava 1997), these systems can be modeled as either single or multi-echelon systems depending on having single or multiple inventory locations.

In the context of the application presented here, the problem to solve can be simply described by the question: How many spare engines and engine spare parts does the company need to own in order to support the completion of the promised flying schedule? Planning for spare engines and its parts is an important problem for the company not only from the financial point of view but also from the operational perspective. On the one hand, engine spares and its repairable components or assemblies are expensive. The cost of a commercial airline aircraft engines usually ranges in the millions of dollars. For example, a CFM56 engine used in Boeing 737 aircraft can cost more than $10 million apiece (CFM International 2013). In the same way, some engine parts can reach the hundreds of thousands of dollars per unit. Thus, from the financial point of view, the ability to avoid an incremental spare engine/part purchase can provide significant benefits. Moreover, among all the maintenance, repair and overhaul (MRO) processes conducted in the airline industry, the engine repair is one of the most expensive operations (Ackert 2011).

On the other hand, accurate calculations of the required ownership of spare engines/parts are required to support maintenance and flight operations when such type of assets are repaired periodically. For instance, during the engine repair process, the company is required to have enough spare parts that will be used to replace the components being repaired. In some cases, the parts are repaired before or by the time these are required, but in many cases the repair times are longer than the time when the parts are needed to build an engine. In those cases, having the right ownership of spare parts is essential to provide an on-time delivery of the engine. If there are not enough engine spares parts available, then the risk of not having available spare engines is increased. Moreover, acquiring more spare engines or spare parts than necessary will help the availability of the equipment but it could also have an unnecessary negative financial impact. Thus, the approach presented in this paper aims to facilitate the estimation of spare engines and parts ownership needed to efficiently support both repair and flying operations while avoiding the excess in spare engine/parts inventory.

There is an important body of research in the field of repairable inventory systems for both single and multi-echelon cases, such as it is discussed in (Guide and Srivastava 1997; Tysseland and Halskau 2007). Although, to best of our knowledge, there are no specific papers describing simulation applications for both engine spares and engine parts in the context of commercial aviation, research has been conducted in the field of aircraft components in general. Interestingly, research in the field of repairable inventory systems is rooted on military applications for aircraft components, such as it is described by the seminal paper by (Sherbrook 1968) which focuses on the mathematical formulation to optimally stocking repairable parts for aircraft in a multi-location setting. The formulation; however, requires to adhere to specific mathematical assumptions. A more recent paper in the context of aircraft components is presented by Simao and Powell (2008), which provides the application of an approximate dynamic programming approach for a multi-echelon problem, where a combination of simulation and optimization techniques are utilized to determining the inventory levels of aircraft components at different locations in the system while maximizing the financial benefits from such decisions. This approach is applied at the higher level considering warehousing, locations, suppliers, but without entering in specific details of the repair process of the assets.

The application presented in this paper considers both a single and two-echelon repairable inventory system, referred hereafter as single-location and multi-location models, respectively. In contrast with previous work, our application is focused on the repair processes of engines and engine parts in the context of commercial aviation, and considers specific modeling details proper of the repair processes of these types of assets. Such features include borrowing or cannibalization of parts, scrapping processes, modeling of capacity constraints for the repair shops in the system, and work conducted pre and post the introduction of engines into the repair shops. In addition, our models consider so-called harvesting processes in which available stored or retired engines with remaining useful life can be advantageously reconditioned at smaller costs and with shorter turn-times than other more expensive repair programs. By using simulation, our application also allows increased flexibility to model complex details proper of engine repair processes as well as additional flexibility to use any probabilistic distribution that could be found to properly fit, e.g., repair times, demand patterns, transportation times. Moreover, our models provide the possibility to assess the impact that the planning of engine parts or “shop pool” can have in the spare engine ownership, and also provide information of other important performance metrics such as Out of Service (OTS) events due to unavailability of spare engines. Currently, our models are being utilized by the company for decision support in the ownership planning process of four different types of engines.

In terms of previous practices in the company, many of the procedures employed in the past to plan for the ownership of spare engines and parts involved the use of manual calculations. These were prone to errors and did not consider the variability of the repair processes. Thus, our application also represents a step forward towards a more accurate and reliable approach to estimate the ownership requirements to support the company’s engine maintenance operations.

It is important to mention that all the procedures, simulation, and algorithmic details of this application are part of an original invention property of American Airlines, Inc., which is patent pending according to the United States patent law, and based on corresponding provisional filing conducted before the United States Patent and Trademark Office.

The organization of this paper is as follows: section 2 provides an overview of the process modeling and simulation. A description of the implementation of the models and corresponding calculation tools is given in section 3. In section 4 we present three case studies to illustrate how our simulation application has been utilized in different analysis and for decision support purposes. Finally, in section 5 we provide some conclusions.

2. PROCESS MODELING & SIMULATION

In this section we provide details of the process modeling and simulation implementation. The application considers single-location models for both spare engines and engine spare parts. Also, a multi-location model is utilized for spare engines when there are several stations with an allocation of spare engines to support the flying operations. An overview of these models is presented next.

2.1. Single-Location Model for Spare Engines

2.1.1. Simulation Process Overview

Figure 1 below illustrates the flow diagram followed by the engine repair process for the single-location case which utilized in the simulation. As depicted in the figure, the process starts with the arrival of engines for repair, where each arriving engine corresponds to a removal from an aircraft due to either failure or a planned maintenance procedure. Actual data indicates that Bernoulli or Poisson process provide good approximation of the arrival or removal process. However, the model also allows the use of other arriving processes if necessary.

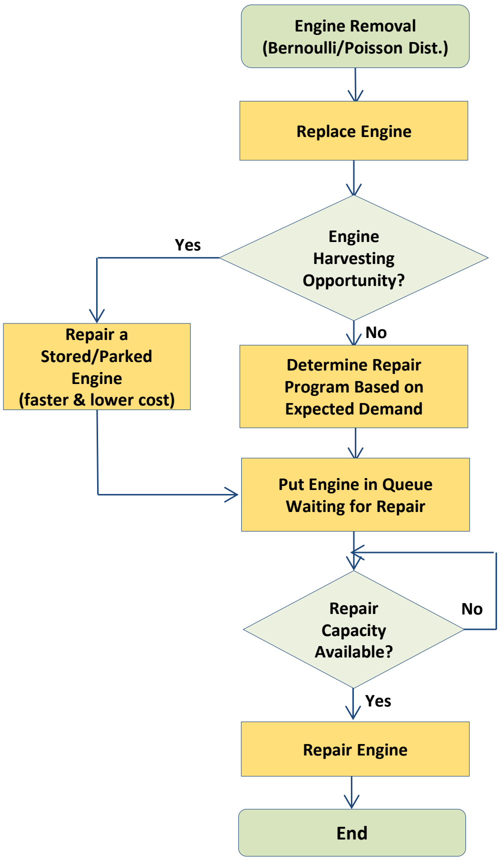


Figure 1: Flow Diagram for the Engine Repair Simulation Process, Single-Location Case.

Once an engine is removed and arrives for repair, the spare ownership level is updated given that each engine removed is replaced by an available spare. Then, a decision is made to determine if a harvesting process can be conducted. Harvesting is the process in which the company can defer the removed engine from being repaired, and instead, advantageously reconditioning a stored or previously retired/parked engine with remaining useful life but with lower repair costs and shorter repair turn-around-time than the removed engine. If harvesting is not an option, then a regular repair program is assigned. Such repair programs can be of a light type in which case the time needed to complete the repair is usually short. Similarly, there are also more extensive repairs usually referred as “heavy” which are more expensive and require longer turn-times than light repairs. The expected number of repairs to be conducted under each repair program is obtained from probabilistic engine removal forecast models developed by the company.

As indicated in Figure 1, the repair process could also have capacity constraints that may limit the number of engines that can be repaired concurrently. Thus, engines that arrive for repair are initially included in a queue where these wait until capacity is available to introduce the engine into the repair shops. While in queue, the engine can undergo a series of paper work and other procedures in preparation for the actual repair. Once an engine is sent into the shops, there is a repair time required to complete the reconditioning which is dependent on the repair program. In our application, repair times are random and are modeled using probabilistic distributions. For instance, we tested normal and gamma distributions by fitting them to various types of engine repair data using the method of moments. The gamma distributions produced more accurate fits overall and we concluded they were appropriate for the repair time distributions. At the end of the process, after completing the repair, the engine becomes available to the system and the engine spare ownership level is properly updated to reflect the addition of a new spare.

Some of the basic parameters utilized for modeling the single-location engine spare model include demand and repair time distributions, desired service level (percent of successfully fulfilled spare requests), and repair shop capacity constraints, e.g., maximum number of engines allowed under repair. Most of these parameters are obtained from historical data available from the company’s information systems and forecasting models. Moreover, there are specific parameters for the simulation runs that need to be defined: number of replications, simulation length, and warm-up period. In the case of this model, a total of 100 replications of 20 years each have been utilized with a warm-up period of one year. The warm-up period was determined by examining simulation outputs (replications) and determining the time required to reach the steady-state condition. Also, the number of replications and simulation length were selected after testing different combinations of these two parameters in order to achieve a desired accuracy in the estimated values. For instance, under the chosen values of simulation length and replications, we obtained a tight 95% confidence interval for the estimated ownership, with a lower and upper bounds that deviated from the average value by around 1%, but without sacrificing in simulation speed for practical purposes.

2.1.2. Model Output

The output generated by the simulation model corresponds to the variation of the engine spare level in time due to events such as engine removals and repair completions. Company proprietary statistical and optimization methods are then utilized to estimate the minimum required levels of total spare ownership such that a pre-specified service level is achieved.

In addition to the estimation of the total spare ownership requirements, the simulation model can be utilized to evaluate the performance of the system under the calculated ownership requirements. For instance, using specific ownership levels our models can be utilized to generate OTS-related metrics including the expected number of OTS events per year and statistics of the duration of such events. Similarly, the simulation output provides statistics of the WIP at the repair shops, average number of engines waiting in queue before being introduced for repair, and average spare count (ASC).

2.2. Single-Location Model for Engine Spare Parts: Shop Pool

In the case of engine spare parts, the modeling and simulation process follows a similar approach to that of the spare engines model. Again, the goal is to estimate the total required ownership of engine parts based on a pre-specified service level. However, in this case it is necessary to consider additional details proper of the engine parts repair process also called "piece-part repair" (PPR) process.

Figure 2 below depicts a general version of the simulation process for the repair of engine parts. Again, the process starts with the removal and arrival of an engine for repair. After arrival, a repair program is assigned and the engine parts are separated from the main engine assemblies. It is important to indicate, that an engine can have multiple parts of the same type, e.g., blades, vanes. Thus, at the time of separating the parts from the engine assemblies, there are components for which multiple units of the same type are sent for repair.

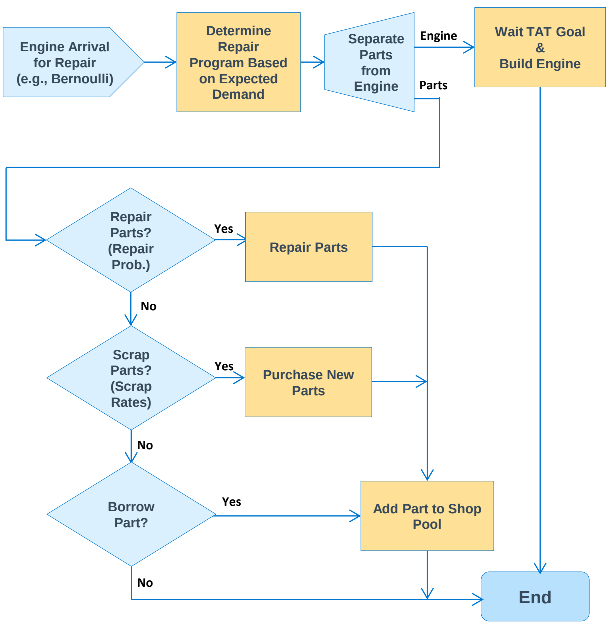


Figure 2: Flow Diagram for the Engine Spare Parts Repair Simulation process, Single-Location Case.

Once separated from the engine, the parts are evaluated to decide if repair is necessary, or there is need for scrapping the parts, or if the part is in good condition and can be borrowed to be used in the building of other engines downstream in the process. When the parts are sent for repair, the delay in the repair process is modeled using a probabilistic distribution, e.g., gamma. After completion of the repair process, the repaired parts are added to the available inventory or shop pool.

In the process there is also the possibility of having the parts scrapped. In that case, a purchasing order is generated and a random lead-time is used to model the delay after which the new purchased parts are back and available in the shop pool. The repair process of the parts and the engine is usually guided by turn-around-time (TAT) goals for the completion of the repair process. Thus, when parts are separated from the engine, the main assemblies wait for a specific amount of time or TAT goal before collecting the required spare parts from the shop pool and continue to complete the repair process. In some cases this goal is much shorter than the time required for completing the repair of parts. Thus, specific level of available inventory in the shop pool is necessary to support the repair process.

The parameters utilized by this model include, among others, demand and repair time distributions, repair probabilities, scrapping rates, capacity constraints, and desired service level (percent of successfully fulfilled spare part requests). As in the case of the spare engine model, most of the parameters are obtained from historical data available from the company's information systems. Other parameters such as the number of expected engine removals (demand) are estimated using probabilistic forecasting models developed by the company. For practical purposes, the model has been designed to be run for individual engine parts. That is, repair time, repair probabilities, and scrapping rates, and other parameters are specified at the part level. Similarly, the model requires specifications for the simulation runs. For example, this model has been run using 500 replications of 15 years each with a warm-up period of one year which is appropriate to obtain enough simulation samples in steady state conditions. In this case; however, more simulation replications were utilized to obtain a tight 95% confidence interval for the estimated ownership (upper and lower bounds with around 1% deviation from average value). The reason for this is the increased variability of the simulation output given there are components with multiple units per engine.

2.3. Multi-Location Model for Spare Engines

The multi-location spare engine model can be seen as a two-echelon system with repairable components in which different locations or stations can have a spare allocation to support the operation. In addition to the stations, the system includes a location with engine repair shops and where engines are sent back and forth after removals and repair completions, respectively.

Figure 3 below illustrates the flow diagram of the simulation process for the multi-location model. As before, the process starts with the removal of engines due to failures or planned maintenance procedures. In this case, the removal process occurs at different stations. To model such removal processes, Bernoulli and Poisson distributions have been found to be a good fit based on actual data.

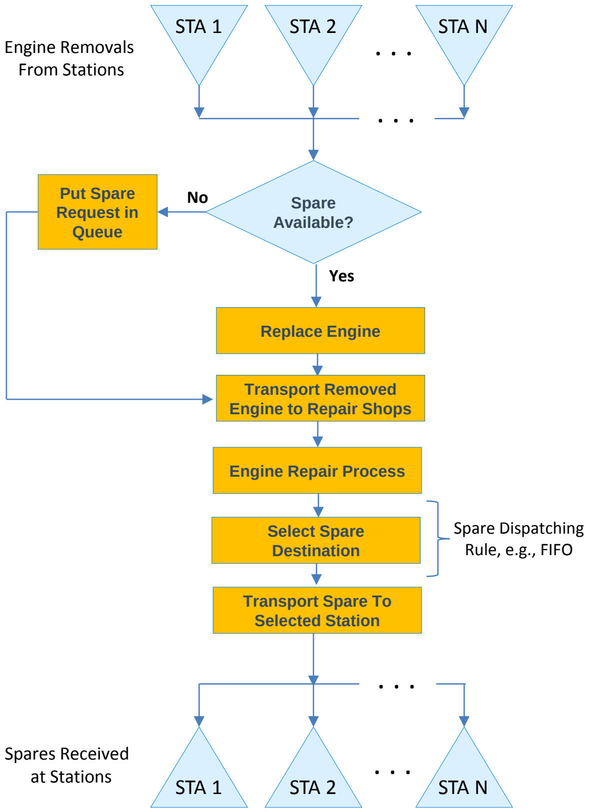


Figure 3: Flow Diagram for the Multi-Location Engine Spare Simulation Process.

As soon as a removal occurs at a station, then a new spare available from the shelf is utilized to replace the removed unit. However, if there are no available spares, then an OTS is generated and the spare request is put in queue until it is fulfilled by using a spare sent back from the repair shops, or in some cases, from another close by location that could allow the borrowing of a spare. All the removed engines are then transported to the repair shops and such transportation is modeled as a delay to arrive to the engine repair location. Once the engines arrived to the repair shops these follow the same engine repair process depicted in Figure 1. After completing the repair, a spare dispatching decision is conducted to select which station will receive the new spare. The selection is made based on the queue of outstanding spare requests from the different stations and by using specific dispatching rules. Some of the rules utilized include static type of rules such as First-In-First-Out (FIFO), and longest transportation time (LTT), as well as dynamic policies based on current inventory levels. After applying one of these rules and selecting the spare destination, then a transportation process is conducted before the engine arrives to the corresponding station.

In this model each of the stations in the system has a defined spare allocation, and each station can be measured in terms of service level and OTS-related metrics. Again, the output generated by the simulation is utilized to determine the total spare engine ownership per station that will meet a desired service level. Moreover, the output generated by the model allows measuring the performance at system-wide level, e.g., system-wide service level and expected number of OTS events per year.

The model utilizes different process parameters, including removal rates by station, desired service level by station, and transportation times between stations and the repair shops. Again, most of these parameters are obtained from actual data available from the company's information systems. Also, parameterization of the simulation runs is required. For instance, the model has been run using 50 replications of 30 years each with a warm-up period of one year.

3. IMPLEMENTATION OVERVIEW

The models described in the previous sections have been implemented within so-called calculation tools for the end-user. Figure 4 below depicts the general architecture followed in the implementation of these tools. As illustrated in the figure, the implementation includes both a user and an external server side. The user side has a Graphical User Interface (GUI) built in MS-Excel and using Visual Basic (VBA), for easy use and portability. The GUI is utilized to facilitate the configuration and specification of the parameters required by the model as well as to execute the calculations (run simulations). In the case of the engine spare model, the GUI offers the possibility of setting multiple scenarios, each with a different set of parameters, which allows the comparison of ownership requirements for different operational conditions.

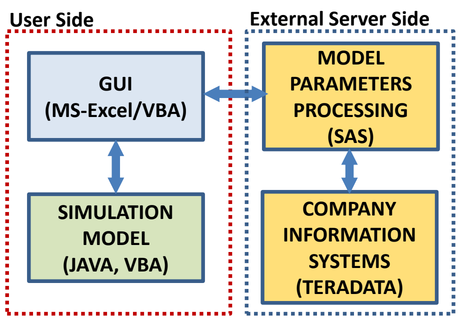


Figure 4: General Implementation Architecture for the Spare Engine and Spare Parts Simulation-based Calculation Tools.

Also within the user side of the implementation is the simulation model which uses the parameters specified by the user in the GUI. In this application the simulation models have been implemented using Java (as independent executable files) and VBA (integrated within MS-Excel). Prototyping of the models were also conducted in Arena (Kelton et al. 2006) before implementing the models in VBA and Java. In particular, Java allows fast simulation runs which is advantageously utilized for the engine spare parts model where the model is run for dozens of parts and speed is a factor. For developing purposes, for example in Java, in addition to the basic programming libraries (Oracle Corporation 2014) we also utilized the Commons Math library (The Apache Software Foundation 2014) for random number generation and statistical functions. Moreover, the discrete event simulation model was built utilizing a fixed time increment simulation clock with a 1 day interval. For each interval, the number of arrivals, processing times, repair completions and other events were simulated using the corresponding probabilistic distributions, and to update the state of the system. The approach taken to simulate the system was to build a non-terminating simulation that allowed us to study the steady state behavior. We utilized Java1.7 to code the model, including a calendar of events, queues counters, and the statistics engine.

On the external server side of the implementation, we have the company information systems, e.g., Teradata (Teradata Corporation 2014), which are used to extract historical information needed to obtain the parameters utilized by the models. The required processing of these data is conducted using specifically designed code in SAS (SAS Institute Inc. 2014). The resulting parameters after this processing are sent back to the GUI and are then made available for simulation and calculation purposes.

As part of the model development, a process of verification and validation (Sargen 2010) of the simulation models was conducted. First, the conceptual models were verified by the business units in charge of the process, e.g., engine repair production control and asset management teams, to verify that the assumptions and modeling details were correct. Then, a process of verification of the programming language code utilized to implement the models was conducted to ensure that the code was properly representing the conceptual model. In addition, verification of the models was also conducted by comparing the simulation output against closed-form formulas utilized to estimate the ownership under more relaxed conditions, e.g., infinite capacity. For validation purposes, historical data of spare requirements based on actual demand was utilized and compared with the estimations provided by our models. That is, under the same actual demand conditions, our models provided close calculations of the ownership requirements, with only small deviations from actual requirements.

Finally, and although a formal accreditation process has not been conducted, the models have been extensively and successfully evaluated by final users in the engine production control and asset management groups in American Airlines. Among the different groups involved in the development of these models, there has been consensus on the fact that simulation was the right approach to provide ownership estimations given the complexity of the process. The positive acceptance of implemented models in the business units has also facilitated the expansion on the application of our models to estimate ownership requirements for other type of aircraft assets such as Auxiliary Power Units (APU).

4. CASE STUDIES

The implemented models described in this application have been utilized for different analysis aimed to support the company's decision process regarding engine spares and parts planning. To illustrate the use of our models, we present next three case studies considering both single and the multi-location scenarios.

4.1.1. Impact of Engine TAT in Spare Ownership & Shop Pool Investment

The first case study is focused on the impact that the engine repair TAT has on both the engine spares ownership and the engine spare parts or shop pool investment. The shop pool investment is directly proportional to the required ownership, and in this case, it represents the aggregated amount from all parts considered in the calculations (more than 200). This case study was part of an analysis conducted with the objective of selecting a feasible and appropriate engine TAT goal for the engine repair shops and compatible with the company's flying operations objectives.

Figure 5 below shows some of the results obtained from this analysis. Please notice that the values of the scales in the chart of Figure 5 are omitted because these correspond to company proprietary information.

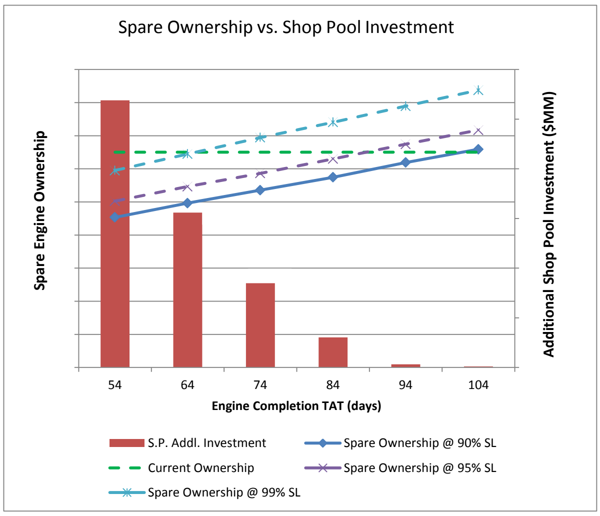


Figure 5: Impact of Engine Repair TAT in Spare Ownership and Shop Pool Additional Investment.

We conducted calculations of the spare engines and engine parts ownership for engine repair TAT's ranging from 54 to 104 days in average. As indicated in the chart, the additional investment in the shop pool (obtained at a 98% service level) is decreasing with the increment in the engine TAT. Conversely, as the engine TAT is increased, the engine spare ownership also increases. These results are expected given that increasing the engine TAT will allow additional time for completing the repair of engine parts, which leads to smaller shop pool requirements. However, increasing the engine TAT will also lead to increments in the engine spare ownership levels.

The chart in Figure 5 also shows the required spare ownership for different service levels (SL): 90%, 95%, and 99%. As indicated in the figure, higher service levels required higher engine spare ownerships. As a reference, the chart also includes the current ownership at the moment of the analysis (dashed line). For example, the results illustrated in the chart indicate that for the current ownership it is possible to obtain up to 99% service level when the engine repair TAT is about 64 days. However, such conditions also require an additional shop pool investment. Thus, the results generated by the models can provide valuable information to evaluate the trade-off between engine spare ownership, shop pool investment, and engine repair TAT.

4.1.2. Impact of Engine Spare Borrowing Between Stations on the Duration of OTS Events

The reduction or mitigation of OTS events is an important task in the planning and management of spare engines. Thus, using the simulation models presented here, the company is able to conduct analyses to estimate the impact that certain planning strategies may have in the occurrence of OTS events.

In this particular case study, the goal was to determine the impact that borrowing of engine spares between stations could have in the reduction of the average duration of OTS events. For this purpose, we considered the possibility of transporting available spare engines from a selected station to other stations in the system that could have outstanding spare requests. This task was conducted by setting a threshold point for the inventory of the station borrowing the spares. Thus, if the inventory level was equal or larger than such threshold, then it was allowed to transport a spare from the borrowing station to another station holding an outstanding spare request. Figure 6 below illustrates the results obtained across the different stations and by using borrowing threshold levels from 0 to 5 spares.

For the experiment conducted, the repair base and station was selected as the location from which spares can be borrowed. The reason for this was that such station not only had the largest spare inventory but it is also located at the same place as the repair location, which provides added flexibility in handling the inventory. Results from this case study indicate that decreasing the borrowing threshold has a positive impact in the duration of the OTS events, such as it is illustrated in the chart from Figure 6. That is, when more spares are allowed to be borrowed from the repair and base location, then the time that stations with outstanding spare requests have to wait is reduced. In particular, notice how stations 2, 3, and 4 have a significant reduction (more than 50%) in the duration of OTS events when borrowing is allowed compared to the case of no borrowing at all. Also notice that the repair base station was not negatively impacted in the duration of OTS events when borrowing of spares was allowed.

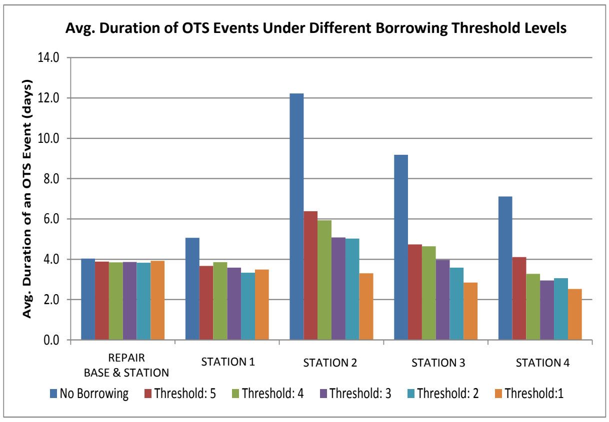


Figure 6: Impact of Engine Spare Borrowing Between Stations in the Duration of OTS Events.

In fact, there is also a slight reduction in the OTS duration observed at that station. Thus, in this case study our models are useful to demonstrate the positive impact that the borrowing of spares between stations has in reducing the average duration of OTS events.

4.1.3. Impact of Engine Harvesting Process on the Service Level

Another type of decisions that the implemented simulation model is able to support, are those of determining the level of engine harvesting required to achieve specific service levels in the system.

As mentioned earlier, the engine harvesting process consists in the reconditioning of stored or retired engines which still have useful working life, but that require less expensive repair procedures and shorter repair turn-around times than other more extensive repair programs. Thus, if the initial plan considers heavy type of repairs on several engines, but there is an opportunity to conduct a harvesting process, then the heavy repairs may be deferred and instead the harvesting candidate engines could be reconditioned and made available as spares.

In this case study we considered the impact that different levels of harvesting have in the service level observed across the different stations. The harvesting level is defined here as a percent of heavy repairs that are deferred to conduct harvesting of stored or retired engines instead. We tested harvesting levels from 0% to 30% of the planned heavy repairs while maintaining the total ownership constant. Figure 7 below illustrates the service level obtained at the stations under the different levels of harvesting. It also shows the impact in the system-wide service level.

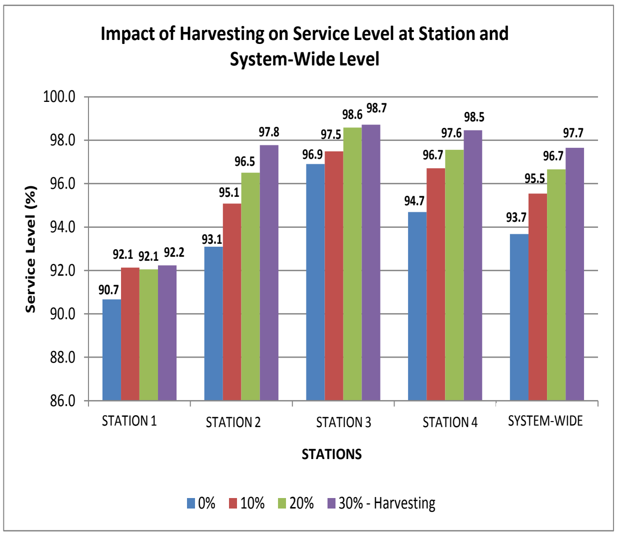


Figure 7: Impact of Engine Harvesting Process on Service Level at Station and System-Wide Level (multi-location case).

The results indicate that the harvesting process has a positive impact in the service level observed not only at station level but also system-wide. As expected, the service level is increasing with the level of harvesting. That is, as more engines can be harvested instead of conducting the heavy repairs that require long turn-times, then the new spares are made available faster to the stations, which in turn, leads to a reduction of number of OTS events and, consequently, to the increase of the service level. Moreover, notice that in this case the harvesting process is advantageously utilized to increase the service level without increasing the total ownership which is constant. Thus, important financial benefits can be obtained for the company by avoiding the need of additional investment in either new spares or expensive heavy repairs while achieving higher service levels. As an example, notice that station 2 increases its service level from 93% to about 98% by allowing a 30% of harvesting in the process.

5. CONCLUSIONS

In this paper we presented the application of a simulation approach for planning the required ownership levels of spare engines and parts. In maintenance operations for the airline industry this is an important problem not only because of its financial significance due to the high value of the engines, but also because of the need to effectively and reliably support the company's flying operations.

The implemented simulation models have demonstrated to be a useful tool not only for planning but also for decision support as illustrated in the different case studies presented here. Moreover, the models implemented have demonstrated its versatility in providing additional performance measures such as OTS-related metrics, which are useful for planning and selecting optimized strategies for engine maintenance operations.

In the context of aircraft maintenance operations, engines are not the only high value assets for which the models can be applied. There are an important number of other repairable aircraft components for which the models can be also applied. Currently, the company is in the process of extending the application of these models to other type of assets.

Finally, there is still opportunity to expand and refine the models described here by including additional and more sophisticated features. For instance, advanced simulation-based optimization approaches could also be considered in the case of the multi-location model in order to better tune the required inventory levels at different stations while meeting specific performance requirements.

ACKNOWLEDGMENTS

The authors thank Matt Pfeifer, Richard Czuchlewski, and Juan Leon from Operations Strategic Planning at American Airlines, as well as the Engine Production Control team at the American Airlines Tulsa Maintenance Base, for all their support and the valuable insight provided in the development and refinement of the models described in this paper. Special thanks to Jim Diamond, Managing Director of Operations Research & Advanced Analytics in American Airlines, for supporting the publication and presentation of this paper in the 13th International Conference on Modeling and Applied Simulation.

REFERENCES

Ackert, Shannon 2011. *Engine Maintenance Concepts for Financier*, Aircraft Monitor. Available from: http://www.aircraftmonitor.com/uploads/1/5/9/9/15993320/engine_mx_concepts_for_financiers___v2.pdf [accessed 20 March 2014]

The Apache Software Foundation 2014. *The Apache Commons Mathematics Library*. Available from: http://commons.apache.org/proper/commons-math/ [accessed 7 July 2014]

CFM International 2013. *2013 CFM56 and LEAP engine orders soar*, CFM Press Center. Available from: http://www.cfmaeroengines.com/press/2013-cfm56-and-leap-engine-orders-soar/722 [accessed 20 March 2014]

Guide, V. Daniel R., Srivastava, Rajesh, 1997. Repairable inventory theory: Models and applications. *European Journal of Operational Research*, 1 (102), 1–20.

Kelton, W. D., Sadowski, R. P., Sturrock, and D. T., 1997. *Simulation with Arena*, 4th ed. USA: McGraw-Hill, 2006.

Oracle Corporation 2014. *Java SE at a Glanse*. Available from: http://www.oracle.com/technetwork/java/javase/overview/index.html [accessed 7 July 2014]

Sargent, Robert G., Verification and Validation of Simulation Models, *Proceedings of the 2010*

Winter Simulation Conference, pp.166-183. December 5-8, Baltimore, MD (USA).
SAS Institute Inc. 2014. *SAS® 9.4*. Available from: http://www.sas.com/en_us/software/sas9.html [accessed 26 March 2014]
Sherbrooke, C. (1968): METRIC: *A multi-echelon technique for recoverable item control*, in Guide Jr., V. Daniel R., and Srivastava, Rasjesh (1997): *Repairable inventory theory: Models and applications. European Journal of Operational Research*, 1 (102), 1-20.
Simao, Hugo, Powell, Warren, 2008. Approximate Dynamic Programming for Management of High-Value Spares. *Journal of Manufacturing Technology*, 2 (20), 1–20.
Teradata Corporation 2014. *Teradata Database.* Available from: http://www.teradata.com/products-and-services/Teradata-Database [accessed 26 March 2014]
Tysseland, B.E., Halskau, H. 2007. Spare Parts Inventory – A Literature Review with Focus on Initial Provisioning and Obsolescence Management. *The 19th Annual NOFOMA Conference Proceedings*, pp. 1075-91. June 7-8, Reykjavic (Iceland).

JOSE A RAMIREZ-HERNANDEZ received the Ph.D. degree in Electrical Engineering from the University of Cincinnati, Cincinnati, OH, in 2010. He has B.S. (1995), Licenciatura (1996) and M.S. (1999) degrees (both with honors) in Electrical Engineering from the University of Costa Rica, San Jose, Costa Rica. From 1996 to 2001, he was an Adjunct Professor in the School of Electrical Engineering at the University of Costa Rica; from 2001 to 2007, he was member of the Systems Modeling and Information Technology Laboratory (SMITLab) at the University of Cincinnati; and since 2010 he works as Senior Operations Research Consultant within the Operations Research & Advanced Analytics and Operations Strategic Planning groups at American Airlines, Inc. His research interests include simulation, Markov decision processes, dynamic programming, and simulation-based optimization methodologies with applications to commercial aircraft maintenance operations. Dr. Ramirez-Hernandez is also member of IEEE, INFORMS, and the Omega Rho honor society.

STEVEN OAKLEY Received the Ph.D. in Systems and Industrial Engineering from University of Arizona (UA), Tucson, AZ, in 1990, his M.S. in Industrial Engineering (UA), 1986, and his B.S. Engineering Mathematics (UA, High Distinction), 1985. In 1990-91, he was Visiting Assistant Professor at the University of Tennessee in Knoxville, TN. Since 1991 he has been a technical consultant for American Airlines in the areas of revenue management and maintenance and engineering. He specializes in applied probability modeling and statistical science. Dr. Oakley's professional affiliations include INFORMS, American Statistical Association, and Tau Beta Pi.

MEI ZHANG received the Ph.D. degree in Industrial and Systems Engineering from Georgia Institute of Technology in 1997, M.S. in Geography from University of Tennessee, Knoxville (UTK) in 1993, B.S. in Mathematics from UTK in 1991. She has worked at Sabre and i2 Technology briefly before joining American Airlines in 2001. Currently, she is Senior Manager of Operations Research and Advanced Analytics at American Airlines supporting Technical Operations (Tech Ops) which includes the Maintenance and Engineering organization. Dr. Zhang and her team apply business process modeling, optimization, simulation and statistical analysis methodologies to help business units to improve efficiencies in their planning and operations. Dr. Zhang is a member of INFORMS and IIE.

ALEJANDRO SCALISE received the Ph.D. in Industrial Engineering from University of Pittsburgh (UP), Pittsburgh, PA, 2001. M.S. in Industrial Engineering (UP), 1999, B.S. in Industrial Engineering, Universidad Nacional de Cuyo (UNC) Mendoza, Argentina, 1986. Between 2001 to 2004 he was Airport Consultant at TransSolutions, Fort Worth, TX. Since 2004 he has been a consultant for American Airlines in the areas of airports and customer service, providing advanced analytics consulting for decision support. Dr. Scalise specializes in simulation modeling, operations research, and computer programming.

ASSESSING RISK AND UNCERTAINTY: A COMBINED APPROACH BASED ON AHP METHOD

Fabio De Felice[(a)], Antonella Petrillo[(b)]

[(a)] Department of Civil and Mechanical Engineering - University of Cassino and Southern Lazio, Italy
[(b)] Department of Engineering - University of Naples "Parthenope", Italy

[(a)]defelice@unicas.it, [(b)]antonella.petrillo@uniparthenope.it

ABSTRACT
A successful organization recognizes that when an effective strategy is properly implemented, it will result in a sustainable competitive advantage. But when you examine the formulation of an organizational strategy, you quickly realize that strategy is really about choice.
In this context project risk management based on Analytical Hierarchy Process (AHP) approach is a systematic process of identifying, analyzing and responding to risks manage. This paper presents the Analytical Hierarchy Process (AHP) as a potential decision making method in project risk management field. The final aim of our work is to define a model for evaluating the performance of product development in order to measure their achievement and activate pathways for improvement.

Keywords: project risk management, , decision support system, AHP, competitive advantage

1. INTRODUCTION
The competitiveness in global and local markets highlights the importance of design, quality, productivity, multi-company collaboration, optimal price levels and production process predictability (Susterova *et al.,* 2012). When an industrial company launches a new product at the market the goal is to obtain a viable business (Longo *et al.,* 2012).

Acceptable levels of product qualities such as safety, performance, reliability and security are critical successful factors. Acceptable levels include maximizing the probability and consequences of positive events and minimizing the probability and consequences of adverse events to project objectives (Dey, 2001). We sought to address the following research question: *How can appropriate performance strategies be selected and integrated into process models that target a specific product quality?*

The success parameters for any project are on time completion, within specific budget and with requisite performance (technical requirement). It is necessary to develop strategies and measures to manage these risks (Chapman, 2006).

As the interval of technical innovation cycles has become shorter, the life cycle of products has been shortened. Due to the diversification of customer needs, the functions and performance of products should be improved quickly.

In the past, a number of systematic frameworks have been proposed for use in the risk-evaluation phase of the risk management process. Kangari and Riggs (Kangari and Riggs, 1989) classified these methods into two categories: *classical models* (i.e. probability analysis and Monte Carlo simulation), and *conceptual models* (i.e. fuzzy-set analysis).

In fact diverse risk factors that occur during product development are obstacles for the successful development of new products. From this point of view, project risk management is faced with decision environments and problems in projects that are complex. Relationships between elements of a problem may be highly nonlinear; changes in the elements may not be related by simple proportionality (Kwak and Anbari, 2009).

Project risk is an uncertain event, feature, activity or situation that can have a positive or negative effect on the outcome of a project. Project risk and opportunity management formally identifies, assesses and plans for uncertainty. Many studies on risk analysis and management have been performed, but systematic research on how a risk management system is built has been rare (Park *et al.,* 2011).

In this paper, a *decision support tool* method is proposed, to help development teams choose appropriate quality performances across the lifecycle.

This is based on three perspectives: *product-quality-risk management, process integration and cost/benefit*.

Decision making is difficult enough as it is, especially when decisions you are making are based on incomplete information, uncertainty, and lack of freely available resources. If you combine those difficulties with an approach to decision making that is unstructured, inefficient, personality driven, and full of analysis paralysis, you will definitely not get the outcomes you want. This quandary is compounded when the types of decisions you are making are predictive in nature since you may not know for years if choices you make today are wise or foolish (De Felice and Petrillo, 2009).

At every stage of the decision making process, misperceptions, biases, and other tricks of the mind can influence the choices we make. Highly complex and important decisions are the most prone to distortion because they tend to involve the most assumptions, the most estimates, and the most inputs from the most people (Poveda-Bautista *et al.*, 2013). Since there are few scientific risk management systems available for new product development to predict risk factors and to prepare for responding activities against each risk factor, in the present paper a combined approach based on a particular multicriteria decision-making method (MCDM) called Analytic Hierarchy Process (AHP), is applied to weigh the degree of importance of the strategies identified.

We present a proposed method and validation from an industry case study. The paper is organized in the following way: Section 2 discusses work related to this research and identifies the gaps of knowledge in the area of project risk management. Section 3 elaborates the methodology for the research based on AHP. Section 4 introduces the conceptual framework of the methodological approach. Section 5 demonstrates the application of the proposed framework. Section 6 provides a detailed discussion and conclusion on the application of the proposed framework.

2. RISK MANAGEMENT AND PRODUCT DEVELOPMENT PERFORMANCE

Product development projects should include also risk assessment, that allows managers to identify and measure the risks associated with resource constraints and then develop appropriate responses (Bruzzone *et al.*, 2008). Many studies on risk analysis and management have been performed, but systematic research on how a risk management system is built has been rare. In particular, there are few systematic studies on the establishment of risk management systems for new product development (Cooper, 2003).

Desired product quality attributes can be achieved by using specific processes. Appropriate techniques, methods and tools can be applied to analyse, avoid, reduce, minimise and eliminate the risks related to products development. Jones *et al.* emphasizes that the risk management process must be an integral part of the quality management system. Management literature from various perspectives contains empirical and theoretical discussions of how firms develop new products. Although differences in emphases exist, especially with regard to how researchers believe firms should generate new product ideas or manage this process, overall, there is a surprisingly wide area of agreement (Cusumano and Nobeoka, 1991).

Product development is view as problem solving that needs to understand user (market) needs and then match these needs with the capabilities of particular technologies, rather than letting technology overly influence the development process (Schmidt and Calantone, 2002).

Figure 1 outlines the framework used in this article to analyze major features for product-development strategy.

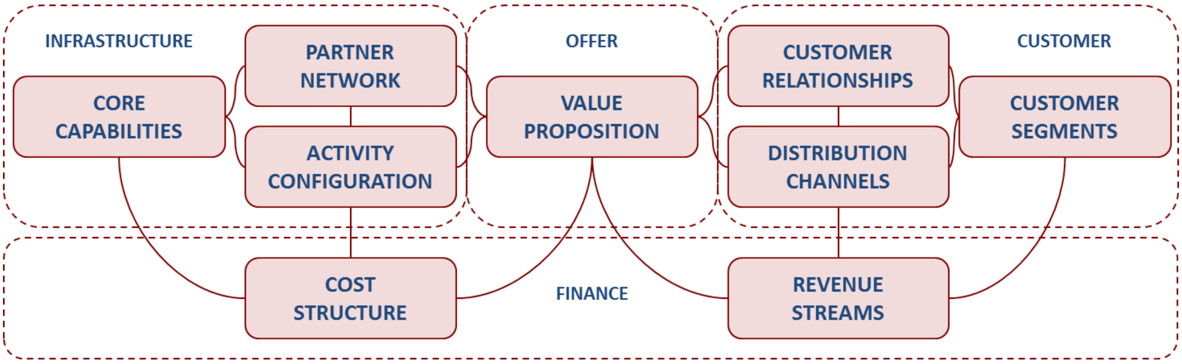


Figure 1: Organizational model

The above model describes the value an organization offers to various customers and portrays the capabilities and partners required for creating, marketing, and delivering this value and relationship capital with the goal of generating profitable and sustainable revenue streams.

3. THE ANALYTICAL HIERARCHY PROCESS (AHP)

A product is a set of benefits offered for exchange and can be tangible (that is, something physical you can touch) or intangible (like a service, experience, or belief) (Karnie and Reich, 2011).

We apply concepts from risk management to accommodate multiple quality attributes.

Numerous multicriteria decision making (MCDM) methods have been developed to help with decision problems by evaluating a set of candidates against pre-specified criteria. Examples of MCDM include Multi-Attribute Utility Theory (MAUT) (Keeney and Raiffa, 1993), AHP (Saaty, 1980), outranking techniques (Roy, 1996), weighting techniques (Keeney, 1999) and fuzzy techniques (Fuller and Carlsson, 1996).

In this research, we apply AHP. It is widely used for many practical decision-making problems in industry and academia. The decision-making process in AHP is based on relative assessment. In AHP, all candidates are evaluated using pairwise comparisons. As a result, the evaluation is less sensitive to judgment errors when compared to other MCDM methods using absolute assignments.

AHP method is based on three fundamental principles: *decomposition of the structure, comparison of judgments and hierarchical composition* (or synthesis) of priorities. AHP is applicable to decision situations involving subjective expert judgments and uses both qualitative and quantitative data. This method creates a priority index for each expert decision or judgment. AHP summarizes these judgments by ensuring their consistency.

The proposed approach involves the AHP method for the paired comparison of the risk factors, which was carried out. AHP allows to:

- Facilitate key decision makers to identify relevant criteria;
- Provide an approach to weight decision criteria and objectives
- Identify the best choices from a set of potential alternatives
- Allocate critical resources to “best-value” projects
- Generate advanced portfolio analysis reports: risk and “what-if” scenarios.

The main results that can be achieved are:

- Rapidly achieve consensus and buy-in to decisions.
- Make more informed decisions that can adjust and you or your environment changes.
- Improve transparency of key decisions by provide a repeatable method of tracking, auditing and improving decisions over time.

The strength of this approach is that it organizes tangible and intangible factors in a systematic way, and provides a structured yet relatively simple solution to the decision-making problems (Al-Harbi, 2001). Then over time, the project portfolio could be optimized as the needs of the business change.

The Analytic Hierarchy Process enables decision makers to structure decisions hierarchically: the goal of the decision at the top, strategic objectives in the higher levels, evaluation criteria in the middle, and alternative choices at the bottom (as shown in Figure 2).

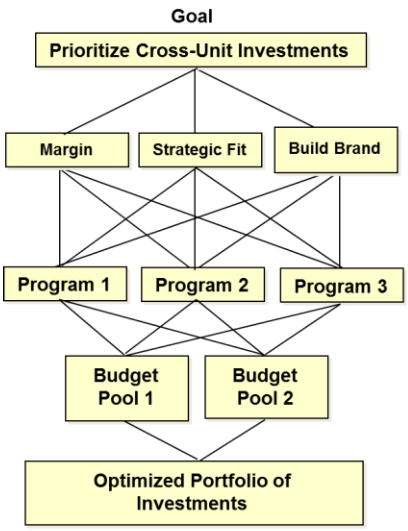


Figure 2: Sample of hierarchy

We establish a relative order of importance for business process improvement projects based on this four-step process:

1. Develop a hierarchy of business drivers and criteria—from high level drivers at the top to more specific criteria that will be used to measure the value of projects.
2. Compare business drivers, then use decision criteria to determine their priorities in helping the organization be successful.
3. Rate projects against the criteria using accurate numerical scales derived through pairwise comparisons.
4. Optimize the allocation of resources (human and financial) by maximizing value for cost based on well-understood business rules (interdependencies, must fund projects, time based allocations). This can be accomplished using linear and integer optimization techniques.

AHP creates a structured baseline for continuously improving decision making processes in an organization, which results in higher levels of efficiency and effectiveness (De Felice and Petrillo, 2013). To properly manage a business process improvement program aligned with an organizational strategy, strategy focused organizations should use AHP (Figure 1).

The most critical factor in AHP is to perform *pairwise comparisons* to develop relative weights on criteria hierarchy.

Participants perform multiple sets of comparisons for each level of hierarchy (as shown in Figure 3). For each judgment, participants determine which criterion is more important and by how much. Judgments are used to form ratios in a matrix; The matrix is used to calculate priorities for the judgment set (eigenvector). The Scale for Pairwise Comparison is a comparison scale of 1–9. The consistency index (CI) for all matrices of judgment are calculated according to: CI = $(\lambda_{max} - n)/(n-1)$ where λ_{max} is the maximum eigenvalue and *n* is matrix dimension.

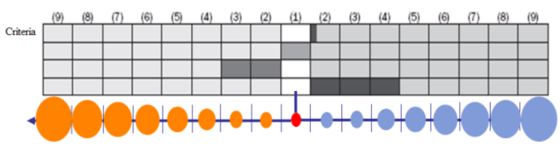


Figure 3: Sample of pairwise comparisons

A sensitivity analysis can be performed to check the sensitivity of the final decisions to minor changes in judgments (Saaty, 2005).

In conclusion, the feature of combining both quantitative and qualitative data and controlling the consistency of expert judgments makes AHP the most applicable to the proposed approach.

4. PROPOSED METHODOLOGICAL APPROACH

The model is based on teamwork and knowledge of multicriteria analysis techniques. It should be noted that multi-criteria analysis is used partly to compare the risk factors, not to compare the risks identified.

Figure 4 shows the general architecture of the proposed methodology. The proposed approach is divided into three phases and each phase is divided into steps. This approach outlines all phases of risk management including: (1) risk identification; (2) risk assessment and (3) actions.

The proposed approach is divided into three phases and each phase is divided into steps:

- *Phase I: Initial State.* The aim of the present phase is to assess the ideal positioning of the company or in other words the *"desidered performance"*. It evaluates the alignment of practices according to the contextual conditions internal and external to the organization. This phase is characterized by the following two steps:
  - *Step 1: Define the problem.* The measure of the complexity of the product-market ratio is determined through the administration of a questionnaire (32 questions).
  - *Step 2: Define Best Practices*. The aim of this step is to identify a set of "best practices" that if used correctly allow the achievement of performance targets. In particular, we investigate 19 best practices. In doing so, we tried to focus on the goals that we had put in terms of *completeness, simplicity and functionality*.

After steps 1 and 2 the answers to the questionnaire are crossed with the best practices and using the method IMS (Independent scoring method) in order to assign all best practices the optimum level (on a scale from 1 to 4), thus reaching the identification of Desideres State.

In particular, a matrix (32 x 19) is built and a correlation coefficient (0 = no correlation , 1= slight correlation, 3 = good correlation, 9= complete correlation) is assigned to each row-column intersection. To calculate for each practice the optimal level L, Equation 1 is used:

$$L_i = \frac{\sum_{h=1}^{32} R_h \cdot d_{ih}}{\sum_{h=1}^{32} d_{ih}} \qquad (1)$$

Where i = 1, ..., 19 h = 1, ..., 32

- *L_i = perfect level for each practice (Desideres State) = R_h values of the answers given in the questionnaire*
- *D_{ih} = correlation coefficients*
- *The value of Li is then normalized in the range 1-4.*

- *Phase II: Current State.* The aim of the present phase is to analyze the current state (the practices actually in use in the company) according to defined best practices. In the phase II a second questionnaire is administered. This phase is characterized by the following step:
  - *Step 3: Gap Analysis.* In this step a gap analysis is conduct between the two profiles (initial state and current state).
- *Phase III: Final State*. The aim of the present phase is to define a degree of importance for all of the good practices identified. This phase is characterized by the following steps:
  - *Step 4: Identify alternative project*. In order to assign weights to the different orientations and then to find the rankings of importance of the strategies Analysis Hierarchy Process approach is used (AHP Matrix).
  - *Step 5: Priority Map.* A two-dimensional mapping of the practices on the Cartesian plane Gap - Significance is defined.

In a similar way as seen in Phase I, is built a second relational matrix whose rows contain the 4 orientations and columns coincide with the 19 best practices. The level of importance of a certain practice is obtained through Equation 2:

$$I_i = \sum_{j=1}^{4} P_j \, C_{ij} \qquad (2)$$

where

- *P_h = values obtained from the AHP matrix*
- *C_{ih} = coefficient of correlation*
- *i = 1, ..., 19 h = 1, .., 4*

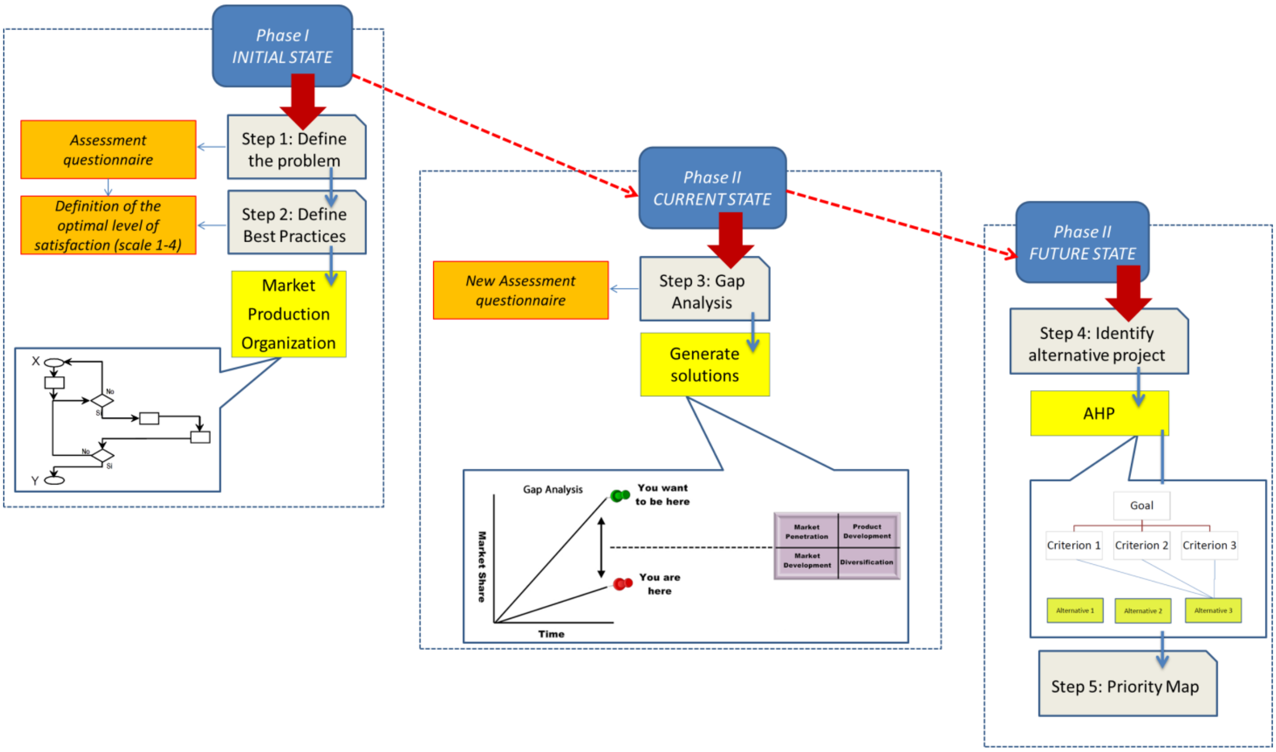


Figure 4: Structure of the methodological approach

5. APPLICATION OF THE PROPOSED APPROACH

In this paragraph the methodological approach is applied to a real case of a multinational company leader in power technologies and automation. In Figure 5 are shown some products.

Figure 5: Products

5.1. Phase I: Initial State

Phase I is the most critical phase because in the present phase it is necessary define the problem and best practices

Step 1: Define the problem. In order to define the "state of the art" a questionnaire is administered to management staff. It consists of 32 statements for which a rating is sought. The degree of agreement or disagreement is expressed in value from 1 to 10 with 1 being the highest and 10 the maximum disagreement agreement. In Table 1 is shown an extract of the questionnaire.

Table 1: Sample of questionnaire

Extract of questionnaire	Average
1. Question 1: The diversity of products is expanding	7,10
2. Question 2: Our products meet customer demands	6,50
3. Question 3: We are competitive	4,20
4. Question 4: We use an appropriate technology	5,15
….	
32. Question 32: Our suppliers are reliable	6,30

Step 2: Define Best Practices. In order to identify best practices 3 dimensions were chosen: market; production and organization.

In Table 2 is shown the Best Practices list.

Table 2: Best Practices List

Best Practices (BP)	
Market	1. New product development strategy 2. Diversity of products 3. Needs of customers 4. Product specifications 5. Marketing new products 6. Eco design
Production	7. Involvement of production personnel 8. Setting goals cost / investment 9. Production strategies 10. Involvement of suppliers 11. Robust Design 12. Production launch new products 13. Integrated product / process
Organization	14. Continuous Improvement 15. Management leadership 16. Project Management 17. R & D 18. New product development team 19. Automated processes

In Table 3 is shown a sample of matrix (32x19)

Table 3: Sample of matrix (32 x 19)

Questions/ Best Practices	#BP 1	#BP 2	#BP 3	#BP 4	…	#BP 19
#Q1	9	0	9	1	…	3
#Q2	3	0		9	…	9
#Q3	1	9	3	9	…	3
#Q4	9	3	1	9	…	1
….	3	9	3	3	…	1
#Q32	1	9	3	9	…	3

For all the best practices a level of satisfaction (scale 1-4) was identified. In Table 4 is shown average level for each practice.

Table 4: Level of satisfaction (L_i)

BP	L_i	BP	L_i
1	3,28	11	2,26
2	3,08	12	2,21
3	2,78	13	2,64
4	2,79	14	3,00
5	2,70	15	3,38
6	2,26	16	2,84
7	2,26	17	2,78
8	3,05	18	3,05
9	2,52	19	3,11
10	2,34		

5.2. Phase II: Current State

Step 3: Gap Analysis. From the comparison between Initial State and Final State differences arise between the two profiles (Table 5).

Table 5: Gap Analysis

BP	L_i	L_{i*}	ΔL
1	3,28	3,16	-0,12
2	3,08	3,02	-0,06
3	2,78	3,13	0,35
4	2,79	3,04	0,25
5	2,70	3,19	0,49
6	2,26	3,17	0,91
7	2,26	3,04	0,78
8	3,05	3,07	0,02
9	2,52	2,93	0,41
10	2,34	3,01	0,67
11	2,26	3,06	0,8
12	2,21	3,12	0,91
13	2,64	2,99	0,35
14	3,00	3,03	0,03
15	3,38	3,10	-0,28
16	2,84	3,05	0,21
17	2,78	3,01	0,23
18	3,05	3,04	-0,01
19	3,11	2,96	-0,15

In Figure 6 is shown the graph on Gap Analysis.

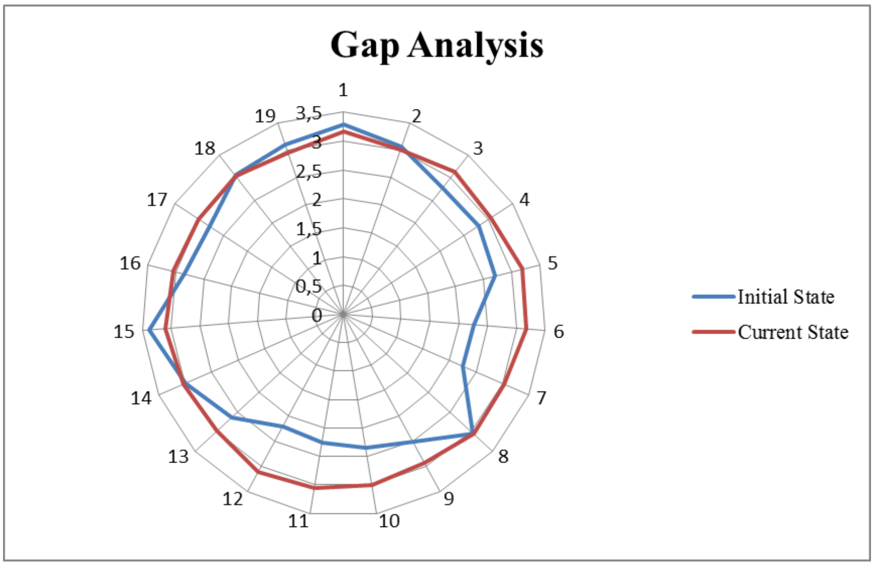


Figure 6: Gap Analysis

5.3. Phase III: Final State

Step 4: Identify alternative project. It is necessary to establish the strategic choices that will guide the development activities by identifying four strategic thrusts: Time to market (C1); Product Cost (C2); Performance/technology (C3) and Quality/reliability (C4). For this purpose AHP Matrix is built.

Table 6: AHP Matrix

AHP Matrix					
	C1	C2	C3	C4	Weight
C1	1	5	5	5	56%
C2	1/5	1	5	5	23%
C3	1/5	1/5	1	1/3	7%
C4	1/5	1/3	1/3	1	14%
tot	1,60	6,53	14,0	9,33	100%

Table 1: Single Line Table Caption

BP	I	BP	I
1	7,57	11	3,00
2	2,00	12	6,42
3	3,42	13	6,42
4	2,25	14	1,56
5	5,50	15	8,47
6	1,86	16	5,51
7	2,87	17	2,62
8	3,29	18	3,75
9	4,05	19	3,00
10	5,18		

Step 5: Priority Map. In Figure 7 Priority Map is built.

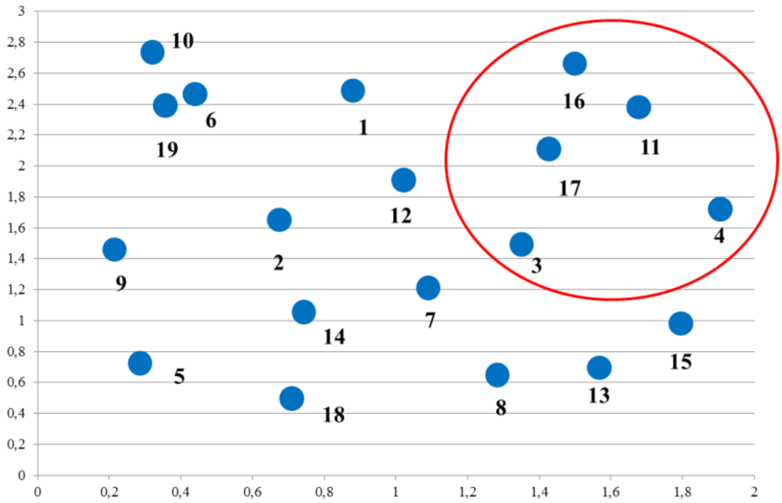


Figure 7: Priority Map

Our study shows that specific risks are being perceived in several projects. These risks might be inspected thoroughly to find ways for structural improvement on them. The framework can also be beneficial for other companies. R&D management can take the framework and customize it for use in their projects. For this, they need to take the model and delete from it what is not relevant and to add what is missing. In this way, the framework can be given a first customisation round. In subsequent use at starts-ups or during different development phases, the model could be refined and customized further.

6. CONCLUSION

We developed an integrated model that was applied to manage the performance of product development in order to measure their achievement and activate pathways for improvement.

The study results indicate the practical feasibility of our integrated model, which includes an innovative design tool and an MCDM framework for innovative and sustainable product development. Of course the proposed model cannot claim to include every risk issue that may appear during a specific development project. The challenge is to have and use an approach that stimulates people involved to identify risks, while there is time to take action to manage them. For this, some kind of formal risk assessment needs to take place during a long period time. Making a judgment on perceived risks, involves the integration of a large amount of information. Therefore, we conclude that it is preferable to add a structured and systematic component to the process of risk identification. The company investigated can use the framework as one of their tools to make people aware at the start and during the development of new products of the risks that are associated with their projects. Further use of the framework for the research company might include efforts to improve their practices.

REFERENCES

Al-Subhi Al-Harbi K. M., 2001. Application of the AHP in project management. I*nternational Journal of Project Management*, 19 (2001) 19±27.

Bruzzone, A., Massei, M., Pierfederici, L., 2008. Advanced models for innovative decision support systems in broadcasting schedule planning & management. *7th International Workshop on Modeling and Applied Simulation, MAS 2008, Held at the International Mediterranean Modeling Multiconference, I3M 2008*, Pages 123-129.

Chapman, C., 2006. Key points of contention in framing assumptions for risk and uncertainty management. I*nternational Journal of Project Management*, 24, 303–313.

Cooper, L.P., 2003. A research agenda to reduce risk in new product development through knowledge management: a practitioner perspective. *Journal of Engineering and Technology Management*, 2003.

Cusumano, M.A., Nobeoka, K., 1991. Strategy, structure and performance in product development: Observations from the auto industry. *Research Policy* 21 (1992) 265-293

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.

Felice, F., Petrillo, A., 2009. Proposal of a new model for the optimization of the organizational process in industrial company through the application of the analytic network process. *8th International Workshop on Modeling and Applied Simulation, MAS 2009*. Puerto de la Cruz; Spain; September 23 - 25 September 2009, Pages 92-100.

Dey, P.K., 2001. Decision support system for risk management: a case study. *Management Decision* 39 (8), 634–648.

Fuller, R., Carlsson, C., 1996. Fuzzy multiple criteria decision making: recent developments. *Fuzzy Sets and Systems* 78 (2) 139–153.

Jones, P.L, Jorgens J. IIII, Taylor, A.R. Jr., Weber, M., 2002. Risk management in the design of medical device software systems, Biomedical Instrumentation & Technology, 36, 237–266.

Karnie, A., Reich, Y., 2011. Managing the Dynamic of New Product Development Processes. A new Product Lifecycle Management Paradigm. Springer. p. 13. ISBN 978-0-85729-569-9.

Kangari, R., Riggs, L.S. 1989. Construction risk assessment by linguistics, IEEE Transactions on Engineering Management 36 (2) 126–131.

Keeney, R., Raiffa, H., 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*, Cambridge University Press.

Keeney, R., 1999. Foundations for making smart decisions, *IIE Solutions* 31 (5) 24–30.

Kwak, Y.H., Anbari, F.T. 2009. Analyzing project management research: perspective from top management, International Journal of Project Management 27 (5) (2009) 435–446.

Longo, F., Massei, M., Nicoletti, L., 2012. An application of modeling and simulation to support industrial plants design. *International Journal of Modeling, Simulation, and Scientific Computing*, Volume 3, Issue 1, March 2012.

Park, S., Kim, J., Choi H-C., 2011. A Risk Management System Framework for New Product Development (NPD). *International Conference on Economics and Finance Research*, vol.4, IACSIT Press, Singapore.

Poveda-Bautista, R., García-Melón, M., Baptista, D.C., 2013. Competitiveness measurement system in the advertising sector. *SpringerPlus*, Volume 2, Issue 1, 2013, Pages 1-14.

Roy, B., 1996. *Multicriteria Methodology for Decision Aiding*, Kluwer Academic Publishers, 1996.

Saaty, T.L., 1980. *The Analytical Hierarchy Process*, McGraw-Hill.

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.

Schmidt, J.F., Calantone, R.J., 2002. Escalation of commitment during new product development.

Journal of the Academy of Marketing Sciences, 30(2): 103-118.

Susterova, M., Lavin, J., Riives, J., 2012. Risk management in product development process. Annals of DAAAM for 2012 & Proceedings of the 23rd International DAAAM Symposium, Volume 23, No.1, ISSN 2304-1382.

AUTHORS BIOGRAPHY

Fabio De Felice, Professor at the University of Cassino and Southern Lazio, board member of several international organizations and responsible for scientific research and training in industrial plants. Holder of the course planning and management of industrial plants.

The scientific activity developed through studies and researches on problems concerning industrial plant engineering. Such activity ranges over all fields from improvement of quality in productive processes to the simulation of industrial plants, from support multicriteria techniques to decisions (Analytic Hierarchy Process, Analytic Network Process), to RAMS Analysis and Human Reliability Analysis. The main courses, both curricular and specialistic for territorial businesses in which he is involved are: Safety of Industrial Plants, Industrial Production Management, Industrial Simulation, Human Reliability Analysis. General Secretary of the Analytic Hierarchy Process – AHPAcademy - International Association for the promotion of multi-criteria decision making methods.

Antonella Petrillo, degree in Mechanical Engineering, PhD at the University of Cassino and Southern Lazio. Now Researcher at University of Naples “Parthenope” (Department of Engineering) where she conducts research activities on Multi-criteria decision analysis (MCDA), Industrial Plant, Logistic and Safety.

A SYSTEMS THEORETICAL PERSPECTIVE ON DATA-DRIVEN MODELING AND SIMULATION

Yilin Huang[(a)], Mamadou D. Seck[(b)], Alexander Verbraeck[(a)]

[(a)] Systems Engineering and Simulation, Delft University of Technology
[(b)] Engineering Management and Systems Engineering, Old Dominion University

[(a)] y.huang@tudelft.nl, a.verbraeck@tudelft.nl [(b)] mseck@odu.edu

ABSTRACT
This paper discusses Data-Driven Modeling and Simulation (DDM&S) from a systems theoretical viewpoint. The focus is on understanding Systems Theory in relation with Modeling and Simulation (M&S) and how data may convey information about a system. Such an understanding could help formalizing the transformation process from data to systems knowledge and in turn offers a possibility to automate the process. In DDM&S, a simulation environment should have the ability to correctly interpret the data provided, extract useful information and automatically incorporate it into the simulation model so that changes in data may change the model. This flexibility could be achieved by representing the system not as a monolithic whole but as a composition of interrelated parts, and by the support of intelligent data analysis and model transformation algorithms. An example of automatic model generation is given to illustrate the concept.

Keywords: data-driven, modeling and simulation, systems theory, model transformation

1. TOWARDS DATA-DRIVEN MODELING AND SIMULATION

The role of data in Modeling and Simulation (M&S) is becoming increasingly important along with the advances in data collection and storage technologies. Despite the discussions about input/output data analysis in literature, little is said about more comprehensive use of data in the M&S process. Data that can be useful for M&S is not limited to measurements or observations from real systems but can also comprise data that is generated from computer-aided tools such as Computer-Aided Design, Engineering and Geographic Information Systems. Data may be used for different purposes such as model construction, configuration and calibration. Fully utilize data in M&S could not only benefit the M&S process but also better integrate M&S with its applications such as design, engineering and training.

Data-Driven is not a new term in M&S but its earlier meaning differs from how it is understood now. Data-driven simulators were once those that allowed users to build models without programming knowledge (Ekere & Hannam, 1989; O'Keefe & Haddock, 1991; Pidd, 1992). Those programs accepted front-end entry data such as flow diagram or GUI input, which have been used since the 1950s. This is not how we conceive of Data-Driven Modeling and Simulation (DDM&S) nowadays. DDM&S discussed in this paper refers to M&S processes where data is used to automatically define, specify and/or modify model properties and logics. The simulation software is designed to correctly interpret the data provided, extract useful information and automatically incorporate it into the model so that changes in data can change the model correspondingly.

In DDM&S, the use of data may range from model construction, configuration and initialization, to model update, estimation and calibration. Simulation models should be developed to allow for such automated data uses. The primary concern here is that we need to create simulation software that provides enough degrees of freedom to represent the set of systems in our domain of interest. This flexibility can be achieved by representing the system of interest not as a monolithic whole but as a composition of interrelated sub-systems (or parts). In this context, we can see the definition of model parts as the "*fixed*" which represents the corresponding domain knowledge, whereas the selection of those parts, their parameterization, initialization, and composition are the "*unfixed*" whose degrees of freedom are constrained by the corresponding domain knowledge. The unfixed affords the flexibility. In such cases, data can be used to specify the unfixed, and changes in data may change the unfixed. The cost of this flexibility is complexity in simulation software. The design choice is then often a compromise between flexibility and complexity.

Although the DDM&S concept is straightforward, developing DDM&S software introduces complexity. The effort can be nonetheless beneficial for large-scale systems that need long-term use of simulation models. These models often take long to develop and modify, and they often need to be changed along with changes in the systems. Many examples can be found in supply chains, transportation and manufacturing among other domains (Kim, 2009; Qiao, Riddick, & McLean, 2003; Tannock, Cao, Farr, & Byrne, 2007; Wang, Chang, Xiao, Wang, & Li, 2011). Some systems may need simulation in (near) real-time so that forecasting can be made responding timely to crucial events. Examples can be found in fire forecasting, water management, critical infrastructure and emergency response systems where sensor data are often used (Darema, 2011; Hu, 2011; Jahn, Rein, & Torero, 2012).

In literature, e.g., Banks, Carson, Nelson, & Nicol (2010); Law (2007), the steps in a simulation study and data analysis within are well discussed. Many discuss data from a statistical perspective. In this paper, as a complement to statistical views, we discuss data from a systems theoretical perspective. We believe this could help formalize transformation processes from data to systems knowledge and could in turn offer a possibility to automate the processes.

We first present the theoretical foundation, namely how knowledge is understood in Information Science (IS) and Systems Theory and how model composition is relevant in this context. Then we discuss DDM&S from a systems theoretical perspective where transformations of data-information-knowledge can be performed at different systems epistemological levels. An example of DDM&S used in component-based automatic model generation in light-rail domain is given to illustrate the concept.

2. THEORETICAL FOUNDATIONS

DDM&S brings together the two fields of IS and M&S. To root its concept on solid theoretical grounds, we revisit two hierarchies that concern knowledge: (1) the *Knowledge Hierarchy* widely recognized by the IS community to contextualize data, information and knowledge (Rowley, 2007), and (2) the *Epistemological Hierarchy of Systems* which explains knowledge from a systems theoretical viewpoint (Klir & Elias, 2003).

2.1. Data-Information-Knowledge Hierarchy

In the knowledge hierarchy, data, information and knowledge are arranged from the low to high levels in a pyramid. They are three fundamental and interrelated concepts still with debates about their nature, definition and relation. We present the basic ideas based on (Checkland & Holwell, 1998; Rowley, 2007; Ulrich, 2001; Zins, 2007).

Data comprises sets of symbols recorded in or on a medium. As such, data is unstructured, unprocessed and invariant. Checkland & Holwell (1998) also introduced the term *capta* in between data and information to refer to the subset of data that we select for attention.

Information is data (or capta) that are enriched with contextual "meaning attribution". It is done in a context that may well be shared by some but may also be unique to an individual. What machines cannot do, in a strict sense, is to generate unequivocal information; what they can do is merely to process capta into useful forms that can imply (or match) certain prescribed categories of information (Checkland & Holwell, 1998). Information can be inferred from data. This process is not viable without semiotic clarifications, namely syntactic clarity (comprehensibility), semantic clarity (meaning) and pragmatic clarity (relevance).

Knowledge is propositional. It is a "justified belief substantiated by compelling reasons" (Ulrich, 2001). Knowledge is structured and organized information expected to have greater longevity than a collection of many items of information that are only ephemerally meaningful and relevant. It is the general understanding and awareness garnered from accumulated information tempered by experience, enabling new contexts to be envisaged. Knowledge exists in the human mind but can be given physical representations.

With data we yet know nothing, with information we know what (and/or who, when, where), and with knowledge we know how (Ackoff, 1989; Zeleny, 1987). Understanding supports the transformation process from a lower level to the next (Rowley, 2007). If we consider a simulation model as a representation of the relevant knowledge of a system, then the transformation from data to knowledge for a given context is what one wants to achieve in DDM&S. For the transformation, we have to understand what data is available and what type of systems knowledge it can represent.

2.2. Epistemological Hierarchy of Systems

In M&S, a system being modeled can be called *referent*. Understanding what we know about the referent is essential in modeling. Simon (1962) roughly defined a complex system as one made up of a large number of parts that interact in a non-simple way. Generally, a system can be viewed as a pair of two orthogonal sets that are extremely rich in content: (1) a set of *things* (thinghood) and (2) a set of *relations* among things (systemhood) (Klir, 2001; Klir & Elias, 2003). Modeling (in computer simulation) is the activity to capture the thinghood and systemhood of interest in a computational form and study their evolvement for the purpose of understanding and/or prediction. Based on different levels of knowledge about the systemhood, Klir & Elias (2003) established an epistemological hierarchy of systems, as shown in Figure 1. Each level in the hierarchy implies the profoundness of our system knowledge.

Level 3 Structure System	*Relations between models at Level 2*
Level 2 Generative System	*Models that generate data at Level 1*
Level 1 Data System	*Observations or desirable states of Level 0*
Level 0 Source System	*A source of empirical data*

Figure 1: Hierarchy of epistemological levels of systems (based on Klir & Elias, 2003)

The bottom level of the hierarchy (Level 0) is the "primitive understanding" of a system, known as a *source system*. At this level, we know which variables and states we are interested in. The source system is (at least potentially) a source of empirical data. At the next level is the *data system* (Level 1) where data of the variables (or states) are obtained from observation or measurement, or are defined as desirable states. This is often the level where a system is accessible from the outside. Above the data system is the *generative system* (Level 2) where we acquire the knowledge to define the translation rules that are able to generate the observed or hypothetical (unobservable and/or internal) data of the system. At Level 3, the *structure system* describes systems knowledge as a set of generative systems that interact with each other in some way. At this level we understand the system as an interactive whole that has generative systems as parts. (Thus generative systems are often referred to as sub-systems.) This level of knowledge is also known as the knowledge of systems or structured sub-systems.

A higher epistemological level entails that the knowledge attained at the lower levels is known and it contains additional knowledge that is not available at the lower levels. The process of climbing the epistemological hierarchy transforms the notion of systems knowledge from primitive to more meaningful in order to specify a generative or a structure system that can reproduce the data at Level 1. Problems of this type fall into the category of *Systems Modeling*, and this category can be further divided into *Systems Inference* and *Systems Design* depending on whether the system is in existence (Klir, 1988; Zeigler, Praehofer, & Kim, 2000). *Systems analysis*, on the other hand, is the process of using the generative or structure system to produce data; computer simulation is an example of this type (Zeigler, Praehofer, & Kim, 2000). In DDM&S, we have data about a system that we want to model at the generative and structure levels. This requires knowledge about system decomposition and model composition that we discuss in the following section.

2.3. System Decomposition and Model Composition

Approaching systems complexity, we habitually divide a system into less complex parts, and analyze the parts and their interrelations that are more comprehensible. This approach is well grounded in systems theory and systems thinking literature (Checkland, 1999; Klir & Elias, 2003; Simon, 1996). Systems decomposition also has another advantage. What one knows about a complex system is often partial, i.e., a subset of the parts and interrelations in a system may be known to some experts. Adequate decomposition of a system can provide a systemic structure to aggregate and merge the knowledge about different parts and interrelations. The decomposition can be recursive until an elementary level of sub-systems is reached (Simon, 1996).

The word "system" in Greek means composition in its literary sense. Model components are basic parts (or building blocks) in model construction. If components and their interactions are designed to represent some distinguishable objects and relations in a real system, modelers and domain experts may easily identify and use the components to construct larger models and to validate them. The domain knowledge can be integrated into the model by terminology and definition of the components (Valentin & Verbraeck, 2002). Using components, changes of the model can be limited to some components instead of the whole model. Domain ontology and mapping between the ontology and simulation components are useful in this context.

In M&S, composability is defined as the capability to select and assemble simulation components in various combinations into valid simulation systems to satisfy specific user requirements (Petty & Weisel, 2003). Although component-based modeling is widely promoted and encouraged for its considerable benefits, it is shown difficult to apply (Szabo & Teo, 2007; Tolk, Diallo, King, Turnitsa, & Padilla, 2010; Yilmaz, 2004). Model composability shows many characteristics that are similar to those of software composability and systems design (Baldwin & Clark, 2000; Braude & Bernstein, 2010; Hofmann, 2004; Pidd & Robinson, 2007; Sommerville, 1996) whose design principles can be learnt from. Designing composable models should start at the conceptualization phase.

To design composable and reusable simulation models, the design goal needs to be at least one level higher than the design of any specific application of the model components. How the components may operate with other components and how other designers can make use of the components are the higher-level goals, which make standardization in model component design particularly important. This is not only meant for interface definition (as one does, e.g., for composable software) but also for model definition at a conceptual level (Tolk, Diallo, King, Turnitsa, & Padilla, 2010; Tolk & Muguira, 2003; Yilmaz, 2004). Systems decomposition determines the model composition and interaction. Designers should aim at conceptually decomposing a system into parts and relations such that they capture the essence in a system to serve the intended simulation goal and at the same time provide the flexibility for other composite combinations to represent a set of systems in the domain of interest. In general, domain knowledge can be separated in two classes, the fixed and the unfixed (as mentioned in Section 1). The fixed can be defined by model components at the generative level. The unfixed can be handled by the selection, configuration and composition of the components. In DDM&S, data is used to infer the unfixed. Systems decomposition is thus critical in the sense that it determines possible model structures and configurations in model composition.

3. A SYSTEM THEORETICAL APPROACH OF DDM&S

For a given M&S purpose, a modeler's understanding of the referent follows an epistemological hierarchy (Section 2.2). The transformation of data-information-knowledge (Section 2.1) in DDM&S is supported by modelers' systems understanding associated to the epistemological levels.

3.1. Data and the Epistemological Hierarchy

Data at Level 1 (L1) is the *empirical data* obtained by observations and/or measurements performed on the referent. It is sets of samples on some observables including input, output and state variables. Having the data as a starting point, the effort of modeling is to specify a generative or a structure system (L2/L3) that can produce some data that agree with the empirical data. Can the L2/L3 knowledge be obtained solely through transformation of L1 data? This can be achieved, e.g., by using a regression model. But if we can obtain domain knowledge about the dynamics and the structure of the referent (L2/L3), then the L2 and L3 knowledge of the model is not necessarily derived directly from the L1 data. In practice, the elicitation of L2/L3 knowledge in M&S relies heavily on (human) domain expertise. The L2/L3 knowledge of the referent often can be obtained from other systems.

Take a public transportation system (say, System T) as an example (the referent). Its L3 knowledge is, e.g., the design of the infrastructure network, and the design is the outcome of some design process, which can be conceived of as another independent system (say, System D). The design process produces design documents (L1 of System D) according to which the infrastructure network (L3 of System T) is built. These documents, therefore, correspond to some knowledge at L3 of System T. When modeling System T, we may use the design documents to determine some L3 knowledge about the model of System T. In other words, we need to find a function that transforms L1 of System D into some L3 of System T.

The knowledge of a referent at a certain level may be generated from more than one system, and similarity, one system may generate knowledge about the referent at different epistemological levels. The knowledge hierarchy (and transformation) in IS can be applied to each of the epistemological levels. A certain type of data after correct analysis and interpretation may deliver certain information of the referent. In DDM&S, designers should explore different possibilities of using various types of data and try to transform them into information and eventually into knowledge at different epistemological levels. The transformation is meant in a computational sense where machines (and software) can only interpret data in a predefined way. When carefully designed, they can efficiently process data into certain structure that can match some prescribed categories of information and knowledge. As modelers we also need to carefully design these prescribed categories in order to satisfy the flexibility needs of the simulation software.

3.2. Using Data in Modeling

Data is used in every simulation study. They can be gathered, e.g., from systems design (e.g., CAD data), planning and operation (e.g., scheduling and ERP data). Simulated data from other models may also be a data source if the results are validated. Some data sources typically contain information that can be mapped into different levels of model knowledge. Making these relations explicit to the modelers help them design the simulation software.

Modeling concerns two worlds, the world of the “real” where our interests are situated, and the world of the “virtual” where the models of the “real” are built. The tough questions in modeling are about constructing viable passages from the real to the virtual. In a strict sense, what is obtainable from a “real system” is only data. When one observes a system of interest (as a referent), one always chooses a perspective and tries to get or give some rational explanations of the system. A modeler tries to express these explanations (many of which are domain knowledge) by means of a model. When the modeler uses data for this purpose, he or she often needs the domain expertise to know how to correctly interpret the data. When manual analysis and interpretation of the data become too cumbersome and time-consuming, we may consider formalizing the existing know-how and encoding this into automated processes.

4. DATA-DRIVEN MODEL GENERATION: AN EXAMPLE

In this section, we discuss a data-driven approach for automatic simulation model generation where prebuilt and validated model components are used as building blocks. We view this approach as an automated reuse of model components. Automation is the execution by a machine agent (usually a computer) of a function that was previously carried out by a human (Parasuraman & Riley, 1997). When we, as human modelers, want to construct a model from model components, we need to know what components to use, and how to configure and structure them together. For an automated process, the same types of information are required which are derived from the provided data.

4.1. Case Description

The modeling case is in the domain of urban public light-rail transportation. It is a long-term project in cooperation with a public transportation company in the Netherlands. At the beginning of the project, the scales of the models developed were relatively small, e.g., modeling an intersection or a specific area in a city to assess the control and operation strategies (Kanacilo & Verbraeck, 2006; Kanacilo & Verbraeck, 2006; Kanacilo & Verbraeck, 2007). The models (i.e., the components and structures) were defined by humans in XML files that were then converted into models. After a number of simulation studies, the organization decided to use the simulation models more extensively. Larger models were needed, e.g., a complete light-rail service line or the network of a whole city. This was when problems arose because manual definition of the model became unmanageable. In a later study (Huang, Seck, & Verbraeck, 2010; Huang, Verbraeck, Oort, & Veldhoen, 2010), the XML definition contained thousands of nodes and dozens of levels and attributes per node on average. The problem lies not only in the amount of effort and time but also in the fact that the manual procedure turned out to be increasingly error-prone which caused difficulties in the debugging process.

Automation seemed to be a solution but it was unclear by then how it could be done. Our approach was threefold. First, the existing models (and model components) were investigated with experiences in how they were constructed. A computational procedure with the data structure and algorithms were designed and the necessary information requirement for model generation was specified. Second, the available data sources that could deliver the required information were identified, and the plan of how to obtain the missing data was made. Third, the model components were adapted and completed so that they can be easily used for automated composition and configuration. These three major tasks were carried out in parallel and the outcome of one influenced another.

4.2. The Model Component Library

The model component library has been gradually developed for the light-rail transportation simulation project. Some components have been adapted and added for the purpose of model generation. The library is called LIBROS (Library for Rail Operations Simulation). It follows the DEVS formalism (Zeigler, Praehofer, & Kim, 2000) for model component specifications. Railway operational elements such as vehicles, tracks, and sensors are specified as atomic models, each of which represents one functional aspect of the rail infrastructure or as required by the simulation model. They can be used to create more complex rail components such as stations and block sections, which in turn can be further composed until a complete representation of the modeled system is formed. The DEVS simulator underlying LIBROS is ESDEVS (Seck & Verbraeck, 2009). It implements the parallel DEVS and dynamic structure DEVS (Barros, 1995) on top of DSOL (Distributed Simulation Object Library) (Jacobs, 2005; Jacobs, Lang, & Verbraeck, 2002) which is a general-purpose event-scheduling based simulator. To enable model identification, configuration and coupling, different types of coupled models are defined in forms of meta-models in the library.

4.3. Model Generator: The Concept

Some model generators used formal model definitions for model generation (Balci, Nance, Derrick, Page, & Bishop, 1990; Foeken & Voskuijl, 2010; Kang, 1997; Son, Jones, & Wysk, 2000; Son, Wysk, & Jones, 2003). We prefer the data-driven approach for the reasons stated in Section 4.1. More recent works are inclined towards this approach (Bergmann & Strassburger, 2010; Jeong & Allan, 2004; Lucko, Benjamin, Swaminathan, & Madden, 2010; Shephard, Beall, O'Bara, & Webster, 2004; Tannock, Cao, Farr, & Byrne, 2007; Wang, Chang, Xiao, Wang, & Li, 2011). Some of those works discussed the concept or built a prototype. Some used data (or data models) that contained logical relations represented the model structure, or the model structure is generated in a parameterized way. In our work, the model structure is generated from data that do not directly contain logical relations. The generator can create the model structure from the data sources. The algorithm constructs models from selecting, structuring and configuring model components. Model selection heuristics that represent the domain knowledge of which components are relevant to the modeling goal are used to guide the component identification and the composite (Lee & Zobel, 1996), i.e., they are used to define the data inferential rules.

In some cases, if the modeled system has a simple structure then the model may be directly generated from the data. However, in many cases, the system is complex and the data that describe the system do not contain the relational logic that can be directly applied to the desired model structure. In such cases, several steps are necessary in the model generation procedure and intermediate data structures are employed to incrementally construct a relational representation of the system structure that in turn can be transformed into the corresponding model composite structure.

As mentioned earlier, different data sources may provide systems knowledge at different levels. The data may come from design, operation scheduling, resource allocation, etc. Starting from these data sources, the data analysis results shall eventually reflect the structural and behavioral preconditions that are the basis of constructing the (initial) model structure and initializing the model state. The transformation from the data source to the model structure is often too complex to be accomplished in one go. It can be generally divided into three steps. Assume that after pre-processing the data can be correctly interpreted describing the system in a primitive format without logical relations, e.g., a list of numerical or textual descriptions.

At the first step, a relational graph is created from the data based on its descriptive content. The data inference may involve some common sense rules or some basic domain knowledge. In both cases, the information needed to create the relational graph is self-contained by the data sources, i.e., no extra supporting information is required. The relational graph may be created incrementally for the convenience of structuring. For example, we may first identify that entities a, b, c are related to form A, and entities x, y, z are related to form B; at a later round of structuring, A and B may be grouped together to form C, and so forth until the desired level of structure is reached. The relational graph represents the structure of the data content. At the second step, we can discover the systems structure it represents with the help of a domain ontology map. By searching the ontology space, a match in patterns of entity attributes or relations ascertains what that part of the data structure represents within the overall systems structure. At the third step, the systems structure is transformed into corresponding model structure according to a model counterpart table. A model counterpart table specifies a mapping relation between a systems entity and its counterpart in the model. The mapping relation maybe one to one or one to more, i.e., a systems entity may have more than one version of models. For instance, a rail vehicle can be modeled with one physical entity or several segmented physical entities. Depending on the available data detail and the desired model detail, one model counterpart can be chosen for model generation. Additionally, the model counterpart table also specifies if any other model parts shall be added to a given systems entity. In railway simulation, e.g., different types of intersections may have different control rules. These rules are not a part of the data but defined as a configurable sub-component in the intersection model. For an identified type of intersection, the model counterpart table specifies which control unit shall be added. With the identified component counterparts and the added components, the entire model structure is fully fledged. The simulation model can then be constructed, configured and initialized accordingly using the available model components in the library.

Limited by the length of this paper, we cannot give detailed examples of the model generation method and algorithms. Readers of interest may refer to Huang, Seck, & Verbraeck (2011) for a short example, or to Huang (2013) for a more complete example.

4.4. Some Remarks

Component-based modeling or component-based engineering in general is founded on a paradigm common to many engineering disciplines: complex systems can be obtained by assembling components (Gössler & Sifakis, 2005). Recursively constructing more complex components from simpler ones is a useful concept because it tackles incomprehensible problems from tangible bases. The simplicity of the concept makes it powerful. Component-based model generation automates the model composition for a given modeling objective. It is useful or may be the only solution if the model scale drastically increases and using simpler models is not an alternative. Once developed, the software has appealing long-term benefits.

The completeness of data used for model generation is important because the model generator cannot deal with unanticipated incompleteness. Such problems need to be solved prior to model generation. A straightforward solution is to complete the missing data when possible. In the light-rail model generation case, some contents that were not indicated in the original data were added manually for model transformation. If it is impossible to complete missing data, implementing rules in the generator, the model counterpart table or in the model components can be alternatives. In the case discussed, some data was not available but according to domain experts, the information may be inferred from the data that was available. Therefore, additional rules were added into the model counterpart table to generate the part of information that was missing.

5. CONCLUSIONS

This paper presented a broad view of DDM&S that is not limited to the use of observed data from the system of interest but encompasses various categories of data such as design and engineering documents. This view is motivated by the fact that more and more data becomes available along with the advances in data collection and storage technologies. Model-based approach in systems development provides different data sources that may be useful for inferring information about systems structure and behavior. To use these data in M&S, a comprehensive approach with understanding the data in relation with Systems Theory is necessary in addition to statistical procedures.

To explain the concept, we reviewed the data-information-knowledge hierarchy in IS and the epistemological hierarchy in Systems Theory. Based on these theories, we showed how multiplicity of data sources can be associated to different epistemological levels of systems, and how data may be transformed to the related systems knowledge. Using data in M&S has heavily relied on human intervention. However, when the manual process is well understood (especially at a systems theoretical level), the relevant domain knowledge and modeling knowledge may be embodied into an automated process. These processes can be beneficial for many systems such as those that have long-term or real-time needs for simulation models. The automation may also better integrate M&S with other technologies or applications such as optimization, design and engineering. Because of the variation in systems knowledge and the diversity in simulation goals, we may decompose the system such that the "fixed" parts (parts that are unlikely to change) are represented by pre-developed model components and the "unfixed" parts to be represented by model configuration and composition. The decomposition of a system into parts and relations should capture the essence in a system to serve the intended simulation goal, and at the same time provide the flexibility for other composite combinations to represent a set of systems in our domain of interest. This flexibility is supported by data analysis and transformation algorithms that infer data with certain structures that can match some prescribed categories of systems knowledge. The example provided in the last section explained an automated process of model generation using different data sources. The research in DDM&S is rich in content. It often requires knowledge in many disciplines. This interdisciplinary nature predestines the application of systems engineering approaches in developing such simulation software.

REFERENCES

Ackoff, R. L. (1989). From data to wisdom. Journal of Applied Systems Analysis, 16, 3-9.

Balci, O., Nance, R. E., Derrick, E. J., Page, E. H., & Bishop, J. L. (1990). Model generation issues in a simulation support environment. In Proceedings of the 1990 Winter Simulation Conference (pp. 257-263).

Baldwin, C. Y. & Clark, K. B. (2000). The Power of Modularity. MIT Press.

Banks, J., Carson, II, J. S., Nelson, B. L., & Nicol, D. M. (2010). Discrete-Event System Simulation (5th). Pearson Education.

Barros, F. J. (1995). Dynamic Structure Discrete Event System Specification: A new Formalism for Dynamic Structure Modeling and Simulation. In Proceedings of The 1995 Winter Simulation Conference (pp. 781-785).

Bergmann, S. & Strassburger, S. (2010). Challenges for the Automatic Generation of Simulation Models for Production Systems. In Proceedings of the 2010 Summer Simulation Multiconference, Ottawa, Canada, 2010 (pp. 545-549).

Braude, E. J. & Bernstein, M. E. (2010). Software Engineering: Modern Approaches (2nd). John Wiley & Sons.

Checkland, P. (1999). Systems Thinking, Systems Practice: Includes a 30-Year Retrospective. John Wiley & Sons.

Checkland, P. & Holwell, S. (1998). Information, Systems and Information Systems - making sense of the field. John Wiley & Sons.

Darema, F. (2011). {DDDAS} computational model and environments. Journal of Algorithms and Computational Technology, 5(4), 545-560.

Ekere, N. N. & Hannam, R. G. (1989). Evaluation of approaches to modelling and simulating manufacturing systems. International Journal of Production Research, 27(4), 599-611.

Foeken, M. J. & Voskuijl, M. (2010). Knowledge-based simulation model generation for control law design applied to a quadrotor UAV. Mathematical and Computer Modelling of Dynamical Systems, 16(3), 241-256.

Gössler, G. & Sifakis, J. (2005). Composition for component-based modeling. Science of Computer Programming, 55(1-3), 161-183.

Hofmann, M. (2004). Criteria for decomposing systems into components in modeling and simulation: Lessons learned with military simulations. Simulation, 80(7-8), 357-365.

Hu, X. (2011). Dynamic Data Driven Simulation. SCS M&S Magazine, 5, 16-22.

Huang, Y. (2013). Automated Simulation Model Generation. Ph.D. dissertation, Delft Univeristy of Technology.

Huang, Y., Seck, M. D., & Verbraeck, A. (2010). The Architecture and Components of LIBROS: Strengths, Limitations, and Plans. In Proceedings of The 2010 European Simulation and Modelling Conference, Hasselt, Belgium, 2010 (pp. 80-87). Eurosis-ETI.

Huang, Y., Seck, M. D., & Verbraeck, A. (2011). From Data to Simulation Models: Component-based Model Generation with a Data-driven Approach. In Proceedings of the 2011 Winter Simulation Conference, Phoenix, AZ, 2011 (pp. 3724-3734).

Huang, Y., Verbraeck, A., van Oort, N., & Veldhoen, H. (2010). Rail Transit Network Design Supported by an Open Source Simulation Library: Towards Reliability Improvement. In Transportation Research Board 89th Annual Meeting Compendium of Papers, Washington, DC, USA, 2010. TRB.

Jacobs, P. H. M. (2005). The {DSOL} simulation suite - Enabling multi-formalism simulation in a distributed context. Ph.D. dissertation, Delft University of Technology.

Jacobs, P. H. M., Lang, N. A., & Verbraeck, A. (2002). D-SOL: A Distributed Java Based discrete event simulation architecture. In Proceedings of the 2002 Winter Simulation Conference (pp. 793-800). IEEE.

Jahn, W., Rein, G., & Torero, J. L. (2012). Forecasting fire dynamics using inverse computational fluid dynamics and tangent linearisation. Advances in Engineering Software, 47(1), 114-126.

Jeong, K.-Y. & Allan, D. (2004). Integrated system design, analysis and database-driven simulation model generation. In Proceedings of the IEEE Annual Simulation Symposium (pp. 80-85).

Kanacilo, E. M. & Verbraeck, A. (2006). Decision Support for Control Design of Rail Infrastructures. In SIMTECT 2006 - Simulation: Challenges and Opportunities for a Complex and Networked World.

Kanacilo, E. M. & Verbraeck, A. (2006). Simulation services to support the control design of rail infrastructures. In Proceedings of the 2006 Winter Simulation Conference (pp. 1372-1379). IEEE.

Kanacilo, E. M. & Verbraeck, A. (2007). Assessing tram schedules using a library of simulation components. In Proceedings of the 2007 Winter Simulation Conference (pp. 1878-1886). IEEE.

Kang, S. (1997). Knowledge based automatic simulation model generation system. IEE Proceedings: Circuits, Devices and Systems, 144(2), 88-96.

Kim, B.-I., Kim, Jeong, S., Shin, J., Koo, J., Chae, J., and Lee, S. (2009). A layout- and data-driven generic simulation model for semiconductor fabs. IEEE Transactions on Semiconductor Manufacturing, 22(2), 225-231.

Klir, G. J. (2001). Facets of Systems Science (2nd). Kluwer Academic/Plenum Publishers.

Klir, G. J. (1988). The Role of Uncertainty Principles in Inductive Systems Modelling. Kybernetes, 17(2), 24-34. Klir, G. J. & Elias, D. (2003). Architecture of Systems Problem Solving (2nd). Kluwer Academic/Plenum Publishers.

Law, A. M. (2007). Simulation Modeling and Analysis (4th). McGraw-Hill.

Lee, C. H. & Zobel, R. N. (1996). Representation of simulation model components for model generation and a model library. In Proceedings of the IEEE Annual Simulation Symposium (pp. 193-201).

Lucko, G., Benjamin, P. C., Swaminathan, K., & Madden, M. G. (2010). Comparison of manual and automated simulation generation approaches and their use for construction applications. In

Proceedings of the 2010 Winter Simulation Conference (pp. 3132-3144).
O'Keefe, R. M. & Haddock, J. (1991). Data-driven generic simulators for flexible manufacturing systems. International Journal of Production Research, 29(9), 1795-1810.
Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230-253. Petty, M. D. & Weisel, E. W. (2003). A composability lexico. In Proceedings of the Spring 2003 Simulation Interoperability Workshop (pp. 181-187).
Pidd, M. (1992). Guidelines for the design of data driven generic simulators for specific domains. Simulation, 59(4), 237-243.
Pidd, M. & Robinson, S. (2007). Organising insights into simulation practice. In Proceedings of the 2007 Winter Simulation Conference, Piscataway, NJ, USA, 2007 (pp. 771-775). IEEE Press.
Qiao, G., Riddick, F., & McLean, C. (2003). Data driven design and simulation system based on XML. In Proceedings of the 2003 Winter Simulation Conference (pp. 1143 - 1148 vol.2).
Rowley, J. E. (2007). The wisdom hierarchy: representations of the {DIKW} hierarchy. Journal of Information Science, 33(2), 163-180.
Seck, M. D. & Verbraeck, A. (2009). {DEVS} in {DSOL}: Adding {DEVS} operational semantics to a generic Event-Scheduling Simulation Environment. In Proceedings of the 2009 Summer Computer Simulation Conference.
Shephard, M. S., Beall, M. W., O'Bara, R. M., & Webster, B. E. (2004). Toward simulation-based design. Finite Elements in Analysis and Design, 40(12), 1575-1598.
Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, 106(6), 467-482.
Simon, H. A. (1996). The Sciences of the Artificial (3rd). MIT Press.
Sommerville, I. (1996). Software Engineering (5th). Addison-Wesley.
Son, Y. J., Jones, A. T., & Wysk, R. A. (2000). Automatic generation of simulation models from neutral libraries: An example. In Proceedings of the 2000 Winter Simulation Conference (pp. 1558-1567).
Son, Y. J., Wysk, R. A., & Jones, A. T. (2003). Simulation-based shop floor control: Formal model, model generation and control interface. IIE Transactions, 35(1), 29-48.
Szabo, C. & Teo, Y. M. (2007). On Syntactic Composability and Model Reuse. In Proceedings of the International Conference on Modeling and Simulation (pp. 230-237). IEEE Computer Society Press.
Tannock, J., Cao, B., Farr, R., & Byrne, M. (2007). Data-driven simulation of the supply-chain -- Insights from the aerospace sector. International Journal of Production Economics, 110, 70-84.
Tolk, A. & Muguira, J. A. (2003). The Levels of Conceptual Interoperability Model. In Proceedings of the 2003 Fall Simulation Interoperability Workshop. IEEE CS Press.
Ulrich, W. (2001). A Philosophical Staircase for Information Systems Definition, Design and Development: A Discursive Approach to Reflective Practice in ISD (Part 1). Journal of Information Technology Theory and Application (JITTA), 3(3), 55-84.
Valentin, E. & Verbraeck, A. (2002). Guidelines for designing simulation building blocks. In Proceedings of the 2002 Winter Simulation Conference, San Diego, CA., 2002 (pp. 563-571). IEEE.
Wang, J., Chang, Q., Xiao, G., Wang, N., & Li, S. (2011). Data driven production modeling and simulation of complex automobile general assembly plant. Computers in Industry, 62(7), 765-775.
Yilmaz, L. (2004). On the Need for Contextualized Introspective Models to Improve Reuse and Composability of Defense Simulations. The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 1(3), 141-151.
Zeigler, B. P., Praehofer, H., & Kim, T. G. (2000). Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems (2nd). Elsevier/Academic Press.
Zeleny, M. (1987). Management Support Systems: Towards Integrated Knowledge Management. Human Systems Management, 7(1), 59-70. Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. Journal of the American Society for Information Science and Technology, 58(4), 479-493.

AUTHORS BIOGRAPHY

Yilin Huang is a Postdoctoral Researcher at the Section Systems Engineering and Simulation, Faculty of Technology, Policy and Management, Delft University of Technology, Netherlands.

Mamadou D. Seck is an Assistant Professor at the Department of Engineering Management and Systems Engineering, Old Dominion University, USA.

Alexander Verbraeck is a Full Professor at the Section Systems Engineering and Simulation, Faculty of Technology, Policy and Management, Delft University of Technology, Netherlands, and a part-time full professor in supply chain management at the R.H. Smith School of Business of the University of Maryland, USA.

ONE APPROACH OF ADEQUATE MATHEMATICAL DESCRIPTION CONSTRUCTION

Menshikov Yu.L.[(a)]

[(a)]Dnepropetrovsk University, Gagarina av., 72, 49010 Dnepropetrovsk, Ukraine

[(a)]Email: Menshikov2003@list.ru

ABSTRACT
Mathematical description of physical process consists from the mathematical model of physical process and vector function of external load. In this work mathematical descriptions of physical processes which are being described by linear systems of ordinary differential equations are investigated. The algorithms of construction of the mathematical description of real process (for example the motion of some dynamic system) are considered which allows receive adequate results of mathematical simulation. Two basic approaches to this problem are selected. Within the framework of one of these approaches some algorithms are offered.

Keywords: adequate mathematical description, inverse problems, regularization.

1. INTRODUCTION

The typical situation which arises by the analysis of new processes or phenomena is investigated. It is supposed that real physical process is observed and the records of experimental measurements of some characteristics of this process are given. It is necessary to develop adequate mathematical description of this process for further use.

The similar questions and problems considered in many other works [1,2,3,4,5].

In the given work the mathematical models of physical processes described only by the system of the ordinary differential equations will be examined [6,7,8,9]. Such idealization of real processes or dynamic systems is widely used in various areas for the description of control systems, as well as of mechanical systems with the concentrated parameters, as well as economic processes, biological and ecological processes etc. It is shown that with the help of such systems in some works even human emotions are simulated [10].

As a concrete example, the dynamic system which describes motion of the main mechanical line of the rolling mill is used.

We will assume also, that the process is open, i.e. it interacts with other neighbouring processes (in the closed processes there is no such the interaction).

For the simplicity, we select the dynamic process with mathematical models in the form of linear system of ordinary differential equations:

$$\dot{x}(t) = \tilde{C}\, x(t) + \tilde{D}\, z(t)\,, \qquad (1)$$

with the equation of observation

$$y(t) = \tilde{F}\, x(t)\,, \qquad (2)$$

where $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$ is vector-function variables characterized the state of process, $z(t) = (z_1(t), z_2(t), ..., z_l(t))^T$ is vector-function of unknown external loads, $y(t) = (y_1(t), y_2(t), ..., y_m(t))^T$; $\tilde{C}, \tilde{D}, \tilde{F}$ are matrices of the appropriate dimension with constant coefficients which are given approximately, $\tilde{F}$ is nonsingular matrix dimension $m \times n$ and rang $\tilde{F} = m$, ($(.)^T$ is a mark of transposition).

If the part of external loads of real process is known, this case can be reduced to one which is examined earlier with using the linearity of initial dynamic system.

We assume that state variables $x_i(t), 1 \le i \le n$ of system (1) correspond to some real characteristics $\tilde{x}_i(t), 1 \le i \le n$ of process which are being investigated and that the vector function $\tilde{y}(t) = \tilde{F}\,\tilde{x}(t)$ is obtained from experimental measurements.

The *problem of synthesis of adequate mathematical description* with the use of system (1) can be formulated as follows: it is necessary to find unknown vector function of external loads $z(t)$ in such a way that the vector function $y(t)$, which are obtained from system (1), (2) under this external load $z(t)$, coincides with experimental data $\tilde{y}(t)$ with a given accuracy of experimental measurements in chosen functional metrics.

Let us consider questions what prospects of adequate mathematical descriptions are valid for further use and what goals should be selected at the creation of adequate mathematical descriptions.

It will be useful to address to classical works in this area. In work [11] the following statement was done: "…the goal of the imitation simulation is the creation of experimental and applied methodology which aimed at the use of its for a prediction of the future behaviour of system".

So the adequate mathematical descriptions first of all are aimed at the forecast of behaviour of real processes. With the help of adequate mathematical description it is possible to predict of behaviour of real process in new conditions of operation. For example, it is possible to test more intensive mode of operations of the real machine without risk of its destruction. Such tool (adequate mathematical description) allows to simulate the characteristics of process in the unconventional modes of operations, and also to determine optimum parameters of real process.

The considered situation requires the formation of some uniform methodological approach to this problem, creation of general algorithms and common criteria of adequacy evaluation [11,12,13].

2. STATEMENT OF A PROBLEM

The main problem of mathematical simulation is the construction (synthesis) of mathematical model (MM) of motion of real dynamic system which in aggregate with model of external load (MEL) gives the adequate to experimental observations the results of mathematical simulation [14].

There exist two approaches to problem of construction of adequate mathematical description [12,13,14]:

1) Mathematical model of process of type (1) is given a priori with inexact parameters and then the models of external loads were determined for which the results of mathematical simulation coincide with experiment [12,13,14];

2) Some models of external loads are given a priori and then mathematical model of process of type (1) is chosen for which the results of mathematical simulation coincide with experiment [15,16,17].

Now we will consider the synthesis of adequate mathematical description in the frame of first approach analysing the process with the concentrated parameters, for which the motion is described by ordinary differential equations of n-order (1).

We assume that vector function $\tilde{y}(t)$ in system (2) is obtained from experiment and presented by graphics. Besides, we suppose that some of functions of external loads, for example, $z(t) = (z_1(t), z_2(t), \ldots, z_l(t))^T$, $l \le m$ are unknown.

According to first approach, it is necessary to develop the construction of vector function of external loads which define such the functions of state $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$ of mathematical model (1), are coincide with experimental measurements $\tilde{y}(t)$ with inaccuracy of initial data in given metrics. Such mathematical model of process behaviour together with obtained vector function of external loads $z(t)$ can be considered as *adequate mathematical description of process*.

Such method of obtaining of mathematical model of external loads (function $z(t) = (z_1(t), z_2(t), \ldots, z_l(t))^T$, $l \le m$) is determined in literature as a method of external loads identification [18,19]. By the way, physical reasons of occurrence of such external loads are not being taken into account. They are only functions which in combination with mathematical model (1) provide results of mathematical simulation, which coincide with experiment with the given accuracy.

Such coincidence is being attained by synthesis of "correct" mathematical model (MM) of the dynamical system and the choice of "good" model of external load (MEL). MM of object the motion of which coincides with experimental measurements with acceptable accuracy under action of MEL which corresponds to real EL ("good" model) is understood as "correct" model. Thus the degree of "correctness" of MM depends directly on the chosen MEL and required accuracy of the coincidence with experiment.

Now we consider the synthesis of model of external loads by a method of identification [12,13].

Let's assume that external load $z(t)$ is unknown and vector function $\tilde{y}(t)$ in the equations (2) is measured by an experimental way.

The part of state variables $\tilde{x}_1(t), \tilde{x}_2(t), \ldots, \tilde{x}_m(t)$ can be obtained by an inverse of equation (2) with function $\tilde{y}(t)$ as $\tilde{F}$ is nonsingular matrix:

$$\tilde{x}_k(t) = N_k(\tilde{y}(t)),\ 1 \le k \le m,$$

where $N_k(\tilde{y}(t))$ is known functions.

Let's consider known state variable $\tilde{x}_k(t)$, $1 \le k \le m$ as two known internal loads $d_k \tilde{x}_k(t)$ and $-d_k \tilde{x}_k(t)$, where d_k, $1 \le k \le m$ is known constants. Such interpretation of state variables $\tilde{x}_k(t)$ allows to simplify initial system. Such transformation will be determined as "k-section" of initial system [12,13].

In some cases after lines of "k-sections" the initial system (1) will be transformed to some subsystem at which one state variable is known, for example, $\tilde{x}_1(t)$ and at which all external loads $z_k(t)$, $k = 2, \ldots, m$, except $z_1(t)$, for example, are known. For the system of type (1) with the help of a number of "sections" can be obtained the subsystem of initial system which movement is described by the differential equations

$$\dot{x}(t) = A_1\, x(t) + B_1\, z_1(t)\,, \tag{3}$$

with the equation of observation

$$\tilde{y}(t) = c_1\, \tilde{x}_1(t)\,, \tag{4}$$

where A_1 is matrix with constant coefficients of the appropriate dimension; B_1 is vector column, c_1 is const. So, it is supposed that the subsystem of initial system has one unknown external load $z_1(t)$ and one known variable of state $\tilde{x}_1(t)$ which is obtained by experimental way.

After simple transformations it can be obtain an integral equation for the unknown function of external load $z_1(t)$:

$$\int_{t_0}^{t} K_1(t-\tau)\, z_1(\tau)\, d\tau = P(t) \tag{5}$$

where $K_1(t-\tau),\ P(t)$ are known functions.

If the initial system (1),(2) does not satisfy the condition, as have been specified above, then this system can be reduced to system (3),(4) by aid of additional measurements [12,13].

With the use of similar transformations it is possible to receive the other integral equations for all unknown functions of external loads $z_k(t),\ k \le m$. After the solving of the integral equations, such as (5), all unknown functions of external loads of system of the equations (1) will be obtained.

The model $z(t)$ which was obtained with the use of such method depends on chosen mathematical model (1).

Further, the model of external loads which are found in such a way together with mathematical model of process (1) gives the adequate mathematical description of process.

Function $P(t)$ in equation (5) was obtained with use of experimental measurements $\tilde{x}_1(t), \tilde{x}_2(t), \ldots, \tilde{x}_m(t)$. So the function $P(t)$ is function which is given as graphic.

Let us present now (5) as

$$A_p z = u_g = B_p\, x_g \tag{6}$$

where A_p is linear operator, $A_p : Z \to U, z \in Z$, $x_g \in X$, $u_g \in U$, x_g is initial experimental data (graphic), z is unknown function, (X, Z, U are a functional spaces). Let's assume, that the operator A_p continuously depend on some vector parameters p of mathematical model of process: $p = (p_1(t), p_2(t), \ldots, p_N(t))^T$. The coefficients of matrices $\tilde{C}, \tilde{D}$ can be chosen as components of such vector parameters p.

It can be shown that in the most practical problems the operator A_p is completely continuous [20].

Thus, a necessary condition for obtaining the equation (5) concerning required external load z_j is the possibility through a number of "sections" to obtain a subsystem of initial system with one unknown external load and one known state variables $\tilde{x}_j,\ 1 \le j \le m$. It is easy to demonstrate an example of system such as (1), in which such opportunity is absent.

3. SPECIFIC FEATURES OF INTEGRAL EQUATIONS SOLUTIONS

We will consider the integral equation such as (5). From the practical point of view it is convenient to take Z as Banach space of continuous functions $C[0,T]$ or Hilbert space $W_2^1[0,T]$, where $[0,T]$ is interval of time on which the functions of external loads are being investigated [20]. As far as the initial experimental data are frequently varying functions, it is convenient to accept U as Hilbert space $L_2[0,T]$ [20].

Further, we shall suppose that the element x_g in the equation (5) is exchanged by function x_δ which approximated given graphic x_g with a known error:

$$\left\| x_g - x_\delta \right\|_X \le \delta\,,$$

where δ is const, δ >0.

Let's denote by $Q_{\delta,p}$ the set of the possible solutions of an inverse problem of identification of external load function (6) with the fixed operators A_p, B_p:

$$Q_{\delta,p} = \{ z : \left\| A_p z - B_p x_\delta \right\|_U \le \left\| B_p \right\| \cdot \delta \}\,.$$

Any function z from set $Q_{\delta,p}$ may be considered as "good" function of external load as far as the function $A_p z$ coincides with $B_p x_g$ with accuracy of approximation.

Thus, the operators A_p, B_p and any function from the set $Q_{\delta,p}$ give the triple which will provide adequacy of results of mathematical simulation with accuracy $\left\| B_p \right\| \cdot \delta$.

We shall name the process of determination of $z \in Q_{\delta,p}$ as *synthesis of function of external load by a method of identification* [12,14].

However set of the possible solutions $Q_{\delta,p}$ at any δ has a number of specific features (it is actually incorrect problem) [20]). First, and main of them is that this set is not bounded at any δ [20].

Let's consider this feature more in detail due to the fact that it leads to a number of unexpected and unusual consequences.

The set $Q_{\delta,p}$ contains infinite number of the solutions like any problem with the use of approximate data. However, the set $Q_{\delta,p}$ contains functions which can differ one from another on infinite value [12, 20]. It is due to the reason that the operator A_p in the equation (6), as a rule, is completely continuous.

Thus, the set $Q_{\delta,p}$ includes the essentially different functions which are equivalent in sense of the solution of the equation (6) (incorrect problems). Therefore, the basic difficulty will be the selection of the concrete solution from infinite set of the various equivalent solutions. For this purpose it is necessary to involve some additional information [20]. As example, the function which is the more convenient for further use can be selected in this quality [13,14].

More over the inaccuracy of operators A_p, B_p with respect to the exact operators can be not taken into account in these synthesis problems [13,14].

4. METHODS OF SOLUTION OF IDENTIFICATION EQUATION

For obtaining of the steady solutions of formulated above incorrect problems it is necessary to use the method of Tikhonov's regularization [20].

Let us consider the stabilizing functional $\Omega[z]$ which has been defined on set Z_1, where Z_1 is everywhere dense in Z [20]. Consider now the following extreme problem:

$$\Omega[z_{\delta,p}] = \inf_{z \in Q_{\delta,p} \cap Z_1} \Omega[z], \; p \in R^N \tag{7}$$

It was shown that under certain conditions the solution of the extreme problem (7) exists, is unique and stable with respect to small change of initial data x_g [20]. The function $z_{\delta,p}$ is named *the stable model of external load* after taking into account the only inaccuracy of approximation. The solution of a problem (7) can be non-unique. For the purposes of mathematical simulation any such solution will be acceptable. Such function of external load can be used for mathematical simulation of initial system (1).

Still there are no basis to believe that the function $z_{\delta,p}$ will be close to real external load z_{ex}. It is only good and steady function (model) of external load [12,13].

However, such approximate solution can be interpreted in other way. The regularized solution can be treated as steadiest with respect to change of the factors which were not taken into account in mathematical model. These factors may include changes in structure of mathematical model of system, the influence, which were not taken into account, change of conditions of experiment etc. We can prove such interpretation of the approximate solution.

At the synthesis of mathematical model of physical process we will first of all take into account the factors which define a low-frequency part of change of state variables. First, it is due to the fact that this part of a spectrum is well observed during the experiment, as measuring devices do not deform it. Secondly, the high-frequency components of external loads as well as equivalent to them insignificant factors not taken into account quickly die away in process of distribution among inertial elements. Thus, factors of interactions, which were not taken into account and equivalent to them influences, change only high-frequency part of the approximate solutions. If the factors, which are not taken into account, change a low-frequency part of the solutions then it means that mathematical model of process is chosen incorrectly. In work [21] has been shown that the regularized solution $z_{\delta,p}$ represents result of high-frequency filtration of approximate solution. The greater degree of smoothing of the solution corresponds to greater error of initial data.

Hence, the regularized solution $z_{\delta,p}$ can be interpreted as the function from set $Q_{\delta,p}$ which is the steadiest with respect to changes of factors, which are not taken into account. Such quality of the regularized solution is very important when it is used in mathematical simulation of real processes when the results of simulation are steady with respect to small changes factors, which are not taken into account and which are naturally present at any mathematical description of process.

The obtained solution of a synthesis problem of external load function $z_{\delta,p}$ requires, as a rule, the additional analysis. It is necessary, first of all, to determine transformation of model $z_{\delta,p}$ with change of operational conditions, for example, change of speed of movement of real dynamical system, change of sizes of various static loads etc. As a result of the analysis, the model z_p will be obtained which can be used for mathematical modeling of real processes and also at study of new prospective modes of operations.

For the numerical solution of an extreme problem (7) the discrepancy method was used [20]. The problem (7) was replaced by following extreme problem:

$$M^{\alpha}[A_p, z_{\delta,p}] = \inf_{z \in Z_1} \{ \| A_p z - B_p x_g \|_U^2 + \alpha \Omega[z] \} \tag{8}$$

where parameter of regularization α was determined from a condition

$$\| A_p z_{\delta p} - B_p x_g \|_U = \| B_p \| \delta . \tag{9}$$

So the obtained function $z_{\delta,p}$ together with operators A_p, B_p give the adequate mathematical description of process which is to stable change of initial data (stable adequate mathematical description).

In this case the analysis of a problem was reduced to the solution of the Euler's equation for functional (8) if the functional space U is Hilbelt space:

$$A_p^* A_p z_{\delta,p} + \alpha \Omega'[z_{\delta,p}] = A_p^* B_p x_g , \tag{10}$$

where A_p^* is the associate operator to A_p; $\Omega'[z]$ is Frechet's derivative.

The approximate solution of the equation (10) on a uniform discrete grid was carried out by a numerical method.

5. EXAMPLE OF PROBLEM SOLUTION OF SYNTHESIS OF ADEQUATE MATHEMATICAL DESCRIPTION

Now we consider in detail the problem of adequate mathematical description synthesis of the main mechanical line of rolling mill [14,22].

The four-mass model with weightless elastic connections is chosen as mathematical model of dynamic system of the main mechanical line of the rolling mill [14,22].

The equations of motion are obtained from the Lagrangian equations of second kind and have the form:

$$\dot{x}(t) = \tilde{C}\, x(t) + \tilde{D}\, z(t) , \tag{11}$$

where $x(t) = (x_1(t), x_2(t), \ldots, x_6(t))^T$ is vector function of state variables, $z(t) = (z_1(t), z_2(t))^T$ is vector function of unknown external loads,

$$\tilde{C} = \begin{pmatrix} 0 & c_{12} & 0 & 0 & 0 & 0 \\ c_{21} & 0 & c_{23} & 0 & c_{25} & 0 \\ 0 & 0 & 0 & c_{34} & 0 & 0 \\ c_{41} & 0 & c_{43} & 0 & c_{45} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{56} \\ c_{61} & 0 & c_{63} & 0 & c_{65} & 0 \end{pmatrix},$$

$$\tilde{D} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ d_{41} & 0 \\ 0 & 0 \\ 0 & d_{62} \end{pmatrix}.$$

The functions $x_1(t), x_3(t), x_5(t)$ are obtained by experimental way and are given as graphics $\tilde{x}_1(t), \tilde{x}_3(t), \tilde{x}_5(t)$ (Fig,1).

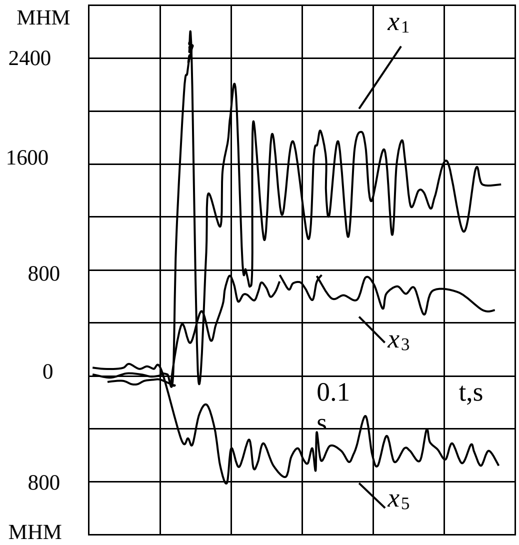


Fig.1.The records of functions $\tilde{x}_1(t), \tilde{x}_3(t), \tilde{x}_5(t)$.

The problem of synthesis of adequate mathematical description can be formulated as: it is necessary to find such functions of external loads $z_1(t), z_2(t)$ that the results of simulation $x_1(t), x_3(t), x_5(t)$ of system (11) with functions of external loads $z_1(t), z_2(t)$ coincide with experiment $\tilde{x}_1(t), \tilde{x}_3(t), \tilde{x}_5(t)$ (Fig,1).

Two integral equations for unknown functions $z_1(t), z_2(t)$ was obtained by method of “sections” [14,22]. The errors of experimental measurements are less than to 7 % in the uniform metrics.

The solutions of identification problem of the stable functions of external loads for experimental data presented in Fig.1 was performed.

The results of calculation are presented in Fig.2.

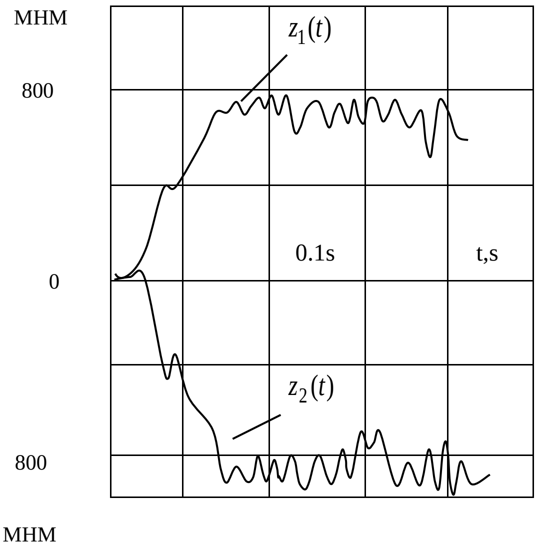


Fig.2. The diagrams of models of external loads $z_1(t)$, $z_2(t)$.

The mathematical model (11) with the found external loads $z_1(t), z_2(t)$ give stable adequate mathematical description of real process.

6. SYNTHESIS OF ADEQUATE MATHEMATICAL DESCRIPTION FOR CLASS OF OPERATORS

It is necessary to apply various methods of simplification in developing the mathematical model, such as taking into account different forces and their impact on the movement of real system. The various models (with different parameters) of real process or system were obtained by the different authors even in cases, when the structures of the mathematical descriptions (models) are similar. So, it is supposed that the vector parameters p is given inexactly. So vector p can get different values in given closed domain $D: p \in D \subset R^N$.

Some operators A_p, B_p correspond to each vector from D. The sets of possible operators A_p, B_p have been denoted as classes of operators $K_A = \{A_p\}, K_B = \{B_p\}$. So we have $A_p \in K_A, B_p \in K_B$. The deviations of operators $A_p \in K_A$ between themselves from set K_A and operators $B_p \in K_B$ from set K_B are given:

$$\sup_{p_\alpha, p_\beta \in D} \left\| A_{p_\alpha} - A_{p_\beta} \right\| \le h_1, \ \sup_{p_\gamma, p_\lambda \in D} \left\| B_{p_\gamma} - B_{p_\lambda} \right\| \le d_1$$

.

Now we transfer to consideration of a more general problem of synthesis of external loads functions in which the inaccuracy of operators A_p, B_p will be taken into account.

The set of possible solution of equation (6) is necessary to extend to set $Q_{h_1,d_1,\delta}$ taking into account the inaccuracy of the operators $A_p, B_p, p \in D$ [20]:

$$Q_{h_1,d_1,\delta} = \{z : \left\| A_p z - B_p x_\delta \right\| \le h_1 \|z\|_U + B_0\},$$

where $B_0 = d_1 \|x_g\| + \|B_p\| \delta$.

Any function from $Q_{h_1,d_1,\delta}$ causes the response of mathematical model coinciding with the response of investigated object with an error into which errors of experimental measurements and errors of a possible deviation of parameters of a vector $p \in D$ are included. A problem of a finding $z \in Q_{h_1,d_1,\delta}$ will be entitled by analogy to the previous one as a *problem of synthesis for a class of operators* [13,14].

It should be noted that the set of the solutions of a problem of synthesis $Q_{h_1,d_1,\delta}$ at the fixed operators $A_p \in K_A, B_p \in K_B$ in $Q_{h_1,d_1,\delta}$ contain elements with unlimited norm (incorrect problem) therefore the value $(h_1\|z\|_U + B_0 \cdot \delta)$ can be infinitely large. Formally speaking, such situation is unacceptable as it means that the error of mathematical simulation tends to infinity, if for the simulation of external load is used the arbitrary function from $Q_{h_1,d_1,\delta}$ as functions of external load.

Hence not all functions from $Q_{h_1,d_1,\delta}$ will serve as "good" functions of external load.

The function of external loads $z(t)$ in this case can be different. They will depend on final goals of mathematical simulation. For example, we can obtain model z_0, z^1 for simulation of given motion of system as solution of extreme problems [12,13,23]:

$$\Omega[z_0] = \inf_{p \in D} \ \inf_{z \in Q_{\delta,p} \cap Z_1} \Omega[z], \ p \in R^N, \quad (12)$$

$$\Omega[z^1] = \sup_{p \in D} \ \inf_{z \in Q_{\delta,p} \cap Z_1} \Omega[z], \ p \in R^N. \quad (13)$$

The function of external loads which is necessary for estimation from below of output of dynamic system (process) can be obtained as solution of the following extreme problem [13,14]:

$$\left\| A_{p_b} z_b \right\| = \inf_{p \in D} \left\| A_p z_{\delta,p} \right\|. \quad (14)$$

where $z_{\delta,p}$ is the solution of extreme problem (7).

Another model for estimation from above of output of dynamic system (process) can be obtained as solution the extreme problem [12,13]:

$$\left\| A_{p_c} z_c \right\| = \sup_{p \in D} \left\| A_p z_{\delta,p} \right\|. \quad (15)$$

As unitary model z_{un} we can call the solution of following extreme problem [12,14]:

$$\left\| A_{p_{un}} z_{un} - B_{p_{un}} x_\delta \right\| = \inf_{p \in D} \ \sup_{p \in D} \left\| A_p z_{\delta,a} - B_p x_\delta \right\|_U \quad (16)$$

where $z_{\delta,a}$ is the solution of extreme problem (7) with $p = a, \ a \in D$.

The triple $\{A_{p_{un}}, B_{p_{un}}, z_{un}\}$ gives the stable adequate mathematical description for class of operators of process as example of possible one.

The method of special mathematical model selection was suggested to increase of approximate solution exactness of extreme problems [14,15,16].

The real calculations of external loads function z_{un} on rolling mill were presented as an example in papers [11,14].

For the case which is shown on Fig.1. the function of external load as solution of extreme problem (16) is presented on Fig.3.

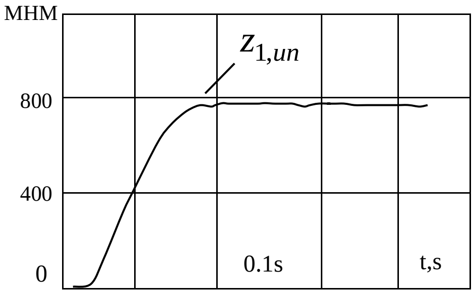


Fig. 3. The diagram of model change.

It is necessary to note that results of synthesis do not vary if to change of the initial data within the limits of accuracy of measurements δ and to change of the initial dynamic system, so that the inaccuracy of operators A_p, B_p would not differ from any operators $A_{p_\alpha} \in K_A$, $B_{p_\alpha} \in K_B$ on value the less then h_1, d_1 accordingly.

The results of calculations show that the estimation from above of accuracy of mathematical modeling with model z_{un} for all $A_p \in K_A$ does not exceed 11 % in the uniform metrics with average error of mathematical model parameters of the main mechanical line of rolling mill equal to 10 % and errors of experimental measurements equal to 7 % in the uniform metrics.

Note that taking into account the error of the operators $A_p \in K_A, B_p \in K_B$ leads to more smooth results of identification (see Fig.2. and Fig.3).

The comparative analysis of mathematical modeling with various known models of external loads and experimental data were presented in work [24]. The model of external load z_{un} better corresponds to new experimental observations.

Now it is possible to give the definition of stable adequate mathematical description of dynamic systems (1).

Definition of stable adequate mathematical description: mathematical description of real process will be considered as stable adequate description with respect to selected variables of state of process model, if under proper limitation on external loads and on the values of variables of state real process with the same additional conditions (initial and boundary) will coincide with experimental measurements of corresponding physical characteristics of real process in given metrics with the accuracy of measurements and exactness of parameters definition of mathematical model. This adequate mathematical description has to be stable with respect to a change of initial data and to a change of experiment's conditions.

Conclusion

The problems of synthesis of adequate mathematical description of real dynamical system are considered in this paper. One of the possible solutions of above-mentioned problems is the choice of function of external loads adapted to dynamical system by the use of identification method. The peculiarities of such approach were investigated. These problems are actually incorrect ones by their nature and that is way for their solution were used the regularization Tikhonov's method. For the case when mathematical model are given approximately different variants of choice of external loads functions which depend on final goals of mathematical simulation are considered. It can be as follows: a copy of given motion of system, different estimation of responses of dynamic system, simulation of best forecast of system motion, the most stable model with respect to small change of initial data, unitary model etc.

REFERENCES

[1]Stepanov, Yu., Shorkin, V. Gordon, V., 2004. Vibratory Working Modelling through Methods of Continuum Mechanics, *Proc. Eleventh International Congress on Sound and Vibration*, pp. 3653-3660. St. Petersburg, Russia.

[2]Soeiro, N., others, 2004. Vibratory Working Modelling through Methods of Continuum Mechanics, *Proc. Eleventh International Congress on Sound and Vibration*, pp. 2821-2828. St. Petersburg, Russia.

[3]Thomas, P., Linden, P., Marah, D., 2004. Modelling Oceanographic Coastal Currents in Small-scale and Largescale Laboratory Experiments, *Proc. XXI International Congress of Theoretical and Applied Mechanics*, p. 5. Warsaw, Poland.

[4]Pop, O., Valle, V., Cottron, M., 2004. Numerical and Experimental Study of the Plastic in the Vicinity of the Crack Tip by the Optical Caustics Method, *Proc. XXI International Congress of Theoretical and Applied Mechanics*, p.4. Warsaw, Poland.

[5]Lino, P., Maione, B., 2008. *Multi-Domain and Control in Mechatronics: the Case of Common Rail Injection Systems*, In-Teh is Groatian Branch of I-Tech and Publishing KG, Vienna, Austria.

[6]Alexik, M,, 2000. Modelling and Identification of Eye-Hand Dynamics, *Simulation Practice and Theory*, 8, 25-38.

[7]Sarmar, I. Malik, A., 2008. Modeling, analysis and simulation of a Pan Tilt Platform based on linear and nonlinear systems, *Proc. IEEE/ASME MESA*, pp. 147-152. China.

[8]Porter, W., 1970. *Modern Foundations of Systems Engineering* (The Macmillan Company, New York, Collier-Macmillan Limited, London).
[9]Krasovskij, N., 1968. *Theory of Motion Control* Science, Moscow.
[10]Breitenecker, F., Judex, F., Popper, N. Breitenecker, K., Mathe, A., Mathe, A., Wassertheurer, S., 2009. Laura and Petrarca - True Emotions vs. Modelled Emotions, *Proc. 6-th Vienna Conference on Mathematical Modelling*, pp.46-69. Vienna, full Papers CD Volume,Vienna Univ. of Technology, ISBM 978-3-901608-35-3.
[11] Shanon, R., 1975. *Systems Simulation - the art and science.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
[12] Menshikov, Yu., 2008. About adequacy of mathematical modeling results, *Proc. Conference "Simulation-2008"*, pp. 119-124. Kiev, Ukraine.
[13] Menshikov, Yu., 2009. Algorithms of construction of adequate mathematical description of dynamic system, *Proc. 6-th Vienna Conference on Mathematical Modelling*, 2009, pp.2482-2485. Vienna - Full Papers CD Volume, Vienna Univ. of Technology, ISBM 978-3-901608-35-3.
[14] Menshikov, Yu., 2009. *Identification of models of external loads*, Dnepropetrovsk, Ukraine.
[15] Stepashko, V., 2008. Method of critical dispersions as analytical apparatus of theory of inductive modeling, *Problems of Control and Information*, 2, pp. 27-32, Kiev, Ukraine.
[16] Gubarev, V., 2008. Method of iterative identification of many-dimensional systems with inexact data. Part 1. Theoretical basises, *Problems of Control and Information,* 2, pp. 8-26, Kiev, Ukraine.
[17]Gukov, O., 2008. Algorithms of iterative identification of many-dimensional systems, *Proc. XV International Conference on Authomatical Control "Authomatics -2008"*, pp.774-777. Odessa: INIA, Ukraine.
[18] Gelfandbein, Ju., Kolosov, L., 1972. *Retrospective identification of perturbations and interferences.* Science, Moscow.
[19] Ikeda, S., Migamoto, S., Sawaragi, Y., 1976. Regularization method for identification of distributed systems, *Proc. IY a Symposium IFAC, Identification and evaluation of parameters of systems,* v.3, pp.153-162. Tbilisi, USSR, Preprint. Moscow.
[20]Tikhonov, A.N., Arsenin, V.Y., 1979. *Methods of incorrect problems solution.* Science, Moscow.
[21]Vilenkin, S., 1968. Application of regularization for evaluation of input signal under realization of output signal, *Automation and Telemechanics,* 21, pp.52-55, Science, Moscow.
[22] Menshikov, Yu., 2011. Synthesis of Adequate Mathematical Description of Rolling Mill Motion, *PAMM, 1*(11), pp.57-58.
[23] Menshikov, Yu., 2011. Inverse Problems in Non-classical Statements, *Int. Journal of Pure and Applied Mathematics, 67*(1), pp.79-96.
[24] Menshikov, Yu., 1994. The Models of External Actions for Mathematical Simulation, *System Analysis and Mathematical Simulation (SAMS)*, *14*(2), pp.139 – 147. New-York.

AUTHORS BIOGRAPHY

Yuri Menshikov. He is working under incorrect problems of identification of external loads on dynamic systems since 1975 year. He has a scientific degree of the Dr. of Science. His list of papers has about 350 scientific works. Together with Prof. Polyakov N.V. the monograph "Identification of Models of External Load" was published by Dr. Yu.Menshikov.

THE DEFENCE OF BRICKS AND MORTAR RETAILING

Roderick Duncan[a], Terry Bossomaier[b], Steve d'Alessandro[c]

[a]School of Accounting and Finance
Charles Sturt University
Panorama Avenue, Bathurst 2795
Australia
[b]Centre for Research in Complex Systems
Charles Sturt University
Panorama Avenue, Bathurst 2795
Australia
[c]School of Management and Marketing
Charles Sturt University
Panorama Avenue, Bathurst 2795
Australia

[a]rduncan@csu.edu.au, [b]tbossomaier@csu.edu.au, [c]sdalessandro@csu.edu.au

ABSTRACT

The traditional retailing sector is under pressure through competition from low-cost online retailers. Certain types of retailers, such as stores selling music and videos, have shrunk dramatically with the introduction of new forms of online retailing in Amazon and iTunes, while other retailers including bookstores are threatened. Yet for other products such as mobile phones, bricks and mortar stores are thriving in this new environment. We develop an agent-based model to study the effect of online retailing on bricks-and-mortar retailers and to predict the types of products most at risk of vanishing from bricks and mortar stores. We develop two factors, immediacy and post-sales service, which help predict the products which will move predominantly to online retailing. This paper also examines possible strategies that bricks and mortar retailers can use to adapt to online competition and the possible use of hybrid channels (a combination of online and offline retailers) by consumers.

Keywords: agent-based model, multi-channels, retailing, online, clicks and bricks, bricks and mortar retailing

1. INTRODUCTION

Norms abound in human society. They are strong determinants of behaviour, but at the same time they are largely arbitrary: the norms of greeting, the exchange of handshakes, hugs and kisses vary dramatically from one culture to another. Young and Burke (2009) show that a norm tends to be stable for long periods and then undergoes tipping points, or sudden transitions (Bossomaier, Barnett, and Harre 2013), to some other norm. A community will thus all behave according to the norm, even though some members of the community will be disadvantaged by it.

The norm which we address in this paper is the shift away from bricks and mortar to online retailing. Although such a shift in norm may be good for profits, it is not necessarily good for consumers. Customers gain through delivery to the door and possibly greater range. They lose through lack of pre-sales advice, after-sales service on a range of timescales, trade-in options and some less tangible sense of community from local shops. The local community loses through reduced jobs and reduced local tax revenues.

2. PREVIOUS RESEARCH ON ONLINE VERSUS BRICKS AND MORTAR RETAILING

The choice between online versus traditional retail bricks and mortar buying behavior has been a topic of much debate over the last decade (Chatterjee 2010; Dawes and Nenycz-Thiel 2014; Keen et al. 2004; Pookulangara, Hawley, and Xiao 2011; Sands, Ferraro, and Luxton 2010; Schramm, Swoboda, and Morschett 2007; Sharma and Krishnan 2002; Toufaily, Souiden, and Ladhari 2013). Essentially the research has focused on the explanation of the migration to online away from traditional retail purchases. Reasons for purchasing online rather than in-store include convenience (Rohm and Swaminathan 2004), lower prices (Junhong, Chintagunta, and Cebollada 2008) and greater choice (Liu, Burns, and Hou 2013). Factors which inhibit online purchasing are; risk of fraud (Huong and Coghill 2008), lack of trust (Toufaily, Souiden, and Ladhari 2013) and the presence of incomplete information about the retailer (Dennis, Jayawardhena, and Papamatthaiou 2010).

Because of the perceived risk of fraud, the need to develop trusting relationships with online retailers in an arena of incomplete and misleading information, consumers rely on word of mouth (WOM) and online reviews more than they do for traditional retailers (Utz, Kerkhof, and van den Bos 2012). Related to WOM, is the role of social norms of behavior. That is, consumers see online retailing as becoming more useful and easier to use, because of the beliefs and actions of others (Činjarević, Tatić, and Petrić 2011; Pavlou 2002; Pookulangara, Hawley, and Xiao 2011).

Consumers do not only decide to use one channel of distribution (online versus brick and mortar retail) for all aspects of decision making. There is emerging evidence that consumers may use some channels to search for information such as online for prices and product availability (often called 'webrooming'), see (Anderson et al. 2010; Sands, Ferraro, and Luxton 2010) and for others, use retail stores for purchases and deliveries (Chatterjee 2010; Tuttle 2013) The deciding factor whether the final purchase is made online or offline, appears to be the expertise and the fulfillment of gratification of consumers (Boyer and Hult 2006; Činjarević, Tatić, and Petrić 2011). Consumers, who use traditional retailing as delivery or purchase points, can have a faster gratification of needs and wants than consumers who have to wait for delivery, and also may have experience less risk since they are purchasing or receiving product or services through more traditional channels. There is also a risk for online retailers that a failure to deliver a product or service within a specified time can lead to greater consumer anxiety and smaller future order sizes (Rao, Griffis, and Goldsby 2011).

It is for these reasons that the death of retail as we know it may be exaggerated. Clearly, though the format of retailing is changing into a hybrid of online and offline channels. Traditional retailers in the U.S for example, have also started to embrace online and mobile marketing approaches such as using text messaging, email and availability of products for pick-up within a half an hour to bring consumers to stores (Byrnes 2007). This means that actions of retailers (online and offline) interact with consumers in a complex system, where for different retail industries different emergent phenomena (for example, the use of hybrid retail models) may form.

3. SIMULATION MODEL

This model simulates the choice of consumers whether to purchase a particular product through a bricks and mortar store or through an online retailer. For simplicity we assume that the customers make such a choice for each type of product. Different products are accommodated by altering parameters in the model to produce a prediction of the social norm for retailing choice for each product.

3.1. Customers

The customers are represented by an agent, denoted *i*. Customers are randomly connected to other customers and exchange information about their retailing experiences through these social networks. The more links within the networks of customers the more effectively information about retailing alternatives can pass through the customers. The probability of agent *i* linking to another agent is given by the parameter η, which is randomly calculated for each agent.

3.2. Customer behaviour

Each time step *t* the customer chooses whether to purchase a product from the bricks and mortar retailer (BMR) or the online retailer (OR). The retailing choice of customer *i* at time step *t* denoted $c_i(t)$ depends on its experience $x_{ij}(t)$ with the *j* being one of the categories of retailer (BMR or OR).

We assume the probability of choosing a given retailer is a logistic function of the customer's levels of past experience with the retailers. The probability of customer *i* choosing BMR at time step *t* is then

$$P\{i\ chooses\ BMR\} = \frac{e^{\beta(x_{i,BMR}(t) - x_{i,OR}(t))}}{1 + e^{\beta(x_{i,BMR}(t) - x_{i,OR}(t))}} \quad (1)$$

This logistic equation is in common use in studying choice in economics (McFadden 1974) and in marketing. The beta parameter controls the degree of noise in the model. When beta is zero, all options have equal likelihood. As beta increases one choice (the higher experience or utility) increases in probability eventually excluding the alternative choice.

The probability function (Eq. 1) arises naturally as the equilibrium solution to a variety of equations, such as the Fokker Planck diffusion equation and classical thermodynamics (Solé 2011). It is of course the Boltzmann distribution which occurs throughout thermodynamics. In thermodynamics, beta is the inverse of temperature. As beta decreases towards zero (temperature becomes infinite), the system becomes hotter and the distribution of possible states flattens.

3.3. The retailers interaction with customers

After the customers have made their choices about which retailing alternative to use, the customers' experience is calculated. We assume that the OR is the base retailing option with the lowest price and the least level of service. The customer experience from an OR is a baseline amount (*V*) which represents the value of the product to the customer less the wholesale price of the product and the OR mark-up

$$x_{i,OR}(t) = V + \varepsilon_i(t) \quad (2)$$

and where ε represents the risk of poor service around an online purchase. This risk includes the choice of inappropriate product because of a lack of pre-sales experience with the product as well as the pre-sales advice which may be provided by a BMR. Customers at a BMR are assumed not to bear this same risk.

Customers have the possibility of a hybrid strategy of purchasing. A customer could visit a BMR to get pre-sales experience and advice about the product, but then make the actual purchase with OR. A BMR will generally not be able to achieve an advantage over an OR with pre-sales advice for this reason, but there may be a possible advantage with pre-sales experience of a product. Even where a customer tries a product at the store, the online-purchased product may not be exactly the same as the one experienced in the store, for example the fit of a pair of shoes. We come back to this hybrid strategy for customers later in the paper.

The BMR offers higher levels of service (both before and after sales), but the BMR will have higher costs of operation which have to be recouped through a higher sales price. The higher level of pre-sales service or advice from the BMR as well as the ability to experience the product before the sale in a BMR context means that the risk of poor product choice is lower than it would be for an online transaction.

The price of a BMR product to a customer (and so also the customer's experience as price negatively affects experience) depends on the choices of the other customers as the overhead of the BMR operations have to be covered by the price premium the BMR charges. It is this network effect that creates a coordination problem for the customers and the social norm aspect of the simulation (Young and Burke 2009; Young 2011). The larger the number of customers who choose BMR, the lower the price charged by the BMR and the higher the experience of those customers.

There are other considerations that customers may take into account when comparing BMR to OR other than price and product risk. One advantage with a BMR purchase is the time between purchase and access to the product. With a BMR the customer can usually walk away with the product or have it delivered that day. Generally delayed gratification has less value to a customer. We call this the immediacy or gratification value, denoted *G*, which adds to customer experience with a BMR. The value of G would be expected to differ across goods. For some products, perhaps mobile phones, walking out of the store with that product right now might be highly desirable for a customer, while it may not be so important for other products.

Another customer advantage with a BMR is the ease and surety of post-sales service for the customer. By post-sales service we mean any interaction with the retailer occurring after the purchase of the product, which may mean replacement of a defective product, purchase of replacement parts sometime in the future or advice with some aspect of the product. The customer can simply go back to the place of purchase to speak to a store representative for a BMR, while the future existence of a website and the ease of online or telephone service for an OR may not be as convenient.

Post-sales service is not as important for some products as for others. The post-sales service for books or compact discs is minimal except for the return of a defective product. The future service and maintenance for cars is essential for the continued use of the car. We denote value of this post-sales service or future service by *F*. The post-sales service might occur much later after the purchase of product, however for simplicity we telescope all the future interactions to the present time step for the purpose of calculating the customer experience.

The customer experience with a BMR is a combination of the net value of the product less wholesale price and retailer mark-up (*V*), as well as the immediacy value, G, and the post-sales service, F. However the BMR price also has to cover the cost of operation of the bricks and mortar presence. We assume that the cost of the presence (overhead or *OH*) is spread across all the customers who purchase at the BMR that time step. The experience of customer *i* at a BMR at time step *t* is then

$$x_{i,BMR}(t) = V + F + G + \frac{OH}{N} \qquad (3)$$

where *N* is the number of customers making the choice of BMR that time step.

After the calculation of all the customers' experiences, the customers then share the experiences across their social networks. To calculate the sharing of information about retailers, each agent calculates a weighted average of their own experience with each type of retailer this time step with the experience of each of their network neighbours. The weight given to the neighbours' experience is $\alpha \in [0\text{-}1]$.

3.4. Industries in transition

The simulation is run for a particular product, such as compact discs, music or high fashion footwear. Each product will have its own values for gratification (*G*) and for post-sales service (*F*). In Table 1 we present some anticipated values of *F* and *G* for particular products.

Table 1: Factors differentiating products in retailing

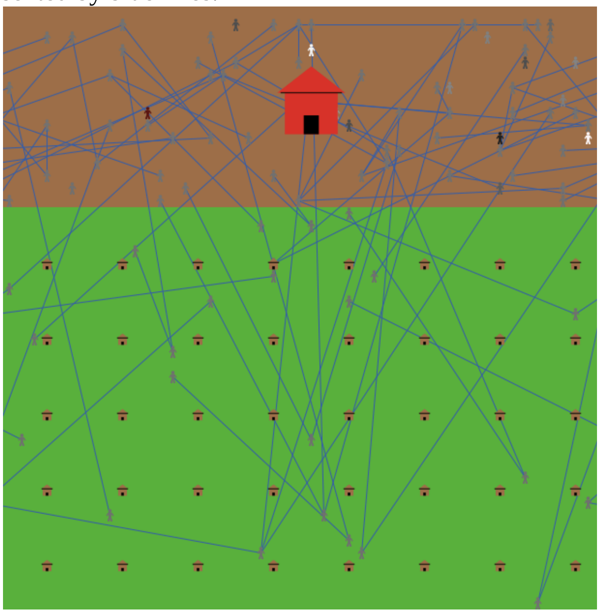
	Factors (out of 0..10)	
Category of product	Gratification (G)	Post-sales service (F)
CDs	2	0
Books	2	0
Hardware	5	1
White goods	4	4
Footwear and clothing	7	4
Mobile phones	8	6
Cars	6	10

For products such as books or compacts discs, the value of immediacy would be low as would be the value of post-sales service. Hardware products would likely be purchased with a particular task in mind, so the immediacy of the purchase is importance, while post-sales service is quite unimportant. For products such as cars the immediacy is not that valued, but the level of after-sales service and continuing maintenance is quite important. The complexity and specificity of car maintenance tools recently suggests that manufacturers may be making use of after-sales service to generate revenue for their retailers – a BMW may need to be brought to a BMW shop to be serviced properly.

3.5. The environment

The map of the simulation is a visual representation of the choice made by the customers. There are two areas on the map: one area with housing representing a choice of online retailing for that customer and one area with a large mall representing a choice of bricks and mortar retailing for that customer.

Figure 1: The environment of the model: the customers represented by figures, social networks represented by blue lines.



3.6. Method

The ABM was created in NetLogo (Wilenksi 1999). In this version the only agents are the consumers, who all buy the same product, but choose between OR and BMR.

The number of consumers is set at 100. The customers' initial levels of experience with the two categories of retails are randomized. The levels of the other parameters for the simulations are presented in Table 2.

Table 2: Model parameters and their values

Parameter	Symbol	Value
Probability of forming links with other customers	η	0.7
Degree of noise in customer decision	β	[0..100]
Importance of social network information	α	0.5
Value of product to consumer net of online purchase price	V	50
Risk of poor choice or poor experience with an online purchase	ε	[-10..10]
Importance of immediate access to customers	G	[0..10]
Importance of continuing service after sales to customer	F	[0..10]

The values of *F* and *G* enter the BMR customer experience equation in a parallel manner, thus we assume that we can examine the value of *F* + *G* for the simulations. We simulate a range of values of *F* +*G*, corresponding to the values in Table 3 which gives putative values for different products. For each such pair of values we run a behaviour space of 100 tests of a given beta parameter and determine the relative proportion of OR and BMRs. We run a series of beta values to determine at what point coordination sets in.

4. RESULTS

Table 3 shows the results of 100 simulations each of the model for values of *F* + *G* between 0 and 20 and values of β between 0 and 100. For each level of (*F* + *G*, *β*) the average number of customers (out of 100 customers in the simulation) for 100 runs is presented.

Table 3: Results of the simulations: average final number of customers for BMR for 100 runs

	β			
F + G	0	1	10	100
0	50.05	46.49	0.05	0.03
5	50.32	47.34	0.05	1.84
10	49.95	50.03	31.1	49.51
15	50.58	50.65	68.36	85.29
20	49.84	52.51	80.12	87.97

We would expect to see that for higher levels of gratification and post-sales service the average number of customers for BMR should increase. This is indeed what we see for values of β, the noise parameter in the logistic choice function, higher than 1.0. For values of β of 1.0 and below, we see that the noise dominates, and the choice between BMR and OR is essentially random.

5. CONCLUSIONS

The results of the simulation presented in Table 3 show that higher customer values for immediate gratification and of post-sales service can push customers towards bricks and mortar retailing. This finding suggests that the disappearance of categories of retailers such as compact disc or books is not a coincidence and that these sectors may continue to shrink in the future. Bricks and mortar retailers of products in these categories will need to search for alternative strategies of attracting customers.

One possibility which bricks and mortar retailers may explore is to search for ways to increase the values of gratification and future service in order to compete better against online retailers. A first thought in response to the entry of an online retailer might be that bricks and mortars should scale back operations to cut costs and thus to compete against online retailers on price, however this is likely to be a limited strategy. It may be better for bricks and mortar retailers to go in the other direction and to invest more in their store presence to improve the store experience for customers or to improve post-sales service to provide a competitive advantage against the virtual retailers.

These simulations assume that customers have a choice of only one retailer, either BMR or OR, to purchase the product, however customers may use a hybrid strategy for purchasing (Tuttle 2013): examine the product and get pre-sales advice at the BMR and then buy the product at the lower OR price. This is, of course, a terrible outcome for the BMR, which is providing the pre-sales advice and experience but then not being able to recoup the costs of that service through a sale.

A customer hybrid strategy however makes a lot of sense for a BMR owned by the manufacturer of the product. The customer can experience the product and get pre-sales advice at the BMR owned by the manufacturer and then purchase the produce at the OR, which is supplied by the manufacturer. This strategy may explain the proliferation of producer-owned stores which we observe currently: the Apple stores, the Sony stores or the Coach handbag stores which are appearing in large cities.

The producer-owned BMR can serve two purposes for the producer. The BMR can be a portal for the provision of pre-sales advice and service to customers, as the producer-owned BMR can recoup the costs of the service through the sales of the product whether online or at the BMR. Just as importantly, the producer-owned BMR serves as a brand signal for the producer. This brand investment may explain the lavishness of BMRs such as the Apple stores, which could not possibly recoup their construction and maintenance cost through store sales.

The gratification advantage of BMRs – the ability to see and immediately purchase the product – has dwindled with the ubiquity of parcel-delivery services shortening the wait for OR customer purchases. Technology innovations such as iTunes or Amazon's Kindle along with greater availability of broadband internet have completely overcome the gratification advantage for the BMR for some types of products, as the products whether books, music tracks or movies can be purchased and downloaded almost immediately, avoiding the possibility of a trip to the BMR.

The coverage of gratification and post-sales service in the current simulation is limited, so much so that the two factors can be compressed into one dimension in these simulations. A planned extension of this research is to expand these factors to allow for a clearer differentiation of the two factors.

Other planned extensions of the simulation are to include the social or joint aspects of bricks-and-mortar retailing, including interactions with other customers, with other retailers and with BMR sales staff. Many major bricks and mortar retailers are co-located with other retailers, which allow for

- the social aspect of customer experience, where customer interaction with other customers may affect customer experience;
- the purchase of multiple products with one retailing experience for the customer; and
- the possibility of interacting with multiple retailers for a customer.

A further possible extension of the model is to examine the pricing plans of the bricks and mortar retailers. Where customers make a single trip for multiple products, bricks and mortar retailers may seek to match online retailers only on major items, but then cover their margins through mark-ups on minor items, such as power cords or on extended warranties for products.

REFERENCES

Anderson, E. T., Fong, N. M., Simester, D. I., and Tucker, C. E., 2010. How sales taxes affect customer and firm behavior: The role of search on the internet. *Journal of Marketing Research* 47(2), 229-239.

Bossomaier, T., Barnett, L., and Harre, M., 2013. Information and phase transitions in socio-economic systems. *Complex Adaptive Systems Modeling* 1(1), 1-25.

Boyer, K. K., and Hult, G. T. M., 2006. Customer behavioral intentions for online purchases: An examination of fulfillment method and customer experience level. *Journal of Operations Management* 24(2), 124-147.

Byrnes, N., 2007. More clicks at the bricks. *BusinessWeek* (4063), 050-052.

Chatterjee, P., 2010. Causes and consequences of 'order online pick up in-store' shopping behavior. *International Review of Retail, Distribution & Consumer Research* 20(4), 431-448.

Činjarević, M., Tatić, K., and Petrić, S., 2011. See it, like it, buy it! Hedonic shopping motivations and impulse buying. *Economic Review: Journal of Economics & Business / Ekonomska Revija: Casopis za Ekonomiju i Biznis* 93-15.

Dawes, J., and Nenycz-Thiel, M., 2014. Comparing retailer purchase patterns and brand metrics for in-store and online grocery purchasing. *Journal of Marketing Management* 30(3/4), 364-382.

Dennis, C., Jayawardhena, C., and Papamatthaiou, E.-K., 2010. Antecedents of internet shopping intentions and the moderating effects of substitutability. *International Review of Retail, Distribution & Consumer Research* 20(4), 411-430.

Huong, H., and Coghill, K., 2008. Online shoppers in australia: Dealing with problems. *International Journal of Consumer Studies* 32(1), 5-17.

Junhong, C., Chintagunta, P., and Cebollada, J., 2008. A comparison of within-household price sensitivity across online and offline channels. *Marketing Science* 27(2), 283-299.

Keen, C., Wetzels, M., de Ruyter, K., and Feinberg, R., 2004. E-tailers versus retailers: Which factors determine consumer preferences. *Journal of Business Research* 57(7), 685.

Liu, X., Burns, A. C., and Hou, Y., 2013. Comparing online and in-store shopping behavior towards luxury goods. *International Journal of Retail & Distribution Management* 41(11/12), 885-900.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In *Frontiers in econometrics*, P. Zarembka, ed. New York: Academic Press 105-142.

Pavlou, P. A., 2002. What drives electonic commerce? A theory of planned behavior perspective. *Academy of Management Proceedings & Membership Directory*A1-A6.

Pookulangara, S., Hawley, J., and Xiao, G., 2011. Explaining consumers' channel-switching behavior using the theory of planned behavior. *Journal of Retailing & Consumer Services* 18(4), 311-321.

Rao, S., Griffis, S. E., and Goldsby, T. J., 2011. Failure to deliver? Linking online order fulfillment glitches with future purchase behavior. *Journal of Operations Management* 29(7/8), 692-703.

Rohm, A. J., and Swaminathan, V., 2004. A typology of online shoppers based on shopping motivations. *Journal of Business Research* 57(7), 748.

Sands, S., Ferraro, C., and Luxton, S., 2010. Does the online channel pay? A comparison of online versus offline information search on physical store spend. *International Review of Retail, Distribution & Consumer Research* 20(4), 397-410.

Schramm, H., Swoboda, B., and Morschett, D., 2007. Internet vs. Brick-and-mortar stores - analysing the influence of shopping motives on retail channel choice among internet users. *Journal of Customer Behaviour* 6(1), 19-36.

Sharma, A., and Krishnan, R., 2002. Clicks only, clicks and bricks, and bricks only: Are retail salespeople an important factor in choice? *Journal of Marketing Management* 18(3/4), 317-336.

Solé, R. S., 2011. *Phase transitions*. Princeton, N.J: Princeton University Press

Toufaily, E., Souiden, N., and Ladhari, R., 2013. Consumer trust toward retail websites: Comparison between pure click and click-and-brick retailers. *Journal of Retailing & Consumer Services* 20(6), 538-548.

Tuttle, B., 2013. Look out, amazon: More shoppers browse online, then purchase in stores. *Time.com*1-1.

Utz, S., Kerkhof, P., and van den Bos, J., 2012. Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. *Electronic Commerce Research & Applications* 11(1), 49-58.

Young, H. P., 2011. The dynamics of social innovation. *Proceedings of the National Academy of Sciences* 108(Supplement 4), 21285-21291.

Young, H. P., and Burke, M. A., 2009. Social norms. In *Handbook of social economics*, A. Bisin, J. Benhabib and M. Jackson, ed. North-Holland, Amsterdam

SIMULATION OF HOT SPOTS GENERATION PROCESS ON SEPARATOR DISC IN MULTI-DISC CLUTCH

Cenbo Xiong [a], Heyan Li [a*], Song He [a], Biao Ma [a], Qingdong Yan [a], Hongcai Li [a]

Huiyu Xu [b], Jianwen Chen [b]

[a] School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China

[b] JiangLu Machinery & Electronics Group Co.,Ltd, Xiangtan City, Hunan province, 411100, China

[*] Corresponding author. Tel.: +86-10-68918085; fax: +86-10-68918085. E-mail address:

Lovheyan@gmail.com (Heyan Li).

Abstract
In multi-disc clutches, the separator discs slide relatively to the friction discs during engagement, and usually display "hot spots" in a sinusoidal shape on both surfaces of a separator disc. Explanations based on "perturbation method" were widely used to describe the generation and development of hot spots in seals, brakes and clutches. In this study, a multi-disc clutch experiment was implemented to observe hot spots phenomenon, based on the results, a scenario of hot spots generation mechanism on thin disc is introduced and a 3D finite element non-uniform contact model of multi-disc clutch was developed to simulate the developing process of hot spots and illustrate the generation mechanism scenario.

Keywords: Multi-disc clutch; Hot spots; Generation mechanism

1. Introduction

Hot spots were commonly observed in brakes and clutches that may cause damage or even failure of these organs, many researches had been done to explain the causes and influences of hot spots on their serving performance. J. R. Barber proposed the concept of "Thermo-elastic Instability" (TEI) to describe the instability of railway braking systems which may fail due to the coupled thermal-elastic and wear effects between the braking block and the wheel[1]. TH. A. Dow and R. A. Burton introduced a perturbation method to calculate the temperature distribution and surface pressure at the interface between a thin blade and half-space[2], this method related the heat transfer process with sliding speed and determined a "critical speed" for the frictional sliding system, the system would become unstable when surpassing the critical speed.

Anderson and Knapp classified hot spots obtained during automotive braking experiments and described them as well as their possible causes in details[3]; Kasem used infrared camera to record the developing process of hot spots on a one-face brake during braking[4]; Burton observed the generation of hot spots in a transparent glass plate and metal cup friction system with temperature probes[5].

In addition to experimental methods, considerable researches had been done theoretically to describe hot spots phenomenon. Zagrodzki used numerical methods to calculate the temperature distribution during the transient engagement of a multi-disc wet clutch and then analyzed it quasi-statically to obtain the stress distribution in the separator discs and friction discs[6]; He established a coupled thermal-elastic model which employed numerical methods to calculate the temperature field, thermal stress field and contact pressure distribution during severe engagement[7]. And also introduced a contact pressure perturbation to solve the coupled problem and studied the transient behavior of the engagement[8,9,10].

Typical structure of multi-disc clutch is shown in Fig. 1a, the clutch embodies alternately assembled separator discs and friction discs as shown in Fig. 1b, and would engage or disengage under the pressure of control oil. Due to the structure and working limitation of multi-disc clutches, it's difficult to observe the generation process of hot spots or even measure the temperature distribution. The generation of hot spots is a coupled thermal-elastic problem, it's important to understand the interaction between temperature variation and out-of-plane deformation of the surfaces. Recently, Panier proposed a new progressive waviness distortion (PWD) theory based on plate theory (Kirchhoff assumption) and thermal-elastic theory to explain that the generation of macroscopic hot spots is related to the energy dissipated in the braking system[11]. In this paper, a similar scenario of hot spots generation process on thin discs named "hot spots propagation mechanism" is proposed; a 3D finite element model of non-uniform contact is established to simulate the developing process of hot spots on the separator disc.

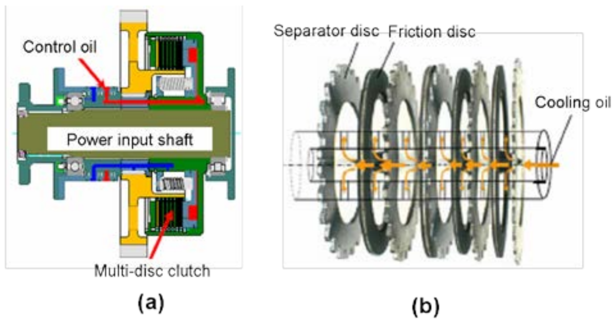


Fig. 1 Multi-disc clutch

2. Experiment

An experimental test bench was set up as shown in Fig. 2 to investigate the generation of hot spots on separator disc, the multi-disc clutch was assembled in a simplified transmission which is driven by electric motor and braked by hydraulic motor.

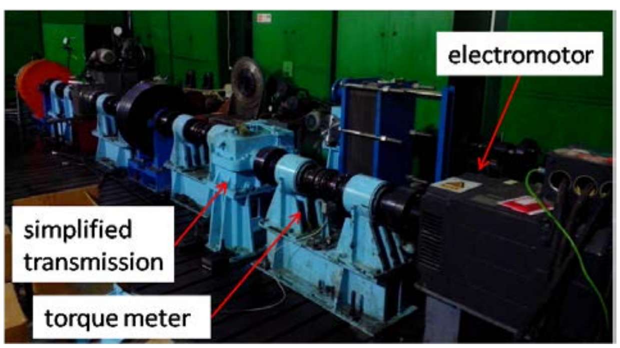


Fig. 2 Test bench of multi-disc clutch

The experiment was implemented mainly in the following steps: First, the control oil pressure was applied on the piston in the clutch and increased from 0 to 1MPa, the discs got into contact with each other statically; Second, hydraulic pressure was applied by the hydraulic motor to simulate the resistance of the vehicle before rotating of the drive shaft; Third, the rotating speed of electric motor started from 0 to 1000r/min, this process lasted for seconds. Hot spots could be found on some separator discs after the experiment, typical photos were taken as shown in Fig. 3.

Some obvious abrasion patches can be seen on the separator discs, they are hot spots produced during the engagement and distributed alternatively on both sides of the separator disc and located on the middle radius of the disc, which means that the high pressure contact area is on the middle radius.

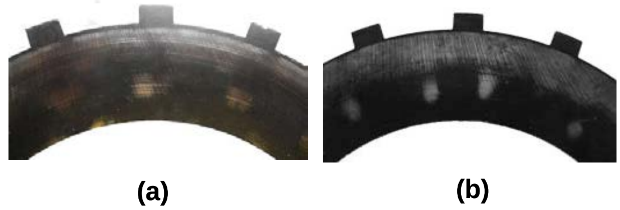


Fig. 3 Hot spots on separator discs

3. Generation mechanism hypothesis of hot spots on separator disc

The well known "Thermal-elastic Instability (TEI)" theory assumed a sinusoidal perturbation of temperature on the surface of a blade or a disc, with boundary conditions, critical speed for these organs could be figured out. While the generation and developing process of hot spots on thin discs may have alternative explanations. It is strongly influenced by the contact status of the two contacting surfaces as has been mentioned by a lot of researches in the past.

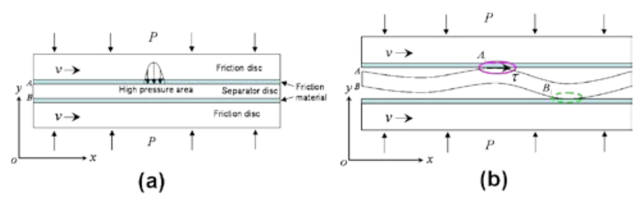


Fig. 4 Generation mechanism of hot spots on the separator disc

Based on the experimental phenomenon above, we assume that a local high pressure area exists on surface A of the separator disc which is caused by the non-uniform initial contact pressure distribution as shown in Fig. 4a. This area will be heated due to frictional heat flux and out-of-plane deflection is expected there. Under the effects of thermal banding, vertical pressure and shear traction on surface A, the disc would deform as shown in Fig. 4b and the first hot spot A_1 emerges. In the down-flow direction of the relative velocity on surface B, the separator disc will bulge until getting into contact with the friction disc and a new hot spot B_1 will be generated there in the same way as A_1.

This coupled thermal-elastic process can be illustrated detailedly in Fig. 5. An initial local high pressure emerges on surface A which inspires the coupled thermal-elastic process of hot spots. The hot spots developing process on one surface can be roughly divided into three steps: 1. Local heating; 2. Local bulging in the heated area; 3. Coupled thermal-elastic process to change the contact pressure.

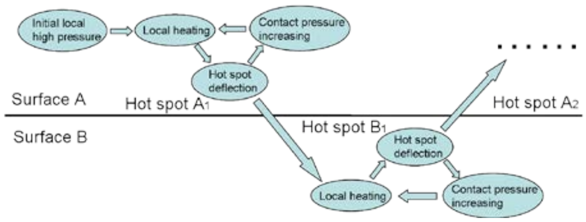


Fig. 5 Coupled thermal-elastic effect and propagation process of hot spots generation

These processes are repeated again and again and passed from one surface to the other on the disc until eventually a dissymmetric form of hot spots distribution is formed on the separator disc, which is similar to what mentioned by Lee and Barber[5], this hot spots generation process can be introduced as "hot spots propagation mechanism".

4. Simulation of hot spots generation

A 3D coupled thermal-elastic model of clutch was established in finite element software ABAQUS in order to simulate the generation process of hot spots during a transient engagement.

4.1 The model

Fig. 6a shows the 3D clutch model, part 1 and 3 are the friction discs, each one consists of a supporting plate (blue) and friction layer (brown), in order to simulate the higher pressure on the middle radius of the separator disc part 2, two ridges are designed in the radius interval of $[0.095m, 0.11m]$ on both friction discs (Fig. 6c shows the enlarged local details). The distance between the steel disc and friction disc is enlarged for the convenience of observation.

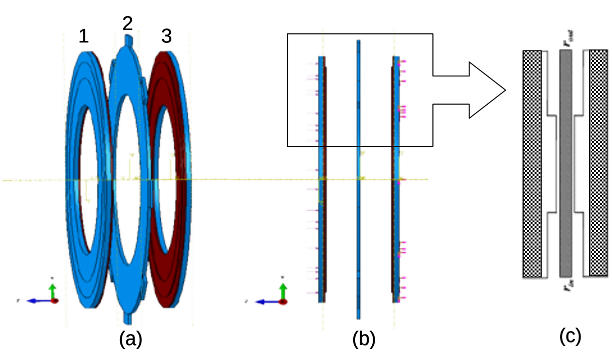


Fig. 6 The 3D non-uniform contact model of the clutch

In Fig. 6b, the back surface of friction disc 3 is fixed and a uniform axial load pressure is applied on the friction disc 1. Wear of the separator disc and friction disc is ignored in the model.

We assume the friction coefficient of the contact surfaces as a constant $\mu = 0.1$, axial pressure in the model was applied before the rotation of the separator disc to avoid engagement impact at most. The relative load pressure and rotating amplitude curves are shown in Fig. 7 and the initial temperature of the whole model was designated to be $20^{\circ}C$. Since the engagement time was so short that the liquid cooling and heat radiation were negligible, thus the model was established as a dry friction system.

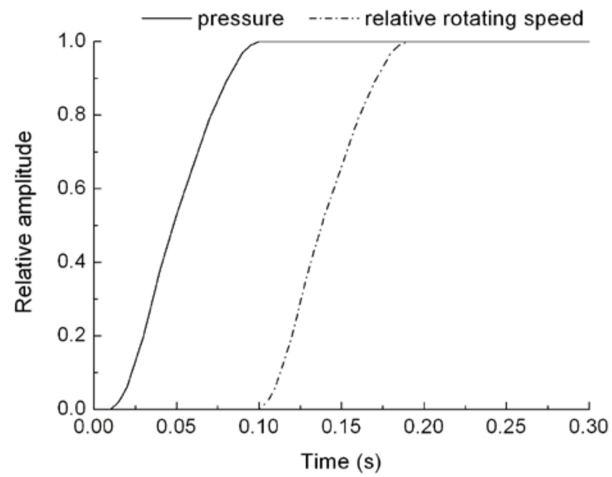


Fig. 7 Relative amplitude curves of pressure applied on the clutch and rotating speed

4.2 Heat generation, separation and conduction

The material and geometric parameters of the separator disc and friction discs in the model are shown in Table 1.

Table 1 Material and geometric parameters of the model

Material Properties	Separator disc	Friction layer
Thermal Conductivity	42	8.5
Density(Kg/ m^3)	7800	5500
Young's Modulus(Pa)	2.1E11	2.26E10
Poisson's Ratio	0.3	0.3
Expansion Coefficient(K^{-1})	5.27E-5	1.27E-5
Specific Heat(J/(kg · K))	452	600
Geometry Parameters	**Separator disc**	**Contacting**
Outer radius(m) (without teeth)	0.12	0.11
Inner radius(m)	0.08	0.095
Thickness(m)	0.002	

On contact surfaces, the heat flux q_f generated by frictional sliding can be expressed as:

$$q_f = \eta \tau v \tag{1}$$

Where η is the conversion factor of friction work to heat, $\tau = p\mu$ is the frictional stress between the contacting surfaces which has relationship with the axial pressure p, friction coefficient μ and also with the temperatures on both interacting faces, $v = r\omega$ is the relative sliding speed between the contacting surfaces.

The heat is separated and conducted to both of the contacting surfaces immediately when it's generated. According to Tien-Chen Jen's research[12], following the continuous conditions of temperature and heat flux on the interface between different materials, the ratio of the separated heat between the surfaces can be expressed as:

$$\frac{q_1}{q_2} = \frac{\lambda_1}{\lambda_2}\sqrt{\frac{\alpha_2}{\alpha_1}} \tag{2}$$

Where q_1 and q_2 denotes the heat flux conducted into the friction disc and separator disc separately as shown in Fig. 8. From the expression, we can see that the heat separation depends on the material properties of the contacting surfaces. For the given materials, the ratio is $\frac{q_1}{q_2} = \frac{3}{7}$ which means that 70% heat is conducted into the separator disc.

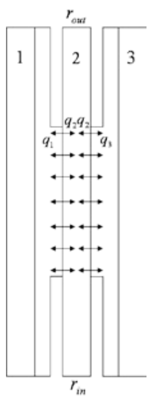


Fig. 8 Heat generation and conduction between the discs

The temperature field in the separator disc can be calculated by the transient heat conduction equation, since there is no internal heat source, the equation can be expressed in cylindrical coordinates as:

$$\rho c\frac{\partial T}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(\lambda_r r\frac{\partial T}{\partial r}\right) + \frac{1}{r^2}\frac{\partial}{\partial \theta}\left(\lambda_\theta \frac{\partial T}{\partial \theta}\right) + \frac{\partial}{\partial z}\left(\lambda_z\frac{\partial T}{\partial z}\right) \tag{3}$$

There is only heat flux boundary condition on the contacting surface S_q, that is

$$\lambda_r\frac{\partial T}{\partial r}n_r + \lambda_\theta\frac{\partial T}{\partial \theta}n_\theta + \lambda_z\frac{\partial T}{\partial z}n_z = q \tag{4}$$

For homogenous materials, $\lambda_r = \lambda_\theta = \lambda_z = \lambda$ is the thermal conductivity, the governing equation and heat flux boundary condition can be simplified as:

$$\frac{\partial T}{\partial t} = a\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 T}{\partial \theta^2} + \frac{\partial^2 T}{\partial z^2}\right] \tag{5}$$

$$\lambda\left(\frac{\partial T}{\partial r}n_r + \frac{\partial T}{\partial \theta}n_\theta + \frac{\partial T}{\partial z}n_z\right)\Bigg|_{S_q} = q \tag{6}$$

where $a = \frac{\lambda}{\rho c}$ is thermal diffusivity of the separator disc material, neglecting the convective and radiation heat dissipation, with weighted residual method, the equivalent integration formation of this problem is:

$$\int_V w\left[\rho c\frac{\partial T}{\partial t} - \frac{\lambda}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) - \frac{\lambda}{r^2}\frac{\partial^2 T}{\partial \theta^2} - \lambda\frac{\partial^2 T}{\partial z^2}\right]dV$$
$$+\int_{S_q} w_1\left[\lambda\left(\frac{\partial T}{\partial r}n_r + \frac{\partial T}{\partial \theta}n_\theta + \frac{\partial T}{\partial z}n_z\right) - q\right]dS = 0 \tag{7}$$

For the arbitrariness of weight function in Galerkin method, we choose

$$w = w_1 = \delta T \tag{8}$$

Using Guassian theorem, the weak form of equation 7 can be expressed as:

$$\int_V \rho c\delta T\frac{\partial T}{\partial t}dV - \int_{S_q}\delta T q dS +$$
$$\int_V \lambda\left[r\frac{\partial T}{\partial r}\frac{\partial}{\partial r}\left(\frac{\delta T}{r}\right) + \frac{\partial T}{\partial \theta}\frac{\partial}{\partial \theta}\left(\frac{\delta T}{r^2}\right) + \frac{\partial T}{\partial z}\frac{\partial \delta T}{\partial z}\right]dV = 0 \tag{9}$$

The temperature T in this model can be approximated by:

$$T = \tilde{T} = \sum_{i=1}^{n_e} N_i(r,\theta,z)T_i(t) \tag{10}$$

Where $N_i(r,\theta,z)$ are shape functions and $T_i(t)$ are the nodal temperature values, substituting equation 10 into equation 9 and choosing the shape function as the weight function results in the discrete finite element form:

$$\int_V \rho c N_i N_j \dot{T}_j dV - \int_{S_q} N_i q dS +$$
$$\int_V \lambda\left[r\frac{\partial N_j}{\partial r}\frac{\partial}{\partial r}\left(\frac{N_i}{r}\right) + \frac{1}{r^2}\frac{\partial N_j}{\partial \theta}\frac{\partial N_i}{\partial \theta} + \frac{\partial N_j}{\partial z}\frac{\partial N_i}{\partial z}\right]T_j dV = 0 \tag{11}$$

Simplify this equation to be matrix form:

$$C\dot{T} + KT = P \tag{12}$$

These are a set of ordinary differential equations and all the matrix coefficients are lamped by elements:

$C_{ij} = \sum_e C_{ij}^e$, where $C_{ij}^e = \int_{V^e}\rho c N_i N_j dV$ is the element specific capacity matrix.

$K_{ij} = \sum_e K_{ij}^e$, Where

$$K_{ij}^e = \int_{V^e} \lambda \left[r \frac{\partial N_j}{\partial r} \frac{\partial}{\partial r} \left(\frac{N_j}{r} \right) + \frac{1}{r^2} \frac{\partial N_j}{\partial \theta} \frac{\partial N_i}{\partial \theta} + \frac{\partial N_j}{\partial z} \frac{\partial N_i}{\partial z} \right] dV$$

is the element heat conduction matrix.

$P_i = \sum_e P_{q_i}^e$, where $P_{q_i}^e = \int_{S_q} N_i qdS$ is the heat flux load boundary condition vector.

This equation can be solved by direct integration method and the backward difference algorithm is used since it is unconditionally stable.

4.3 Thermal-elastic problem

The thermal deformation and stresses in the model can be calculated since the temperature field had been figured out. Using the variational principle, the finite element equilibrium equation for the discs can be expressed in matrix form as

$$KU = P + P_T \tag{13}$$

Where

$K = \sum_{e=1}^{n_e} k^e$, $k^e = \int_{V_e} B^T DBdV$ is the element stiffness matrix. $P = \sum_{e=1}^{n_e} \int_S N^T FdS$ is the load vector on the surfaces, N is the shape function and F is the force vector applied on the surfaces. $P_T = \sum_{e=1}^{n_e} \int_V B^T D\varepsilon_0 dV$ is the equivalent load caused by temperature variation in the domain where B is the strain matrix, D is the elastic matrix and $\varepsilon_0 = \alpha\Delta T$ denotes the strain result from the temperature variation.

The displacement U of the discscould be calculated from equation 13, the explicit central-difference integration rule is used to obtain the mechanical response with a lumped mass matrix, the heat transfer and mechanical solutions are obtained simultaneously by an explicit coupling, because both the forward-difference and central-difference integrations are explicit, no iterations or tangent stiffness matrices are required for explicit integration, it can be less expensive computationally and the contact problem can be simplified[13].

5 Simulation and results

In the simulation, we applied uniform pressure on the friction disc that induced 1MPa nominal pressure on the contact interface, the friction discs didn't rotate and the separator disc rotated from 0 to 1000 rpm in 0.3s following the rotating speed amplitude curve in Fig. 7. The whole simulation period was 0.32s with 0.02s for applying the axial pressure and 0.3s for rotating.

There are 30000 elements in the model, due to the tiny deformation in the hot spots area, it is reasonable to adopt the linear coupled "solid temperature-displacement" element C3D8RT which is an 8-node trilinear displacement and temperature element, reduced integration with hourglass control. This type of element doesn't include geometrical nonlinear, so we used 8 elements along the thickness of the separator disc to increase the bending stiffness.

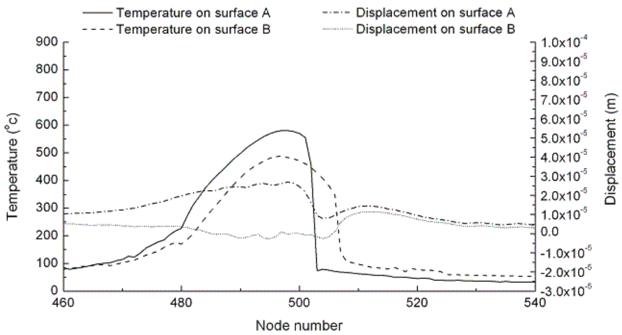
There were 750710 time increments in the whole simulation process and the stable time increment was 4.01237e-7s. With the simulation results, we can see the generation process of hot spots on the surfaces of the discs.

5.1 The generation of hot spots

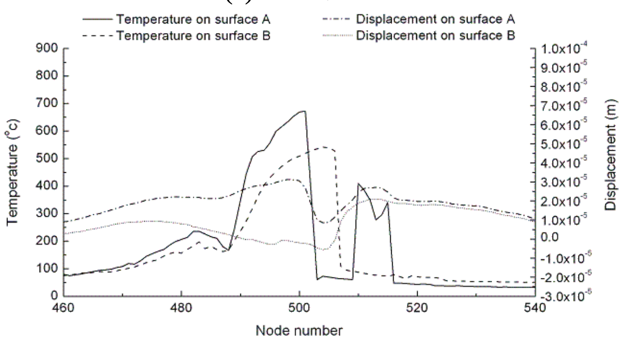
The surfaces of the discs were flat at the very beginning of the simulation in this ABAQUS model, but when the axial pressure was applied on the friction disc and rotating torque was applied on the separator disc respectively, the surfaces of the separator disc had some tiny initial non-uniform axial deflection, this could be considered as the initial perturbation on the surface just as roughness on practical discs. Because of the coupled thermal-elastic effect, the initial axial displacement fluctuation was enlarged until the first local hot spot formed. Extracting the axial displacement and temperature data from all 560 points evenly along a circle path through all the hot spots on the upside and downside surfaces of the separator disc, we can reveal the generation and developing process of hot spots.

Fig. 9 shows the axial displacement and temperature variations along the circumference of the separator disc at three moments. In Fig. 9a, an initial bump formed between the area from node number 460 to 500 because of the non-uniform heat input on the upside and downside surfaces at time $t = 0.064$s and the temperature there is accordingly much higher than other places which means the first hot spot had formed there. The temperature distribution there is not symmetric about the middle plane, the temperature on the downside surface is lagged than the upside surface. In the area between node 503 to 507, the temperature on the downside surface is much higher than the upside surface so that thermal stress and moment emerged there and forced the plate to bend towards the upside direction.

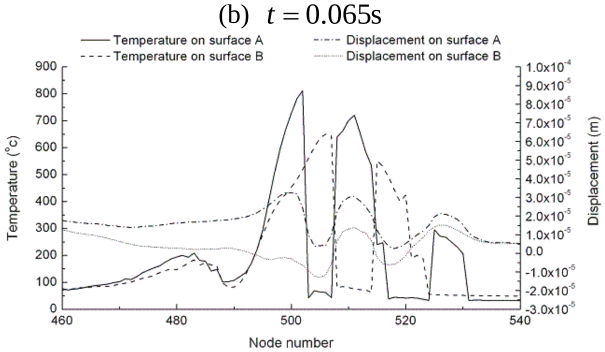
When comes to the moment $t = 0.065$s, because the upside surface contacts with the frictional disc and be heated, the displacement and temperature increase sharply until the second hot spot emerges there as shown in Fig. 9b, and the temperature there raised sharply from the initial temperature 20^{o}C to 300^{o}C on average.



(a) $t = 0.064\text{s}$



(b) $t = 0.065\text{s}$



(c) $t = 0.066\text{s}$

Fig. 9 The axial displacement and temperature variation along the circle on the upside and downside surfaces of the separator disc

When comes to the moment $t = 0.066\text{s}$ as shown in Fig. 9c, there are already two hot spots on the upside and downside surfaces, following the same process, a third hot spot is forming on the upside surface. Hence, this process can be conclude as: because of the non-uniform heat input, once an initial local hot spot been generated on one surface, under the thermal stress and axial pressure, the nearby area on the other surface in the downstream direction of the velocity will get into contact with the friction surface and another hot spot will develop on this surface, the same process will repeat again and again between the two surfaces until the hot spots have been distributed along the whole circumference, finally the hot spots distribute alternatively on the two surfaces of the separator disc in a sinusoidal shape. All the hot spots emerged in a short time that depends on the energy flows into the separator disc. The deformed separator disc with hot spots is shown in Fig. 10 where the axial displacement is enlarged 30 times for the convenience of observation.

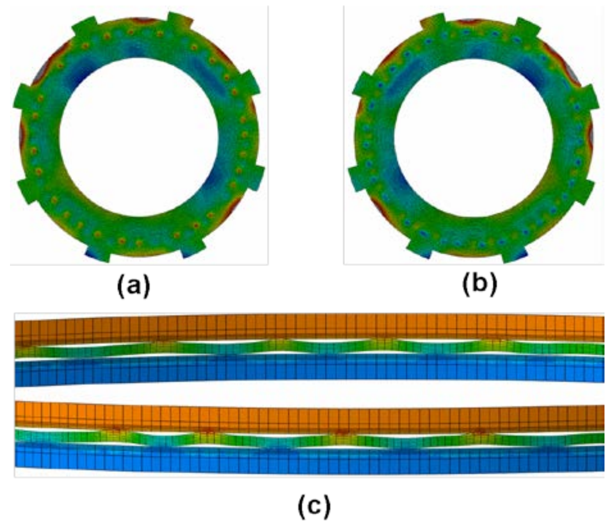


Fig. 10 Hot spots and axial deformation of the separator disc

Fig. 10a and Fig. 10b shows the hot spots on the upside and downside surfaces of the separator disc respectively, Fig. 10c shows the sinusoidal axial deformation of the separator disc, similar distributions of hot spots were also observed by Burton[11]. Since the number of hot spots and their distribution on the separator disc is strongly influenced by the working conditions, material properties and dynamic processes of the clutch, they may not exactly evenly distributed along the circle, hence we can define average wavelength $\lambda = \frac{L}{n}$ to study the influence of some factors, where $L = 0.644$ is the perimeter of the circle, the more hot spots on the circle the short the wavelength would be. Here in this work, several simulations were done with different contact pressure, rotating speed, Young's modulus and thermal expansion coefficient to investigate their influence on the hot spots wavelength.

5.2 The effect of contact pressure

The number of hot spots is directly related to the engagement pressure, relative rotating speed and consequent relationship with the heat flux generated by friction.

The engagement pressure, applied by the hydraulic piston, will not only influence the heat generation of the steel disc but will also suppress the axial displacement of the hot spots. Fig. 11 shows the relationship between the hot spots number and the contact pressure, the Young's modulus of the steel disc was kept at a constant $200GPa$, the thermal expansion coefficient was $5\times10^{-5}\text{K}^{-1}$ and the rotating speed was 1000r/min.

From the trend line we can conclude that as p increases from 0.2MPa to 2.6MPa, the hot spots number decreases linearly. There are 32 hot spots corresponding to 0.2MPa and 24 hot spots corresponding to 2.6MPa. When the contact pressure is small, the deformation remains in a "hot band" mode, which eventually becomes separated hot spots, this evolution occurs because of the small heat production and insufficient expansion. Since the tangential frictional force is insufficient to cause the steel disc to buckle and generate subsequent hot spots, it is only when the heat had been accumulated to a certain level that an initial hot spot can be generated.

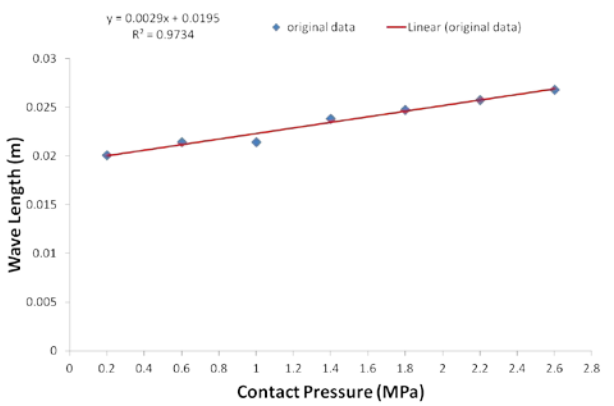


Fig. 11 Influence of contact pressure on the hot spots wavelength

5.3 The effect of rotating speed

According to the experimental results of Panier[17], only when the load had reached a certain level can the hot spots be produced even if the relative rotating speed had achieved a certain value, this result reveals that the rotating speed is not a significant factor of hot spots. The relative sliding speed of the separator disc directly affects the heat flux, and thereby affects the initial hot spots wavelength. Since the final distribution of hot spots in the circumferential direction is generated by propagation, all hot spots wavelengths would be affected by the rotating speed.

Fig. 12 shows that when the separator disc rotates under the pressure 0.2MPa, 0.6MPa and 1.0MPa, the hot spots wavelength linearly increases during the rotating speed between 200r/min to 4000r/min.

The lines show that rotating speed affects the hot spots wavelength in a wide range. Regular hot spots distributions could be obtained even at a slow rotating speed such as 200r/min.

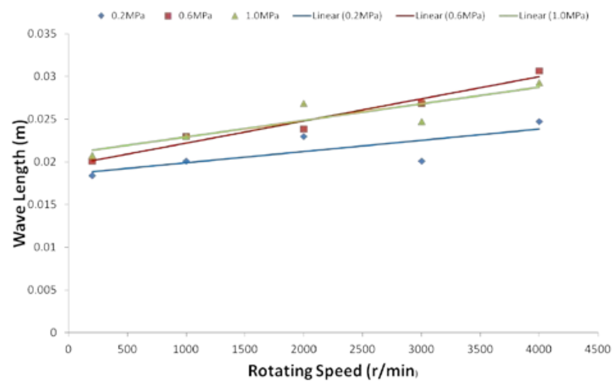


Fig. 12 Influence of rotating speed on the hot spots wavelength

5.4 The effect of Young's modulus

Commonly used material for practical clutches and experimental studies has fixed physical properties, the present study fixes the friction disc material parameters but changes the separator disc material parameters. Fig. 13 shows the relationship between the steel's Young's modulus E and hot spots wavelength λ, the wavelength variation under the axial pressure of 0.2MPa, 0.6MPa and 1.0MPa was obtained.

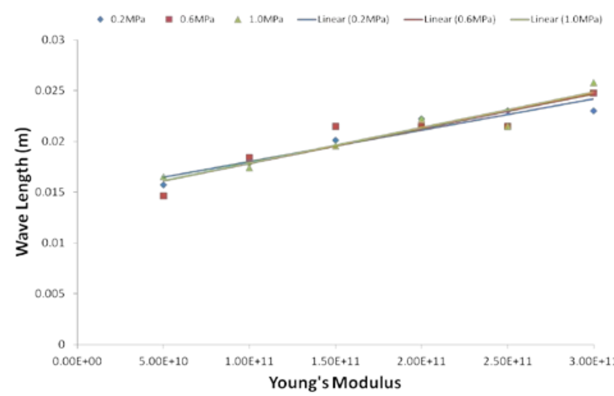


Fig. 13 Influence of steel's Young's modulus on the hot spots wavelength

From Fig. 13 we can see that the hot spots wavelength increases when the steel's Young's modulus increases, which means that the deformation of the separator disc becomes increasingly difficult. When E reaches 3×10^{12} GPa, The number of hot spots will drop dramatically to only 13 and the corresponding wavelength reaches 0.05m, when the Young's modulus decreases, the number of hot spots will increase until there is no individual hot spot.

5.5 The effect of thermal expansion coefficient

Since the thermal expansion coefficient α of commonly used steel is about 5×10^{-5} GPa, studies with wider range of α as shown in Fig. 14 indicate that when α becomes larger, the wavelength of hot spots becomes longer, and the minimum number of hot spots is only 1 which is similar to Burton's conclusions[11] ; when α becomes smaller, the wavelength of hot spots becomes shorter until a hot band formed.

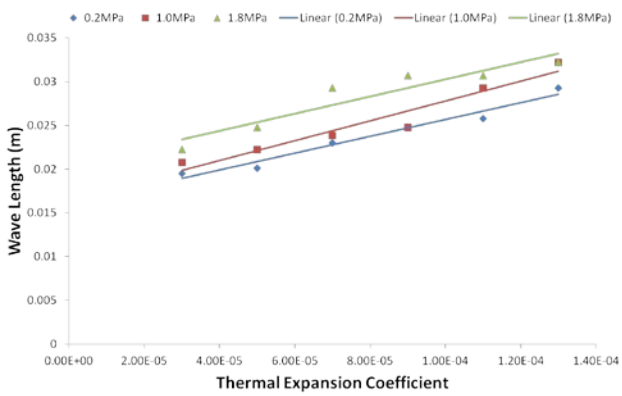


Fig. 14 Influence of thermal expansion coefficient on the hot spots wavelength

6. Conclusions and Discussions

By using finite element method, the developing process of hot spots on the separator disc during the engagement of a practical multi-disc clutch was studied, and a propagation mechanism scenario of hot spots was introduced. This mechanism is divided into three phases: First, initial local heating causes local hot expansion and the first hot spot is formed; Second, due to thermal stress, another hot spot is produced alternatively on the other side of the separator disc; Third, more hot spots form on the disc following the same way of propagation until a sinusoidal shape deformation with hot spots over the entire circumference is achieved.

A serial of simulations show that when the load pressure, rotating speed, Young's modulus and thermal expansion coefficient of the separator disc increase, the hot spots wavelength will increase almost linearly.

These results were calculated by reliable FE code, they can illustrate the hot spots generation mechanism scenario correctly and forecast the trend of hot spots wavelength's dependency on some working conditions, and also provide insights for clutch design. However, due to the complicacy of 3D model, load pattern, dynamic impact and ignorance of wear, the axial displacement and the wavelength of hot spots are not so accurate, in the future, more accurate model need to be set up to obtain more accurate results of this problem.

7. Acknowledgements

The authors would like to acknowledge the financial support by National Natural Science Foundation of China (Grant No.51175042 and No.51005021).

8. References

[1] J. R. Barber, Thermoelastic instabilities in the sliding of conforming solids, Proc. Roy. Soc. A. 312, (1969) 381-394.
[2] TH. A. Dow and R. A. Burton, Thermoelastic instability of sliding contact in the absence of wear, Wear, 19 (1972) 315-328.
[3] A. E. Anderson, R. A. Knapp, Hot spotting in automotive friction system, Wear, 135 (1990) 319-337.
[4] H. Kasem, Thermal levels and subsurface damage induced by the occurrence of hot spots during high-energy braking, Wear, 270 (2011) 355-364.
[5] Ralph A. Burton, Thermal deformation in frictionally heated contact, Wear, 59 (1980) 1-20.
[6] P. Zagrodzki, Numerical analysis of temperature fields and thermal stresses in the friction discs of a multidisc wet clutch, Wear, 101 (1985) 255-271.
[7] P. Zagrodzki, Analysis of thermomechanical phenomena in multidisc clutches and brakes, Wear, 140 (1990) 291-308.
[8] P. Zagrodzki, Nonlinear transient behavior of a sliding system with frictionally excited thermoelastic instability, Journal of Tribology, Vol. 123 (2001) 699-708.
[9] P. Zagrodzki, Samuel A. Truncone, Generation of hot spots in a wet multidisk clutch during short-term engagement, Wear, 254 (2003) 474-491.
[10] P. Zagrodzki, Thermoelastic instability in friction clutches and brakes – Transient modal analysis revealing mechanism of excitation of unstable modes, International Journal of Solids and Structures, 46 (2009) 2463-2476.
[11] S. Panier, P. Dufrenoy, J. F. Brunel and D. Weichert, Progressive waviness distortion: A new approach of hot spotting in disc brakes, Journal of Thermal Stresses, 28 (2005) 47-62.
[12] Tien-Chen Jen, Daniel James Nemecek, Thermal analysis of a wet-disk clutch subjected to a constant energy engagement, International Journal of Heat and Mass Transfer, 51, 1757-1769, 2008
[13] ABAQUS Theory manual 6.11

STUDY OF THE INTERACTIONS BETWEEN STAKEHOLDERS BY A MULTI-AGENTS SYSTEM: APPLICATION TO THE GOVERNANCE OF NATURAL RESOURCES IN MIARINARIVO DISTRICT (MADAGASCAR)

Aurélie Gaudieux [a], Yassine Gangat [b], Joël Kwan [c], Rémy Courdier [d]

[a] Center of Economics and Management of The Indian Ocean, University of La Réunion, France
[b] Laboratory of Computer Science and Mathematics, University of La Réunion, France
[c] Higher School of Engineering of Réunion and Indian Océan, University of La Réunion, France
[d] Laboratory of Computer Science and Mathematics, University of La Réunion, France

[a]aurelie.gaudieux@univ-reunion.fr, [b]yassine.gangat@univ-reunion.fr, [c]joel.kwan@univ-reunion.fr , [d]remy.courdier@univ-reunion.fr

ABSTRACT
This article proposes to study the interactions between the stakeholders through the study of management transfers in favor of the Common Pool Resources (CPR). The SIEGMAS model (Stakeholders Interactions in Environmental Governance by a Multi-Agent System) shows the interest of using a multi-agent simulation to build new alternatives to existing approaches in the context of decision support to implement Common Pool Resources governance policies. Agent Based Simulation (ABS) that were used in community complex systems offer an innovative approach to this field of economic and environmental study where non transfer modelling transfers of management has yet been made for all the island territories in the Indian Ocean. This article argues that carries extend the application of this method to the study of interactions between stakeholders in relationship to transfers of management in the others parts of Madagascar and others islands in the Indian Ocean.

Keywords: Interaction in Common Pool Resources, ABS

1. INTRODUCTION
The model SIEGMAS (Stakeholders Interactions in Environmental Governance by a Multi-Agent System) uses an ABS approach in order to establish a model for interactions of stakeholders in the management's transfers in Miarinarivo district of the Itasy region (in Madagascar). In the context of this project, the economic and classic steps being limited enough for the conceptualization of the individuals' behaviors and the process of common pool resources, the ABS approach allows some improvements in the study of these phenomenon and the creation of a sales tool for the decision. This establishment of a model must check if the management's transfers are efficacious via the application of penalties. It should allow us to study the actors' strategies as well as their behaviors.

2. MAJOR HEADINGS

2.1. Context of the Common Pool Resources (GCRN)
The Common Pool Resources points out the efficient application from the strategies of sustainable developments to common goods thanks to the decentralization of powers to the regional authorities. This approach is applied in various countries like the developing countries which, facing the inexorable destruction of the organic diversity, seeks an efficacious and local management, (Wade 1987) by adopting an environmental regulation in favor of the management of resources (Alden Willy 2004). The management's transfers allow the governments to decentralize the management of natural resources to the defined actors in this process. In 1990, Elinor Ostrom establishes a theory of common natural run in common by the stakeholders for more performance. In addition, for Ostrom (Ostrom, 2010), the local governance comes under the interrelations between a beam of discipline; it comes as a political, social, legal and economic problem. Ostrom establishes a list of eight principles of governance to achieve an optimal management of common property:

- To regulate access to resources;
- To make the costs and benefits commensurate with this approach;
- To make decisions jointly with the whole of the stakeholders;
- To coordinate and monitor the governance;
- To apply penalties for transgressions;
- To establish methods of dispute resolutions;
- To promote this management by state organs;
- To act as the supra-regional infrastructure.

With regard to shared resources and those in open access, (Ostrom, 1994) considers that sanctions and monitoring highlight good practices.

2.2. State of art

The structuring of environmental models constitutes a starting point unavoidable for the development of our establishment of a model (Müller and Aubert 2012) and (Bousquet and al. 2001).

Several cross-disciplinary and economic methods are conceivable for the study of natural resources and interactions. However, they hardly model the interactions between the stakeholders. The methods of the market's efficiencies and general economy focus on the relative analysis to the market's value and the theories of the growth and the models of Calculable general balance.

- The environmental methods used, focus on the analysis, mono or several criteria, the restorative or compensatory measures and the Eco-certification as well as the Eco-potentiality.
- The methods relative to the ecological economy come under some indicators of biodiversity, the calculation of the ecological loan and some indicators relative to the national accounting. Moreover, the incorporation of biodiversity in the methods of economic calculations remains controversial and limited (Nunes, Van den Bergh and Nijkamp, 2001). The economic analysis *ex ante* and *ex post* of biodiversity focus on either rationality of the preservation (analysis, cost-profit and the social well-being earned for each invested euro) or the effectiveness from the strategy of retained preservation.
- The cross-disciplinary methods focus on law economy and the ABS. We use ABS to optimize economics theories such as individual and collectives preferences (Arrow, 1951) to improve their performances.

The MAS have been choosed because they provide directly some pioneer tools for the problem.

3. THE ESTABLISHMENT OF A MODEL

We use an architecture, which allows us to describe the phases of our simulation (Ralambondrainy 2009). To create an ABS model for behavior and interactions between stakeholders, we use the method explained by (Gangat, 2013) that can be resumed in **Erreur ! Source du renvoi introuvable.**.

3.1. The conceptual establishment of a model

SIEGMAS models the interactions of the stakeholders in the context of the management's transfers in Miarinarivo District, the capital of the Itasy region (in Madagascar) thanks to the ABS approach. Itasy constitutes the smallest of the 22 regions from Madagascar in term of surface area with 6570 km2.

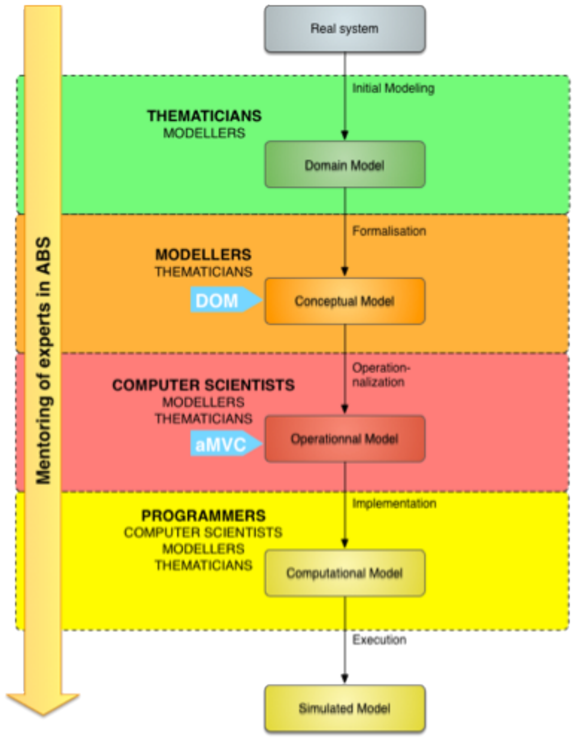


Figure 1 - Core Curriculum to the methodologies conception ABS

The Common Pool Resources emerged from the ninth Century in Malagasy environmental law and became intensified with the translation of French law under the French colonial era (1864-1960). The decade 1990 marks a real turning point in law of Malagasy environment with the apparition of the laws in favor of the safeguarding of common goods. Thus, the Charter of Environment in 1990, Law N° 90-033, the Gélose law (Secured Local Management) and the Contracted Management of Forests mark the emergence of a real consensus for Environment in Madagascar by the establishment of the management's transfers.

The Gélose Contract, Law N° 96-025 of 30 September 1996, for the local management of natural renewable of resources establishes a general framework for the preservation of the territory resources and its ecosystem. The Gélose Contract is composed of a contract of a management transfer, which contains a bill of specifications, the Dina controlling the basis community also called the COBA and a simplified land inventory known under the name of a collect Relative Security Land. The whole of stakeholders (Wade 1987) participates in the creation of the contract about the management's transfer signed by the COBA, the town and the local government representatives.

The Contracted Management of Forests, CMF, established by the decree N° 2001/122 of 14th February

2001, simplifies the procedure of the Gélose Contract but its duration is the same as one Gélose Contract: three years for the initial contract which can be renewed after a control for 10 years. We can count two parts for this contract: the forestry administration and the COBA. The community, as the secondary actor raises awareness and informs the COBA.

Since the coming into force of the management transfer contracts in 1999, more than 2000 contracts have been signed in Madagascar. These transfers have been converged by a strengthening of awareness and the application of sanctions. The good application of the transfers depends on the interactions between the actors and the controls.

This model has been conceptualized in order to study the deviances of the farmers against the forester and natural resources protected by the management's transfers or another text of law. The farmers know some sanctions granted in case of transgression by the government. The farmer (deviant or not) aims to make his land the more profitable as possible given that the spending he must use for his activity his personal needs, the costs generated by his cultivation (the costs of maintenance and investments) and the discounted profits. Thus, a farmer can choose to spread legally his cultivation via the purchase of new plots or the use allowed / controlled of natural resources. In fact, the management's transfers grant the farmer through the basis community some right of cultivation of some natural resources. However, the farmer can be a deviant by farming illegally some plots of natural and forester resources.

3.2. The working establishment of a model

The realization of the Ontology allows us to point out the concepts (Grüber, 1993) relative to the entities of the Common Pool Resources and the current links between the concepts. The use of ontologies in order to model the multi-agents simulations facilitate the conceptualization of the officers' characteristics within the platform of retained (Ralambondrainy, 2009).

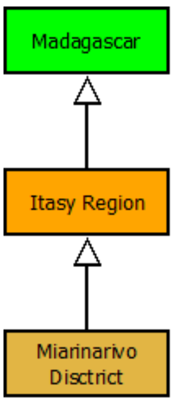


Figure 2 - Ontology of the reviewed territory

The arrows represent the relations of subsomption, « is a/an», on the **Erreur ! Source du renvoi introuvable.** to **Erreur ! Source du renvoi introuvable.**.

Four stakeholders compose the model (**Erreur ! Source du renvoi introuvable.**):

- The citizens, working via the natural resources, are the farmers and the other citizens.
- The Malagasy government which the ministries, the decentralized instances (Decentralized regional authorities and Services, Regional management of Environment and the Forests), the basis communities.
- The firms
- The sponsors and the stranger governments.

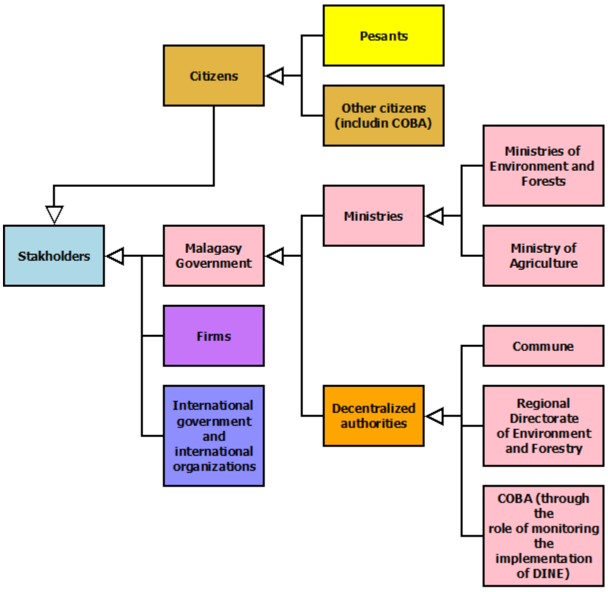


Figure 3 - Ontology of the stakeholders

However, only the first two actors interest our model: the farmer and the Malagasy government. We chose to study the actors two by two in the different establishments of a model.

This study is focused in the forester and Common Pool Resources as well as the resources produced on the farms (Figure 4).

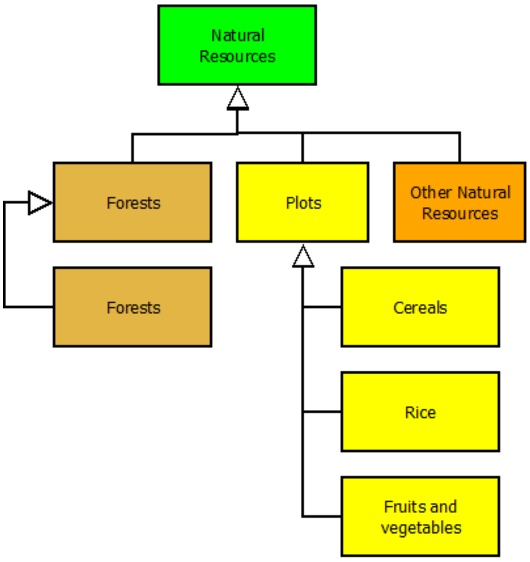


Figure 4 - Ontology of natural resources

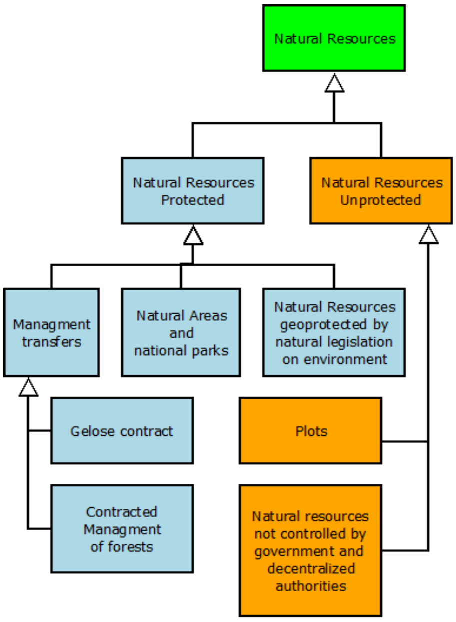


Figure 5 - Ontology of protected and not protected natural resources

4. IMPLEMENTATION

The model SIEGMAS has been implemented under the MABS platform called NetLogo (Wilensky, 1999).

4.1. Overview

The whole of elements of the model represents the different entities and variables clarified in the table.

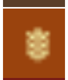
4.1.1. Temporal and space scales

An iteration corresponds to a month in the model SIEGMAS. In this simulation, a patch represents an agricultural plot, or a plot of forest or natural resources.

Our environment of simulation is composed of 1936 plots, that is to say 44x44 patches. The surface area of Miarinarivo is 2818 km^2. The district is composed for the best part of natural resources and farms. The average surface of a plot hold by a farmer in the district rises to 1.4 hectares, which corresponds to about 1990 plots per a farmer in the model.

Table 1 : Our model and its implementation

	Establishment of a model	Implementation
Agents	**PEASANT** Some farmers work on some exploitations: to cultivate and exploit the resources that are available (those which belong to him).	**Peasants peasant** - nbPeasants (peasants)
	DEVIANT Some of the farmers are some deviants. They can exploit the neighbouring plots not belonging to other farmers, a forest or a plot that they consider farmable. It is possible to fix a percentage of deviants.	**Deviants deviant** - deviantsPercentage (deviants)
	EXPLOITATION The farmers have several agricultural practices : - bio : bio - conventional : con - On burnt land : snb A number that determines the order of creation identifies the owners. Different colours also identify the owners and their farming. We consider that the farming belong to the farmers.	**Exploitations exploitation** Each type of agriculture represents a certain percentage. Other variables: - Colour (the colour of the peasant to be identified visually) - owner

Environment	**PLOTS** By default, the plots are all some natural resources. We consider that it exists a percentage of the deterioration level of these natural resources.	- deteriorationLevel
	OTHERS PLOTS We consider that the areas that are not farming plots are either natural resources or forests that do not belong to any farmers. The environment (composed of cells) has a dynamic size, because his dimension could be changed through the values of width and height configurable via the interface. We chose 18 pixels for the size of a patch in order to be able to appear on screen. It is a matter of an arbitrary choice for the visual comfort.	 Plots - IsOwned brownn : Ressources Naturelles autres que forêts.
	FOREST The forester surface area is defined by a percentage of parameter. The forests are some protected areas by the contracts of the Contracted Management of Forests and the environmental legislation. The other natural resources (others than the farming plots) are supervised by the same plans of protection.	 Forest - isForest - ForestPercentage -Pcolor (a variant of green for the forest)
	AGRICULTURAL PLOTS A farming plot is a cell that belong to the farmer.	Pcolor : three colors  brown Very deep agriculture at dominance on burn land  Orange: The bio agriculture  Purple: The conventional agriculture

However, we count about 1936 plots according to its size 44x44. A farmer can have several plots, from one to five.

4.1.2. Organization of the model

The land is created according to a certain number of obligations given by the interface (Figure 6) like the size of the land, in number of patches, and the percentage of land, formulated in number of patches, and the percentage of land covered by the forest. We also distribute a determined number of farmers (from the interface) randomly on the in a way that they are not superimposed. Then, their farming are created around them in a radius going until 8 patches for now all by avoiding to use a patch already exploited or forbidden like the forest. For the farming of natural resources, the farmers exploit or try to exploit as much as possible resources neighboring their farming than the resources separated geographically.

4.1.3. The sanctions

The farmers can respect or transgress the legislation towards the protection of natural resources.

The controls or the rupture can identify the transgressions. The authorities grant some individual sanctions to the deviants. Thus, no collective sanction affects the respectful farmers of the legislation. Besides, the state laws for Environment, the laws and traditional rules like the Dina permit the excise of sanction based on the habits et traditions as well as the respect of nature.

4.2. Payment mandate of the process

4.2.1. Concept of elaboration

In our establishment of a model, the level of detail depends on the link between the number of patch and the actual size. Knowing that the smallest indivisible unity under the NetLogo is the patch, this latter will correspond to one type of land. A patch can point out either plot of forest, or a plot of farming and natural resources. Besides, a farmer adopts a unique mode of agriculture per a plot: organic, conventional or on burnt land. It does not exist any segregation about the plots.

The farmers have some information relative to the localization of the agar areas, under the GCF, some other resources under the environmental legislation and some farms of environment from the simulation. The farmers also know the farming, which belong to them. The government is not directly represented by an agent. It acts through the checks and the sanctions applied on the patches and the farmers.

In this simulation, we consider, basing upon the theory of games (Von Neuman and Morgenstern, 1944) that environment can be random and interdependent. In our simulation, the officers will exploit the neighboring plots or the forest illegally in order to maximize their profits. Thus, on one part, it is a matter of a random environment because the chance and the behavior of an officer in order to maximize his profits. Also an environment of interdependence because the collective behaviors and the individual behavior of an officer, i.e. his rationality, influence the maximization of these profits (Cavagnac 2006).

In the game where each of some n officers of the game chooses the behaviors maximizing his profit in the whole of the possible behaviors. In our establishment of a model, the games can be static or dynamic and the complete or incomplete information. In a static game, an officer acts (an action is a strategy) without knowing the behavior of another officer whereas they act simultaneously in a dynamic game. The strategies can be dominant or dominated. A whole of strategy is a balance of Nash (1950). When we maximize his expectations of profits basing upon his beliefs, the whole of strategies that he deploys is a «

Bayesian » balance (Kreps and Wilson 1982). The plays can be also be repetitive or mixed (Kuhn and Tucker 1950).

4.2.2. Adaptation

The farmers set yourself in order to increase their productivity and to perpetuate the trust that the state authorities grant them (or to avoid the sanctions for everything by having a nearby context in favor of their activity. (Wade, 1987; Grüber, 1993; Bontems and Rotillon, 2007; Brandouy, Mathieu and Venryzhenko, 2012).

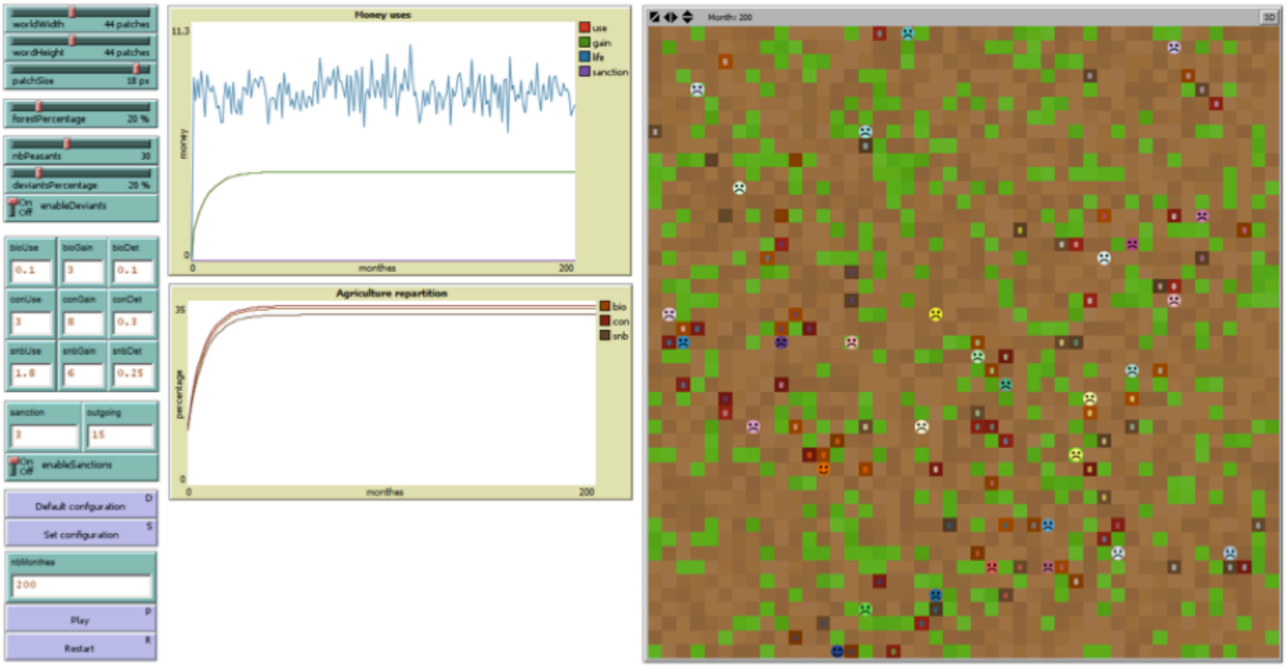


Figure 6 - Intitial conditions of simulation

4.2.3. Perception and interaction

The arbitrary data can be distinguished in this simulation.

Being a matter of different types of agriculture, the variable « Use » corresponds to the investments that the farmers make according to the type of agriculture adopted. Iteration allows getting the cost for 100 % of agriculture. A farmer makes no investments if he chooses the organic agriculture in iteration. However, if he either adopts conventional agriculture or burnt land, these investments are equivalent in an iteration.

Concerning the profits relative to the variable « Gain », they fluctuate in accordance with the type of agriculture chose by the farmer. For this simulation, the income that 100% of three types of agriculture bring will be given in some fields intended to this effect on the graphic interface. Likewise, the variable « Det » determines the variable « Det » shapeless of the deterioration level from the grounds of farming plots revealed in percentage in an iteration.

Being a matter of sanctions, at each iteration, we puncture some money in the form of fee to the farmer at each iteration for 100 % of damage created by an illegal farming on a patch. The rules and sanctions of the State are applied during the month through the check of some patches). The sanctions are applied by collecting money to the farmer when we calculate his income available at the end of the month via the variable « outgoing » which points out the maximal money spent by a farmer in order to « live » in an iteration (random).

The deterioration is made at random for the patches non exploited, « update-rawPatches ». Moreover, a random and natural deterioration «update-exploited Patches » applied for each farm according to the percentage and the type of agriculture. We also obtain, according to the percentage and the type of agriculture, the money spent for the farming as well as the money earned by the farmer. The sum, which a farmer has, is updated each month according to the results of their farming. This available sum of money can lead to a change of the types of agriculture being able to be random and dependent on the money from the money remaining to the farmer.

The deviant, « update-deviants », remarks the neighboring plots to his ones. Then, he will or not exploits a land that does not belong to him. If during a made control, the « government », acting through a check, notices that a non-exploited land normally can be in the major of case illegally exploited, it punishes the deviant responsible for the transgression by a fee to pay every month in accordance with the degradation of the patch run by the illegal farming.

4.3. Results

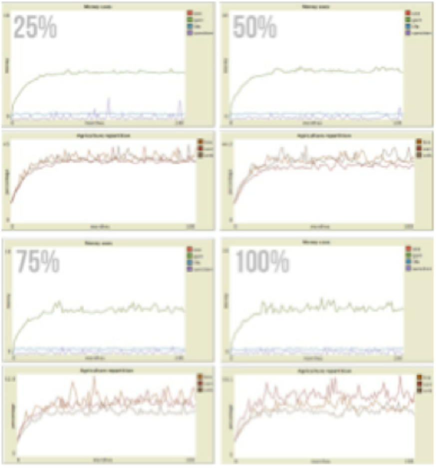


Figure 7 - Curves of evolution at 25 %, 50 %, 75 % and 100 % of deviants from the distribution of the use of money by the farmer and the distribution of the agriculture

4.3.1. Manipulation of a model

Firstly, the initial conditions of simulations are fixed (Figure 6). The simulation contains thirty farmers and takes place on a hundred iterations, that is to say hundred months. Then, the farmers can be some deviants. The following step consists in choosing some proportions of conventional and bio farming and on burnt land as well as some rate of deterioration for these three modes of agriculture. Thereafter, it must fix some values for the variable « Use », corresponding to the investments that the farmers make according to the types of agriculture. For these simulations, the values « gains », « use », « Det » and the percentage of deviants vary thanks to the changes of the slider « deviants' Percentage ». It is a matter of favoring the impact of the deviants on Environment.

For the protections, when a land of natural resources hold by the State and which does not belong to the farmers arrives at more than 75 % of « deterioration Level », the land must be considered as strongly damaged and set under high protection by the government. Thus, during the lapse of time when the land is considered as protected by a management's transfer or a plan of emergency. The land will be able to see only its deterioration level reduce up to come back to 25% all by preserving at this stage, the maintaining of status of increased protection. The deterioration level random is from 0 to 0.2 at each iteration.

For the simulations, with 25 % of deviants on the territory, Oscillations are enough stables (Figure 7)

Thus, the area is poorly degraded. For 50% of the oscillations, we note light peaks of oscillation in Figure 7 are observed. To 75% and 100% of the deviant oscillations increases. This reflects a strong or very strong natural resource degradation Miarinarivo.

For the sanctions, the deviants want to exploit randomly the lands being around his farming, even those, which they have been created illegally. At each iteration, the government comes under a random number of lands exploited illegally and seeks the one who created the farming. The government calculates a sum (value of the sanction x level of damages brought by the farming x ratio) like a fee to pay by the farmer before destroying his farming. If at the time of the sanction the land is protected, a ratio is added to the fee. This ratio equal to x2 for a protected land and still at x1.5 for a forest). This ratio also equal to x1 for a protected ordinary land, to x2 for a not protected forest, to x3 for a protected forest (x1.5 + x2).

4.3.2. Results of simulations

After launching several times the simulation with different values of configurations, it appears by observing the variation of the rate of deviants that if the government controls the management's transfers correctly, the sanctions are well applied (Figure 7). The presence of the deviants engenders more deterioration of the lands (Figure 8). Some deviants avoid the sanctions and run some non-taxable income. However, the application of the sanction to the deviants rebalances the system so that the deviants could not get any profits in the illegal practice of agriculture or the exploitation of natural resources. Thus, the individual behaviors will affect the collective behaviors if a farmer knows the gestures of his neighboring. The farmer grants his trust to his neighbor when it is not a matter of a deviant. However, if he knows the acts of deviance of his neighbor, he will be more suspicious against him and he will denounce him.

Figure 8- Representation of a situation of deviance

By fixing the rate of deviants at 50 %, we notice that the use of money which the farmer is going to decrease then increase strongly whereas the individual profits of the farmer and the lifespan of the farming will have tendency to remain stable. This intensive farming is going to decrease the lifespan of the farming. The type of agriculture fluctuate and the conventional agriculture and on burnt land dominate. The sanctions are also less applied (Figure 10).

Figure 9 - Representation of a situation of deviance with a deviants ratios of 50%

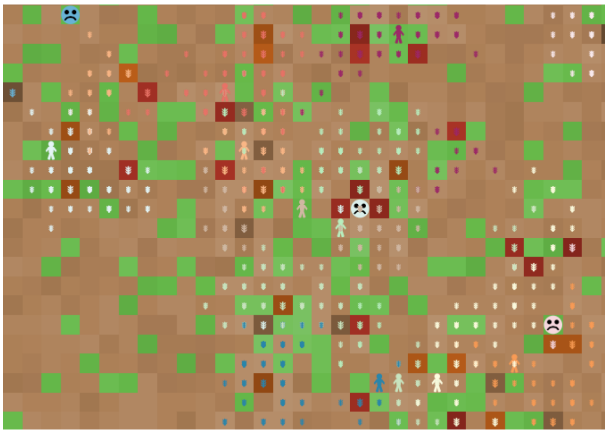


Figure 11 Representation of a situation of deviance with 75% of deviants

By fixing the rate of deviants at 75 %, which would correspond to a situation of strong degradation from environment, we can notice that the individual profits, the lifespan of some farming remain stable notwithstanding, the money that he spends increases very much (Figure 10 and 11). It is a matter of situation when the controls are rare even missing.

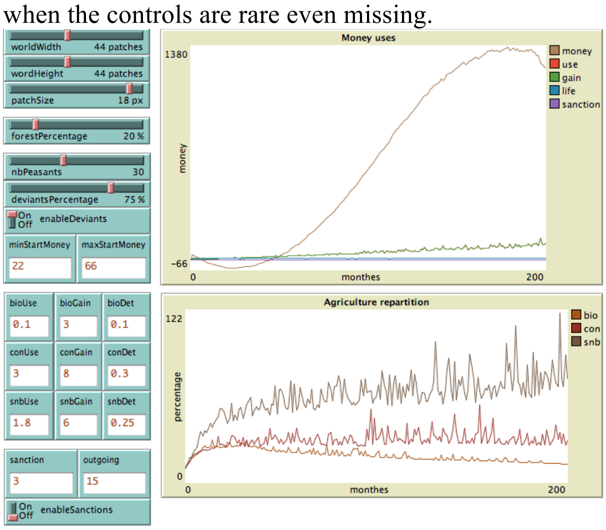


Figure 10 - Simulation of 75 % of deviants

4.3.3. Supply of this simulation in economy

The use of the MAS for the creation of a sales tool for the decision in economic sciences and of environment in order to study the interactions opens the disciplinary field of the MAS to the study of the management's transfers and environment in a developing country which is also an island territory. Thus, it is easier to study the process of communication and cognition of the individuals are difficultly educable by the current economic methods.

5. CONCLUSION

The realization of a sales tool for the multi-disciplinary decision in order to study the interactions of the actors in the Common Pool Resources in Miarinarivo district brings a pertinent response to the study of the interactions, being difficultly a model by the economic methods. This simulation allowed to identify the interactions between the actors of the governance taking part in the management's transfers and to propose some targeted measures in agreement being able to reduce the merging anomalies. Thus, this model reveals that the knowledge of the sanctions by the farmers does not prevent them to infringe the legislative measures of protection of natural resources below the management's transfers or located in some few protected areas.

Further to this simulation, firstly, the realization of a simulation with the incorporation of the identified agents will permit the acquisition of a global simulation for other regions from Madagascar and the island territories of Indian Ocean (Reunion Island, Mauritius, Seychelles, Mayotte and the Comoros). In fact, it is a matter of aspect which has not been treated in the isles of Indian Ocean and that it would be flourishing to spread in order to check the transposability of the model. We are working on a tool of extraction about knowledge from the geographic maps of reference in order to reach a largest pertinence in our simulations in order to show the transposability of the model.

Thus, we can mutualize the positive synergies and to make the forecasting in order to be proactive in the environmental management in the decades to come.

REFERENCES

Alden Willy, L., 2004. Can we really own the forest? A critical examination of tenure development in community forestry in Africa. *Tenth Biennial Conference, International Association for the Study of Common Property (IASCP)*, paper 251d, p.15.

Arrow K., 1951, *Social choice and individual values", New York*, John Wiley; Chapman and Hall.

Bontems, P., and Rotillon, G., 2007. *L'économie de l'environnement*. Repères.

Bousquet, F., Barreteau, O., Mullon, C., and Weber, J., 2001. Modélisation d'accompagnement : système multi-agents et gestion des ressources renouvelables. *Actes internationales: Quel environnement au 21ème siècle ? Environnement, maîtrise du long terme et démocratie. Germes*.

Brandouy, O., Mathieu, P., and Venryzhenko, I., 2012. Optimal Portfolio Diversification? A multi-agent ecological competition analysis. *PPAMS (Special Sessions)*, 2012: 323-331.

Cavagnac, M., 2006. *Théorie des jeux*. Gualino éditeur.

Gangat, Y., 2013. Thesis in Computer Science, *Architecture Agent pour la modélisation et simulation de systèmes complexes multidynamiques : une approche multicomportementale base sur le pattern "Agent MVC"*.

Grüber, T. R., 1993. «A translation approach to portable ontologies.» Knowledge acquisition, 5(2), pp.199-220.

Kreps, D. M., and Wilson, R., 1982. Reputation and imperfect information, *Journal of Economic Theory*. Elsevier, vol. 27(2), pages 253-279, August.

Kuhn, H. W., and Tucker., A. W., 1950. Contributions to the Theory of Games, Volume I, *Annals of Mathematics Studies, 24*, Princeton: Princeton University Press.

Morgenster, O., and Von Neumann, J., 1944. *Theory of games and Economic Behavior*, PUP, 1ère edition.

Müller, J.-P., et Aubert, S., 2012. Formaliser les rôles et les territoires par les systèmes multi-agents institutionnels. *JFSMA 2012*, pp. pp. 13-22. Cépaduès.

Nash, J., 1950. *Equilibrium points in n-person games.* PNAS, vol.36, n°1.

Nunes, P. A., Van den Bergh, J. J., and Nijkamp, P., 2001. Ecological-Economic Analysis and Valuation of biodiversity. *FEEM Working paper*.

Ostrom, E. *Gouvernance des biens communs*. Boeck, 2010.

Ostrom, E., Gardner, R., and Walker, J., 1994. *Rules, games and common-pool resources.* T. U. Press, Ed.

Ralambondrainy, T., 2009. Thesis in Computer Science *Observation de simulations multi-agent à grande échelle*.

Wade, R., 1987. The management of common property resources: collective action as an alternative to privatisation or state régulation. *Cambridge Journal of Economics*, 11:95-106.

Wilensky, U., 1999. Netlogo home page. https://ccl.northwestern.edu/netlogo/.

AUTHORS BIOGRAPHY

Aurélie Gaudieux is specialize in environmental law in Europe and Indian Ocean. Since 2011, she is PhD in economics at University of Reunion Island (France) and Catholic University of Madagascar. Currently, she is preparing a thesis on strategies and interaction between stakeholders in the common pool resources in Indian Ocean islands by a multi-agents system after a master degree in economics (specialized in sustainable development and territorial management) and a master degree in law.

Joël Kwan is engineer in IT specialty of ESIROI of the University of Reunion Island, France. He manages the technical.

Yassine Gangat is a part time Assistant professor in the Department of Computer Sciences of the University of La Réunion in France. He hold a Bsc and PhD in Computer Sciences from the same University. His research interest are in finding ways to reuse and co-build models in Agent Based Simulation. His thesis was focused on an agent's architecture that will help in this field. Current works involves an investigation in the robotic field as well as in the modelling field.

Rémy Courdier is full professor in the Computer Science and Mathematics Research Laboratory (EA2525-LIM/IREMIA) of the University of Reunion Islands, France. His research interests include design of simulation applications using multi agents systems and infrastructures required for their use. Rémy Courdier originally studied in Nice Sophia-Antipolis University (France) and then worked in the IT industry as a Research & Development engineer. After working at Digital Equipment Corporation (DEC) in Geneva, he managed several R&D innovative Projects at Schneider PLC Laboratories in Sophia-Antipolis and received the Ph.D. degree in computer science from Nice Sophia-Antipolis University (France) in 1993. He returned to a full academia activity in 1996 as assistant Professor at the University of Reunion Island's.

A SERIUOS GAME FOR TESTING BUSINESS CONTINUITY PLANS IN A MANUFACTURING FIRM

Del Vescovo Arturo [(a)], Rella Serena[(a)], Vox Carmine[(a)]; Gnoni Maria Grazia [(b)], Mossa Giorgio [(c)],

[(a)]Bosch Tecnologie Diesel Bari, Via degli Oleandri, 8/10, 70026 Modugno (Bari), Italy
[(b)]University of Salento, Via per monteroni, 73100, Lecce, Italy
[(c)] Polytechnic of Bari, Viale Japigia 182, 70126 Bari, Italy

(a) arturo.delvescovo@it.bosch.com; serenarella@gmail.com, carmine.vox@it.bosch.com;
(b) mariagrazia.gnoni@unisalento.it; (c)g.mossa@poliba.it

ABSTRACT

Disaster management and Business continuity are now new emerging topics for companies all over the world. Several international standards have been defined in order to support companies in developing their own Business Continuity Plans. The aim of this paper is to propose a framework for designing Business Continuity Games (BCGs) in the manufacturing sector. The proposal is based on applying lean concepts in the BCG design process. An application in an automotive firm supplier is proposed to validate the framework.

Keywords: Business Continuity Management, serious game, game design, manufacturing firm

1 INTRODUCTION

Companies after the disruptive event of Japanese earthquake are now facing with new strategies for analysing impact on their business due to unexpected disruptive events. According to an organizational point of view, potential effects on business activities usually regard (Business Continuity Institute, 2010):

- a loss of productivity, which could easily causes a loss of revenues;
- an increased cost of working (e.g. overtime costs to recover lost production);
- a loss of company reputation: which could influence the overall business of a company.

Business continuity management (BCM) is now a new emerging topic for companies all over the world; thus, the BCM could be defined as a process which aims to actively ensure operational continuity, for preserving competitive advantage (Herbane et al., 2004). Applying an effective BCM strategy could support both a faster recovery process and an efficient reputation protection; furthermore, developing an efficient BCM could also provide a more effective knowledge of critical processes in an organization.

The operational tool for applying BCM in an organization is the Business Continuity Plan (BCP): it aims to identify potential threats of an organization and build internal capabilities to respond to these threats in order to safeguard the interests of key stakeholders, reputation, brand and value-adding activities.

Designing an effective BCP is a complex task due to several factors: one most important is that accidental disruptive events may cause serious threats against the business of a single firm; however, impacts could increase as the whole supply chain may be involved. Furthermore, different sources have to be evaluated in the BCP: sources generating a disruptive event could be very different, i.e. from natural causes to human actions. Wing Lam (2002) suggests a classification of threats based on the specific source type; the three proposed categories are:

- *Technology threats*: this category refers to an unavailability of such a technology due to natural disasters (such as flooding) as well as plant failures, such as fire, power, systems and network failure, systems and network flooding, vandalism, and sabotage;
- *Information threats*: this category includes threats derived from a lack of information due to several conditions including hacking, natural disaster, fire, etc.;
- *People threats*: this category mainly refers to personnel unavailability due to illness, recruitment shortfalls, resignation, weather, transportation, etc..

Developing effective BCPs allow organizations to recover quickly and to reduce impacts of disruptive events (Herbane et al., 2004); international standards have been recently defined for designing and applying BCPs in companies. The most recent standard for BCM is the ISO 22301 (2012): similarly to other management systems, this framework is based on the well know PDCA (Plan, Do, Check, Act) cycle. It affects all step for designing and managing a BCP in an organization; one critical phase is the verification phase where the designed BCP has to be "verified" continuously by companies.

The paper aims to support company managers in both developing and testing its own BCP by applying a serious game as an "arena" for testing company BCP by a "real" simulation experience. Furthermore, the game will also provide to company managers constructive feedbacks for improving the overall efficacy of their BCP.

Thus, after a brief review about the application of serious games in disaster management introduced in section 2, a methodology for developing a Business Continuity Game is proposed in section 3. Finally, an application in a real case study is proposed in section 4.

2 SERIOUS GAMES IN DISASTER MANAGEMENT: A BRIEF ANALYSIS

Serious games are widely applied since several years for testing and communicating such a topic in the business environment: the recreational element raises the degree of involvement, thus allowing players to act rather genuinely (Boyle et al., 2011). Modern theories suggest that learning is most effective when it is active, experiential, situated, problem-based and provides immediate feedback. serious games are effective since they potentially improve the acquisition of knowledge and cognitive skills (Wouters et al., 2013).

The approach of serious gaming combines amusing methods and notions as well as game technologies (Ruppel, Scatz, 2011) in order to interlace two requirements which may seem apparently incompatible: the need of entertaining- typical of ordinary games - and the need of training and enhancing the behavior of players in the long haul. The latter need is disclosed by the use of the adjective "serious" to stress the learning objective underlying the game (Crookall, 2010).

Serious games are widely applied for training purposes, such as to describe, or demonstrate an issue, a situation or a process in complex environments (van der Zee, Holkenborg, Robinson, 2012). However, few applications of serious games could be outlined for BCM: in this context, serious games may provide a test environment to guarantee a proper performance in crisis circumstances, which often require dynamic decisions under time-pressure. Responding in a timely and effective way can reduce deaths and injuries, prevent secondary disaster, and reduce the economic losses (Massaguer, et al., 2006, Di Loreto et al., 2012).

Thus, several previous games developed for disaster management are based on simulation models (Connolly et al., 2012). One example is the "Serious Human Rescue Game (SHRG)" developed by Ruppel and Schatz (2011). This game applies simulation models to provide a virtual environment for developing rescue tests in burning buildings: the main purpose is to evaluate the effect of building condition on human behavior during the evacuation process. The "DRILLSIM game" applies a multi-agent simulation for assessing each role in the disaster response process for a building evacuation (Massaguer et al., 2006). Furthermore, the game called "Zero Hours", developed through a collaboration between the Chicago Health Department, the U.S. Centers for Desease Control and Prevention and the University of Illinois, allows to manage by a simulation game a mass anthrax attack; the scenario game called "Hazmat: Hotzone", rolled out by Carnegie-Mellon University's Entertainment Technology Center, simulates a fire with the aim to test the first response of firefighters in Pittsburgh and New York City; "Ground Truth", developed by the Department of Energy's Sandia National Laboratory, simulates a chemical tanker truck leak.

BC24 - an online incident simulation game - consists of a single scenario game involving a flood, a supply chain failure and impending reputational consequences (Business Continuity Institute, 2013).

As BCM is now applied in several different contexts, serious game for BCM have been developed in both production and service sectors. One example in the service sector is the so called "CODE RED: Triage!" developed by Wouters, van der Spek, Oostendorp (2009): it aims to train medical first responder emergency personnel to perform a triage after an explosion in the subway.

As described previously, simulation based models are also applied to compare different scenarios in disaster management aiming to train organizations. One example is the Flu Pandemic Game, a training tool developed by the UK Department of Health (2009) to help staff, managers and voluntary organizations to develop their own BCPs. Its main purpose is to train players about the impact of a possible cough pandemic on their organization by simulating the consequences this flu could have on the firm.

3 DESIGNING A BUSINESS CONTINUTIY GAME: A PROPOSAL

A framework for developing a Business Continuity game (BCG) in the manufacturing sector is proposed as follows. The aim is to provide managers of manufacturing firms with guidelines for building their own BCG for supporting an effective BCM process.

Before defining the game design process, a characterization of firm features according to the BCM process has to be outlined. As BCM is a relatively new topic, not all firms are nowadays provided with a reliable BCP in order to promptly react to unexpected events. Thus, three starting conditions which could affect a firm have been outlined as follows:

- class A: firms which have already adopted and implemented a BCM system based on international standards;
- class B: firms which have developed their own BCM process, usually based on unstructured processes;
- class C: firms which have not yet applied any BCM system.

One critical phase for Class A firms regards the full verification process required by the international standard: firm managers have to check procedures designed in previous phases: thus, for these companies, the prosed BCG could be an effective tool to control the effectiveness of their BCP.

For Class B firms, the proposed game could be used as a bottom-up approach to improve their BCP effectiveness, rather than a top-down one which is usually more rigid and less effective. One relevant result provided by the BCG application is the knowledge sharing about recovery procedures between all firm managers thus allowing an effective analysis of their coordination degree. Feedbacks provided by the BCG development will allow to harmonize the current procedures in a unique BCP.

Finally, class C firms can also effectively apply the proposed methodology as a support tool for developing their own BCP: in this case, the game application could be more problem-solving oriented, as the firm has to start to develop its BCM strategy.

After analyzing the type of organization, the general framework for designing BCG is proposed in Figure 1.

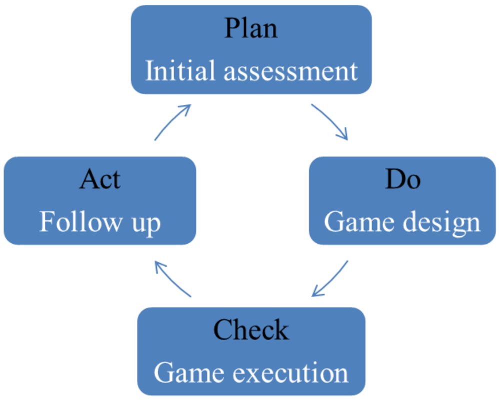


Figure 1: The PDCA application for the BCG design process

The framework is based on three main processes: initial assessment, game design, game execution; a detailed description of each step is proposed as follows.

3.1 The Initial assessment process

At first, game designers have to carry out a detailed analysis of the current situation affecting the organization from a BCM perspective. The main purpose of this activity is to evaluate and share the knowledge of firm BCM process within the whole organization. The overall game complexity will vary according to firm type (i.e. based on the previous classification). Some tools that could support game designers in developing this step could be:

- Study of organizational chart, to verify the presence of determined roles and responsibilities, from the BCM point of view (business continuity manager, business continuity teams or emergency response teams);
- Direct interviews and questionnaires, to submit to a sample of the whole staff;
- Assessment of documents concerning BCM, i.e. business continuity plan, emergency procedures, certification of the implementation of a standard BCM system; certification of the accomplished training.

Then, other two activities have to be carried out aiming to highlight most critical events for the firms:

1. Developing a documented risk assessment that identifies and evaluates the risk of disruptive events. The risk is weighted according to two dimensions: the probability of occurrence and the seriousness of the event, on the basis of financial loss, loss of reputation, costumer complaints and delay in product release. The outcome of risk assessments provides evidence of the order of priorities of threats to be overcome.

2. Conducting a Business Impact Analysis to identify the products and analyze the functions and the effects that the business disruption may have upon them.

3.2 The BCG design process

The proposed BCG is a role play where each player covers the actual role he/she has within the firm, however he/she has to make decisions in a fake environment. The BCG design process is detailed in Figure 2. The first activity for the BCG design is to define the triggering events – i.e. incidental (or accidental) events starting the crisis -: events are deducted from the previous risk assessment activity.

The output of this step is the so called EVENT CARD processing. The managers of the game could use this tool to outline operational scenarios to the players; it also indicates the typology of accident, date, time and shift and place of area in the area in analysis the involved department. After event definition, game designers, should quantify consequences due to the accidental event on the firm activities, the interruption working activities since all activities are recovered. For this purpose, the game designer has to map the most relevant activities for the emergency management process. For the class A firms, the involved activities and actors are traceable by the BCP; thus, the game designer will set the layout of this process starting from the activities planned within the existing BCP and improving them. For class B and C firms – where BCPs are not fully developed – critical activities have to be evaluated by game designers, based on their knowledge and management skills. Thus, the use of structured tools (e.g. value stream maps) could be effective for carrying put these activities.

Outputs of this phase allow to design the so called SOLUTION CARDS process. These cards represent how each player contribute to solve its specific task during emergency and recovery processes. Inputs are all information concerning the emergency scenario; expected outputs are decisions developed by each player to face with the unexpected event. Suggestions could be introduced by the game designer to help players in their decision making process. Next, escalation parameters are introduced in the game design process. Two parameters have been introduced:

- Escalation factors;
- Random events.

The escalation factor F is defined based on a level of criticality (pi) of each event according to equation 1:

$$F = \sum_{i=2}^{f} pi \qquad (1)$$

where the variable p_i is a binary variable which can have value 1 if the impact of the event is low or value 2 if the impact is high; f is the total number of introduced events. The i-th event (i = 2, 3, …, f) is an unexpected event occurred after an accidental one. The insertion of a set of events f >1 entails the iteration of all the steps pertaining to the process of game design for each event f. Hence, even the escalation factor itself f could be dynamically updated in each iteration.

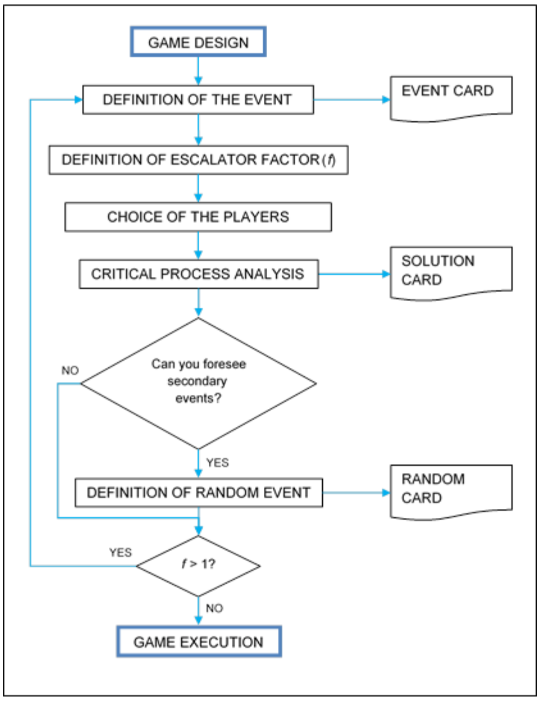


Figure 2: The proposed game design process

The game designer could use the escalation factor to control the game criticality level: thus, by introducing a high number of escalation events, players are forced to manage more severe conditions (e.g. in terms of production disruptions) thus increasing the criticality level of the simulation game.

Random events are casual events triggered by the primary accidental event. They can be defined as positive or negative externalities caused by an event, whose effect is either the increase of the level of criticality (with negative impacts) or the facilitation of the emergency management (with positive impacts). Differently from the escalation events, they do not directly modify the operation interruption time.

Escalation factor and random events introduction have been depicted in Figure 2.

For each f main event, the game designer could introduce j random events (j = 1, 2,…, r), whose weight on the impact of the paramount event i is due to the factor R_i:

$$\mathrm{Ri} = \sum_{j=1}^{r} \pm ej \qquad (2)$$

where the variable e_j is equal to 1 if the impact of the j-th random event is low; or it is equal to 2 if the impact is high; ± refers if a negative or positive impact is evaluated.

Hence, if random events are introduced, the level of the thorough complexity of the game is defined by equation 3:

$$C = \sum_{i=2}^{f} pi + Ri \qquad (3)$$

For each random event the game designer ought to set the correspondent random card, later delivered to each player, without any notification, randomly, indeed, during the performance of the game.

3.3 The Game execution

In order to test a first feasibility of the BCG, a trial run, which is a pre-simulation of the game not fully working, could be carried out. The trial run should not be performed by the same players selected for the full scale BCG execution; however it would be suitable to identify different players within the firm sectors for developing this step. During the trial run the game designer has to observe: his/her own task is properly the evaluation of the outcomes. If the trial run is reckoned to be positive, than the actual game will be eventually performed. Otherwise, if the test has not elicited the hoped results, then the game designer will have to, in accordance with the gathered feedback, bring about the necessary changes. Hence, he/she has to iterate all the activities pertaining to the process of game design.

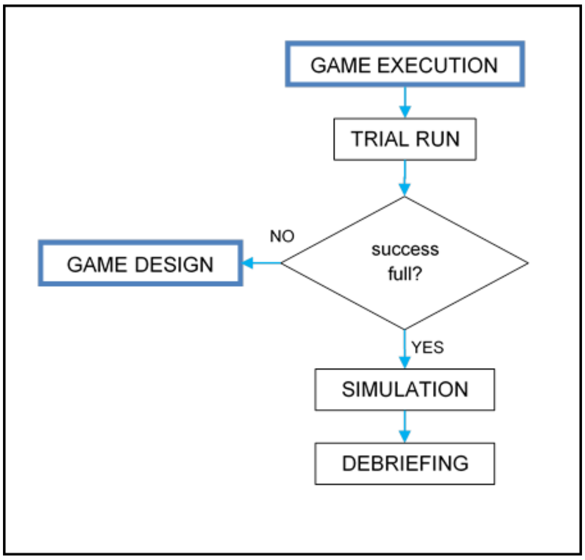


Figure 3: The Game Execution Process

Next, the game execution phase could be carried out: a key role is the game manager team which has to support players in responding to the specific game dynamic. Furthermore, each player is provided with the game kit such as solution cards, and other supporting materials (e.g. BCPs and/or emergency procedures).

The simulation session is structured in *f* round, each one for each main event. Every round is bordered with a deadline to convey the solution.

Within each round, the facilitator is allowed to subject, randomly, the players to one or more random cards, which can be different for each player.

Finally, it would be preferable to plan a debriefing session at the end of the game performance, in order for all the players to deal with their personal suggestions and, moreover, to draw the relative conclusions, concerning the level the firm shows in terms of readiness to tackle such unpredictable circumstances.

4 THE CASE STUDY AT THE BOSCH BARI PLANT

The aim of this paragraph is that of presenting how the framework, deeply illustrated in the previous paragraph, has been applied to build the Bosch Business Continuity Game. The theoretical model has been applied to a manufacturing firm of Bosch Group. The plant is located in South Italy (Bari) and it produces components for the automotive market. Its supply chain is very complex and interconnected; thus, a variation in production rate at Bosch Bari plant could heavily affect the overall company performance (Gnoni et al., 2003; Gnoni et al., 2013).

The production system is based on lean manufacturing principles thus all the manufacturing processes are smooth with an optimal stock of goods which are sufficient just for a very narrowed period of production interruption. Strictly bounded to this premise, the need of adopting an efficient BCM system is necessary.

4.1 The Initial assessment step

The documentation assessment at our disposal shows that only some departments are provided with emergency plans, even though these do not take into consideration the whole of external circumstances that may elicit the interruption of the business, however they are mainly focused on the threats against the IT system. Besides, the direct interviews with some spokespeople of the staff have disclosed that the discipline BCM is only acquainted by top managers and not by members at the operational level. Currently, any suggestion oriented towards the implementation of BCM system has not undertaken yet. This can be traced by the organizational chart where nobody responsible for BCM is included.

Currently the plant is endowed with an emergency procedure focused on the emergency management rather than on the business continuity.

The goal of the game is to integrate the emergency management procedures with ones regarding the business continuity process.

Based on previous proposed classification (see section 3), the Bosch Bari plant belongs to class B companies. Thus, the game application mainly aims:

- Evaluating how management runs an emergency situation and make decisions;
- Evaluating how the organizational structure reacts in order to reduce the weight of the critical business interruptions;
- Highlighting potential deficiencies in the existing procedures;
- Developing a most structured approach towards the Business Continuity Management;
- Harmonizing and complementing the existent procedures in a unique BCP.

Next, following activities have been developed such as:

1. A preliminary risk assessment, where unforeseen events affecting business continuity have to be analyzed. The focus was on external threats derived from natural disasters. By analyzing External Emergency Plan defined by the municipality where the plant is located, game designers have outlined earthquake as the most critical event in terms of both probabilities and consequences (essentially on firm productivity).
2. Business Impact Analysis (BIA), where we were taking into account the several business units traceable in the establishment which are three. In order to ponder the business impact, we verified whether the Plant of Bari was the only one supplier for each brand, as far as Bosch is concerned. This analysis shows that Bari plant is the production site of these products, which are characterized by the lowest MAO (Maximum Acceptable Outage) value.

4.2 The Game design step

The Bosch BCG has been developed based on the previous proposed framework. As earthquake with a 5.4 magnitude. Next, event cards have been defined (see figure 4); they include following information:

- Starting conditions: e.g. date, time, shift when the accident will occur;
- The specific event description;
- The Post-event conditions: in the event of earthquake, we are providing information concerning the condition of the machineries and of the escaped out staff.

Round 1 - Starting Situation

It's 6 p.m., 9th December - The second shift is normally running

Event

5.4 magnitude earthquake

Post event context

- Emergency situation of "class 4" (according to the Procedure Bar-02-04)
- The ERT organizes the mobile accident unit and puts on the alert
- All of the workers escape out of the plant and go to the meeting points situated outside the plant without their personal effects
- The ERT checks that the Emergency Procedure has been strictly followed and that the first aiders provide assistance to the injured associates
- The buildings have withstood the earthquake without having suffered any apparent structural damage

Figure 4: Event Card

4.2.1 The Escalation factor definition

In this phase, game designers have fixed the total number of rounds based on the criticality level to be reached: the total number of run is 3 starting from an earthquake (event 1). Two following events have been introduced such as:

- Event 2: the earthquake has determined the failure of some components of the compressed air line that supplies a determined assembly line
- Event 3: explosion in the proximity of a critical assembly line during equipment restarting.

The event 2 has a low impact on the business interruption (p_2 = 1) as equipment faults are located at the top of the air compressed line which provides service fluids to a not critical assembly line. This assembly line is not critical since other plants of Bosch Group in the world are configured to manufacture that specific products. Furthermore, the time necessary for the maintenance and restoration of the line is quite short (about 48 hours), so the inventory level available in the plant could compensate the temporary production unavailability.

The following round (due to event 3) is more critical as it determines a high impact (p_3 = 2) since the explosion involves the assembly line of a critical item: the time necessary to restart the line is high (at least 1 week) because machines have been damaged both in their hardware and software (e.g. PLC) components. The figure 5 illustrates the timeline of the occurrence of the events and the time lapse they cause.

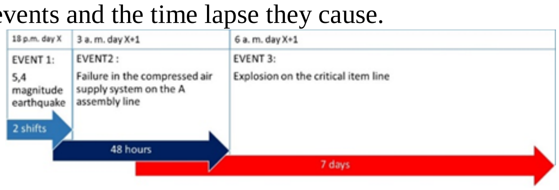


Figure 5: The proposed event timeline

4.2.2 The Player involvement process

The firm top management has decided to fully participate to the game as they are directly involved in carrying out Bosch emergency plan. In detail, during an emergency, the firm top management coordinates the so called Emergency Coordination Committee to manage crisis conditions. During the game, the intervention of this committee is simulated. Main functions involved are: the firm Top manager, Technicians, Human resources, Health Safety Environment, Production, Logistics and Security.

4.2.3 The critical process evaluation in the BCG

Since Bosch belongs to class B, thus lacking a BCP, we mapped the potential activities to be performed in order to tackle the crisis triggered by the earthquake. We used a swimming lane tool pictured in figure 6.

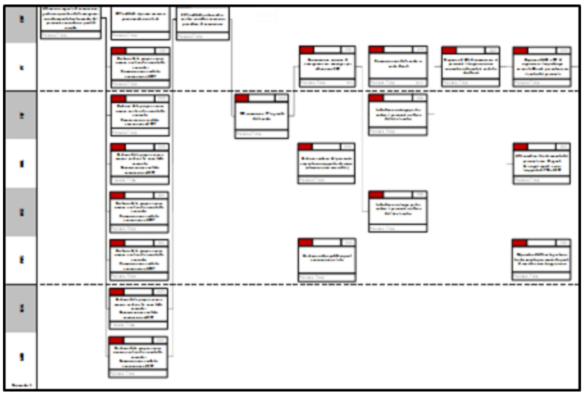


Figure 6: Swimming Lane Tool Map

Game players are outlined in the swim lane, and activities are depicted in the boxes. Each activity corresponds to a valued process time; moreover, links represents the information flows among the actors.

PLAYER	PLANT MANAGER
INPUT	1. The building of the plant has withstood the earthquake (no clear damages, collapses, etc.) 2. The workers gathered in the meeting points (< 1000 persons) have not got their personal effects and therefore they cannot come back home 3. It is necessary to minimize the production losses (max 1 shift) with the aim to resume the production not later than 6 a.m. of the next day
EXPECTED OUTPUT	Coordinate the activities of emergency management through : elaboration of a Recovery Production Plan; maintenance of a proper flow of internal/ external information, interfacing with the administrative organs and police departments
SUGGESTIONS	1. Call the ECC for the emergency coordination 2. Organize the internal/external communication 3. Manage the relations with administrative organs and police departments

Figure 7: Solution Card

Example of solution cards developed for the case study are in Figure 7. While input are the same for all players, the expected output and the suggestions are personalized and formulated according to each player role.

4.2.4 The Random events definition

One random event has been introduced for impacting on event 1: the arrival of press officers at our plant to make some interviews and to document the crisis taking place. This random event has a negative impact with value 1 since it caused the slowing down of the recovery process. The example is shown in Figure 8.

Jolly Event - Starting Situation

It's 4 p.m., 10th December – In the plant there is a state of emergency due to the earthquake

Event

A troupe of a public television's reporters arrives at the plant

Post event Situation

- Having heard the news of the earthquake that involved Bar, the journalists want to report the event and to hold interviews with the top managers
- For this reason they ask the authorization to realize some video to document the emergency situation resulting from the earthquake

Figure 8: Random Card example

After the outline of the equipping assets for the Bosch BCG, the Game execution has been carried out by a trial run: a trial run has been carried out by involving a more restricted group of players. After the trial run development such considerations have been deducted:

- The players were not aware of the procedures;
- The players did not carefully read the solution cards, especially the hints;
- The players promoted a joined work, rather than an autonomous work, in order to find the fittest solution;
- Such solutions suggested by the players, was different by that one hypothesized by the game designers, during the value stream design step.

Thus, few modifications have been carried out such as:

- The mapping of the activities has been modified according to suggestions provided by the players;
- The solution card format has been modified: each player will be given a personal blackboards, where input information and required output have to be pointed out.

Furthermore, a set of key performance indicators have been also introduced to monitor the global effectiveness of the BCG such as: time required by each player in providing output; occurred delay of the overall team player in providing solution for each round; number of suggestion deducted by the game development not yet introduced in the forecast preliminary analysis.

5 CONCLUSIONS

The paper proposes a guideline to design BCGs in the manufacturing sector based on common activities characterizing business continuity plans. The framework proposed has been validated in a real case study regarding an automotive firm: the firm supplies components to final customers as well as to other production sites. Furthermore, the firm applies intensively lean management strategies; thus, business continuity has been outlined as a relevant issue.

The application of the proposed framework has revealed effective as modifications and new emerging issues have been outlined to support firm management in designing a more efficient business continuity plan.

REFERENCES

Boyle, E. A., Connolly, T. M., & Hainey, T., 2011. The role of psychology in understanding the impact of computer games. *Entertainment Computing*, 2, 69–74.

Crookall, D., 2010. Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & Gaming*, 41 (6), 898–921.

Massaguer, D., Balasubramanian, V., Mehrotra, S., Venkatasubramanian, N., 2006. Multi-Agent simulation of Disaster Response. *Proceeding AAMAS '06 Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems,* 1469 – 1470. May 8-12, Hakodate (Hokkaido, Japan).

van der Zee, D-J., Holkenborg, B., Robinson, S., 2012. Conceptual modeling for simulation-based serious gaming. *Decision Support Systems*, 54: 33-45

Di Loreto, I., Mora, S., Divitini, M., 2012. Collaborative serious games for crisis management: an overview, IEEE 21st International WETICE Conference.

Wouters, P., van der Spek, E., van Oostendorp, H., 2009. Serious game for crisis management: What can we learn from research on animations? Chapter XIV in Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices, IGI global Editor, 231-250.

Cowley, B., Mountinho, J. L., Bateman, C., Oliveira, A., 2011. Learning principle and interaction design for 'Green My Place': A massively multiplayer serious game. *Entertainment Computing*, 2: 103-113.

Department of Health, 2009. *The Flu Pandemic Game: a business continuity training resource.* Available from: http://www.continuitycentral.com/flupandemicgame.pdf [accessed 24 March 2014]

Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M., 2012. A systematic literature review of empirical evidence on computer games and serious games. *Computers & education*, 59: 661-686.

Ruppel, U., Scatz, K., 2011. Designing a BIM-based serious game for fire safety evacuation simulations. *Advances Engineering Informatics*, 25: 600-611.

Business Continuity Institute, 2013. Guide to play the game. Available from: http://www.bcifiles.com/BC24GuideFeb2014.pdf.

St-Germain, R., Aliu, F., Lachapelle, E., Dewez, P., 2012. *ISO 22301 Social Security Business Continuity Management Systems*. PECB. Available from: http://www.pecb.org/iso22301/iso22301_whitepaper.pdf [Accessed 28 January 2014]

Lam, W., 2002. Ensuring Business Continuity. *IT Pro*, May|June: 19-25.

Herbane, B., Elliot, D., Swartz, E.M., 2004. Business Continuity Management: time for a strategic role?. *Long Range Planning*, 37:435-457.

Jarvelainen, J., 2013. IT incidents and business impacts: Validating a framework for continuity management in information systems. *International journal of information management*, 33: 583-590.

Business Continuity Institute, 2010. *The Business Case for BCM.* Available from: http://www.bcifiles.com/BCIBusinessCaseforBCMMarch2010FULLREPORT.pdf

Gnoni, M.G., Iavagnilio, R., Mossa, G., Mummolo, G. , 2003, Modelling dynamics of a supply chain under uncertainty: A case from the automotive industry, International Journal of Automotive Technology and Management, 3(3-4), 354-367.

Gnoni, M.G., Andriulo, S., Maggio, G., Nardone, P., 2013, "Lean occupational" safety: An application for a Near-miss Management System design, Safety Science, 53, 96-104,

MODEL-BASED PREDICTION OF IT SERVICE AVAILABILITY – A LITERATURE REVIEW

Sascha Bosse[(a)], Matthias Splieth[(b)], Klaus Turowski[(c)]

Very Large Business Applications Lab, Faculty of Computer Science
Otto von Guericke University Magdeburg
P.O. Box 4120, 39016 Magdeburg

[(a)]sascha.bosse@ovgu.de, [(b)]matthias.splieth@ovgu.de, [(c)]klaus.turowski@ovgu.de

ABSTRACT

In today's world, more and more systems are supported by IT services. High availability of these IT services is often crucial to customers and – in order to avoid penalty costs for service-level violation – also for providers. Decisions affecting IT service availability are mainly made in the service design stage and may be costly to correct afterwards. Model-based prediction models can support decision making in this stage. However, the suitability of the existing approaches is still discussed in literature. In this paper, a quantitative evaluation of the state of the art in IT service availability prediction is performed based on a structured literature review. Results indicate that no standard approaches do exist to this moment in time. One reason for this fact is the limited comparability of these models due to the lack of quantitative and comparable evaluations.

Keywords: IT Service Management, Availability Management, Reliability Analysis

1. INTRODUCTION

The continuing service-orientation of IT has led to an increased use of IT services in both the private and the enterprise sector. This trend is not at least catalyzed by the concept of cloud computing, which allows consumers to obtain computing resources as elastic IT services (Mell and Grance 2011). In order to compensate the lack of control on external IT service providers, service level agreements (SLAs) are concluded between providers and consumers. In addition to legal aspects, these agreements specify quantitative guarantees for non-functional quality metrics. Among these metrics, the availability of a service is the most crucial one to IT service consumers (Franke, Johnson and König 2013).

Availability is defined as "the ability of a service [...] to perform its agreed function when required" (Hunnebeck 2011). An IT Service's availability (A) is usually measured as the ratio of the time an IT service was able to perform its function (uptime) and the time it was required (time), cf. equation 1 (Shooman 2002). Related to availability is the attribute reliability which is often quantized by the failure rate of the service.

$$A = uptime\ /\ time \qquad (1)$$

Since violations of the guaranteed availability levels of SLAs may cause penalty costs and loss of reputation (Emeakaroha et al. 2012), IT service providers have to manage IT service availability with respect to the costs of fault-tolerance mechanisms in order to achieve high availability. The challenge hereby is that new technologies and changing business requirements lead to a continuous demand for changes in the underlying data centers (Eckert et al. 2007), making assumptions for future IT service availability even harder (Miroslaw Malek 2008).

In order to cope with this challenge, good-practices in IT service management such as the IT Infrastructure Library (ITIL) recommend the application of prediction methods in the service design phase (Hunnebeck 2011). Since crucial decisions are made in that stage which may be expensive to correct in the subsequent service transition and operation phase, prediction models are intended to serve as a decision support tool for availability management. However, the industrial application of these approaches is rather low (Hunnebeck 2011). One reason for this fact is the lack of reliable information about the suitability and the accuracy of the existing approaches (Chan, Lyu and M. Malek 2006; Gokhale and Trivedi 2006). A systematic and quantitative overview of these prediction approaches may provide this information, but is missing in literature to this date.

Therefore, in this paper, the current state of the art of IT service availability prediction in the design phase is presented, based on a structured literature review, e.g. defined in (Webster and Watson 2002). On this basis, fields for future research can be identified in order to increase the applicability of prediction approaches in industrial contexts.

In the remainder of this paper, first the setup of the literature review and the material collection is described in section 2. After that, identified literature reviews related to this work are presented in section 3, before the results of the descriptive analysis are discussed in section 4. Section 5 provides insights into the defined analysis categories and their values while section 6 presents the results of the material evaluation. Section 7 concludes this contribution by summarizing the paper and discussing future work in the field of IT service availability prediction.

2. LITERATURE REVIEW PROCESS AND MATERIAL COLLECTION

The structured literature review process that was performed for this paper is derived from (Seuring and Müller 2008), which consists of four steps:

1. Material collection,
2. Descriptive analysis,
3. Category selection and
4. Material evaluation.

The material collection aims to identify the relevant contributions in the research field. In the descriptive analysis, the formal aspects of the identified papers are analyzed in order to characterize the result set. In order to evaluate the material in the last step, content-related categories are identified and selected in step three.

In order to be considered as relevant, a paper should provide an approach for predicting IT service availability using an architectural model and should furthermore be applicable in the service design phase. Hence, measured data about the service's availability from the past should not be required by the prediction model.

The first step of the material collection was performed by keyword search in major publication databases of the IT domain:

- *IEEE Xplore Digital Library* (http://ieeexplore.ieee.org/),
- *SpringerLink* (http://link.springer.com/),
- *ACM Digital Library* (http://dl.acm.org/),
- *Wiley Online Library* (http://onlinelibrary.wiley.com/),
- *Google Scholar* (http://scholar.google.com/) and
- *AIS Electronic Library (*http://aisel.aisnet.org/).

Only double-blind reviewed contributions written in the English language were considered. For the keyword term the following phrase was used: *"(Reliability OR Availability OR Dependability) AND (IT OR Service OR Computing OR System) AND (Prediction OR Estimation OR Forecast)".*

Using the keyword search, around 6,500 contributions were found. After the results from the databases were filtered by checking title and abstract for relevance, 121 papers remained. Once these papers were downloaded, content filtering could be carried out which resulted in 79 relevant publications and three related literature reviews.

3. RELATED WORK

Based on the keyword search described in the previous section, three papers are identified that aimed to investigate the state of the art in availability prediction by analyzing the literature.

In (Goseva-Popstojanova and Trivedi 2003), the authors give an overview about quantitative methods to assess the reliability of software systems. They distinguish state-based, path-based and additive approaches and present examples from literature for each category. State-based approaches model all possible states of the system while path-based approaches focus on execution paths. Additive models consider the system reliability as a random variable and try to map it to stochastic processes. Furthermore, the authors define requirements for reliability prediction in the software domain and identify the most important problems in the field to that date: the estimation of the model parameters, the validity of the Markov assumption, the incorporation of usage profiles, the ability to map inter-component dependencies and the prediction of multiple quality attributes.

The authors of (Gokhale and Trivedi 2006) continue this research stream, developing a unification framework for state-based models. Therefore, seven continuous- and seven discrete-time Markov chains were analyzed and harmonized.

In (Immonen and Niemelä 2008), the authors provide a survey of reliability and availability prediction methods from the viewpoint of software architecture. They develop a framework for comparing different methods which includes criteria from the context, users, contents and validation of the models. They compare six exemplary approaches from literature using the previously described framework and conclude that none of the investigated methods satisfies all defined criteria.

Although the three identified literature reviews provide interesting insights and enhance the research field, none of them evaluates the state of the art in availability prediction quantitatively. In addition to that, the last literature review to be found was published in 2008. Due to the ongoing discussion of that field, the insights of these papers should be updated with respect to recent publications.

4. DESCRIPTIVE ANALYSIS

In this section, the formal attributes of the publications are assessed. Therefore, for each contribution, the publication year and type are considered.

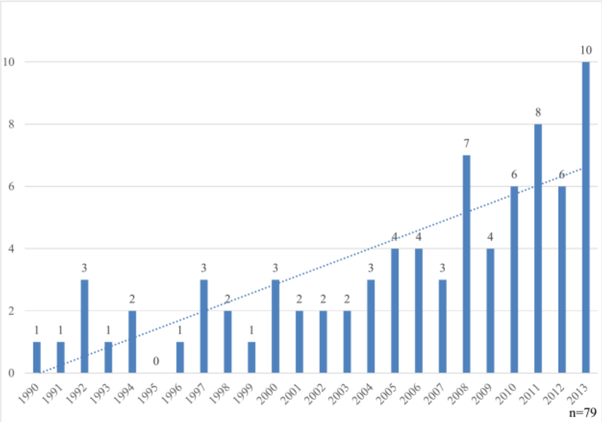


Figure 1: Number of Relevant Publications per Year

Figure 1 presents the number of publications per year, ranging from 1990 to 2013. Although the service-orientation of IT started at the earliest in 1996 with the CORBA standard (OMG 1998), some identified papers had earlier dealt with the development of reliability prediction models for integrated hardware and software systems, e.g. (Laprie et al. 1991). This advancement can be seen as the basis for the following relevant publications, since the developed models can also be applied to IT service availability prediction. Therefore, these papers are included in this literature review.

In the late 1990s, IT systems grew more and more complex, leading to a need of novel prediction models, which can be determined by the increasing number of publications during these years. With the breakthrough of service-oriented computing in the middle of the 2000s, even more approaches have been published in recent years. In figure 1, the linear regression model of the paper number per year (represented by the dotted line) shows that the number of publications increased year by year.

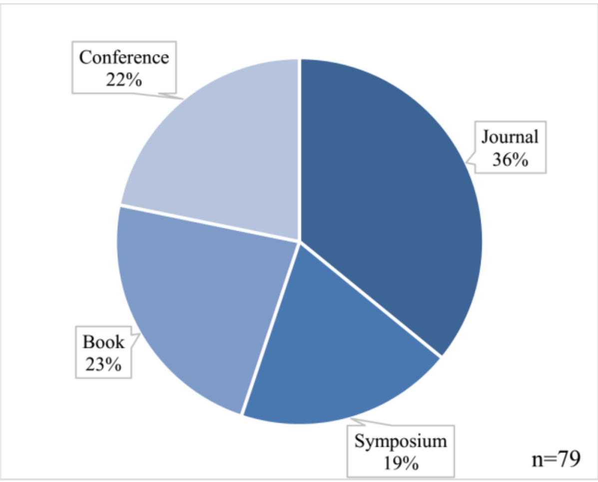


Figure 2: Publication Types of Relevant Contributions

On the one hand, this illustrates the importance of IT service availability prediction. On the other hand, it can be concluded that the evolution of prediction approaches is not finished yet and will receive further attention in the scientific community.

The identified papers were published in four types (cf. figure 2): journal articles (28), book chapters (18) as well as conference (17) and symposium (15) papers. The high number of journal articles indicates that completed research projects in this field do exist, promising validated approaches for IT service availability prediction. However, even more papers were published in symposium or conference proceedings, which shows the lively discussion of the topic. Book chapters often provide a more detailed theoretical foundation of the developed approaches than other publication types due to more available space. The most frequent publication outlets are the *Reliability and Maintainability Symposium* and the *IEEE International Conference on Systems, Man, and Cybernetics* (both four publications), followed by the journals *IEEE Transactions on Reliability*, *IEEE Transactions on Software Engineering*, and the *Journal of Software and Systems Modeling* (each three publications).

Considering both publication year and type, it can be concluded that the field of IT service availability prediction is still heavily discussed in many papers and different contexts. This indicates that current research still lacks of approaches that are widely approved by scientist and practitioners.

5. CATEGORY SELECTION

Since this paper aims to investigate approaches for predicting the availability of IT services, the categories for material evaluation are derived from the identified approaches. These are the followed approach, the used method, the computation technique and the evaluation type and result.

Approaches for predicting IT service availability quantitatively base on measured data (black-box approaches) or on architectural information (white-box approaches), cf. figure 3. Since the investigated approaches should be applicable in the design stage where measured data is not available, only white-box approaches were considered here. These approaches model the components and their interactions that are required for service provisioning. They can be classified further into combinatorial, state-space-based and hierarchical methods (Trivedi et al. 2008). Path-based approaches using information about execution paths are not considered here as a special class because their solution methods are in general either combinatorial or state-space-based.

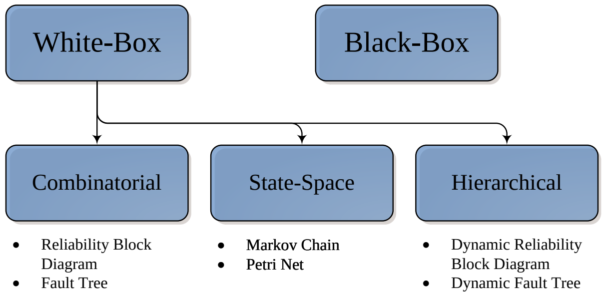


Figure 3: Classification Frame for Availability Prediction Approaches with Example Methods

In combinatorial approaches, an IT service is modeled as a combination of independent components, for instance, in serial, parallel or k-out-of-n systems (Milanovic and Milic 2011). Examples for combinatorial methods are reliability block diagrams (RBD) and fault trees (FT). Due to the assumption of independent components, these models allow simple modeling and fast results' computation. However, complex relations between components such as maintenance and standby configurations cannot be modeled, which is why the accuracy of these approaches for IT service availability prediction is limited (Trivedi et al. 2008).

State-space-based approaches, on the other hand, are able to model inter-dependencies between the components by mapping all possible states of the system and their transition probabilities into a single model (Callou et al. 2012). Markov chains or processes are often applied methods in this class of approaches since the assumption of the Markov property simplifies the evaluation of these models, for example, in discrete- or continuous-time Markov chains (DTMC, CTMC).

However, restricting the state transition time to exponential distributions conflicts with reality where especially recovery times are non-exponentially distributed (Chellappan and Vijayalakshmi 2009). Therefore, other techniques such as semi-Markov processes (Huang, Lin and Kong 2011) or phase-type distributions (Distefano, Longo and Scarpa 2010) are proposed to overcome this assumption.

In comparison to combinatorial approaches, state-space-based approaches not only increase the modeling power drastically, but also the model's complexity. The problem of exponential complexity growth is known as state-space-explosion, leading to challenges in constructing, storing and solving complex models (Trivedi et al. 2008). Some approaches address this problem by encoding the state-space in a Petri net, e.g. in generalized stochastic Petri nets (GSPN) (Kanoun and Ortalo-Borrel 2000) or stochastic reward nets (SRN) (Muppala, Ciado and Trivedi 1994).

Another option to reduce model complexity and increase usability is the use of hierarchical approaches where service and component availability are modeled separately. Often combinatorial models are used for modeling service availability due to their simplicity, while state-space approaches are applied to model components and their dynamic behavior. Examples for methods in this class are dynamic reliability block diagrams (DRBD) or dynamic fault trees (DFT) (Distefano 2009).

A further important aspect of a prediction model is the computation technique, which can be a mathematical or a simulation technique. In the former, closed formula are derived from the model, which are normally solved numerically for complex problems. Since combinatorial approaches are very simple, these models are solved mathematically. However, in state-space approaches, the state-space can grow so complex that numerical methods become error-prone (Sachdeva, D. Kumar and P. Kumar 2008). As an alternative, state-space-based models can be simulated (Xia et al. 2011). On the one hand, this increases the scalability of the approach (Sachdeva, D. Kumar and P. Kumar 2008), on the other hand, only approximate results can be obtained.

Simulation methods also enable the dynamic analysis of availability, providing not only mean values but also information about the availability's variance. This leads to better assumptions about the probability of SLA violations in comparison to steady-state approaches (Franke 2012). Some approaches also provide both computation techniques for maximum flexibility.

Since a developed prediction model is a design artifact, it should be evaluated according to the design science research methodology in order to determine its suitability (Hevner et al. 2004). Therefore, the evaluation type and the prediction accuracy are identified as review categories. In this analysis, the evaluations are classified according to the following scheme:

- Example evaluation would be conducted, if a hypothetic scenario was modeled and availability was computed to show the basic applicability of the developed model. In this evaluation type, results can only be compared to those of other approaches.
- Experimental evaluation would be conducted, if a realistic scenario was modeled and its parameter were varied for experiments which shows the ability of the model to compare different configurations. The results can be evaluated against lab experiments or simulations.
- Case-study evaluation would be conducted, if a real scenario was modeled and analyzed. This proofs the applicability of the model to real-world problems. Evaluation can take place by comparing predicted values to observed ones.
- No evaluation would be conducted, if none of the former evaluation types was carried out.

If a comparison was performed, the accuracy of the approach is computed according to equation 2 where A_m

stands for the measured and A_p for the predicted availability.

$$Acc = |A_m - A_p| / A_m \quad (2)$$

In table 1, the defined categories and their common values found in the results are summarized.

Table 1: Categories and Common Values for Material Evaluation

Category	**Common Values**
Approach	State-Space, Combinatorial, Hierarchical
Method	RBD, FT, CTMC, DTMC, GSPN, SRN
Computation	Mathematical, Simulation
Evaluation Type	None, Example, Experiment, Case-Study
Accuracy	Quantitative

6. MATERIAL EVALUATION

After the categories have been identified, the quantitative analysis of the 79 selected papers can be carried out. Figure 4 presents the classification of the found approaches.

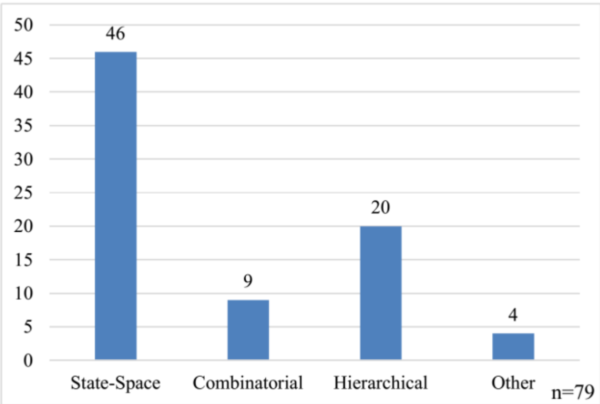


Figure 4: Number of Contributions per Approach

Nearly all approaches could be classified into the three categories state-space, combinatorial and hierarchical. Only four publications dealt with other approaches, e.g. hybrids of state-space- and path-based techniques (Mansour and Dillon 2011). The state-space approaches are most popular because of their modeling power. Nevertheless, some combinatorial approaches were also developed to provide better usability than state-space approaches. Hierarchical approaches are the second most frequent approaches, however, most of the publications presenting hierarchical methods have been published recently. This indicates that hierarchical approaches become more popular since they combine modeling power and usability. Nonetheless, the correlation factor between year and approach is no more than 0.071, which means no trend could be identified over the years. Hence it can be said that no approach is predominant in scientific literature.

By analyzing the applied methods, it could be identified that the combinatorial approaches use reliability block diagrams (RBD) and fault trees (FT) alike. On the other hand, the service availability in hierarchical approaches is mainly modeled with RBDs. This can be explained by the fact that in FTs, failure events are modeled and the component level is not described explicitly. On the contrary, the RBD method model defective components directly, hence these models can be easily connected to the formulated component level state-space models. These models are built in Markov chains and Petri nets in literature without any preference for either of the methods.

In models that are purely state-space-based this is not the case. As figure 5 shows Markov chains and especially continuous-time Markov chains (CTMC) are the dominant method in this approach. Discrete-time Markov chains (DTMC) are used in nearly a quarter, Petri net-based methods in nearly a fifth of the publications. Other used methods are mainly based on Bayesian networks (Roshandel, Medvidovic and Golubchik 2007) or other probabilistic methods.

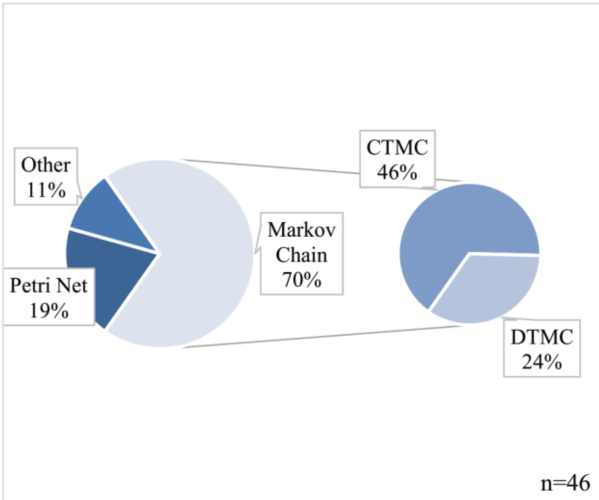


Figure 5: State-Space Approaches classified by the Used Method

The popularity of Markov chain methods is principally founded in the formal specification and the standardized solving algorithms. However, the definition of the state-space is not intuitive and hence error-prone. Petri net models are more intuitive due to the implicit mapping of the state-space (Kanoun and Ortalo-Borrel 2000) and allow the modeling of concurrency. On the other hand, some high-level Petri nets can only be solved via simulation.

This also explains the relatively high number of simulation computation techniques for Petri net approaches as presented in table 2. While a third of the Petri net models can be solved by simulation, for Markov chains and all approaches in total the mathematical computation technique is more dominating.

Table 2: Computation Technique Frequency

	Markov Chains	Petri Nets	Total
Mathematical	84.0 %	66.7 %	81.0 %
Simulation	16.0 %	22.2 %	12.7 %
Dual	0.0 %	11.1 %	6.3 %

Analyzing the evaluations carried out in the identified publications it stands out that only 8.9 % of the papers spare the evaluation, cf. figure 6. In the other contributions, no evaluation type is dominating. Simple numerical examples are used in 27.8 %, experimental setups in 25.3 % of the cases. In nearly 38 % of the papers, a real-world case-study was performed for evaluation. However, just a very small number of publications (7.6 %) provide a quantitative comparison with measured values.

In (Boudali, Sözer and Stoelinga 2009), the authors use fault-injection in their case-study to produce comparable availability results. Their developed CTMC model achieves an accuracy of 0.8 %. (Janevski and Goseva-Popstojanova 2013) evaluated their hybrid prediction model combining state-space- and path-based methods against real data with an accuracy of 1.91 %. Other comparisons were carried out in experimental environments using, for example, testbeds (Reussner, Schmidt and Poernomo 2003). The accuracy achieved in these evaluations varies from 0.36 % to 2.1 %.

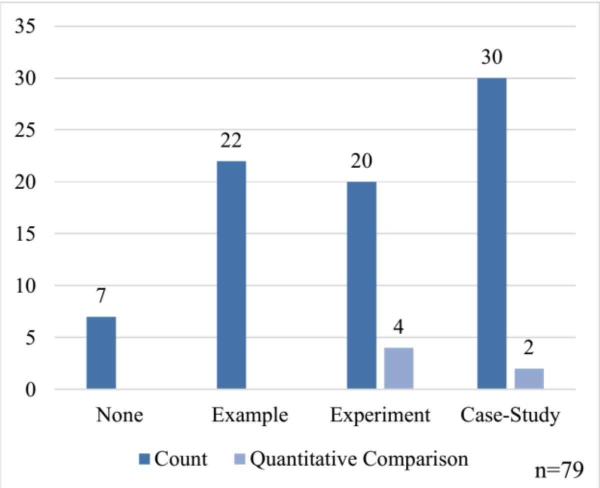


Figure 6: Number of Publications by Evaluation Type

7. CONCLUSIONS

In this paper, a structured literature review was carried out in order to analyze the state of the art in model-based IT service availability prediction quantitatively. Therefore, 79 relevant publications were identified by keyword search. After the descriptive analysis pointed out that the scientific discussion of this research field is still ongoing, the categories for the content analysis of the identified papers were defined. On that basis the material evaluation could be performed. It revealed that approaches based on the state-space of the availability problem are very popular in literature, although hierarchical approaches combining combinatorial and usable service level with state-space-based component level availability models also come into focus.

The detailed analysis of the state-space approaches revealed that continuous- and discrete-time Markov chains are mostly used for modeling the state-space. Petri net-based approaches are developed in nearly a fifth of the identified publications, although the encoding of the state-space decreases the model complexity. Together with simulation methods for solving these models, Petri net approaches provide better scalability than the mathematical evaluation of Markov chains (Sachdeva, D. Kumar and P. Kumar 2008).

While modeling power, applicability and usability of the selected approaches are discussed in the publications, the scalability and even the accuracy of the approaches are evaluated hardly. This is because the evaluation of these models is often not conducted quantitatively in comparison to real-world values. From the 79 identified publications, only six ones provided a quantitative evaluation and compared its result with the real world. Furthermore, only two of these six papers used a real-world case to validate their prediction model. Nevertheless, the evaluation results in these cases could only be produced by manipulating the experiment, e.g. with fault injection, reducing the confidence in the evaluation results.

This lack of comparable evaluation leads to the situation that the accuracy, scalability and suitability of most approaches cannot be assessed exactly. One reason for this might be the lack of accessible data for IT service availability for a constant system configuration over a long period of time. Therefore, it is hard to decide which approach is the most feasible one for a specific scenario. Further research could provide a benchmark where configurations were observed during a long period of time so that the measured availability data can be used for the evaluation of prediction models.

Besides the discussion about approaches there are two major barriers that make the application of design phase prediction models difficult. First, the model creation process requires deep knowledge of the data center. If human resources are involved in this process, it may be costly and error-prone. Some papers have started to address this problem with automatic model generation procedures using monitored infrastructure data, for example in (Milanovic and Milic 2011).

The second problem is the parameterization of the prediction models in order to quantify component availability. In case of hardware components, manufacturers' data is often over-optimistic (Pinheiro, Weber and Barroso 2007) and field-test data must be generated. In the case of software components, this problem is even more crucial due to the lack of reliable failure data, especially for novel software components that are developed in the design phase or require frequent software updates.

Consequently, the low industrial penetration of availability prediction models may be caused by these two problems in addition to the fact that the existing approaches are barely comparable due to lacking evaluation. In order to ease SLA management by providing assumptions about availability in the design phase, these drawbacks need to be addressed in future research.

REFERENCES

Boudali, H., Sözer, H. and Stoelinga, M., 2009. Architectural Availability Analysis of Software Decomposition for Local Recovery. *Proceedings of the 3rd IEEE International Conference on Secure Software Integration and Reliability Improvement*,

Callou, G., Maciel, P., Tutsch, D., Araújo, J., Ferreira, J. and Souza, R., 2012. A Petri Net-Based Approach to the Quantification of Data Center Dependability. *Petri Nets - Manufacturing and Computer Science*, 313–336, InTech.

Chan, P.P.W., Lyu, M.R. and Malek, M., 2006. Making Services Fault Tolerant. *Proceedings of the 3rd International Service Availability Symposium*, 43–61, Lecture Notes in Computer Science, Berlin, Heidelberg: Springer.

Chellappan, C. and Vijayalakshmi, G., 2009. Dependability modeling and analysis of hybrid redundancy systems. *International Journal of Quality & Reliability Management*, 26, 76–96.

Distefano, S., 2009. Reliability and Dependability Modeling and Analysisof Dynamic Aspects in Complex Systems. *8th IEEE International Conference on Dependable, Autonomic and Secure Computing*, 43–48, Chengdu, China: IEEE Computer Society.

Distefano, S., Longo, F. and Scarpa, M., 2010. Availability Assessment of HA Standby Redundant Clusters. *29th IEEE International Symposium on Reliable Distributed Systems*, 265–274, New Delhi, India: IEEE Computer Society.

Eckert, J., Repp, N., Schulte, S., Berbner, R. and Steinmetz, R., 2007. An Approach for Capacity Planning of Web Service Workflows. *Proceedings of the 13th Americas Conference on Information Systems (AMCIS)*,

Emeakaroha, V.C., Netto, M.A.S., Calheiros, R.N., Brandic, I., Buyya, R. and Rose, C.A.F.D., 2012. Towards autonomic detection of SLA violations in Cloud infrastructures. *Future Generation Computer Systems*, 28, 1017–1029.

Franke, U., 2012. Optimal IT Service Availability: Shorter Outages, or Fewer? *IEEE Transactions on Network and Service Management*, 9, 22–33.

Franke, U., Johnson, P. and König, J., 2013. An architecture framework for enterprise IT service availability analysis. *Software and Systems Modeling*. Available at: http://dx.doi.org/10.1007/s10270-012-0307-3.

Gokhale, S.S. and Trivedi, K.S., 2006. Analytical Models for Architecture-Based Software Reliability Prediction: A Unification Framework. *IEEE Transactions on Reliability*, 55, 578–590.

Goseva-Popstojanova, K. and Trivedi, K.S., 2003. Architecture-Based ApApproach to Software Reliability Prediction. *Computers and Mathematics with Applications*, 46, 1023–1036.

Hevner, A.R., March, S.T., Park, J. and Ram, S., 2004. Design science in information systems research. *MIS Quarterly*, 28, 75–105.

Huang, J., Lin, C. and Kong, X., 2011. Modeling and Analysis of Dependability Attributes of Service Computing Systems. *2011 IEEE International Conference on Services Computing*, 184–191, Washington D.C., United States: IEEE Computer Society.

Hunnebeck, L., 2011. *ITIL Service Design 2011 Edition*. Norwich, UK: The Stationery Office.

Immonen, A. and Niemelä, E., 2008. Survey of reliability and availability prediction methods from the viewpoint of software architecture. *Software and Systems Modeling*, 7, 49–65.

Janevski, N. and Goseva-Popstojanova, K., 2013. Session Reliability of Web Systems under Heavy-Tailed Workloads: An Approach Based on Design and Analysis of Experiments. *IEEE Transactions of Software Engineering*, 39, 1157–1178.

Kanoun, K. and Ortalo-Borrel, M., 2000. Fault-Tolerant System Dependability—Explicit Modeling of Hardware and Software Component-Interactions. *IEEE Transactions on Reliability*, 49, 363–376.

Laprie, J.-C., Kanoun, K., Béounes, C. and Kaâniche, M., 1991. The KAT (Knowledge-Action-Transformation) Approach to the Modeling and Evaluation of Reliability and Availability Growth. *IEEE Transactions of Software Engineering*, 17, 370–382.

Malek, Miroslaw, 2008. Predictive Algorithms and Technologies for Availability Enhancement. *5th International Service Availability Symposium (ISAS)*, 17–19, Lecture Notes in Computer Science, Tokyo, Japan: Springer Verlag Berlin Heidelberg.

Mansour, H.E. and Dillon, T., 2011. Dependability and Rollback Recovery for Composite Web Services. *IEEE Transactions on Service Computing*, 4, 328–339.

Mell, P. and Grance, T., 2011. *The NIST Definition of Cloud Clomputing*. National Istitute of Standards and Technology.

Milanovic, N. and Milic, B., 2011. Automatic Generation of Service Availability Models. *IEEE Transactions on Service Computing*, 4, 56–69.

Muppala, J.K., Ciado, G. and Trivedi, K.S., 1994. Stochastic Reward Nets for Reliability Prediction. *Communications in Reliability, Maintainability and Serviceability*, 1, 9–20.

OMG, 1998. *CORBAservices: Common Object Services Specification*. Object Management Group (OMG).

Pinheiro, E., Weber, W.-D. and Barroso, L.A., 2007. Failure Trends in a Large Disk Drive Population. *Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST)*,

Reussner, R.H., Schmidt, H.W. and Poernomo, I.H., 2003. Reliability prediction for component-based software architectures. *The Journal of Systems and Software*, 66, 241–252.

Roshandel, R., Medvidovic, N. and Golubchik, L., 2007. A Bayesian Model for Predicting Reliability of Software Systems at the Architectural Level. *Proceedings of the 3rd International Conference on the Quality of Software-Architectures*, 108–126, Lecture Notes in Computer Science, Springer Heidelberg.

Sachdeva, A., Kumar, D. and Kumar, P., 2008. Reliability analysis of pulping system using Petri nets. *International Journal of Quality & Reliability Management*, 25, 860–877.

Seuring, S. and Müller, M., 2008. From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, 16, 1699–1710.

Shooman, M.L., 2002. *Reliability of Computer Systems and Networks – Fault Tolerance, Analysis, and Design*. New York, NY, USA: John Wiley & Sons New York.

Trivedi, K.S., Ciardo, G., Dasarathy, B., Grottke, M., Matias, R., Rindos, A. and Vashaw, B., 2008. Achieving and Assuring High Availability. *5th International Service Availability Symposium (ISAS)*, 20–25, Lecture Notes in Computer Science, Tokyo, Japan: Springer Verlag Berlin Heidelberg.

Webster, J. and Watson, R.T., 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quaterly*, 26, xiii–xxiii.

Xia, Y., Dai, G., Li, J., Sun, T. and Zhu, Q., 2011. A model-driven approach to predicting dependability of WS-CDL based service composition. *Concurrency and Computation: Practise and Experience*, 23, 1127–1145.

ANALYSIS OF SIMULATION TOOLS FOR DETERMINING THE ENERGY CONSUMPTION OF DATA CENTERS FOR CLOUD COMPUTING

Matthias Splieth[(a)], Sascha Bosse[(b)], Klaus Turowski[(c)]

Otto von Guericke University Magdeburg
Faculty of Computer Science, Very Large Business Applications Lab
P.O. Box 4120, 39016 Magdeburg

[(a)]matthias.splieth@ovgu.de, [(b)]sascha.bosse@ovgu.de, [(c)]klaus.turowski@ovgu.de

ABSTRACT

Cloud Computing has become an important paradigm for providing IT services. An important aspect that data centers which offer cloud services need to deal with is the reduction of their energy consumption. A suitable way for reducing the energy consumption and thus the energy costs is using energy efficient load distribution algorithms. These are commonly evaluated by simulation. Up to now, there are only few simulation tools for clouds that are freely disposable and those that do exist have hardly been compared in terms of their applicability for determining the energy consumption of clouds. Therefore, this contribution presents an analysis of simulators that investigates their suitability for determining the energy consumption of cloud data centers. The findings of the analysis will be supported by a unified experiment. The results demonstrate that existing work has limitations, for example in terms of components that are not considered to be energy-consuming.

Keywords: cloud computing, energy consumption, simulation, survey

1. INTRODUCTION

In recent years, Cloud Computing has been established as a new paradigm for the dynamic provisioning of IT services. In spite of the advantages of cloud services for both customers and providers, there are also some challenges that need to be addressed. One of the most important challenges for providers is the reduction of the overall energy consumption that continuously increases due to the rapidly expanding demand for computational power (Zhang, Cheng and Boutaba 2010). Current studies have shown that the energy consumption of data centers increased by 56% from 2005 to 2010 (Koomey 2011). Taking into account that prices and demand for energy will continue to rise and therefore become the dominant factor in the total cost of ownership (Orgerie, De Assuncao and Lefevre 2014), it is desirable for operators of (cloud) data centers to even slightly decrease the energy consumption since this can have significant impacts on their profitability. A possible way to save energy is load distribution (Orgerie, De Assuncao and Lefevre 2014) which is also one of the major challenges regarding cloud computing (Zhang, Cheng and Boutaba 2010). Besides the fact that load distribution in terms of placing or migrating virtual machines (referred to as virtual machine scheduling) is an NP-hard problem (Zhang, Cheng and Boutaba 2010), additional aspects such as the free scalability of the resources according to the current workloads aggravate the task of distributing load in cloud data centers.

In recent years, numerous algorithms for load distribution in clouds have been developed, many of them with the goal to increase energy efficiency. However, their effects on energy consumption in a specific scenario cannot be predicted trivially (Nehru et al. 2013). This is because the possible improvements of an algorithm strongly depend on the parameters of the specific data center, such as the offered services, the configuration of servers, the architecture of the network, the network topology and others. Hence, applying the same algorithm in different scenarios can lead to varying results in terms of energy consumption (Liu et al. 2013). Therefore, there is a need for a solution that is able to exactly predict the energy consumption of cloud data centers and hence to also evaluate the effects of load distribution algorithms on the energy consumption of cloud data centers.

In addition to testbeds and mathematical models, simulation is often used in order to investigate clouds (Sakellari and Loukas 2013). Since tools for the simulation of grids (or other similar paradigms) are not applicable to clouds (Buyya, Ranjan and Calheiros 2009), the number of simulation tools that are available for public use is limited. The existing simulators differ in several aspects, for example in terms of components of a data center that can be modeled. As an example, cooling components (which are an important energy consumer) cannot be modeled in *CloudSim* (Calheiros et al. 2011) but in *CReST* (Cartlidge and Cliff 2013).

In order to be able to investigate the effects if load distribution algorithms on the energy consumption of data centers, a simulator is needed that is able to exactly predict the energy consumption of clouds. Therefore, a survey on and an analysis of simulation tools for cloud data centers is conducted in this paper. It aims to examine their suitability in terms of determining energy consumption in clouds. The survey provides a decision basis for choosing simulation tools for energy efficient cloud computing. Furthermore, a detailed comparison of existing simulation tools can reveal existing flaws that may be addressed in future research. Finally, a unified experiment is defined and implemented with each simulator. The respective results of each simulator will be compared in order to verify the findings of the analysis of the single simulators.

2. RELATED WORK

Cloud Computing is a hot topic and is thus subject of recent research. Since the research area is emergent, research is conducted in many directions. One of the most important areas of this research is the energy efficiency of clouds (Zhang, Cheng and Boutaba 2010). For the investigation of energy aspects, simulation is often used. But until today, an analysis and comparison of cloud computing simulators has only been carried out on a simple level focusing on general aspects of the simulators.

For example, Zhao et al. provide a brief overview of existing cloud simulators without analyzing each simulator in depth (Zhao et al. 2012). In fact, the analysis is limited to only three criteria (underlying platform, programming language and whether the simulator is either based on software or based on software and hardware), while aspects related to energy efficiency are not considered.

The analysis of Sakellari and Loukas presented in (Sakellari and Loukas 2013) is more detailed and also takes energy efficiency into account. However, the analysis in terms of energy is mainly limited to the question of whether energy aspects can be taken into account at all. Furthermore, general aspects such as the programming language are investigated. Details on energy consumption, such as the question which energy-consuming components can be modeled with the simulator are not part of the survey.

The same applies to (Kumar and Rai 2014), wherein the analysis is less detailed as in (Sakellari and Loukas 2013) since this work does not mention how energy efficiency is meant to be considered.

A recent survey on cloud simulators is presented by Ahmed and Sabyasachi in (Ahmed and Sabyasachi 2014). In the analysis of the simulators, their features as well as their respective shortcomings are discussed. The analysis captures, among other things, whether energy models are part of the simulator. However, the analysis does not cover which energy models are included for which components. In addition, the final comparison of the tool is only based on the attributes of the analysis. A detailed comparison of the capabilities of simulation tools in terms of energy aspects is not carried out.

3. ANALYSIS OF CLOUD SIMULATORS

Before starting the analysis, suitable simulators must be identified. Therefore, scientific literature is reviewed towards simulation tools for cloud computing. In order to identify relevant literature, *ACM Digital Library*, *IEEE Xplore Digital Library*, *Springer Link* and *Google Scholar* have been queried for cloud simulators using the search term "*cloud computing & simulat(e|or|ion|ing)*". Emulation environments, test beds and mathematical-analytical approaches are excluded from the survey. Mathematical-analytical approaches are not included since they only depict parts of clouds and are not suitable for modeling complex communication (Sakellari and Loukas 2013). The problem with testbeds or other real-world environments is that experiments are expensive, time-consuming and not repeatable (Zhao et al. 2012). In this paper, the focus is on tools that have been explicitly designed for the simulation of cloud computing systems. Therefore, simulation tools or frameworks that address data centers in general or areas similar to cloud computing (such as grid computing) are not included since such tools are unsuitable for simulating clouds (Buyya, Ranjan and Calheiros 2009). In order to be able to analyze the tools in depth (in terms of their source code) and in order to be able to carry out the unified experiment with each tool, simulators are only analyzed if they are available to public and open source. In order to keep the effort for the experiments low, other tools that do not provide crucial changes to the original are excluded from the analysis.

In total, the search term defined before resulted in 31 different simulators that have been mentioned in scientific literature. Most of them are proprietary implementations and not available for public. On the basis of the aforementioned aspects, four simulators have been identified that are open source and also disposable:

- *CloudSim* (Calheiros et al. 2011), version 3.0.3
- *GreenCloud* (Kliazovich, Bouvry and Khan 2012), version 2.0.2
- *iCanCloud* (Núñez et al. 2012), version 0.9
- *CReST* (Cartlidge and Cliff 2013), version 0.5

These simulators are analyzed in depth in the following.

3.1. Criteria for Analysis

The analysis of the aforementioned simulators consists of three parts: first, general aspects such as the evaluation method or the applied programming language are investigated. In the second part, the simulation tools are examined in terms of their suitability for modeling physical components of a cloud data center. In the last part of the analysis, the simulators are analyzed with respect to issues that are important for simulating load distribution and its effects on energy consumption but not directly consume energy, such as energy models that define how energy is consumed by the components of a data center.

3.1.1. General Aspects

The general criteria are loosely based on (Smit and Stroulia 2013) and are summarized and described in table 1.

Table 1: General Aspects of the Analysis

Criterion	Description
Objective	What was the simulator designed for?
Evaluation	How is the simulator evaluated?
Type of Simulation	Of which type is the simulation? (discrete-event, continuous, …?)
License	License under which the simulator was published
Language	Language the simulator is written in
QoS	Is it possible to measure effects on quality of service?
Services	Supported cloud-layer
Basis	Implementation basis of the simulator

"Objective" captures the intention of the creators, such as estimating the energy-efficiency of clouds or determining their availability. The criterion "Evaluation" describes the evaluation method that has been applied in order to verify the simulator. This can for example refer to evaluation methods such as proof-of-concept or case-study. "Type of Simulation" covers the type of the simulation, such as continuous simulation or discrete-event simulation. The license under which the simulations is released is covered by the criterion "License". This is important due to the possibility for e.g. making changes to the source code. "Language" refers to the programming language that is used for the implementation of the simulator.

"QoS" depicts whether it is possible to investigate indicators that are important for the quality of service, such as availability or response time. The criterion "Services" refers to the cloud-layer supported by the simulator as defined in (Mell and Grance 2011). "Basis" indicates the basis on which the simulator has been implemented, as far as it is based on another tool or framework.

3.1.2. Physical Components

With respect to more advanced criteria, it is necessary that all physical components of a cloud data center that consume energy can be modeled. A rough overview of the relevant components (confer for example (Brown et al. 2007; Jing et al. 2013)) is given in table 2.

For the analysis, these criteria are even more detailed as depicted in table 2. The criterion "Server" represents a physical server that has various energy-consuming components such as CPU, memory, motherboard, peripheral slots, fans and maybe a GPU (Fan, Barroso and Weber 2007; Greenberg et al. 2008).

Table 2: Physical Components of the Analysis

Criterion	Description
Server	Is it possible to model servers in detail?
Cooling	Can cooling components be modeled?
Storage Systems	Is it possible to model storage systems?
Support Systems	Is it possible to model support systems such as UPS or lightning?
Network	How is the network simulated? Can devices and topologies be modeled?

Furthermore, this criterion also includes racks as well as servers in terms of blades (Barroso and Hölzle 2009). This criterion is fulfilled if all components can be modeled; otherwise it is partly fulfilled or not fulfilled (if none of the mentioned components of a server can be modeled).

"Cooling" refers to the ability of a simulation tool to model cooling units that are common in a data center. This can for example refer to CRAC units ("computer room air conditioning"), free cooling methods or in-rack cooling as described in (Barroso and Hölzle 2009). This criterion is fulfilled if at least one cooling system can be modeled.

"Storage Systems" depicts if it is possible to model central storage system since these can be important regarding the way the migration of virtual machines (VMs) is performed (Mishra et al. 2012). This criterion is fulfilled if storage systems can be modeled. The criterion "Support Systems" states whether support systems can be modeled. This refers to uninterruptible power supplies (USP) and lighting and is fulfilled if both of these can be modeled with the simulator.

The criterion "Network" addresses several aspects. The first aspect is concerned with the question how the network is simulated. This can be done using a flow model or a packet model, while the latter is more accurate (Velho et al. 2013). Further aspects concerning this criterion are dealing with the question whether network topologies can be modeled (since there are different types of topologies which have different side-effects (Barroso and Hölzle 2009)) and whether network devices as energy-consuming components are included in the simulator.

3.1.3. Non-Physical Components

In addition to the physical components, there are also non-physical aspects which are relevant for investigating load distribution and energy consumption. An overview of all the relevant non-physical aspects is given in table 3.

In order to be able to evaluate the effects of load distribution algorithms on energy consumption, certainly "Load" has to be considered in a simulator. Load distribution in clouds can either refer to virtual machine scheduling or to the distribution of the workload (Beloglazov, Abawajy and Buyya 2012).

Table 3: Non-Physical Components of the Analysis

Criterion	Description
Load	Can workload and live migration be simulated?
Virtualization / Cloud	Are virtualization and rapid elasticity part of the simulator?
Software	Can software be modeled?
Energy Models	Are energy models provided for each component?
Power Saving Techniques	Are power saving methods a component of the simulator?

VM scheduling refers to the placement or the migration of virtual machines. The distribution of workload means that tasks that are generated by user requests are assigned to applications that run in virtual machines. In order to fulfill this criterion, both aspects must be considered by the respective simulators.

Since this research focuses on the simulation of clouds, it is obvious that the relevant aspects regarding cloud computing need to be taken into account. This is covered by the criterion "Virtualization / Cloud". First, the concepts of virtualization – which is a key technology of cloud computing (Zhang, Cheng and Boutaba 2010; Mishra et al. 2012) – need to be part of the simulator. Furthermore, key aspects such as rapid elasticity and resource pooling (Mell and Grance 2011) must be considered. Resource pooling refers to the possibility to model multi-tenancy aspects with the simulator, which means that it is possible to model software that is for example capable to serve different customers independently from one another. An example for single-tenancy and multi-tenancy is given in figure 1. Rapid elasticity means that provided computing resources are scaled to the actual demand of users. If all these aspects are considered, the criterion is fulfilled.

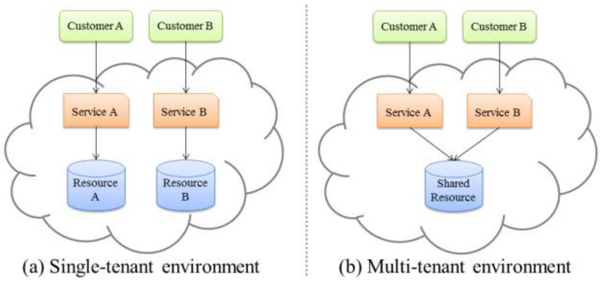


Figure 1: Example for (a) Single-Tenancy and (b) Multi-Tenancy in Clouds

An important aspect are software components in the data center since software determines the workload (Barroso and Hölzle 2009). Since applications usually do not work in isolation but frequently interact with other software components, it is also important to model such dependencies. As an example, this becomes important when a VM that runs a specific application cannot be moved to another physical host because another application would fail due to their dependence or other restrictions, such as legal issues. Therefore, software-dependencies must be taken into account. This criterion is fulfilled if both aspects – meaning the possibility to model software components as well as dependencies between software-components – are covered by the simulator.

Certainly, energy models need to be part of the simulation tool in order to determine the energy consumption of components. Thereby, the simulator should include energy models that cover all physical components. As an example, this can be archived by providing energy models for all components or by providing an energy model that combines all components in one model. This criterion is fulfilled if the simulator provides energy models for all energy-consuming components (such as in (Barroso and Hölzle 2009; Fan, Barroso and Weber 2007)), partly fulfilled if at least some energy models are provided, and not fulfilled if no energy models are provided.

Finally, power saving techniques should be considered by the simulators since they are crucial for data centers in order to save energy (Ge, Sun and Wang 2013). Such techniques can be applied at the level of servers or at the level of networking elements in a data center. Since a vast amount of techniques exists (confer (Ge, Sun and Wang 2013)), this criterion is fulfilled if at least one technique is included in the simulator.

3.2. Analysis of General Aspects

In this section, the analysis of the general criteria is presented. Table 4 summarizes the results of the analysis for each of the investigated simulation tools.

Table 4: General Aspects of the Analysis

Criterion	CloudSim	GreenCloud	CReST	iCanCloud
Objective	General Purpose	Energy-Efficiency	General Purpose	General Purpose
Evaluation	Proof-of-Concept	Proof-of-Concept	Proof-of-Concept	Proof-of-Concept
Type of Simulation	Discrete Event	Discrete Event	Discrete Event	Discrete Event
License	GPLv3	GPLv2	GPLv3	GPLv3
Language	Java	C++, OTcl	Java	C++
QoS	Yes	Yes	Yes	No
Services	IaaS	IaaS	IaaS	IaaS
Basis	-	ns-2	-	OMNet++

Except for *GreenCloud*, the tools were not implemented for a particular purpose, but can be generally used for the simulation of clouds, while *GreenCloud* was designed for determining the energy consumption in clouds. Each simulator is evaluated by a proof-of-concept. *CReST* is evaluated by conducting several experiments identified in literature. The results of the experiments are then compared to the results found in literature. *iCanCloud* is evaluated by comparing its results with results gained form different instance-types from Amazon's EC2. Kliazovich, Bouvry and Khan conduct several experiments in order to demonstrate the suitability of *GreenCloud* (Kliazovich, Bouvry and Khan 2012), but do not provide data for comparison. Also *CloudSim* is evaluated by conducting several experiments (Calheiros et al. 2011). Although Calheiros et al. mention that *CloudSim* is also used in real-world scenarios (Calheiros et al. 2011), evidence is not provided for this statement.

As expected due to the nature of clouds, each tool is a discrete event simulation. However, the fact that all simulators mostly generate deterministic results is surprising. Fluctuating demands for resources in the workload of a user or a random number of users are rarely taken into account. This is surprising since fluctuating and unpredictable demands are a characteristic of clouds (Cartlidge and Cliff 2013).

All tools have been released under the GPL license. The applied programming languages are C++ and Java, while *GreenCloud* uses a mix of C++ and OTcl for defining experiments since it is based on the network simulator *ns-2* which also uses C++ as well as OTcl. Except for *iCanCloud*, all simulators allow investigating impacts on quality of service, although this is limited on performance degradations. Impacts of load distribution on quality of service cannot be measured in terms of common metrics, such as defined in (Li et al. 2012). This can be a disadvantage in so far that too frequent migration of virtual machines on the one hand may significantly reduce the energy consumption, while on the other hand leading to an increasing response time because the load on the physical servers it too high. This can in turn lead to contractual penalties and thus unpick the cost savings. The only tool that considers other aspects than performance degradations is *CReST* since it is able to investigate the availability of physical components. Overall, the options for investigating QoS-aspects in clouds using simulation are limited.

Regarding the criterion "Services", every simulator addresses the infrastructure-layer (Mell and Grance 2011) of the cloud stack.

3.3. Analysis of Physical Components

Table 5 summarizes the results of the analysis regarding the physical components.

Table 5 already suggests that apparently, there are differences between the simulators in terms of which components can be modeled. For example, the model of a "Server" should consist of a CPU, main memory, hard disk drives, a motherboard, one or more peripheral slots and fans since these are the major energy consumers of a server (Greenberg et al. 2008).

Table 5: Analysis of Physical Components

Criterion		**CloudSim**	**GreenCloud**	**CReST**	**iCanCloud**
Server		Partly	Partly	Partly	Partly
Cooling		No	No	Yes	No
Storage		Yes	No	No	No
Support		No	No	No	No
Network	*Model*	Flow	Packet	Flow	Packet
	Topologies	Partly	Yes	Partly	Yes
	Devices	Yes	Yes	Yes	Yes

As mentioned before, also racks and blades are included in this criterion since these are common in data centers. Taking this fact into account with respect to simulators, it can be stated that *CloudSim* can only model CPU, memory, and disk. *GreenCloud* can model everything except for fans, motherboards, and peripheral slots. Nevertheless, main memory, disks, and network cards are not entities but attributes of a server. Furthermore, memory and disk are static values that are not varied during the runtime. The same elements as in *CloudSim* can be modeled with *CReST*. In the context of a direct comparison, only networking cards are missing in the latter. However, main memory and disks can be explicitly modeled as entities in *CReST*. Furthermore it can be stated that *iCanCloud* can also only partly model the energy consuming components of a server: CPU, main memory, and disks are included as well as servers in terms of racks and blades. Other components are not included in the simulator. Therefore, it can be concluded that none of the simulators can model servers according to their energy consuming components.

"Cooling" is only taken into account by *CReST*, whereas none of the other simulators takes this into consideration. This is surprising, since cooling accounts for a substantial part of the energy costs in data centers – confer for example (Pelley et al. 2009; Barroso and Hölzle 2009) – and is important for other aspects as well, such as the availability of data centers since these cannot operate without cooling (Barroso and Hölzle 2009). Therefore, cooling should be taken into account when investigating the energy consumption of clouds.

The only simulator that can model storage systems is *CloudSim*. Since such systems may be important for migrating virtual machines between physical servers (Mishra et al. 2012), they should also be included in a simulator; especially since the energy consumption is different when applying central storage systems in comparison to servers with integrated storage (Minas and Ellison 2009).

Regarding the network criterion, the analysis revealed that *GreenCloud* and *iCanCloud* are packet-level simulators. Packet-simulation is widely used to study network protocols and is very accurate, but also expensive. An alternative is to implement network simulation by flow-models (Velho et al. 2013) as it is done in *CloudSim* and *CReST*. Since *GreenCloud* as well as *iCanCloud* are based on famous open source network simulators (*ns-2* respectively *OMNet++*), they also provide the possibility to model different types of topologies as well as network devices, such as different types of switches. Also *CloudSim* and *CReST* provide the possibility to model topologies as well as network devices, although less detailed in comparison to *GreenCloud* and *iCanCloud*. In *CloudSim*, different types of topologies do not lead to different results, which is due to a bug in the implementation. The modeling of topologies is thus possible, but is not correctly implemented. The criterion is therefore only partially fulfilled for *CloudSim*. Support systems are not considered by any investigated simulator, so that the criterion "Support Systems" is not fulfilled.

Summing up, it can be said that none of the examined simulators covers all energy-consuming components of (cloud) data centers. Therefore, there is a high probability that the energy consumption for an entire data center estimated by one of the simulators will be inaccurate since various components are left out.

3.4. Analysis of Non-Physical Components

Despite the fact that the physical components are crucial for estimating energy consumption and the effects of load distribution algorithms on energy consumption, also non-physical components are highly relevant. Regarding the determination of the energy consumption by simulation, energy models are especially important. But also other components are of importance for this purpose, as described in section 3.1. The results of the analysis concerning the non-physical components are summarized in table 6.

Table 6: Non-Physical Components of the Analysis

Criterion	CloudSim	GreenCloud	CReST	iCanCloud
Load	Yes	Yes	Yes	Partly
Virtualization / Cloud	Partly	Partly	Partly	Partly
Software	Partly	Partly	Partly	Partly
Energy Models	Partly	Partly	Partly	No
Power Saving Techniques	Yes	Yes	No	No

Except for *iCanCloud*, all simulators fulfill the criterion "Load" since they provide functionalities to model different types of workloads as well as taking virtual machine scheduling into account, even though most included workloads lead to a deterministic output. Only *iCanCloud* does not include a functionality for VM scheduling, which is why "Load" is only partially fulfilled.

Regarding the criterion "Virtualization / Cloud", it is to say that no simulator can meet all of the aspects that are relevant to this criterion (confer section 3.1). *CloudSim* does not consider resource pooling. The same applies for *CReST* and *iCanCloud*. *GreenCloud* provides all the functionalities required for this. However, rapid elasticity is only partially fulfilled, as the demand is not subject to stochastic influences. Thus, rapid elasticity is taken into account in principle, but there are no adjustments to the demand at runtime and all information is thus known a priori, which is why the features for rapid elasticity are not used. With respect to the criterion "Software", only *iCanCloud* allows to model both software components and dependencies between software components. The other three simulators lack the feature to model dependencies between software. *GreenCloud* is also not able to model software. It is only possible to model tasks which somehow represent software in *GreenCloud*.

Energy models are provided by *CloudSim*, *GreenCloud* and *CReST*, whereas *iCanCloud* does not include any energy models. But none of the simulators provides energy models for all energy-consuming components: *CloudSim* and *CReST* do only provide an energy model for servers and this model only bases on the CPU respectively its utilization, whereas all other components are left out. Regarding an energy model for servers, the same applies to *GreenCloud*. Additionally, *GreenCloud* provides an energy model for networking devices.

The last criterion for the non-physical components are power saving techniques. Since *iCanCloud* does not even provide energy models, also power saving techniques are not included in the simulator. Also *CReST* currently does not include such techniques. However, *CloudSim* and *GreenCloud* include such techniques. *CloudSim* provides an implementation for Dynamic Voltage and Frequency Scaling (DVFS), whereas *GreenCloud* provides implementations of DVFS and Dynamic Power Management (DPM). Therefore, the criterion is fulfilled for both *CloudSim* and *GreenCloud*.

4. EXPERIMENT

The analysis presented in section 3 revealed that there are significant differences between the simulators regarding the components and concepts they take into account. Accordingly, their results in terms of energy consumption should also differ significantly. Therefore, an experiment that aims to verify this assumption is carried out. First, a hypothetical scenario is defined. This scenario is implemented with each simulator (as far as possible). Finally, the results gained from each simulator are compared in order to verify the assumption mentioned before.

4.1. Experimental Setting

In order to keep the effort for implementing the single experiments low, a simple scenario based on a three-tier data center architecture is defined. An example for a typical three-tier architecture as illustrated in figure 2.

A three-tier architecture is common in modern data centers (Kliazovich, Bouvry and Khan 2012). It consists of different types of switches: core layer, aggregation layer and access layer. The core layer connects the data center to the Internet, the aggregation layer provides several functionalities (such as SSL or firewall) and the access layer connects the servers that are partitioned in a racks (Ge, Sun and Wang 2013).

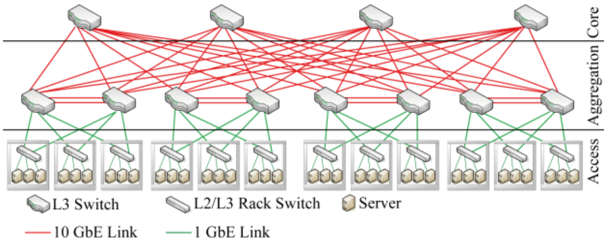


Figure 2: Typical three-tier data center architecture according to (Kliazovich, Bouvry and Khan 2012)

The scenario consists of one core layer switch, two aggregation layer switches and four access layer switches. In total, 40 physical servers are partitioned in four racks. Therefore, each rack holds ten single servers (blades). The servers are homogenous in terms of their provided resources:

- 100.000 MIPS (Million instructions per second)
- 8 GB main memory
- 250 GB hard disk drive
- 1 GbE (Gigabit Ethernet) network adapter

Additionally, a single user is part of the experiment. This user requests a unique service which is always served by a virtual machine or an application within a virtual machine. The user's requests results in 40 different virtual machines. According to the physical servers, the virtual machines have the same resource demands:

- 50.000 MIPS
- 2 GB virtual memory
- 3 GB virtual disk
- 1 GbE virtual network controller

The workload that is generated by the user requests leads to a mean utilization of 50% in terms of the CPU load of each physical server. Each VM is placed on a single server. Live migration of running virtual machines is deactivated, so that the processing of each user request will have to be finished on the server where its serving VM was initially placed. Furthermore, power saving techniques are deactivated since not every simulator provides such techniques. Additionally, the linear energy models of the respective simulators are used. Two cooling units are also part of the scenario in order to manage the temperature of the data center.

4.2. Results

The scenario described in section 4.1 has been implemented in all simulators that had been analyzed before. As already stated in the analysis conducted in section 3, not every detail of the experimental scenario can be modeled with each simulator. For example, it was defined in the previous section that a linear model is used for determining the energy consumption. For instance, such a linear model for the consumption of servers is included in *CloudSim* and in *GreenCloud*. Although both models are based on the utilization of the CPU, they are not identical (compare (Beloglazov and Buyya 2012) and (Kliazovich, Bouvry and Khan 2012)). While *CloudSim* and *GreenCloud* (as well as *CReST*) at least provide an energy model, this is completely missing in *iCanCloud*.

Another example is that, except for *CReST*, none of the simulators is able to model the cooling units (confer section 3). The implementations of the experiments are therefore not completely identical. The simulation results in terms of throughput or response times would probably only be partially comparable – for example since *CloudSim* cannot correctly simulate different topologies, which will lead to errors in the results regarding response times. However, these aspects are not critical for comparing the energy consumption determined by each simulator. Therefore, the similarity of the various implementations of the experiment is sufficient. Indeed, components that maybe cannot be modeled with a simulator, such as specific types of switches, are especially important to the outcome of the experiment. However, profound adjustments of the respective simulation tools would be necessary for other purposes than the comparison of differences in energy consumption in order to get to comparable results.

The simulation results in terms of the energy consumption determined with each simulator are depicted in figure 3. It shows the mean energy consumption determined by each simulator for a time frame of 24 hours.

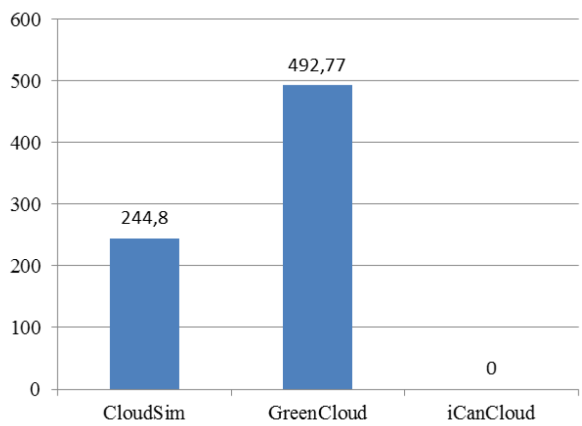


Figure 3: Mean Energy Consumption for each Simulator in kWh

Figure 3 does not contain any results regarding *CReST*. Although the experiment could be modeled in great detail with *CReST*, a program error prevents the experiment from being successfully completed. The same error also occurs in the sample scenarios provide by in *CReST*, which is why it can be foreclosed that the experiment is implemented incorrectly. Therefore, results are missing for this simulator.

As expected, *iCanCloud* is not able to determine the energy consumption of the scenario defined in the previous section. Although many components of the scenario can be modeled, energy models are not included in *iCanCloud* as already stated above. Therefore, the amount of energy consumed by the single components cannot be determined, which is why the energy consumption is 0 as shown in figure 3.

Significant consumption values could thus only be determined for *CloudSim* and *GreenCloud*. As shown in figure 3, the values differ greatly from one another. This is mainly due to the fact that *GreenCloud* simulates the energy consumption of the network devices as well. Since the consumption of both simulators regarding the servers is very similar (244,8 kWh in *CloudSim* and 256,9 kWh in *GreenCloud*), the influence of the network components – although depending on the particular network architecture and topology (Bilal et al. 2013) – is high. As an example, three-tier architectures usually consume plenty of energy. The differences in the determined consumption of the servers are the result of the functioning of the simulators. For example, there are minor differences in the workload profiles: *GreenCloud* reaches a mean utilization of 50% over all servers, but the utilization can be higher or lower for single servers. Since in all simulators, the energy consumption of servers is based on the utilization of the CPU, the determined energy consumption is not exactly the same due to differing utilization factors, although the consumption values are very similar. Since the confidence intervals of both results do not overlap, the differences in the results are significant. In fact, the results are deterministic. Thus, a significant difference exists (confer figure 3).

The assumption made at the beginning of this section that the amount components and concepts taken into account as presented in section 3 have a significant impact on the simulated energy consumption can thus be confirmed.

5. DISCUSSION

The analysis conducted in this paper has revealed some differences between the analyzed simulators.

There is a certain homogeneity between the simulators in terms of which components are included in the servers. Referring for example to the energy-consuming components defined in (Fan, Barroso and Weber 2007), the amount of considered components is not sufficient for providing an accurate prediction of the energy consumption and is hence also not sufficient for determining the effects of load distribution algorithms on energy consumption. The experiment carried out in section 4 supports this assumption, although only a simple scenario was modeled and only two simulators were able to provide results. Considering the proportioning of the energy consuming components of a data center as illustrated in figure 4, it is likely to expect that the results of a more comprehensive simulator will differ even more from the analyzed simulation tools since these do not even completely include the IT equipment (confer the analysis of physical components in section 3.3), not to speak of other consumers such as cooling.

The fact that many components are not considered by the respective simulators is thereby not surprising since appropriate energy models for the single components are often hard to find (Kansal et al. 2010). As the energy models of servers provided by the analyzed simulators are limited on the CPU, the simulators are also not applicable for all purposes.

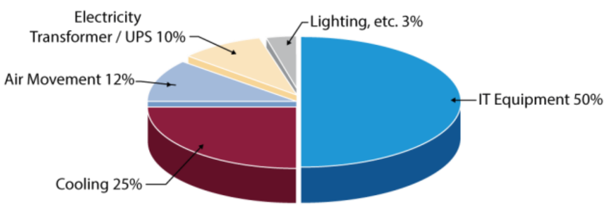


Figure 4: Sources of Energy Consumption in Data Centers (Source: (Power 2007))

For example, if storage systems are investigated, these will generate a lot more load on the disks than the on the CPU (Orgerie, Lefevre and Gelas 2010). Due to the higher load, the energy consumption of the disks would rise, but that cannot be determined by the current energy models. Another type of load would not even require a storage system; also an I/O-intensive workload would possibly increase the load of the disks and therefore their energy consumption. Especially memory is to become a dominant factor in the energy consumption of servers (Minas and Ellison 2009). Regarding an exact prediction of the energy consumption, such aspects must be taken into account.

Surprisingly, cooling is almost ignored by existing simulators, which is astonishing considering their importance for data centers and their energy consumption – cooling accounts for about 25% of the energy consumption of data centers (Orgerie, De Assuncao and Lefevre 2014; Power 2007). Only *CReST* implements a module for cooling, which however does not determine any energy consumption since only a thermal model is implemented. However, the integration of an energy model for cooling has already been prepared in the source code.

Just like cooling, also storage systems and support systems are hardly considered by the simulators. However, these are crucial both for the operation of the data center as well as for its energy consumption (confer figure 3). Especially regarding the physical components, there is a lot of potential for improvements on the side of the simulators in terms of their accuracy of predicting the energy consumption of data centers.

There is also potential for improvement in terms of the non-physical components of simulators. For example, there is a lot of backlog regarding the criterion “Virtualization / Cloud”, especially in terms of multi-tenancy aspects (confer section 3.4). Furthermore, the consideration of software within the simulators needs to be improved. Especially with regard to non-functional properties such as availability, it is important to be able to model dependencies between applications. As the analysis revealed, this aspect is excluded by most simulators. Particularly aspects that are relevant for quality of service, such as the availability of a service, are ignored by most simulators.

It is also worth mentioning that none of the analyzed simulators was actually evaluated on the basis of a complete data center. Due to the fact that relevant data for this purpose is hard to get, this is not surprising. However, an evaluation that captures a complete data center is important in order to provide evidence for the validity of the simulation tool. Such an evaluation would be a significant asset with regard to the proof of the suitability of a simulator for determining the energy consumption of data centers. Furthermore, this can help when it comes to determining the effects of load distribution on other components of a data center, such as effects on response time or availability.

Regarding the simulation of clouds or cloud data centers, several issues are still unsolved and should be addressed in future work.

6. CONCLUSION

In recent years, cloud computing has become an important paradigm for providing IT services. Besides the advantages for cloud service providers, cloud computing also comes along with several challenges, among which reducing the overall energy consumption is one of the most important ones. A common method for reducing energy consumption is using load distribution in terms of VM scheduling (Zhang, Cheng and Boutaba 2010). In the past years, numerous algorithms for energy efficient load distribution in clouds have been proposed. Such algorithms are often evaluated by using simulation approaches (Sakellari and Loukas 2013). In order to investigate the applicability of existing simulators for determining the energy consumption in clouds and the effects of load distribution algorithms on energy consumption, an analysis of existing simulators was conducted in this contribution.

The four simulators analyzed in this paper have similar objectives, but diverse designs and implementations. Regarding their suitability for determining the energy consumption of cloud data centers, the simulators have several limitations when it comes to modeling the relevant energy consuming components and also non-physical concepts that are relevant in the context of cloud computing, such as resource pooling (Mell and Grance 2011) or availability. The additionally performed experiment confirms the assumption that the consideration of other components of a data center can strongly affect the outcome of the simulation in terms of energy consumption.

In order to get to more accurate results in terms of energy consumption, a more comprehensive simulator needs to be designed that addresses the shortcomings identified in the course of the analysis.

REFERENCES

Ahmed, A. and Sabyasachi, A.S., 2014. Cloud computing simulators: A detailed survey and future direction. *2014 IEEE International Advance Computing Conference (IACC)*, 866–872.

Barroso, L.A. and Hölzle, U., 2009. The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis Lectures on Computer Architecture*, 4 (1), 1–108.

Beloglazov, A., Abawajy, J.H. and Buyya, R., 2012. Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. *Future Generation Computer Systems*, 28 (5), 755–768.

Beloglazov, A. and Buyya, R., 2012. Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers. *Concurrency and Computation: Practice and Experience*, 24 (13), 1397–1420.

Bilal, K., Khan, S.U., Malik, S.U.R., Khalid, O., Hameed, A., Alvarez, E., Wijaysekara, V., Irfan, R., Shrestha, S., Dwivedy, D., Ali, M., Khan, U.S., Abbas, A. and Jalil, N., 2013. A taxonomy and survey on Green Data Center Networks. *Future Generation Computer Systems*.

Brown, R., Koomey, J., Masanet, E., Nordman, B., Tschudi, B., Shehabi, A., Stanley, J., Sartor, D., Chan, P., Loper, J., Capana, S., Hedman, B., Duff, R., Haines, E., Sass, D. and Fanara, A., 2007. *Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431*. United States Environmental Protection Agency.

Buyya, R., Ranjan, R. and Calheiros, R.N., 2009. Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities. *International Conference on High Performance Computing & Simulation (HPCS ’09)*, 1–11, IEEE.

Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A.F. and Buyya, R., 2011. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience*, 41 (1), 23–50.

Cartlidge, J. and Cliff, D., 2013. Comparison of Cloud Middleware Protocols and Subscription Network Topologies using CReST, the Cloud Research Simulation Toolkit. *Third International Conference* *on Cloud Computing and Services Science (CLOSER '13)*, 58–68.

Fan, X., Barroso, L.A. and Weber, W.-D., 2007. Power Provisioning for a Warehouse-sized Computer. *ACM International Symposium on Computer Architecture*, 13–23.

Ge, C., Sun, Z. and Wang, N., 2013. A Survey of Power-Saving Techniques on Data Centers and Content Delivery Networks. *IEEE Communications Surveys & Tutorials*, 15 (3), 1334–1354.

Greenberg, A., Hamilton, J., Maltz, D.A. and Patel, P., 2008. The cost of a cloud: research problems in data center networks. *ACM SIGCOMM Computer Communication Review*, 39 (1), 68–73.

Jing, S.-Y., Ali, S., She, K. and Zhong, Y., 2013. State-of-the-art research study for green cloud computing. *The Journal of Supercomputing*, 65 (1), 445–468.

Kansal, A., Zhao, F., Liu, J., Kothari, N. and Bhattacharya, A.A., 2010. Virtual Machine Power Metering and Provisioning. *ACM Symposium on Cloud Computing*, 39–50, New York, New York, USA: ACM.

Kliazovich, D., Bouvry, P. and Khan, S.U., 2012. GreenCloud: a packet-level simulator of energy-aware cloud computing data centers. *The Journal of Supercomputing*, 62 (3), 1263–1283.

Koomey, J., 2011. *Growth in Data center electricity use 2005 to 2010*. Oakland, CA: Analytics Press.

Kumar, P. and Rai, A.K., 2014. An Overview and Survey of Various Cloud Simulation Tools. *Journal of Global Research in Computer Science*, 5 (1), 24–26.

Li, Z., O'Brien, L., Zhang, H. and Cai, R., 2012. On a Catalogue of Metrics for Evaluating Commercial Cloud Services. *ACM/IEEE 13th International Conference on Grid Computing (GRID)*, 164–173, IEEE.

Liu, H., Jin, H., Xu, C.-Z. and Liao, X., 2013. Performance and energy modeling for live migration of virtual machines. *Cluster Computing*, 16 (2), 249–264.

Mell, P. and Grance, T., 2011. *The NIST Definition of Cloud Computing*. Gaithersburg, MD, USA: National Institute of Standards and Technology (NIST).

Minas, L. and Ellison, B., 2009. *Energy Efficiency for Information Technology: How to Reduce Power Consumption in Servers and Data Centers*. Intel Press.

Mishra, M., Das, A., Kulkarni, P. and Sahoo, A., 2012. Dynamic Resource Management Using Virtual Machine Migrations. *IEEE Communications Magazine*, 50 (9), 34–40.

Nehru, E.I., Venkatalakshmi, B., Balakrishnan, R. and Nithya, R., 2013. Neural load prediction technique for power optimization in cloud management system. *2013 IEEE Conference on Information & Communication Technologies (ICT '13)*, 541–544, IEEE.

Núñez, A., Vázquez-Poletti, J.L., Caminero, A.C., Castañé, G.G., Carretero, J. and Llorente, I.M., 2012. iCanCloud: A flexible and scalable cloud infrastructure simulator. *Journal of Grid Computing*, 10 (1), 185–209.

Orgerie, A.-C., De Assuncao, M.D. and Lefevre, L., 2014. A Survey on Techniques for Improving the Energy Efficiency of Large Scale Distributed Systems. *ACM Computing Surveys*, 46 (4), 1–35.

Orgerie, A.-C., Lefevre, L. and Gelas, J.-P., 2010. Demystifying energy consumption in Grids and Clouds. *Green Computing Conference*, 335–342, IEEE.

Pelley, S., Meisner, D., Wenisch, T.F. and VanGilder, J.W., 2009. Understanding and abstracting total data center power. *Workshop on Energy-Efficient Design*,

Power, E.N., 2007. *Five Strategies for Cutting Data Center: Energy Costs Through Enhanced Cooling Efficiency*. Emerson Network Power.

Sakellari, G. and Loukas, G., 2013. A survey of mathematical models, simulation approaches and testbeds used for research in cloud computing. *Simulation Modelling Practice and Theory*, 39, 92–103.

Smit, M. and Stroulia, E., 2013. Simulating Service-Oriented Systems: A Survey and the Services-Aware Simulation Framework. *IEEE Transactions on Services Computing*, 6 (4), 443–456.

Velho, P., Schnorr, L.M., Casanova, H. and Legrand, A., 2013. On the validity of flow-level tcp network models for grid and cloud simulations. *ACM Transactions on Modeling and Computer Simulation*, 23 (4), 23.

Zhang, Q., Cheng, L. and Boutaba, R., 2010. Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1 (1), 7–18.

Zhao, W., Peng, Y., Xie, F. and Dai, Z., 2012. Modeling and simulation of cloud computing: A review. *IEEE Asia Pacific Cloud Computing Congress*, 20–24, IEEE.

CFD SIMULATION OF A RETORT PROCESS FOR FOOD VEGETABLE PRODUCTS

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

This work aims at simulating a retort sterilization process of flexible packaging, performed in "autoclave". The retort sterilization is one of the most adopted processes in order to reduce the microbial load of packaged food.

Ansys CFD© software has been used in this study in order to simulate the heat transfer inside the retort room. Flexible packaging filled with a particular vegetable soup have been modelled. 3D geometrical configuration of the equipment has been also modelled in order to evaluate the sterilization level reached by each flexible packaging inside the retort room. In particular, the internal temperature of the most disadvantaged simulated product has been monitored during 5 minutes of the process in order to control the time/temperature trend.

Finally the model has been validated by comparing the simulation results with those obtained by experimental tests.

Keywords: *CFD; Food sterilization; Retort packaging; Heat treatment; autoclave, vegetable soup*

1. INTRODUCTION

Nowadays retort treatment using pressured saturated steam is used especially for food vegetables or soups, which are hardly sterilized by means of continuous processes.

The water is boiled at high pressure in order to create saturated steam, which must not condense on packaging surfaces of the treated products also at temperature higher than 120°C. The heat transmitted by the steam has to pass through the packaging in order to reach the inner part of the product. Moreover, air has to be removed from the sterilization chamber, avoiding formation of colder zones where the temperature may remain below the sterilization setting. At the end of the process it is necessary to dry the product to minimize the condensate content. A high presence of condensate steam may in fact damage the package and it may promote the growth of bacterial substrates. Metallic containers, jars, or some special carton beverages, such as Tetra Recart® or Combisafe®, could be adopted in this process.

Two types of heat transfer occur during this process: convection, and conduction. The convection rate can be considerably increased by inducing a forced convection by means of mechanical agitation of the trays. For this reason many retort rooms are designed to provide axial rotations or longitudinal movements of the product (Dwivedi and Ramaswamy 2010; Ramaswamy and Dwivedi, 2011). Thanks to these movements, it is possible to reduce the treatment time and to obtain higher quality products.

Literature widely explored the issue of retort process for vegetables and other food products (Durance, 1997; Teixera and Tucker, 1997), with the aim to understand the better setting of this technology (Simpson et al., 2007). Some authors tried to approach the problem experimentally (Dwivedi and Ramaswamy, 2010 b), but this approach requires an high cost of equipment and the impossibility to forecast the process behaviour, if something change. In recent years the development of the numerical simulation helped to overpass this problem and to identify the best thermal setting of the process (Miri et al., 2008). Among the different numerical approaches, Computational Fluid Dynamics CFD appeared as the most suitable in order to understand the evolution of the temperature inside the product during a retort process (Abdul Ghani et al, 2001; Abdul Ghani et al, 2003; Kızıltaş et al., 2010; Augusto and Cristianini, 2011). All of these letter studies simulated a 3D process considering the real configuration of the packaging, but none of them aims at considering the flow of vapour reaching the packaging inside a complex geometry like those of an "autoclave". Similar Studies have been performed in other food sector, like pasta (Bottani et al., 2013; Armenzoni et al, 2013) or refrigerated rooms (Ambaw et al., 2014), showing how is important the knowledge of the air/vapour flows for food processes.

Aim of this article is then to analyse the retort process considering a specific section of an Autoclave with a simultaneous treatment of 320 food packages. The product tested in this work is a pumpkin soup (a homogeneous product).

2. MATERIALS AND METHODS

2.1. Equipment

The handling system for the bricks loading and unloading interacts with the retort sterilizer. Using a coupling system, the baskets are inserted or extracted in a few minutes from the retort.

The retort system considered in this work (Figure 1) can contain until 6 baskets each composed from sixteen trays which contain 70 cartons per tray. This system is able to sterilize up to 6720 brick of 500 ml or 8640 containers of 390 ml per cycle.

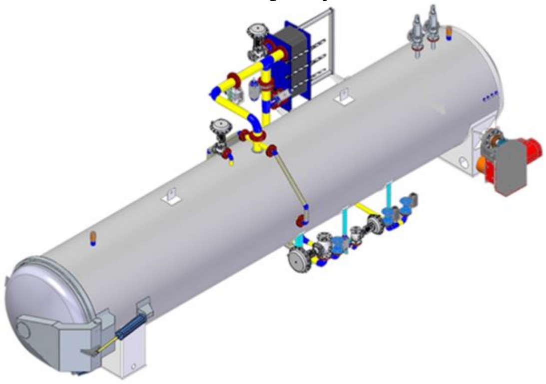


Figure 1: retort system considered in this study

The thermal cycle of the pumpkin soup lasts approximately 2 hours and it is composed by the following steps:

- Preheating phase: the internal temperature of the sterilizer rises up to 50°C at 2 bar;
- Heating phase: the internal temperature of the sterilizer rises up 110°C with an increased pressure up to 4 bars;
- Sterilization phase: inside the sterilizer the temperature reaches 130°C and the vapour is inserted with a pressure of about 4 bars; the product is handled in these conditions for a defined sterilization time;
- Pre-cooling phase: the sterilizer temperature drops to 105°C with a decreased pressure to 2 bar;
- Cooling phase: the whole system achieves an ambient temperature and pressure close to 1.5 bar;
- Discharge phase: when the system reaches the atmospheric pressure, the retort system can be opened to discharge the products.

At the end of the sterilization, all the cartons are discharged on a conveyor belts system and conducted to the secondary packaging operation.

From a technological point of view, the most critical phase is the sterilization one, especially in the 5 minute after the reaching of 85°C inside the product (usually after 1200s of treatment). This is the time range in which the simulations were performed (1200s-1500s).

2.2. Materials

The product considered in this study is a pumpkin soup, packaged in TetraRecart brick having a volume of 500ml. This packaging allows to be treated with a retort technology, having a particular heat resistance. The product is filled by means of a piston filler, which ensures a good accuracy and a limited head space (on average 13ml).

In order to perform CFD simulations, two main materials were considered: saturated steam water and pumpkin soup. Properties of the first one are well known and can be easily retrieved from the software library. Conversely, for the food product, it was necessary to find (or compute) the correct values for each physical property.

In particular, for the pumpkin soup, the following properties were considered: density, dynamic viscosity, thermal conductivity and heat capacity. The values of these properties were provided by a supporting company which, during experimental tests, has measured the corresponding data.

In order to perform the simulation, another material has been considered. In fact the thermal exchange became between vapour and the food liquid through a thin food package multilayer. In this type of packaging, paper is the material with greatest impact on heat transfer. For this reason only the paper layer was considered. All the features of each material are reported in Table 1.

Table 1: features of steam water, pumpkin soup and packaging paper

Saturated steam water		
Thermodynamic state	Mixture liquid/gas in equilibrium	
Molar Mass	18.02	kg/mol
Density	1.91	kg/m3
Reference pressure	4	bar
Reference Temperature	130	°C
Heat Capacity	1901	J/kgK
Thermal Conductivity	0.016	W/mK
Dynamic Viscosity	0.228	cp

Pumpkin soup		
Thermodynamic state	Liquid	
Density	1030	kg/m3
Heat Capacity	3350	J/kgK
Thermal Conductivity	0.48	W/mK
Dynamic Viscosity	0.085	Pa·s

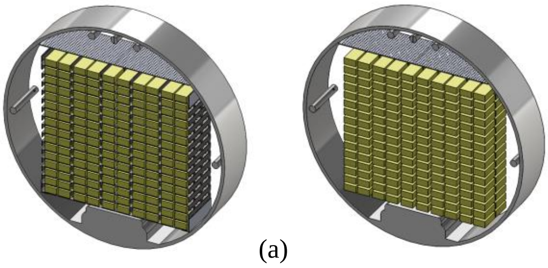
Food packaging paper		
Thermodynamic state	Solid	
Density	700	kg/m3
Heat Capacity	1321	J/kgK
Thermal Conductivity	0.21	W/mK

2.3. Mathematical modelling

2.3.1 Simulation setting: geometry and mesh

Given the large size of the retort sterilizer, the simulations were conducted considering a section of the whole system. Due to the complexity of the domain, it has been decided to approach the issue considering both the presence of the trays under the carton bricks, and a domain without trays (figure 2 a) and b)).

For each configuration, 320 bricks were loaded, considering however only 1 brick for the thermal analysis. As shown in Figure 3, the considered brick is located in the centre of the third layer starting from the bottom of the vessels.



(a) (b)

Figure 2: Configuration with (a) and without trays (b)

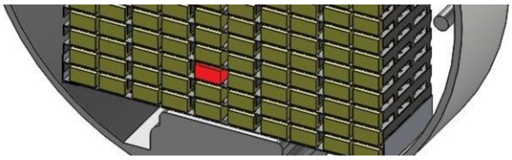


Figure 3. Focus on the analyzed brick

The fluid domain was obtained for both the two configurations using ICEM CFD, the modeller associated with ANSYS CFX. The volumes are divided into a finite number of cells, on which the analysis is carried out.

The meshes were created following a gradient that respects the Courant number. This number is of fundamental importance for transient flows. For a one-dimensional grid, it is defined as:

$$Courant = u\frac{\Delta t}{\Delta x}$$

where u is the fluid speed, Δt is the timestep and Δx is the mesh size. The Courant number calculated in Ansys CFX is a multidimensional generalization of this expression where the velocity and length scale are based on the mass flow into the control volume and the dimension of the control volume (Löhner, 1987). To allow a correct CFD simulation, the timestep must be chosen with the purpose of keep the Courant number sufficiently small.

The number of cells used in the simulations was determined starting from a coarse mesh, and gradually refined, evaluating the changes in the results. The mesh setting started from the definition of the external surface mesh (Figure 4) as suggested in Ansys solver modelling guide (Ansys, 2011). Table 2 reports the values of the surface mesh for each part.

Table 2: values of the surface mesh

Part	Size [mm]	Height [mm]	Height Ratio	Tetra Ratio
Wall	32	4	1.3	1.5
Open	16	4	1.3	1.5
Inlet	1	4	1.3	1.5
Brick	2	4	1.3	1.5
Trays	4	4	1.3	1.5
Pipe	4	4	1.3	1.5

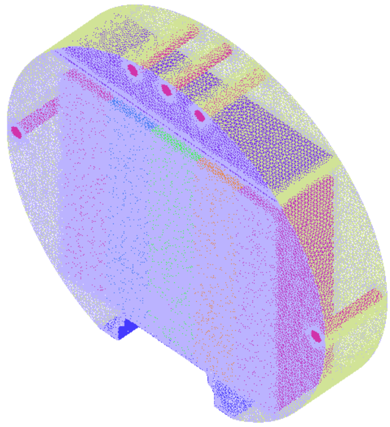
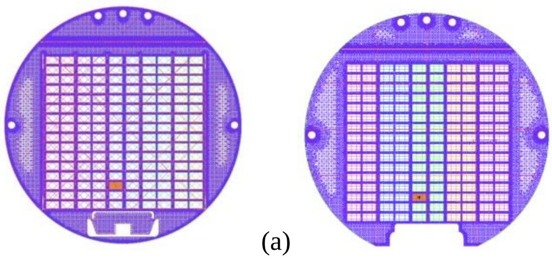


Figure 4. External mesh surface

The volume mesh was initially set by creating a uniform subdivision, and then thickened in the critical areas of the fluid volume. In particular, a finer mesh was used near the outlet section of the nozzle, where it can be expected that the shear rates would be higher and closer to the sterilizer walls in order to accurately simulate the flow boundary layer.

Figure 5 shows the generated mesh used for the calculation: an unstructured tetrahedral meshing scheme was used for each configuration.



(a) (b)

Figure 5: volume mesh of the configuration with trays (a) and without trays (b)

The final meshes were determined when the increase in quality of the mesh did not provide any significant improvements in the results. The overall number of cells created for the first mesh (Figure 5a) is about 12,500,000, while, for the second one (Figure 5b), the overall number of cells created is about 11,200,000.

2.3.2 Simulation setting: domain equations and boundary conditions

Three-dimensional, multiphase, two-fluid model simulations were developed to investigate the temperature trend inside the carton package. ANSYS CFX 14.5 software was used to solve the governing continuity, momentum and energy equations for the defined geometry and associated boundary conditions. The generalized transport equations solved are:

The continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho V) = 0 \quad (1)$$

The momentum equation:

$$\left(\frac{\partial \rho V}{\partial t} + \nabla \cdot (\rho V \otimes V)\right)$$
$$= \nabla \cdot (-p\partial + \eta \cdot (\nabla V + (\nabla V)^t)) + S_M \quad (2)$$

In this work, according to the materials used, three fluid domains were created: “autoclave”, “packaging paper” and “pumpkin soup”. ANSYS CFX uses the concept of domains to define the type, properties, and region of the fluid, porous, or solid. Domains are regions of space in which the equations of fluid flow or heat transfer are solved. This includes selecting the 3D bounding regions and choosing appropriate physical models for each domain. Packaging paper is created as an interface domain, between the “autoclave” and the “pumpkin soup” domains, to connect the two different domains with different properties and conditions. The interface between the surfaces is inserted as a solid layer. In this case a paper layer with thickness of 0.5 mm was created.

A “Thermal Energy” model is used to predict the temperature inside the pumpkin soup domain, i.e.:

$$\frac{\partial(\rho h_{tot})}{\partial t} + \nabla \cdot (\rho U h_{tot}) = \nabla \cdot (\lambda \nabla T) + \nabla \cdot (U \cdot \tau) \quad (3)$$

where h_{tot} is the total enthalpy, which can be expressed as a function of the static enthalpy h (T, p) as follows:

$$h_{tot} = h + \frac{1}{2} U^2 \quad (4)$$

The term $\nabla \cdot (U \cdot \tau)$ in eq. 3 is the work due to viscous stresses and it is known as the viscous work term. Inside the “autoclave” domain an isothermal heat transfer model was set. The isothermal model requires a uniform temperature for the fluid in absolute temperature terms. The temperature was fixed at 130°C. Being also the steam flowing in the “autoclave” domain at high velocity, a turbulent model has been adopted. One of the main problems in turbulence modelling is the accurate prediction of flow separation from a smooth surface. For this reason, the model adopted was the Shear Stress Transport (SST). The SST model, proposed by Menter (1994), is an eddy-viscosity model which is a combination of a k-ω model and k-ε model. The first is used in the inner boundary, while the second in the outer region and outside of the boundary layer. The SST model has been used in order to overcome the problems of both the methods. These features make the SST model more accurate and reliable for a wider class of flows than the standard k-ω and k-ε models. For the “pumpkin soup” domain the food product is instead in laminar flow.

The boundary conditions (Figure 6) are related to 10 inlet holes and 2 opening section; in particular, a uniform orthogonal velocity input and a relative pressure for outlet are set. The external wall was considered as adiabatic while the other brick not considered for the internal thermal process, were kept at 85°C. Table 3 resumes the boundary and initial conditions for the “autoclave” domain.

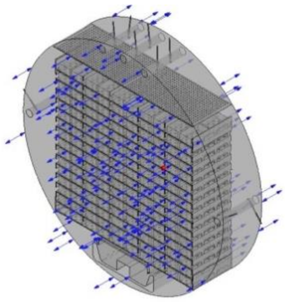


Figure 6. Boundary condition in CFX-Pre Solver

Table 3: Boundary and Initial conditions

Boundary conditions	
INLET	25 m/s
OPENING	3 relative bar
WALL	No slip wall and Adiabatic thermal condition
BRICK	No slip wall and initial temperature at 85°C

Initial conditions (t = 0)	
Pressure	3 relative bar
Velocity	0 m/s
“Autoclave” initial temperature	130°C
“packaging paper” initial temperature	85°C
“pumpkin” initial temperature	85°C

All the simulations were carried out in transition state, to evaluate the trend of the sterilization temperature inside the brick in function of the time. In particular for both configurations, a total time of 5 minutes with 0.5 seconds time-step were performed.

2.4. Experimental method

The tests have been performed to validate the simulations and to obtain a heat treatment able to ensure a healthy and safe product for the consumer. The tests have been developed according to the following phases:

1. Analysis of the product
2. Collecting samples
3. Positioning the temperature probes inside the package
4. Partially loading the retort
5. Controlling and analysing the samples by opening the cartons.

In the first phase, the chemical and physical parameters like the brix degree, pH, acidity, colour and texture were controlled. Then, some samples are collected to check the weight and the suitability of the welding performed by the machine. Then the thermocouples probes were inserted inside two containers and connected to a ELLAB data logger (Figure 7). The tests were carried out using temperatures probes located in two bricks previously considered as the most disadvantaged packages inside of the retort.

Figure 7. Thermocouples and ELLAB data logger

The most disadvantaged brick has been identified on the third tray in the center positions. After the probes connection, the retort sterilizer has been loaded with only a few hundred cartons, disposed in two sections as supposed in the simulation model.

At the end of the process, the carton analyzed were removed and opened to verify the real treatment obtained by the product.

3. SIMULATIONS RESULTS

The CFX-Post Solver was used to analyse velocities, temperatures and pressures fields inside the retort sterilizer and pumpkin soup domains. Different section planes were used to view the results of the process. Furthermore, a point located in the centre of the brick (the same point where the probe was put) was identified to monitor the temperature in the most disadvantaged zone of the whole system (Figure 8).

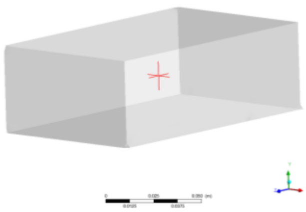
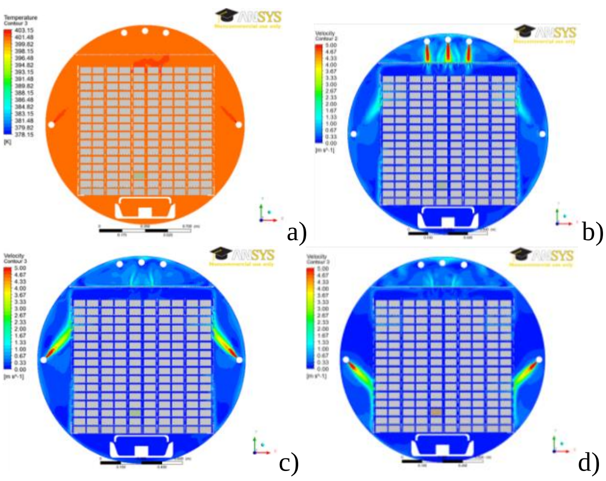


Figure 8. Point analyzed inside the pumpkin domain

For both simulations, the autoclave domain (Figure 9a) shows a constant temperature at 130°C. As regards the steam velocity, in the figures 9b), 9c) and 9d) the flows is in a range between 0-5 m/s.



a) b) c) d)

Figure 9: Temperature inside the sterilizer a); velocity inside the sterilizer at 3 planes: b) c) and d)

For each simulation, the temperatures inside the bricks were reported in separate paragraphs.

3.1. Simulation results of the configuration with trays

Analysing the temperature of the product in the configuration with trays, it was observed that the product heats up with the increase of the time. The figure below shows the temperature inside the brick during the sterilization phase at different time interval (between 0 to 5 minutes) in a range between 85 and 130°C.

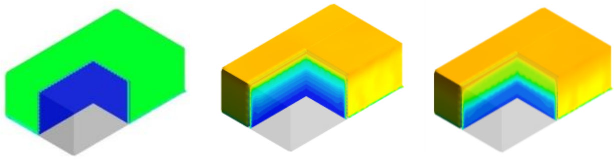


Figure 10. Temperature inside the brick at t=0 min, t=2,5 min, t=5 min

The temperatures were also analysed in function of the time in the previously defined point inside the product. Starting by 85°C, the heart of product reaches about 92.5°C after 5 minutes. This temperature trend is described in Figure 11

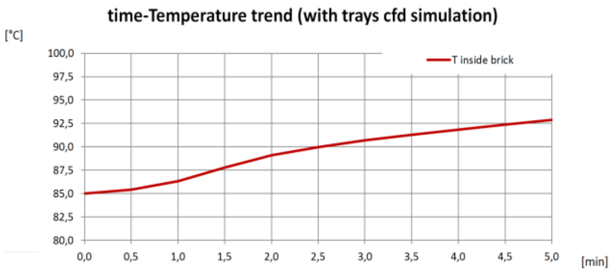


Figure 11 Time-Temperature trend in the previously defined point inside the brick

Thanks to Microsoft Excel elaborations, all the temperatures in the brick volume were obtained. With those data, we were able to see the temperature trend in the product volume and not only in a specific point.

The percentage of volume increase was reported in Table 4. Values equal to 100% mean that the entire volume is located at a temperature higher or equal to the reference temperature set.

Table 4% of volume with T greater or equal of Tref (at t = 2.5 min and t=5 min)

Tref (t = 2.5 min)

Tref	°C	K	%Volume
T0	85	358.15	100
T1	90	363.15	66.06
T2	95	368.15	54.20
T3	100	373.15	42.51

Tref (t = 5 min)

Tref	°C	K	%Volume
T0	85	358.15	100
T1	90	363.15	76.21
T2	95	368.15	64.37
T3	100	373.15	57.66

3.2. Simulation results of the configuration without trays

The same increasing of the temperature is shown by the food product with the configuration without trays. The figure below shows the temperature inside the brick during the sterilization phase at different time interval (between 0 to 5 minutes) in a range between 85 and 130°C. In the case without trays, the temperatures of the external layer are higher than before (Figure 12).

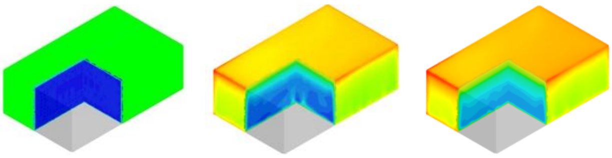


Figure 12. Temperature inside the brick at t=0 min, t=2,5 min, t=5 min

As before, the temperatures were analysed in function of the time in the previously defined point inside the product. Starting by 85°C, the heart of product reaches about 95°C after 5 minutes. Figure 13 describes this temperature trend.

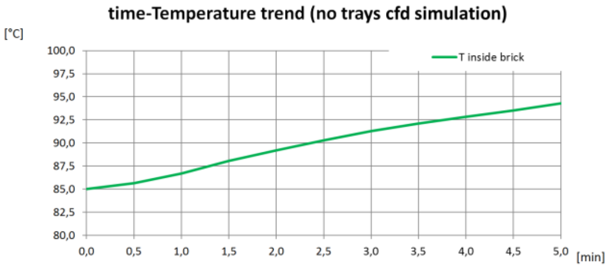


Figure 13 Time-Temperature trend in the local point inside the brick

As before thanks to Microsoft Excel elaborations, it was possible to calculate all the temperatures in the brick volume. As reported in Table 5, an increasing of percentage volume with higher temperature was observed respect the case without trays. Values equal to 100% mean that the entire volume is located at a temperature higher or equal to the reference temperature set.

Table 5 % of volume for T greater or equal in respect of Tref (at t = 2.5 min and at t = 5 min)

t = 2.5 min

Tref	°C	K	%Volume
T0	85	358.15	100
T1	90	363.15	79.76
T2	95	368.15	59.90
T3	100	373.15	46.31

t = 5 min

Tref	°C	K	%Volume
T0	85	358.15	100
T1	90	363.15	100
T2	95	368.15	78.66
T3	100	373.15	60.54

4. EXPERIMENTAL VALIDATION

4.1 Experimental tests

Following the prescription reported in section 2.4, an experimental test has been performed. Figure 14 reports the temperature trend in function of the time for the sterilizer room and inside the brick in the same point evaluated by the simulation during all the process.

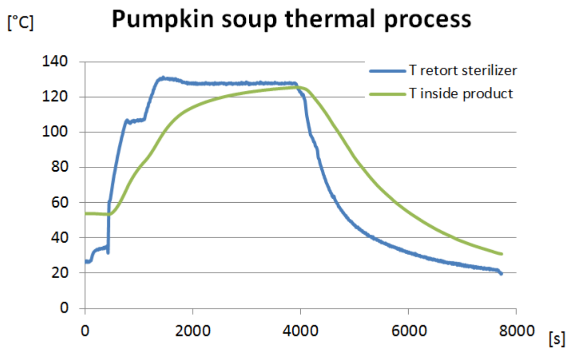


Figure 14: Temperature trend of the retort room and inside the brick

Figure 15 show instead the pumpkin soup and retort temperatures in the reference time (after 120s of treatment).

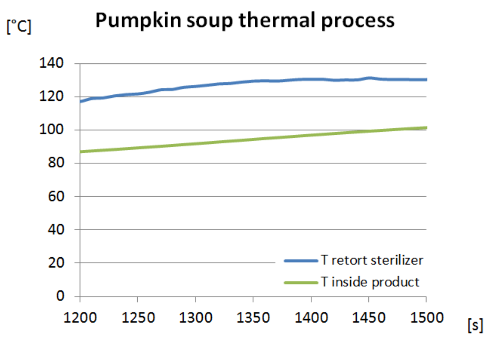


Figure 15 Temperature trend of the retort room and inside the brick in the reference time (1200-1500s)

During the reference time (1200-1500s), the temperature values in the retort sterilizer and inside the brick at step of 0.5min are reported in Table 6. The product temperature has been taken in the same point before analysed in the CFD simulation.

Table 6: experimental temperature inside the retort room and inside the brik (step of 0.5min)

Time [min]	T retort sterilizer [°C]	T inside brick [°C]
0,0	130,3	84,96
0,5	129,3	86,01
1,0	129,1	87,03
1,5	129,5	88,04
2,0	128,3	89,04
2,5	128,5	90,01
3,0	128,2	90,95
3,5	128,3	91,88
4,0	128,1	92,78
4,5	127,6	93,65
5,0	127,4	94,51

4.2 Comparison with simulation results

Table 7 and 8 compare the experimental values with them provided by the software simulations. The following values refer to the ones measured and calculated at the product core.

Table 7 Comparison between experimental values and those from the simulation with trays

Time [min]	Experimental Data [°C]	CFD simulations with trays [°C]	ΔError [°C]
0	84.96	84.96	0
1	87,03	86.33	0,70
2	89,04	89.08	-0,04
3	90,95	90.70	0,25
4	92,78	91.83	0,95
5	94.51	92.89	1,62

Table 8 Comparison between experimental values and those from the simulation without trays

Time [min]	Experimental Data [°C]	CFD simulations without trays [°C]	ΔError [°C]
0	84.96	84.96	0
1	87,03	86.71	0,32
2	89,04	89.17	-0,13
3	90,95	91.30	-0,35
4	92,78	92.81	-0,03
5	94.51	94.30	0.21

Figure 16 shows the comparison trend. From the figure, it is possible to observe a slight deviation between the experimental results and those calculated by the software with the configuration with trays. This could be due the absence of heat exchange between the trays and the bricks. Removing the trays the heat exchange happened on the whole package allowing a very good agreement between experimental and simulated data.

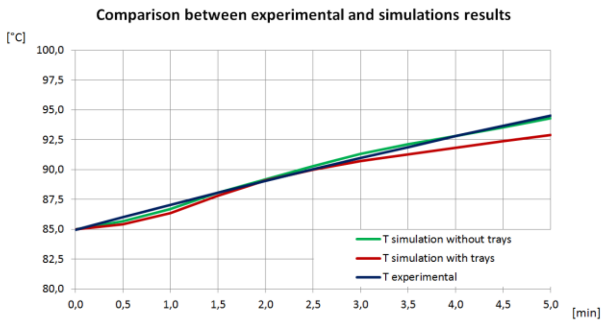


Figure 16. Comparison between the experimental test and simulation results

5 CONCLUSIONS

This work aimed to simulate a retort sterilization process of pumpkin soup packaged in flexible packaging. Ansys CFD© software has been used in this study in order to simulate the heat transfer inside the retort room. 3D geometrical configuration of the equipment has been modelled in order to evaluate the sterilization level reached by each flexible packaging inside the retort room. Until now any studies aims at considering the flow of vapour reaching the packaging inside a complex geometry like those of an “autoclave”. In this study, in particular, the internal temperature of the most disadvantaged point has been monitored during 5 minutes of the process in order to control the time/temperature trend. The simulations have been performed considering two configurations: one with trays and another without them.

The experimental validation has shown the better results obtained by the simulation without trays, having it the ability to better understand the behaviour of the heat exchange during the considered process.

Future research will be then address to better simulate the process considering also the heat exchange between the stainless steel trays and the paper packaging, in order to make the simulations results still more adherent to the real ones.

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, pp 147-156.

Abdul Ghani A.G., Farid M.M., Zarrouk S.J. 2003. The effect of can rotation on sterilization of liquid food using computational fluid dynamics. *Journal of Food Engineering 57*, pp 9-16.

Ambaw A., Verboven P., Delele M.A., Defraeye T., Tijskens E., Schenk A., Verlinden B.E., Opara U.L., Nicolai B.M. 2014. CFD-Based Analysis of 1-MCP Distribution in Commercial Cool Store Rooms: Porous Medium Model Application. *Food Bioprocess Technology* 7, pp 1903-1916.

ANSYS. CFX-solver modeling guide, release 14.0. ANSYS, Inc. Southpointe 275 Technology Drive, 2011.

Armenzoni M., Solari F., Marchini D., Montanari R., Bottani E., Vignali G., Ferretti G. 2013. Advanced design of the pasta drying process with simulation tools. *International Journal of Simulation and Process Modelling*, 8 (2-3), pp. 104-117

Augusto P.E.D., Cristianini M. 2011. Numerical evaluation of liquid food heat sterilization in a brick-shaped package. *Procedia Food Science* 1, pp 1290-1294.

Bottani, E., Ferretti, G., Manfredi, M., Vignali, G. 2013. Modeling and thermo-fluid dynamic simulation of a fresh pasta pasteurization process. *International Journal of Food Engineering* 9 (3), pp. 327-339.

Dwivedi M., Ramaswamy H. S. 2010. Comparison of heat transfer rates during thermal processing under end-over-end and axial modes of rotation. *Food Science and Technology* 43, pp 350-360.

Dwivedi M., Ramaswamy H. S. 2010b. An empirical methodology for evaluating the fluid to particle heat transfer coefficient in bi-axially rotating cans using liquid temperature data. *Food and Bioprocess Technology* 3 (5), pp. 716-731

Kiziltas S.,Erdogdu F., Palazoglu T. K. 2010. Simulation of heat transfer for solid–liquid food mixtures in cans and model validation under pasteurization conditions. *Journal of Food Engineering* 97, pp. 449-456.

Löhner, R., 1987. An adaptive finite element scheme for transient problems in CFD. *Computer Methods in Applied Mechanics and Engineering* 61 (3), pp. 323-338.

Miri T., Tsoukalas A., Bakalis S., Pistikoupoulos E.N., Rustem B., Fryer P.J. 2008. Global optimization of process conditions in batch thermal sterilization of food. *Journal of Food Engineering* 87, pp 485-494.

Ramaswamy H. S., Dwivedi M., 2011. Effect of Process Variables on Heat-Transfer Rates to Canned Particulate Newtonian Fluids During Free Bi-axial Rotary Processing. *Food Bioprocess Technology* 4 pp 61–78

Simpson R., Teixeira A., Almonacid S. 2007. Advances with intelligent on-line retort control and automation in thermal processing of canned foods. *Food Control* 18, pp 821-833.

Teixera A. A., Tucker G. S. 1997. On-line retort control in thermal sterilization of canned foods. *Food Control* (Vol. 8 No 1), pp 13-20.

Timothy D. Durance, 1997. Improving canned food quality with variable retort temperature processes. *Trends in Food Science & Technology* (Vol. 8), pp. 113-118.

AUTHORS BIOGRAPHY

Matteo FOLEZZANI is scholarship holder at the Interdepartmental Center CIPACK of the University of Parma. In 2011 he has achieved a master degree in Mechanical Engineering for the Food Industry. His main fields of research concern food process modelling and simulation, with a particular focus on the CFD simulation for the advanced design of food plants.

Michele Manfredi is scholarship at the Interdepartmental Center CIPACK, and a Ph.D student at the Department of Industrial Engineering of University of Parma. In 2011 he has achieved a master degree in Mechanical Engineering for the Food Industry. His main fields of research concern food process modelling and simulation, and LCA of food processes and packaging

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. 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 60 scientific papers, some of which appear both in national and international journals, as well in national and international conferences.

TEMPERATURE ANALYSIS OF THE WATER SUPPLY SYSTEM OF A DAIRY COMPANY BY MEANS OF A SIMULATION MODEL

Davide Marchini[(a)], Marta Rinaldi[(b)], Roberto Montanari[(c)], Eleonora Bottani[(d)] and Federico Solari[(e)]

[(a)] Interdepartmental Centre SITEIA.PARMA, c/o Department of Industrial Engineering, University of Parma, Parco Area delle Scienze, 181/A – 43124 Parma (ITALY)

[(b)] Interdepartmental Centre CIPACK, c/o Department of Industrial Engineering, University of Parma, Parco Area delle Scienze, 181/A – 43124 Parma (ITALY)

[(c), (d), (e)] Department of Industrial Engineering, University of Parma, Parco Area delle Scienze 181/A - 43124 Parma (ITALY)

[(a)]davide.marchini@unipr.it, [(b)]marta.rinaldi@unipr.it, [(c)]roberto.montanari@unipr.it, [(d)]eleonora.bottani@unipr.it, [(e)]federico.solari@unipr.it

ABSTRACT

This paper describes the second part of a research project that examined, by means of simulation, the water distribution system of a dairy company located near Parma (Italy) and active for decades in the production of *Parmigiano Reggiano* cheese and butter. The first part of the research was described by Marchini et al. (2013) and focused on finding opportunities for recycling water in the distribution system of the dairy company considered. In this work, we go ahead by analyzing the thermal properties (such as, primarily, the temperature) of the water used inside the distribution system, as well as of that discharged. Specifically, the aim of the present study is to estimate the water temperature into each tank of the plant. To achieve this aim, we build upon the MS Excel™ simulation model developed by Marchini et al. (2013) and add the related thermal equations. As a result, we define the temperature trend of water inside all the tanks of the distribution system. Situations where the water temperature is higher than the boundary defined by the company (i.e., approx. 25°C) are also highlighted.

Keywords: water distribution system; thermal analysis; time-temperature trends; dairy industry; simulation.

1. INTRODUCTION

Nowadays, the world faces a wide range of ecological and human health crises related to inadequate access to, or inappropriate management of, water. The amount of water consumption in industrialized countries has almost doubled in the last two decades, making the rational use of water resources is a key issue for sustainable growth (Gleick, 1998).

Water consumption faces significant competing forces for change, which include decreasing water resource availability, developing stricter water quality regulations, decreasing federal subsidies, increasing public scrutiny, decreasing financial health, and increasing infrastructure replacement costs (Rogers and Louis, 2009). Consequently, the reuse of water has become an important issue within the process industry.

Water in food industries is primary used for two basic purposes: to be incorporated to specific products or sub-product (i.e., as a raw material) or to eliminate certain undesirable components, e.g. during cleaning operations (Poretti, 1990; Almatò et al., 1999). Recycling the amount of water used for cleaning operations can be an opportunity to combine a reduction in the costs of industrial water with an improved control of water management and a better environmental impact on natural resources (Centi and Perathoner, 1999). Indeed, the cleaning operations, besides the economic impact, have environmental impacts. In those processes, the water requirements and wastewater generation are closely tied to the sequence and schedule of the different production tasks. Discontinuous operations of the equipment generate frequent cleaning and preparation tasks, especially when product changeover takes place. Those operations are often carried out using water of different qualities at different temperatures and flows; the wastewater generated in these tasks can represent a considerable part of the total wastewater originated in a food plant (Almatò et al., 1997).

For 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). Some authors have developed solutions for water savings, taking into account both the contamination level and the thermal properties (i.e., primary, the temperature) of the water discharged. The main purpose of those solutions is to generate savings, thanks to lower procurement and energy cost, by decreasing the amount of freshwater required, as well as to the lower cost for the treatment and discharge of waste water. Almatò et al. (1999) developed a methodology for the optimization of water use in the batch process industries. Al-Redhwan et al. (2005) address the problem of uncertainty in optimizing wastewater networks in the process industries. They start from the assumption that waste water flow rates and levels of contaminants can vary widely as a result of changes in operational conditions and/or feedstock and product specifications; therefore, an optimal water network design should be robust against those uncertainties. De Faria et al. (2009) address a similar topic, i.e. the optimization of water flows in industrial processes, by means of a non-linear programming model, whose objective function is to minimize the freshwater consumption. Kim (2012) introduces a system-wide analysis of water networks, grounded on the graphical design method, to help consider waste water charges in the water network design and making decisions about water regeneration and recycling. Looking at the food industry, some further contributions can be found, addressing both the general problem of water consumption in that industry (Casani and Knochel, 2002; Casani et al., 2005), as well as examining specific industrial segments, e.g. the sugar industry (Bogliolo et al., 1996) or the dairy one (Marchini et al., 2013).

This paper describes the second part of a research project, addressing the problem of optimizing the water consumption in the water supply system of a dairy company, whose main products are *Parmigiano Reggiano* cheese and butter. The first part of the research was described by Marchini et al. (2013) and focused on modelling, by means of a simulation tool, the water flows of the plant, so as to gain a precise knowledge of that system. Once the simulation model was validated, it was used to identify opportunities for recycling water in the distribution system of the company considered. In this work, we go ahead by analyzing the thermal properties (such as, primarily, the temperature) of the water used inside the distribution system, as well as of that discharged. Specifically, the aim of the present study is to estimate the water temperature into each tank of the plant. To achieve this aim, we build upon the MS Excel^{TM} simulation model developed by Marchini et al. (2013) and add the related thermal equations.

The remainder of the paper is organized as follows. In the next section we provide an overview of the water distribution system examined in this study. Section 3 describes the model developed to assess the thermal properties of water inside the distribution system. The model is validated in Section 4. Section 5 provides the results of the simulation runs, in terms of the time-temperature curves of water at different points in the distribution system. Countermeasures for critical situations, i.e. those situations where the temperature turned out to be excessively high, are also discussed. Section 6 summarizes the contributions of this study, the related limitations and implications, and describes future research directions.

2. WATER DISTRIBUTION SYSTEM AND PROBLEM STATEMENT

2.1. Overview of the water distribution system

As mentioned, the research described in this paper targets 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* producers of that area, with a production volume of about 60 wheels of cheese and more than 10 tons of butter per day.

The factory is equipped with two independent water circuits, one for the cheese and the other one for the butter production line. In the study by Marchini et al. (2013), the initial (AS IS) operating conditions of the plant lines were examined, through simulation, and their performance (in terms of water consumption) evaluated. On the basis of the findings from the AS IS analysis, the authors developed two reengineered (TO BE) scenarios for the plant, with the purpose of decreasing the overall consumption of water in the plant. One TO BE configuration, in particular, offered interesting savings in terms of water consumption. In that configuration, the authors proposed to increment the storage capacity of the central tank of the plant, called TK901, by adding a new tank, named TK902, working in a complementary manner to the first one.

The introduction of the new tank would avoid wasting a significant amount of water, by 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, the tank would work as an inter-operational buffer, with the purpose to decouple water recovery and water withdrawals from the various plant lines. The optimal capacity of the new tank was estimated to be approx. 30 m^3. By introducing the new tank, the factory could save a total of 211.58 m^3 of water over two weeks (14 days) of functioning. If compared to the daily water consumption of the plant, this saving appears as particularly relevant, accounting for approx. 7.2% of the total amount of water used in the plant. In this work, the TO BE configuration of the water supply system described above (as it was derived from the previous study) is taken as the reference scenario. A scheme of that configuration is reported in Figure 1. The reader is referred to Marchini et al. (2013) for further details.

2.2. Problem statement

In this study, the reference scenario described in the previous section is subject to further analyses, with the purpose to evaluate the water temperature at different points in the water supply system and its trend in time. Additional aims of the analyses carried out include:

- The evaluation of the amount of freshwater and recycled water used in the plant;

- The identification of those points in the plant where the water temperature is excessively high and should be lowered.

With respect to the aims listed above, it should be mentioned that each tank if the plant can be filled:

- by means of freshwater, picked from the wells, depicted in the left side of Figure 1. Usually, the temperature of that water is approx. 16°C;
- by water flows arriving from other tanks in the plant, with the related temperature;
- by recycled process water. That water has been already used in the plant for some processes, and discharged, but is free of microorganisms and can be reused in subsequent processes. Compared to the remaining flows, recycled water is, usually, at high temperature.

According to the indications of the dairy company investigated, the water temperature in the plant should not exceed 25°C, because of safety reasons as well as to limit energy consumption. This point is particularly relevant for cleaning in place (CIP) operations. Therefore, the analysis described in this paper is particularly focused on those tanks that provide water to the CIP process, i.e. TK601 and TK605 (cf. Figure 1); indeed, these tanks should provide cool water, used to wash the plant components, with the purpose of avoiding the growth of microorganisms.

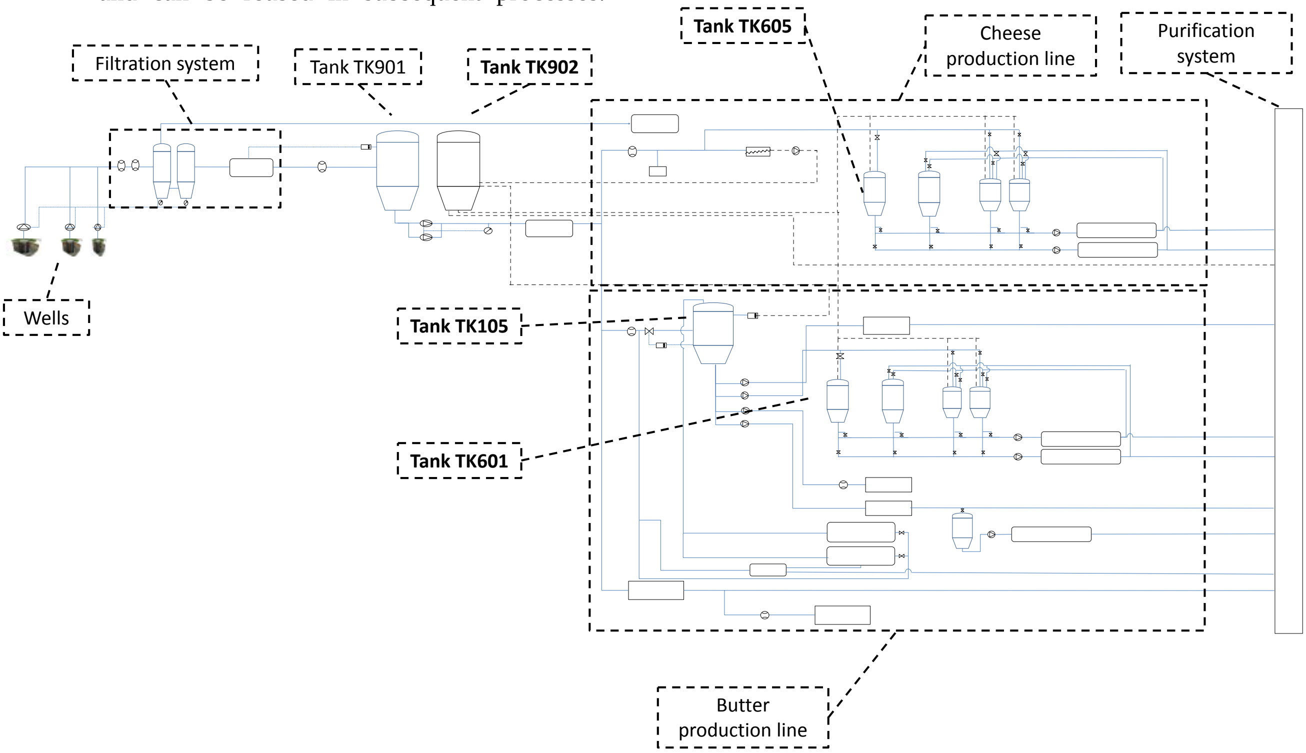


Figure 1: Reference scenario of this paper: the water distribution system of the factory after the introduction of the tank TK902 (source: Marchini et al., 2013).

3. THE THERMAL MODEL

3.1. Data input and main assumptions

As mentioned, in the previous study, the overall water flow and consumption of the dairy company were reproduced using a simulation model, developed under MS Excel™. More precisely, discrete-event simulation has been used to reproduce the flows of the water supply system.

Starting from this model, a new simulator is developed in this work to evaluate the water temperature inside all the tanks of the plants, even though this study is particularly focused on two specific tanks.

As it can be easily realized, the input data of the thermal simulation model are the output data of the previous one. In particular, for each tank, the thermal model needs all the flows of water which entered it; examples of those flows can be the water flow from the well (usually at 16°C), the flow arriving from other tanks, or other flows from other processes (with the related temperature). Moreover, for each tank, the model takes into account all the water flows exiting it, i.e. the water required by the processes and for the CIP operations. All the values related to the water flows obtained during the first study were recorded in a database, which details the final results obtained from the simulation at each time step.

For the sake of clarity, we describe, as an example, the water flows in the main tank of the butter production line (TK105). As inputs of this tank, we consider the water flows from TK901, and the reusable water coming from the pasteurizer and the whey cooler. With respect to the outputs, TK105 provides water to several plant processes and to other tanks, including, e.g.: (1) sanitary water, (2) tanks of the butter CIP line, (3) process mixture, (4) faucets, (5) tank used from CIP 3 and (6) the discharge of water by TK105 to TK902 when it reaches its maximum capacity.

As a further input to be included in this analysis, the temperature of each flow is required. The temperature from both the well and all the processes involved was directly provided by the plant's manager of the company, and is shown in Table 1.

Process	Temperature [°C]
Well	16
Pasteurizer	40
Whey cooler	25
Fermenters	25

Table 1: Water temperature at different point in the plant (source: data provided from the company)

Obviously, the model starts to calculate the temperature in those tanks which do not receive contributions from other tanks. Operating in this way, the temperature of water flows arriving from other tanks will be calculated by the model.

Starting from those data, the thermal model is able to derive the water temperature into each tank of the plant, at different time steps. It should be mentioned that the thermal model was developed under the main assumption of "whole and immediate miscibility" of the fluids in the water distribution system. Considering a tank containing a given amount of water, such assumption means that any water flow entering this tank would generate a new water amount (obtained as the sum of the two original flows) with a new (average) temperature in any point of the water mass. As a further assumption, we did not consider the heat transfer by thermal convection between the wall of the tank and the external environment, since all the water tanks are sealed. For the same reason, the thermal conduction between the insulating material and the wall of the tank, as well as the thermal exchange between the wall of the tank and the fluid inside, can be neglected.

Applying the assumptions described above, the thermal model become quite simple, as it is based on the well-known physical relation that describes the final temperature resulting from two different fluid flows, at different temperatures, when those flows are mixed. That relation is as following:

$$M_1{*}Cp_1{*}(T_f\text{-}T_1) = M_2{*}Cp_2{*}(T_2\text{-}T_f) \qquad (1)$$

Where:

M_1 = water mass of the first flow

M_2 = water mass of the second flow

Cp_1 and Cp_2 = specific heat coefficients of the fluids (in our model, Cp_1 = Cp_2 = specific heat coefficient of the water)

T_1 = temperature of flow 1

T_2 = temperature of flow 2

T_f = final (common) temperature of the mixture.

Solving the equation above by T_f, it is easy to find that:

$$T_f = (m_1{*}T_1 + m_2{*}T_2) / (m_1+m_2) \qquad (2)$$

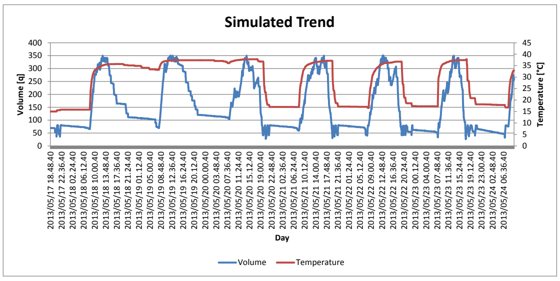
By means of the last formula, the final temperature of the water in each tank can be easily derived. By including that formula in the simulation model, and by computing the water temperature at different points of the plant, it is also immediate to identify those tanks whose temperature is higher than about 25°C, which was indicated as upper bound by the company examined.

4. MODEL VALIDATION

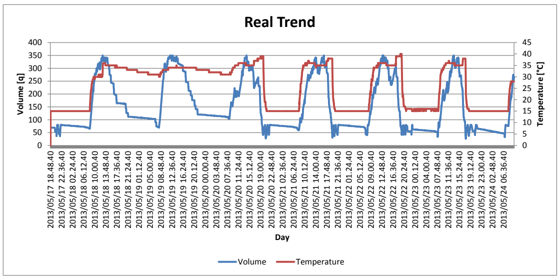
Once the model was developed, we identified a method to validate it, before its use to measure the water temperature inside each tank.

Specifically, for validation purpose, tank TK105 was equipped with a specific instrumentation able to measure the flow of water and the current temperature each day, with a time step of one minute. The real flow of that tank was then used as input of the thermal model, in order to compare its results to the real values provided by the measurement tool.

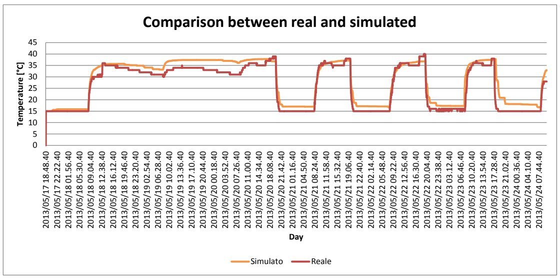
As shown in Figure 2, considering the same flow rate, the time-temperature trends obtained from real measurements and those simulated are quite similar.



(a)



(b)



(c)

Figure 2: simulated time-temperature trend (a), real temperature trend (b) and comparison (c).

The main difference between the simulated temperature trend and the real one is due to the fact that the simulated values are affected by the assumption of whole and complete miscibility of fluids. Because of this assumption, the resulting temperature profile is more regular than the real one. Anyway, taking into account that the simulation is expected to provide an estimate of the real result, the model could be considered validated.

5. SIMULATIONS AND RESULTS

The thermal model was launched to obtain the main outcomes in terms of the water temperature.

As previously mentioned, the plant includes two main critical elements, used during the rinsing phases of both the CIP lines, i.e. TK605 for the cheese production line and TK601 for the butter one. However, we started with the analysis of two other tanks, which supply water to the critical tanks and contain a relevant amount of recycled water: TK105 and TK902. As described earlier, tank TK105 reuses water coming from the pasteurizer and the whey cooler, plus the incoming water from the well when its total volume is lower than its safety level.

Figure 3 shows both the trend of the water flow (blue line) and its temperature (red line) over a two-week period (14 days).

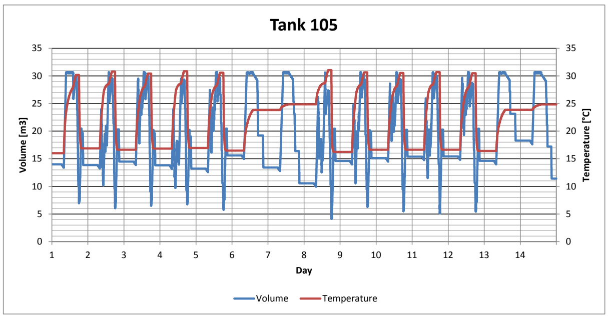


Figure 3: trend of the water flow (blue line) and related temperature (red line) for tank TK105.

As we can see from Figure 3, from Monday to Friday, the temperature increases in the morning up to reach about 30°C and it decreases in the second part of the day. This trend is due to the recycling of hot water arriving from the 2 processes of pasteurization and cooling. Indeed, from 8 a.m. to 2 p.m. both processes are working, providing, overall, 18,000 liters of water at 40°C and 54,000 liters at 25°C. Moreover, from 2 p.m. to 4 p.m., the pasteurizer continues to supply 6,000 liters of water, causing a sudden increase in temperature, up to its maximum. On Saturday and Sunday, the butter production line is not operating (only some washing cycles are carried out); therefore, the water requested is considerably lower compared to the other weekdays. Moreover, only the cooler operates; thus, the water temperature is always lower than 25°C.

Tank TK902 recovers the outputs from the fermenters of the cheese production section and the discharge of water by the TK105 when it reaches its maximum capacity. The corresponding water flow is proposed in Figure 4.

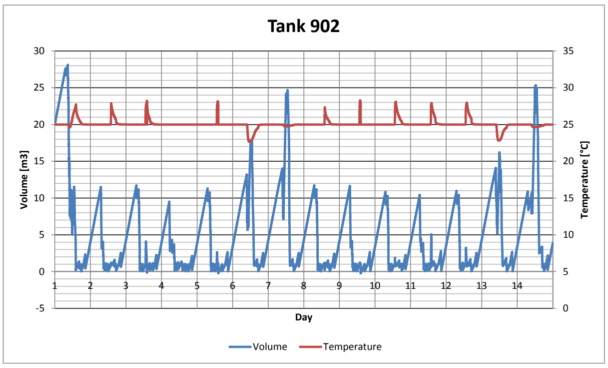


Figure 4: water flow (blue line) and related temperature (red line) for tank TK902.

From Figure 4, it emerges that the temperature is almost always higher than 25°C, with a daily peak of about 27°C, corresponding to the discharge of TK105. This is justified by the high temperature of the water flows that fill this tank. Obviously, on Saturday and Sunday, the scenario described for TK105 affects also the temperature of TK902.

As mentioned, the most important problem for the company examined is to control the water temperature inside two main elements which provide pure water to the CIPs. The corresponding results, which reflect the core outcomes of this work, are discussed below.

Looking at the cheese production line, TK605 can be filled from both the well and TK902. The corresponding temperature profile and water consumption are presented in Figure 5.

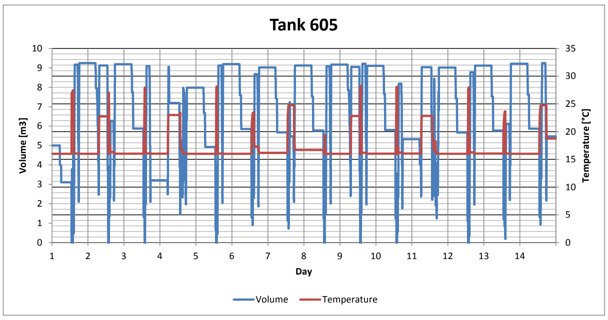


Figure 5: water flow (blue line) and related temperature (red line) for tank TK605.

From Figure 5, one can immediately see that the water temperature reaches unacceptable values, and that almost all the working days include at least one criticality. It is easy to identify the critical points looking at two specific days (i.e., Monday and Tuesday), which are particularly representative of the temperature trend across the two simulated weeks.

Figure 6 shows the behavior of the two valves (green and blue one) which regulate the CIP flow of pure water for the cheese line (1= valve open; 0=valve closed).

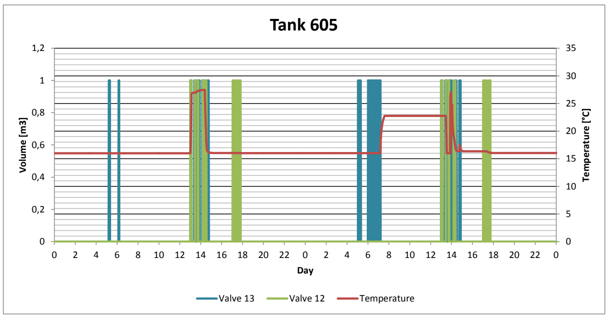


Figure 6: opening/closure of the CIP valves (blue and green lines) and corresponding water temperature (red line) for tank TK605.

Figure 6 highlights a main critical point, corresponding to the first day, from 12.30 p.m. to 2.30 p.m. If the company would like to avoid numerous washing cycles with hot water, the second critical point could be acceptable. Indeed, at about 2 p.m. of the second day, there is only a peak up to 25°C which could be ignored, considering also that the subsequent washing is at 15°C.

Tank TK601 is the last tank analyzed (Figure 7). This tank is slightly more problematic than the previous one, because it picks water only from TK105 and TK902, already studied.

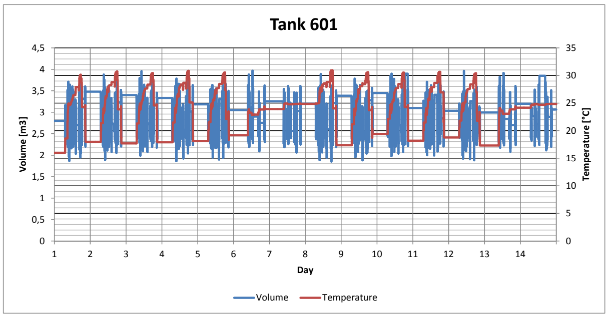


Figure 7: trend of the water flow (blue line) and related temperature (red line) for tank TK601 over a two-week period.

As Figure 7 shows, during all the working hours the temperature inside TK601 is higher than the acceptable limit of 25°C. Moreover, the time-temperature trend is very similar across the different weekdays. Therefore, we focused on a specific working day (i.e., Tuesday), obtaining the trend depicted in Figure 8.

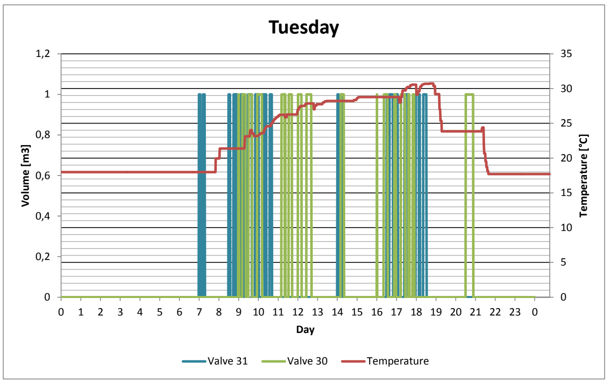


Figure 8: opening/closure of the CIP valves (blue and green lines) and corresponding water temperature (red line) for tank TK601 for a given day (i.e., Tuesday).

As it could be expected, all the washing cycles from 11 a.m. to 7 p.m. are characterized by too high temperature compared to the acceptable limit. From the theoretical perspective, only the washing cycle carried out in the early morning could exploit water from TK605; however, this solution is not feasible from the practical point of view.

6. DISCUSSION AND CONCLUSIONS

In the food industry, the issue of optimizing water consumption, by enhancing its reuse, is becoming more and more important. In this regard, this paper described the second part of a research activity, whose general aim was to optimize the water consumption in a cheese manufacturing company. In this work, a particular attention was paid to the analysis of time-temperature curves of water at different point of the water distribution system of the company, with the purpose of:

- estimating the amount of freshwater and recycled water used in the plant and the related temperature;
- identifying possible critical situations, i.e. those points in the plant where the water temperature is excessively high and should be lowered.

To achieve the aims listed above, in this paper we started by the discrete-event simulation model developed by Marchini et al. (2013), that was used in the first part of the research to describe the water flow in the plant. We added to this model the relevant thermal equations, to derive an estimate of the water temperature in the plant. Once the model was validated, subsequent simulation runs allowed to derive the water temperature at different points in the distribution system. Some critical situations, related, for instance, to water tanks TK605 were also highlighted.

This work has both theoretical and practical implications. From the theoretical point of view, our study is a typical example of how discrete-event simulation can be exploited as a powerful tool for the analysis of the water distribution system of the company investigated, and of the related performance.

Indeed, by means of discrete-event simulation, we were able to predict the operating conditions of the water supply system, taking into account the circulating flow rates, the on/off periods of pumping systems, the filling levels of the storage tanks.

From the practical point of view, we have mentioned that the whole study grounds on the analysis of a real cheese production company. Therefore, results have also a practical usefulness for that company. The identification of critical situations, or the analysis of the current performance of the water distribution system, are typical examples of practical results the simulation model can provide to the company investigated.

Future research directions that can be undertaken starting from this study include, primarily, the identification of countermeasures for the criticalities highlighted in our analysis. In this respect, the simulation model developed in this paper could represent a useful tool to investigate alternative layout configurations without the need of realizing them in practice, and to assess their performance against the criticalities identified.

REFERENCES

Almatò, M., Espuna, A., Puigjaner, L., 1999. Optimisation of water use in batch process industries. *Computers and Chemical Engineering*, 23, 1427–1437.

Almatò, M., Sanmartì, E., Espuna, A., Puigjaner, L., 1997. Rationalizing the water use in the batch process industry. *Computers and Chemical Engineering*, 21 (supplement), S971-S976.

Al-Redhwan, S.A., Crittenden, B.D., Lababidi, H.M.S., 2005. Wastewater minimization under uncertain operational conditions. *Computers and Chemical Engineering*, 29, 1009–1021.

Bogliolo, M., Bottino, A., Capannelli, G., De Pedro, M., Servida, A., Pezzi, G., Vallini, G., 1996. Clean water recycle in sugar extraction process: Performance analysis of reverse osmosis in the treatment of sugar beet press water. *Desalination*, 108, 261-271.

Casani, S., Knochel, S., 2002. Application of HACCP to water reuse in the food industry. *Food Control*, 13, 315-327.

Casani, S., Rouhany, M., Knochel, S., 2005. A discussion paper on challenges and limitations to water reuse and hygiene in the food industry. *Water Research*, 39, 1134-1146.

Centi, G., Perathoner, S., 1999. Recycle rinse water: problems and opportunities. *Catalysis Today*, 53, 11-21.

De Faria, D.C., Ulson de Souza, A.A., Guelli Ulson de Souza, S.M.d.A., 2009. Optimization of water networks in industrial processes. *Journal of Cleaner Production*, 17, 857–862.

Gleick, P.H., 1998. Water in crisis: paths to sustainable water use. *Ecological Applications*, 8(3), 571–579

Kim, J.K., 2012. System analysis of total water systems for water minimization. *Chemical Engineering Journal*, 193–194, 304–317.

Lee, K.P., Cheng, T.A., 2012. A simulation approach for energy efficiency of chilled water system. *Energy and Buildings*, 54, 290-296.

Marchini, D., Rinaldi, M., Montanari, R., Bottani, E., Solari, F., 2013. Performance analysis of a water supply system of a dairy company by means of advanced simulative methods. In: Affenzeller, M., Bruzzone, A.G., De Felice, F., Del Rio Vilas, D., Frydman, C., Massei, M., Merkuryev, Y. (eds.), *Proceedings of the 12th International Conference on Modeling and Applied Simulation (MAS 2013)*, ISBN 978-88-97999-17-1 (paper), ISBN 978-88-97999-23-2 (pdf), pp.103-113. September 25-27, 2013, Athens (Greece).

Poretti, M., 1990. Quality control of water as raw material in the food industry. *Food Control*, 1, 79–83.

Rezaei, E., Shafiei, S., Abdollahnezhad, A., 2010. Reducing water consumption of an industrial plant cooling unit using hybrid cooling tower. *Energy Conversion and Management*, 51, 311-319.

Rogers, J.W., Louis, G.E., 2009. Conceptualization of a robust performance assessment and evaluation model for consolidating community water systems. *Journal of Environmental Management* 90, 786–797.

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”.

Marta Rinaldi is Ph.D. student at the University of Parma (Italy), and scholarship holder in Industrial Engineering at the Interdepartmental Center CIPACK. She graduated (with distinction) in Industrial Engineering and Management in 2011, with a thesis titled ‘Analysis and evaluation of energy efficiency of a shrinkwrap-packer’. She is author of approx. 15 papers in the field of logistics, supply chain management, inventory management, discrete event simulation and its application to industrial plants and manufacturing systems.

Roberto Montanari graduated (with distinction) in 1999 in Mechanical Engineering at the University of Parma (Italy), where he is currently employed as Full Professor of "Industrial Plants". His research activities concern equipment maintenance, power plants, food plants, logistics, supply chain management, supply chain modelling and simulation, inventory management. He has published his research papers in several qualified international journals, as well as in national and international conferences. He acts as a referee for numerous scientific journals and is editorial board member of two international scientific journals.

Eleonora Bottani graduated (with distinction) in 2002 in Industrial Engineering and Management at the University of Parma (Italy), where she is Lecturer (with tenure) in Mechanical Industrial Plants since 2005. In March 2006, she got her Ph.D. in Industrial Engineering at the same university. She authored or co-authored more than 100 papers in the fields of logistics, supply chain management, supply chain modelling, performance analysis, industrial plants safety and optimisation. She acts as a referee for about 60 international scientific journals and for several international conferences. She also owns editorial responsibilities for several international journals.

Federico Solari is research fellow at the Department of Industrial Engineering at the University of Parma (Italy). He got a Master degree in Mechanical Engineering of the Food Industry at the same university, discussing a thesis related to the design of a plant for the extraction of volatile compounds. He received his Ph.D. with a thesis entitled "Advanced approach to the design of industrial plants by means of computational fluid dynamics". He attended several conferences related to food process and plants, and published several papers on the same topics in international conferences.

SIMULATE ENTERPRISE CONCEPTUAL MODELS EXECUTION BY COUPLING ENTERPRISE SERVICES AND DEVS MODELS USING WORKFLOW

Hassan Bazoun[a,* **], Judicael Ribault[b, *], Gregory Zacharewicz[c, *], Yves Ducq[d, *], Hadrien Boyé[e, **]

(*) University of Bordeaux, Lab. IMS, UMR CNRS 5218, 33405 Talence, France.
(**) Hardis / Hardis Conseil, 3 rue Guglielmo Marconi, 44800 Saint Herblain, France

[a]hassan.bazoun@hardis.fr, [b]judicael.ribault@ims-bordeaux.fr, [c]gregory.zacharewicz@ims-bordeaux.fr,
[d]yves.ducq@ims-bordeaux.fr, [e]hadrien.boye@hardis.fr

ABSTRACT

The service process modeling is emerging in the enterprise. Nevertheless, business modelling can be very complex, lying at the heart of many business decisions and demanding a lot of time and effort. A well-designed, well-built business model can lower risk and make enterprises more successful in their objectives. The SLMToolBox is an Eclipse RCP that proposes to transform conceptual models of service processes coming from business level to BPMN models (OMG 2011) and then from BPMN into DEVS models in order to simulate the behavior of the entire conceptual model. For a better integration and deployment of service models in the enterprise, we propose in this paper to test first - thanks to simulation - services freshly modeled (yet non-existing) coupled with existing enterprise services. This paper is a work in progress that recalls the MDSEA methodology and presents the key concept of the transformation of BPMN concepts into executable workflows within the SLMToolBox software, where unavailable enterprise services are simulated using DEVS. The interoperability between real and simulated services will be handled by the tool Taverna workflow and HLA RTI. This step is one step further in the MDSEA development loop.

1. INTRODUCTION

Enterprises develop Business Models to ensure they can produce services in the most effective and efficient manner possible. Business oriented people need a common and explicit graphical notation for business process modelling. They need to describe and communicate high level business processes involving enterprises resources with the help of a simple and explicit formalism. In (Bazoun et al 2013) authors proposed the Extended Actigram Star (EA*) language as a business process modelling language that facilitates the modeling of business process in enterprise offering a dynamic view of the process being modeled.

Then the authors have integrated a Model Driven approach: MDSEA to support the model transformation from conceptual level into more technical oriented models. The software developed is entitled SLMToolBox. This tool is an Eclipse RCP that proposes to transform EA* business level models to BPMN models (Bazoun et al 2013) and then from BPMN into DEVS models (Bazoun et al 2014). This sequence of transformation favors the simulation of the entire conceptual model's behavior. Although this solution allows testing and verifying of the conceptual model's behavior before its implementation and development, but it lacks continuous integration and interoperability with the enterprise services.

This position paper recalls the MDSEA methodology and the existing software bricks to be reused from the SLMToolBox which already proposes a set of transformation. Nevertheless this tool is missing one last step to embrace the full MDSEA methodology. This last step will consist in generating concrete web service calls from the SLMToolBox models thanks to a Workflow Engine. The missing interoperability and time management interoperability will completed by HLA Run Time Infrastructure (RTI) In that objective, we propose a connector between SLMToolBox-DEVS-simulator and the workflow engine Taverna.

2. BACKGROUND

This section presents briefly the EA* language and BPMN, then we describe the SLMToolBox and the Taverna Workflow engine. Finally we introduce DEVS and HLA-RTI.

2.1. Process Modeling Languages: EA*, BPMN

In (Bazoun et al 2013) the Extended Actigram Star (EA*) language was described as a business process modelling language addressed to business users responsible of the creation of the first model, business people responsible of the management, and to technical developers responsible of the development of business process modeling tools. As a graphical modeling language, EA* provides to business users and analysts a standard to visualize business processes in an enterprise, and thus with a comprehensible and easy way to handle these processes. EA*relies on a reduce set of graphical objects and focus on the "business" aspects of enterprise processes. By its simple and accessible syntax, EA* intends to reduce the gap between the ideation and the design of business process.

EA* models need to be enriched with IT elements so that models can be interpreted by developers and technical teams. The choice of modeling languages is BPMN (OMG 2012). It permits to represent the process with activity sequence flow, message, events and resources. It prepares the model to integrate the implementation architecture. Then these models are transformed to DEVS simulation models (Zeigler 2000) and validated thanks to simulation transformations between these models should be framed by a well specified methodology. For this purpose, section 2.2 detailed the model driven service engineering architecture (MDSEA) as a proposed a methodology.

2.2. Methodology MDSEA

The Model Driven Service Engineering Architecture (MDSEA) is inspired from MDA (OMG 2005)/MDI (Bourey et al, 2007). This methodology is proposed in the frame of the MSEE project (FP7 2012) that defines its first Grand Challenge as making SSME (Service Science, Management and Engineering) evolving towards Manufacturing Systems and Factories of the Future. MDSEA provides an integrated methodology dealing with modeling languages at various abstraction levels to support Service models and Service System design and implementation. The relationship between the MDSEA modeling levels (Business Specific Model, Technology independent Model, and Technological Specific Model) and the Service System lifecycle phases (user-requirements, design and implementation) is established. One of the important innovations in MDSEA is to define the integration between domain components (IT, Organization/Human and Physical Means) at the BSM level in order to ensure that these integration aspects will be spread out at other levels. In this sense, this is therefore considered as an adaptation and an extension of MDA/MDI approaches to the engineering context of product related services in virtual enterprise environment. On the basis of MDA/MDI, the proposed MDSEA defines a framework for service system modeling around three abstraction levels: BSM (Business Service Model), TIM (Technology Independent Model) and TSM (Technology Specific Model).

2.2.1. Business Service Model (BSM)

BSM specifies models at a global level, describing the service running inside a single enterprise or inside a set of enterprises as well as the links between these enterprises. The models at the BSM level must be independent from future technologies that will be used for the various resources and must reflect the business perspective of the service system. In this sense, it’s useful not only as an aid to understand a problem, but also it plays an important role in bridging the gap between domain experts and development experts. The BSM level allows also defining the link between Products’ production and Services’ production.

2.2.2. Technology Independent Model (TIM)

TIM delivers models at a second level of abstraction independent from the technology used to implement the system. It provides detailed specifications of the structure and functionality of the service system without including technological details. More concretely, it focuses on the operational details while hiding specific details of particular technology in order to stay technologically independent. At TIM level, the detailed specification of a service system’s components are elaborated with respect to IT, Organization/Human and Physical means involved within the production of the service. This is important to mention that in comparison to MDA or MDI or SOMA (Service Oriented Modeling and Architecture) (Bazoun et al. 2014), the objective of MDSEA is not only IT oriented and this requires enabling the representation of human and physical resources from the BSM level. At TIM level, these representations must add some information in comparison to BSM models.

2.2.3. Technology Specific Model (TSM)

TSM enhances the specifications of the TIM model with details that specify how the implementation of the system uses a particular type of technology (such as, for example IT applications, Machine technology or a specific person). At TSM level, the models must provide sufficient details to allow developing or buying suitable software applications, hardware components, recruiting human operators / managers or establishing internal training plans, buying and realizing machine devices. For instance for IT applications, a TSM model enhance a TIM model with technological details and implementation constructs that are available in a specific implementation platform including middleware, operating systems and programming languages (e.g. Java, C++, EJB, CORBA, XML, Web Services, etc.). Based on the technical specifications given at TSM level, the next step consists of implementing the designed service system in terms of IT components (Applications and Services), Physical Means (machine or device components or material handling), and human resources and organization.

2.2.4. Proposed Modeling Languages

Based on the described modeling levels, MDSEA proposes to associate relevant modeling languages at each level in order to represent confidently the existing system, future service product and future service system. For choosing modeling languages, the required abstraction level is important.

It is obvious to say that the first specification step of a service to be established between two partners is crucial. At the BSM level, the modeling language must be simple to use, powerful and understandable by business oriented users. Moreover, this (or these) language(s) must cover process and decision with coherent models. The choice is affected by the capacity of the language to propose a hierarchical decomposition (global view to detailed ones). Indeed, business decision-makers often have a global view of the running system and need languages allowing this global representation with few high level activities (process or decisions). This global view must be completed by more detailed activities models elaborated by enterprise sector responsible. These models are connected to top level models in a hierarchical and inclusive way. These are the principles of systemic and system theory which must be taken into account in the choice of the languages. But it is also obvious that the choice of modeling languages is subjective, depending on the experience of the languages' practitioners and on the dissemination of these languages within enterprises.

As for process modeling at business level, several languages exist. Extended Actigrams Star (EA*), extended from GRAI extended Actigram (Grangel et al 2008), that was itself de-rived from IDEF0 (NIST 1993), was chosen to model processes at BSM level due to its independence regarding IT consideration, its hierarchical decomposition and the fact it can model three supported resources: material, human and IT. It has been developed as an answer to previous issues encountered with GRAI extended actigram language regarding its interoperability. It intends to capture business process models at a high semantic level, independently from any technological or detailed specifications. Service Oriented Modeling and Architecture principles (Bell M, 2008) developed by IBM were also considered, but these languages are more IT oriented and thus were far away from our requirements. Moreover, GRAI Grid (Doumeingts et al 1998) was selected for modeling governance in a service system. GRAI Grid aims at proposing a cartography of company's decisions which controls business processes, as proposed for instance in the ISO 9000-2008 standard (Goult R, 2008). The interest of GRAI Grid is to represent all decisions and their coordination, from the strategic to the operational levels. This representation is very important for business users because the results of decision making are also at the origin of performance evolution and achievement.

At the TIM level, BPMN 2.0 (OMG 2012) was chosen in particular because this language offers a large set of detailed modeling construct, including IT aspects and benefits from the interoperability of many BPM IT platforms allowing the deployment and automated transformation to execution of BPMN processes. Moreover, BPMN enables also to represent human and technical resources which are required in the MDSEA principles of representation. BPMN has also the advantage to provide a meta-model developed by OMG which facilitates the implementation of the language. Other process modeling languages coexist, for instance GRAI nets are proposed in order to detail the decision processes in coherence with the decisions identified in the GRAI Grid but with adding technical and organization information as the decision rules, the decision makers, and the decision support modules.

2.3. SLMToolBox

SLMToolBox is a software tool developed by Hardis (Hardis). The SLMToolBox is an implementation of the BSM and TIM levels of MDSEA. It will be used by enterprises willing to develop a new service or improve an existing one, within a single enterprise or a virtual manufacturing enterprise (Thoben et al, 2001).

2.3.1. Modeling

MDSEA Meta models and Languages

MDSEA defines a set of constructs and relationships (described with "templates") which are specific to the domain of service system modeling, at 3 modeling levels: BSM/TIM/TSM. For each abstraction level, MDSEA suggest a set of references to standard or former graphical modeling languages (which are domain agnostic), in order to extend and complete the representation of the system to be modeled, under different perspectives (e.g.: decision structure; process; use cases; …). This type of modeling architecture is based on a "view model" pattern (or "viewpoints framework") (ISO 2011) as it defines a coherent set of views to be used, in the construction of a manufacturing service. The purpose of views and viewpoints is to enable humans to comprehend very complex systems, to organize the elements of the problem and the solution around domains of expertise and to separate concerns. In the engineering of physically intensive systems, viewpoints often correspond to capabilities and responsibilities within the engineering organization.

Both BSM (Business Service Models) and TIM (Technology Independent Models) are structured in the same manner. A "core" model gathers a set of generic (meta-) data in order to qualify the service to be modeled (specified / designed) ; this "core" model refers to external graphical modeling languages (e.g. : UML (OMG 2011) so that certain aspects of the service model can be elaborated in more details with the help of graphical languages.

This structure allows to map "view specific" modeling languages (e.g.: GraiGrid, UML Class Diagram) with "domain specific" constructs (i.e.: MDSEA BSM) without introducing modifications or restrictions to the MDSEA metamodel. From the user point of view, it allows the possibility to edit core information, independent from any specific modeling language, and to retrieve and reuse this data under different views, accomplished with the help of several graphical diagrams. With this approach, MDSEA Core Constructs remain agnostic from any representation formalism. Their implementation is realized by a core model, which acts as domain specific (Service System Modeling) "glue" between several modeling languages. Thus, we can reuse standard modeling languages without introducing modifications to their metamodels (e.g.: BPMN, UML…). Graphical languages such as "Extended Actigram Star" or "GRAIGrid" can continue to evolve, with (almost) no impact on MDSEA Core metamodels (i.e.: BSM and TIM).

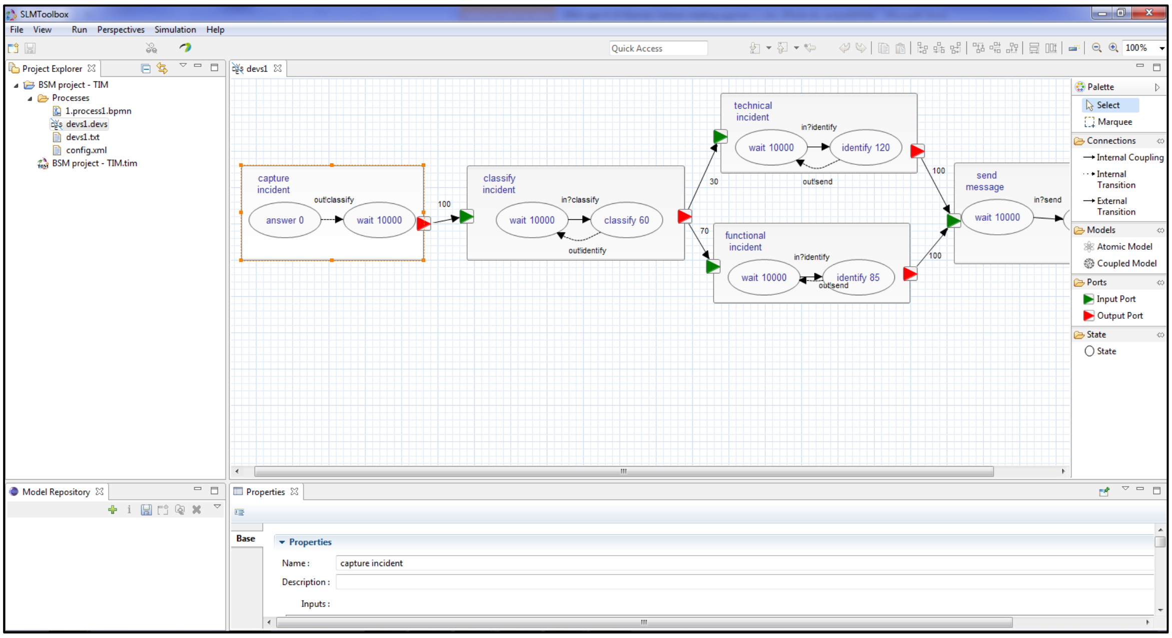


Figure 1 SLMToolBox

Modeling editors

The modeling environment will support service system modeling activities by providing editors for domain specific models (BSM, TIM) and related modeling languages to enhance the description of the BSM and TIM models. A set of language specific modeling editors is provided for each modeling language. These editors are either the result of a Hardis's specific development (BSM templates, EA* and TIM templates editors) or open source plugins integrated within the environment (UML and BPMN editors, DEVS editor).

Model transformation

SLMToolBox supports specific model transformations, mostly to support the continuity between the service concepts & requirements phase to the service design phase. In addition model transformation aims to save effort and reduce errors by automating the development of models when possible. Information and requirements collected at BSM level are reused at TIM level. SLMToolBox supports the transformation of BSM data models into TIM data models, and the transformation of EA* model into BPMN process and collaboration diagrams (Bazoun et al 2013). In addition, it supports the transformation from BPM models to DEVS (Bazoun et al 2014) for simulation purposes.

2.3.2. Integration with external tools

Up to now, the SLMToolBox is preparing the models to be ready for use at TIM level; the models can be simulated. Nevertheless these models miss technical details and so cannot be connected directly and immediately with other software or material components to support a concrete service implementation. The idea proposed in this paper is to reuse the experience acquired when connecting DEVS models with service calls. The interoperability can gain from the service architecture and distributed interoperability simulation architecture. The missing element in the workflow chain is the component that links the model with the service call. DEVS, HLA and Taverna will help for that.

2.4. Workflow Engine Orchestration with HLA

In (Ribault 2014), authors describe how to facilitate interoperability between real services and DEVS simulation using workflow of services and HLA/RTI.

2.4.1. Workflow of services

Workflow was first designed to improve the business process. A production workflow is a set of steps required for developing a product until it is put on the market (Weske, 2012). The workflow steps are based on observing a number of steps that are usually repeated manually and formalizing them. The Workflow Management Coalition (WMC) standard group (WMC 2009) proposes a WF reference model in which the WF is in the center and interacts with other surrounding applications or WF components.

Several surveys have compared different workflow management systems. In (Deelman et al., 2009), the authors analyzed and classified the functionality of workflow system based on the needs of scientists who use them. In (Yu and Buyya, 2006), the authors focused on the features to access distributed resources. In (Curcin and Ghanem, 2008), four of the most popular scientific systems were reviewed. 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. Taverna was chosen to demonstrate the feasibility of the methodology because Taverna eases the interoperability with other services and the data flow modelling compare to other workflow management system we studied.

2.4.2. Taverna

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.

A Taverna workflow can contain several services including: Java code, Remote application via the REST protocol, SOAP/WSDL protocol, Workflow nested within another hierarchically and the use of local tools within a workflow.

In Taverna, a service can take input and produce output. The workflow input can be part of the workflow or can be given prior to the execution of the workflow. For example, the Taverna RESTful service takes in input various data, and it returns a status code and a response. A WSDL Taverna service will find automatically the number and type of input and output thanks to the WSDL file. Taverna offers the possibility to automatically format the input and output based on the type of parameters required by the Web service.

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. Taverna offers several user interfaces with this purpose:

- Ask: opens a box so the user can enter text.
- Choose, Select: lets the user select a value among a list of values.
- Select File: lets the user select a file in the system.
- Tell, Warn: gives a message to the user.

A Taverna workflow can contain nested workflows. Thus, it is possible to create a parent workflow that contains several workflows. Several workflows can be combined together to obtain more complex workflows that do not need the external inputs and are fully automated.

2.5. DEVS & G-DEVS

Discrete EVent Specification (DEVS) was introduced by (Zeigler et al., 2000). 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 changes the state of the model. The model also takes into account the elapsed time while firing an external state transition triggered by an event received from outside the considered model.

The behavioral 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 et al., 2000]. Thus, G-DEVS defines the coefficient-event as a list of values representing the polynomial coefficients that approximate the input–output trajectory. Therefore, an initial DEVS model is a zero order G-DEVS model (the input–output trajectories are piecewise constants). In fact G-DEVS was the pioneer DEVS extension proposing a multi value event.

On the simulation side, G-DEVS models employ an abstract simulator (Zeigler et al., 2000) 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 instead of a unique value in DEVS.

Zacharewicz et al. proposed in (Zacharewicz et al., 2008), an environment, named DEVS Model Editor (LSIS_DME), to create G-DEVS models that are HLA compliant and simulating them in a distributed fashion. In LSIS_DME, a G-DEVS model structure can be split into federate component models in order to build a HLA federation (i.e. a distributed G-DEVS coupled model). The environment maps DEVS Local Coordinator and Simulators into HLA federates and it maps Root Coordinator into RTI. Thus, the “global distributed” model (i.e. the federation) is composed of federates intercommunicating.

2.6. HLA

The High Level Architecture (HLA) (IEEE, 2000) (IEEE, 2003) is a software architecture specification that defines how to create a global software execution composed of distributed simulations and software applications. This standard was originally introduced by the Defense Modelling and Simulation Office (DMSO) of the US Department of Defense (DOD). The original goal was the reuse and interoperability of military applications, simulations and sensors.

In HLA, every participating application is called federate. A federate interacts with other federates within a federation (i.e. a group of federates). The HLA set of definitions brought about the creation of the standard 1.3 in 1996, which then evolved to HLA 1516 in 2000 (IEEE, 2000) and finally to 1516 Evolved (IEEE, 2010).

The interface specification of HLA describes how to communicate within the federation through the implementation of HLA specification: the Run Time Infrastructure (RTI). Federates interact using the proposed services by the RTI. They can notably "Publish" to inform on the intention to send information to the federation and "Subscribe" to reflect information created and updated by other federates. The information exchanged in HLA is represented in the form of classical object-oriented programming. The two kinds of object exchanged in HLA are Object Class and Interaction Class. The first kind is persistent during run time, the other one is just transmitted between two federates. These objects are implemented with XML format. More details on RTI services and information distributed in HLA are presented in (IEEE, 2000) and (IEEE, 2010). In order to respect the temporal causality relations in the execution of distributed computerized applications; HLA proposes to use classical conservative or optimistic synchronization mechanisms (Fujimoto, 2000). In HLA 1516 Evolved (IEEE, 2010) the service approach is demanded as core feature. Nevertheless no software addresses completely that goal at the moment (Tu et al., 2013).

2.7. Time Management

In (Ribault 2014), authors addressed the problem of time synchronization management between real world and the simulated part.

The time is not progressing with the same dimension in the simulated part and was not taken into account in the workflow of service approach. The HLA can be an issue; authors in (Zacharewicz et al., 2008) have proposed to handle time related message exchange between the workflow components and others thanks to RTI. A specific service calls have been specified and adapted for HLA in (Ribault, 2014) making service able to be the bridge between in the real world and the simulated components.

3. CONTRIBUTION

To improve the integration and the validation of the conceptual model, we propose to compose a test that reuses existing enterprise services while simulating non-existing or unavailable enterprise services. This approach should support the progressive involvement of new components to be added to the existing system by adopting the System of Systems (SoS) paradigm. The test, thanks to simulation, is confronting the future components to their future environment. This proposition should anticipate problems that can be faced at final implementation. In particular the causality relations of events and calls to services that are planned to be chained are here tested within the time constraints of the real future system.

More concretely, the first problem is the matching between the concepts announced in the enterprise BPMN models of services with technical services. This issue can be addressed by transforming BPMN concepts into executable workflow of services as described in the figure 2 with the dashed link going from BPMN 2.0 Diagram to Workflow Engine Orchestration. The second problem is to deal with non-existing or unavailable services in the enterprise. This issue can be resolved using DEVS models to mimics the behavior of enterprise services completed with HLA envelop to make them interoperable with other distributed and heterogeneous components. This is represented on the figure 2 with dashed link from DEVS Model to RTI.

Currently, the SLMToolBox exclusively proposes to transform all BPMN components to DEVS models. The idea proposed in this paper is to adapt the tool in order to propose users to select, on one side, the part of the BPMN model that will be transformed into workflow of service in order to call the existing enterprise web services. For this part of the model the tool will prepare the service calls by configuring the service query and locating the server to be called. On the other side, the other part of the model will be automatically transformed into DEVS models and simulate the behavior of the part of the system including the time to respond and the state that memorizes the process status. Some previous work has already put the first stones in this domain. This step will go further in the MDSEA lifecycle by starting to generate real calls to services and external systems.

3.1. From BPMN to executable workflow

The first problem is the matching between the concepts announced in the enterprise models of services with technical services. The SLMToolBox is preparing the model at TIM but the model is not yet extended with primitives to connect with existing systems and in particular servers that provide services. This step is supposed to be assumed only at TSM. Nevertheless it is interesting to test by simulation the system in its future real environment.

This new step in MDSEA approach will start from the BPMN model produced by the SLMToolbox, it will gain in interoperability by coding, from the BPMN communication actions, the primitives for service calls. This will be facilitated by the model structure already saved with the XML format in the Eclipse standard. Concretely a resulting executable WF will be generated from the BPMN model. This model could be played with Taverna to sequence the service calls and answers. It will facilitate the interoperability with existing services and will close the loop of service verification and validation.

After the edition of the BPMN model, the user will be able to annotate the model and select the part that will be transformed to simulation models or Workflow of services. Concerning the transformation of BPMN parts to executable workflow, the Meta model approach will be preferred. For instance, Taverna saves the workflows in XML, so we could transform BPMN message and data flow part that link lanes or pools into Taverna workflow abstract service calls. The user will fulfill the query detail for the service call.

On the other side, unavailable services will be transformed to DEVS models to be simulated. This transformation is highlighted in the next section (3.2).

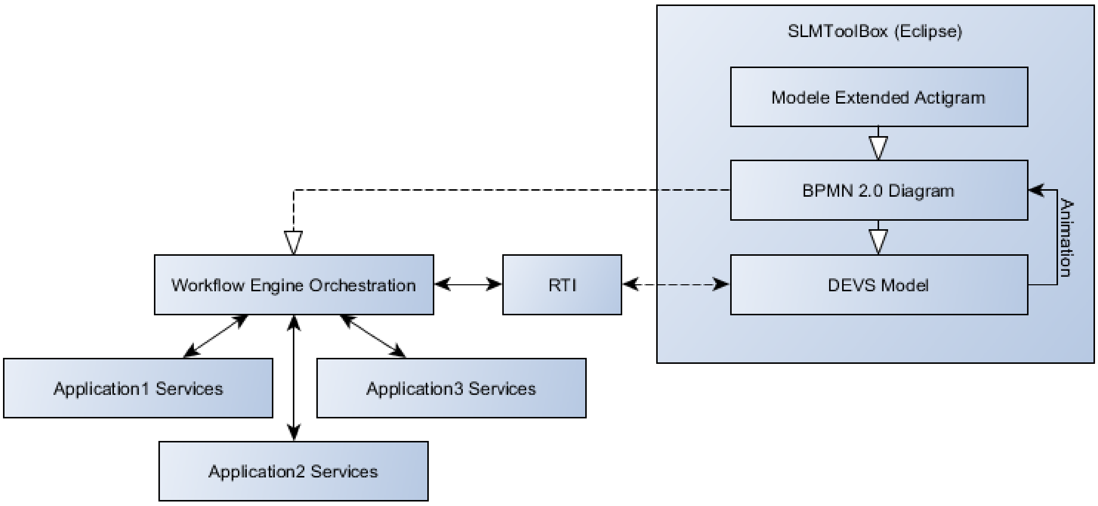


Figure 2 From conceptual model to enterprise execution.

3.2. From BPMN to DEVS Simulated Enterprise Services

The second problem is to deal with, at the same time, on one side existing and/or legacy system and on the other side non-existing or unavailable services in the enterprise. The idea is to mix simulated parts with real environment of web services.

In that objective, the authors of (Ribault 2014 proposed an architecture that is dedicated to compose DEVS models with workflow of service tools. In detail, the workflow is orchestrated with a tool able to call and orchestrate the answers of web services and for services that are not defined already and when the behavior of external actors is required, the system is calling a DEVS component that is timed with a local behavior.

SLMToolBox generates DEVS Models out of BPMN models. This mechanism is based on an implemented transformation from BPMN concepts to DEVS concepts. In (Bazoun 2014) authors explained this transformation and how certain BPMN concepts are mapped to DEVS concept. After the transformation to DEVS Models the process can be simulated depending on cost and time. A link should be established to use the obtained simulation results by Taverna. A possible solution would be to develop web services that can return this simulation results so that it can be used by Taverna.

DEVS models resulting should be extended to be able to communicate with the Taverna Workflow. In that objective the works proposed in (Ribault 2014) and (Zacharewicz 2008) can be adapted and reused. They proposed to embed DEVS models into HLA “federates”. These “federates” gain interoperability properties to communicate within a distributed simulation, thanks to HLA. The “federate” can publish messages and subscribe to information within a time synchronized environment. Nevertheless, these models should, even synchronized, still need to be enriched to have the primitives for “real world” services calls and communication including how to form the query and how to reach the URL of the service.

4. CONCLUSION

This paper has proposed the principle of a mechanism to generate from the conceptual model a continuous integration facility and a testing and validation platform within the existing enterprise services system. This contribution faces two major problems: (1) conceptual models are just blueprint and they contain components with behaviors that need to be simulated in order to test and verify their correctness in the global behavior in the enterprise system, and (2) simulated conceptual models must be able to interact with existing enterprise services. The first problem has been previously tackled up to TIM level thanks to model transformation proposed in the SLMToolBox architecture. The second problem has also been addressed in composing workflow and DEVS. Nevertheless no works were proposed to compose these two questions.

In this paper we proposed a solution to go one step further in the direction of a - as much as possible - fully automated MDSEA approach. We described the basement for this new approach. In particular we have proposed a method for decomposing models, originally based on a BPMN description, into on one side simulation and, on the other, real world interaction. These two sides are then coupled again in distributed testing system that composes simulation models with concrete service calls. This work will be extended in order to generate the technical workflow of services calls.

REFERENCES

Bazoun, H., Zacharewicz, G., Ducq, Y., Boye, H., 2013. Transformation of Extended Actigram Star to BPMN 2.0 and Simulation Model in the Frame of MDSEA. *Proceedings of the Symposium on Theory of Modeling & Simulation 2013*. April 7-10, (San Diego, USA).

Bazoun, H, Zacharewicz, G, Ducq, Y, Boye, H., 2014. Business Process Simulation: Transformation of BPMN 2.0 to DEVS Models. *Proceedings of the Symposium on Theory of Modeling & Simulation - DEVS Integrative M&S Symposium*, April 13-16, Tampa. (Tampa, USA)

Bell, M., 2008. *Service-Oriented Modeling: Service Analysis, Design, and Architecture*. Wiley publishing.

Bourey, JP, Grangel, R, Doumeingts, G, Berre, A. 2007. Deliverable DTG2.3 from the INTEROP project. *Report on Model Driven Interoperability*. Available from: http://interop-vlab.eu [accessed 15 May 2014]

Curcin, V. and Ghanem, M., 2008. Scientific workflow systems can one size fit all? *Biomedical Engineering Conference*, pp. 1-9. December 18-20, Cairo (Cairo, Egypt).

Deelman, E., Gannon, D., Shields, M., and Taylor, I., 2009. Workflows and e-Science: An overview of workflow system features and capabilities. *Future Generation Computer Systems*, 25 (5) 528–540.

Doumeingts, G., Vallespir, B., Chen, D., 1998. GRAI grid, decisional modeling. *In International Handbook on Architecture of Information System*. Springer Verlag, 313-337.

IEEE std 1516TM-2010, 2010. *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) -- Framework and Rules*, New York: Institute of Electrical and Electronic Engineers

IEEE std 1516.2-2000, 2000. *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) - Federate Interface Specification*, New York: Institute of Electrical and Electronic Engineers.

FP7 – FoF-ICT-2011.7.3. 2012. *Manufacturing SErvice Ecosystem Project- Annex 1 description of work*; July 29, Available from: http://msee-ip.eu/ [accessed May 15 2014]

Goult, R. *Introduction to ISO 9001.2008 Handbook*. Available from : http://fr.scribd.com/doc/171647046/ISO-9001-2008-Handbook [accessed 18 October 2013]

Grangel Seguer, R., Cutting-Decelle, A.F., and Bourey, JP., A UML profile for transforming GRAI Extended Actigrams into UML. *In Proceedings of the 17th IFAC World Congress, Session: Architectures and Software Tools for Enterprise Integration and Networking in Manufacturing*, Paper ThB24.3, Séoul, Korea; 2008

Giambiasi, N., Escude B., Ghosh, S., 2000. GDEVS A Generalized Discrete Event Specification for Accurate Modeling of Dynamic Systems. *SCS Transaction of the Computer Society*, 17 (3), 120-134.

Hardis Company Website Available from: http://www.hardis.fr/eng/jsp/site/Portal.jsp [accessed 18 October 2013]

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.

ISO/IEC/IEEE 42010:2011, Systems and software engineering — Architecture description.

National Institute of Standards and Technology (NIST). *IDEF, Integrated DEFinition Methods.* Available from http://www.idef.com/IDEF0.htm [accessed 18 May 2014].

OMG 2011. *OMG Unified Modeling Language (OMG UML), Infrastructure Version 2.4.1*. Available from: http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF/. (accessed 18 October 2013)

OMG. Business Process Model and Notation (BPMN) Version 2.0. http://www.omg.org/spec/BPMN/2.0/PDF/ (accessed 18 October 2013).

OMG. MDA Guide Version 1.0. 2005 http://www.omg.org/cgi-bin/doc?omg/03-06-01 (accessed 18 October 2013).

Ribault, J., & Zacharewicz, G. (2014) Time-Based Orchestration of Workflow, Interoperability with G-DEVS/HLA, *Journal of Computational Science*, available online august 2014, DOI: 10.1016/j.jocs.2014.07.002

Missier Tan, W.,, 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.

Tu Z., Zacharewicz G., Chen D., 2014. Building an HLA Federated Interoperable Framework from Legacy Information Systems, *In International Journal of Computer Integrated Manufacturing*, 27 (4), 313-332.

Weske, M. 2012. *Business Process Management Architectures*. Business Process Management, pp 333-371.

Yu, J., and Buyya, R., 2006. A taxonomy of workflow management systems for grid computing. *Journal of Grid Computing*, 3 (3-4), 171–200

Zacharewicz, G., Frydman, C., Giambiasi, N., 2008 G-DEVS/HLA Environment for Distributed Simulations of Workflows, *Simulation*, 84 (5), 197–213

Zeigler, B. P., Praehofer, H., & Kim, T. G., 2000. *Theory of modeling and simulation: Integrating discrete event and continuous complex dynamic systems.* Academic Press.

ACKNOWLEDGEMENT

This work has been partially supported by the FP7 Research Project MSEE "Manufacturing Service Ecosystem" ID 284860. http://www.msee-ip.eu/

BIOGRAPHY

Hassan BAZOUN is a Software Engineer at Hardis and PhD Student at IMS Lab. His research domain includes business process modeling, model transformation, and Discrete Event Simulation. He is involved in the FP7 European project (MSEE) and has participated in the development of the modeling tool SLMToolBox.

Judicael RIBAULT is Postdoctoral fellow at the University of Bordeaux (IUT MP) Lab. IMS. He received his Ph.D. at INRIA Sophia Antipolis, France. His research interests include Component-based software engineering and simulation, Discrete Event Modelling (e.g. DEVS, G-DEVS), Distributed Simulation, Large-Scale Simulation, Simulation Interoperability and Workflow. He has published several papers in Conferences and is frequent Reviewer in Conferences (SpringSim, TMS/DEVS, etc.). He is involved in European projects.

Gregory ZACHAREWICZ is Associate Professor at University of Bordeaux (IUT MP) with both competences in enterprise engineering and computer sciences. His research interests include Discrete Event modelling (e.g. DEVS), Distributed Simulation, HLA, and Workflow. He has been involved in FP7 European projects (e.g. MSEE and FITMAN). He is recently focused on Enterprise Modelling and Semantic Interoperability. He has published more than 50 papers in international journals and conferences.

Yves DUCQ is Full Professor at University of Bordeaux from September 2008. He is working on Performance Measurement, Enterprise Modelling and Production Management since fifteen years and has published more than 20 papers in books and international journals and more than 70 papers in international conferences. He is responsible of the research group dedicated to production engineering sciences (GRAI research Group) at IMS laboratory. He has been involved in many European and French projects for twenty years and has also act as research engineer on several contracts with industry on performance improvement. He is also quality manager for University of Bordeaux.

Hadrien BOYE is a Project Manager at HARDIS. He is in charge of the e-solution skills center in the western offices of the company. His main fields of interest include web technologies, software architecture and enterprise information systems.

INTEGRATION OF SIMULATION DATA FOR CONSTRUCTION PROJECT PLANNING USING A PROCESS-BASED SIMULATION FRAMEWORK

Ali Ismail[(a)], Raimar Scherer[(b)]

[(a)(b)] Technische Universität Dresden, Germany

[(a)]Ali.Ismail@tu-dresden.de, [(b)]Raimar.Scherer@tu-dresden.de

ABSTRACT

This paper presents the recent developments of the process-based simulation toolkit, which aims to support planning of construction projects through rapid development of simulation models and efficient integration of simulation input data from various data models. The paper focus on the integration of simulation input data for the basic project data, namely: product model based on IFC standards (ISO 16739), process models based on BPMN notation, planning and resources data. It discusses and presents the latest research work and the prototype implementation through two study cases.

Keywords: construction process management, simulation data integration, reference process modelling

1. INTRODUCTION

Simulation methods have been used to optimize and improve the planning quality in many industry sectors like manufacturing and logistic since long time, and they are integrated as an essential part of the whole design and planning process. However, in the construction industry, the use of simulation methods for planning has not been widely adopted and it is usually limited to analyse and solve problems in an ad-hoc manner (simulation study or simulation-on-demand) at certain time point of the project life. The application of simulation techniques for construction projects is a very promising but also a challenging field of research(Lucko et al. 2008). Creating reliable and reusable simulation models is very complex, combined with high costs of software licenses and personal training for in-house simulation. While great advances have been made in construction simulation over the past few decades, adoption by industry has lagged, for three potential reasons: simulation is not accessible, it cannot handle the complexity of modern construction projects, and the benefits are not immediately obvious (AbouRizk 2011).

Therefore, providing a convenient simulation tools and collaborative platforms to integrate the huge and complex projects data with low-cost entry for construction industry is crucial to promote a wider adoption of simulation in construction industry.

As a part of the national German leading research project MEFISTO (www.mefisot-bau.de) a Construction Simulation Toolkit (CST) has been developed. It aims to accelerate the process of creating simulation models for construction and logistic operations during the planning and operation phases of construction projects.

In general it aims to support the planning process in order to reduce the total duration and cost of construction projects and avoid any conflicts during the construction phase by improving the planning quality and utilization rates of resources. The simulation model should help the project manager to verify the feasibility of a given schedule with a combination of different resource constraints or different building design alternatives or construction methods.

The toolkit is created on top of the simulation software Tecnomatix Plant Simulation from Siemens PLM Software. It provides a set of reusable simulation components with a simple user interface for rapid building of process-based simulation models for planning projects with resource and activity calendar constraints. The CST framework has a modular structure and consists of a set of simulation reusable components. Figure 1 shows some of simulation components in CST which can be used to create and customize simulation models rapidly.

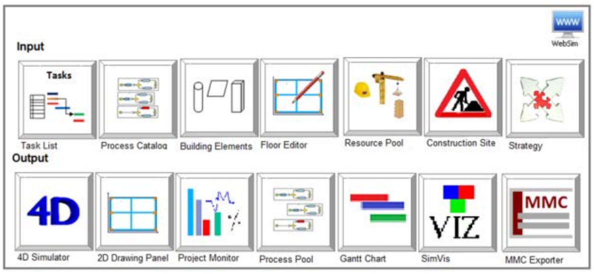


Figure 1: The modular structure of simulation models built with CST

The implementation of the CST toolkit is generic and can be used for different kind of projects and domains. This is a result of separating the core simulation components which are applicable for any simulation domain and the domain specific components (product model, resource, and process models). CST uses formal BPMN process models to define the logic of construction operations. It integrates also most of the important construction project information and can be used for a rapid and effective development of simulation models. The concept of using a formal process models and mapping them to simulation model in addition to a description of the system structure has been presented in (Wagner et al. 2010) and (Ismail et al. 2011). This paper focuses mainly on the data integration between the simulation framework and other project data models. Section 2 gives an overview about data integration for Building Information Model (BIM), process model and planning data. The modeling of resources is discussed briefly in section 3. Finally, two study cases are presented in section 4 in order to illustrate and highlight some of the framework applications. The main contributions and innovative aspects of this research are:

1. Using formal business process models based on BPMN notation to capture and organize the knowledge in construction domain and providing the ability to transform these models into a simulation process directly.
2. Addressing interoperability problems through integration of most important simulation input data from CAD/BIM design and planning software.
3. Coupling the logistic and production operations in one unified simulation environment in a mixed micro and macro levels of details.

2. DATA INTEGRATION BETWEEN BASIC MODELS

One of the top reasons to keep using simulation inside the construction industry for real projects limited is the huge amount of information which must be collected for each simulation study in a certain degree of quality to ensure realistic results. The one of a kind character of construction projects leads to a one of a kind simulation study. A big effort was made in our work to solve the problem of data integration and maximize the reuse of available project data. This was achieved through: (1) Adopting the most used standards for both of building information modeling (product model in AEC) and process modeling, namely the Industry Foundation Classes (IFC) and Business Process Model and Notation (BPMN), (2) Implementing import and export interfaces to integrate the simulation study input and output data within the planning workflow.

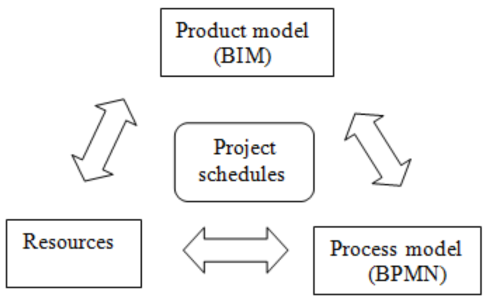


Figure 2: The basic data models for a simulation study

The three basic models or source of information for construction project planning are:
Product model: what is going to be built?
Process model: how it will be built?
Planning model: when?
Resource model: which resources will be used?

2.1. Integration of BIM data model

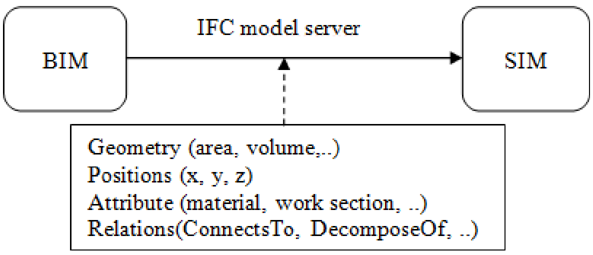
The integration of building information data for any simulation study is one of the most challenging and time consuming problem for real construction projects. Some reasons for this problem are:

1. The quality and level of details of BIM models do not usually match the simulation requirements(e.g. missing information about the base quantities and materials of building elements, or the hierarchy structure is not defined correctly)
2. The freedom of modeling in practice without considering the constructability aspects especially in the architectural models. It usually leads to common problems, for example having some walls extended over many stories, or having a wrong type of building elements.
3. The wide range of BIM authoring tools with inefficient support of IFC standards.

CST as a generic simulation toolkit allows the user to define the data structure of the product model in very flexible way; on the other hand it adopts the IFC data model (ISO 16739) as the most supported format among BIM authoring software for automatic integration between BIM and SIM model.

The data model server project IFCWebServer (Ismail, 2011) was developed in parallel to CST toolkit in order to enable extracting all the necessary information from BIM models effectively. The data extraction and mapping to simulation model were achieved by using predefined queries/filters and data validation rules. The results of the queries can be saved as XML files and imported easily into the simulation model.

Figure 3: Information flow between BIM and SIM



models

2.2. Integration of process model data

In our approach, Business Process Model and Notation (BPMN) is used to create semantic and graphic process models for various construction operations. BPMN models consist of simple diagrams constructed from a limited set of graphical elements. This makes it easy to learn and understand for both, business users as well as developers.

Flow Objects are the main elements to define and control the behavior of a business process.
The basic modeling elements of BPMN which can be used inside the simulation process templates are:

- Start/End events
- Task and sub-process
- Sequence flow
- Gateways: Parallel Fork/Join , Data-Based XOR
- Conditional and default flow.

Process models will be imported into a process catalog as 'ready to use' Reference Process Models (RPM). The process catalog includes reference process models for the best practices of various construction processes that are reusable across different construction projects in addition to their data definitions of resources and productivity factors.

The process catalog includes models for the best practices of various construction processes that are reusable across multiple projects with their data definitions of resources and productivity factors. The knowledge accumulation enables a continuous improvement of the catalog.

The process templates can be as simple as having one single activity or a set of serial activities which run sequentially without any kind of control flow or very complex ones including loops and conditional gateways which may result in skipping/repeating some activities. In order to create reliable simulation model, one needs to carefully examine every detail of the construction process and identify the major events and processes that will be presented in the simulation model (Akhavian et al. 2011).

In CST the user has the freedom to move the logic and the dependencies between different tasks to be a part of the input data for each simulation model or to be included inside process templates. However this may lead to one of the following extreme situations:

1. A very detailed project schedule with a simple process templates for each task
2. A very simple project schedule consisting of few tasks with a high level of abstraction combined with very complex process templates

Both cases must be avoided to get the most out of the simulation model in the sense of ease of use and the flexibility of answering as many as "what if" scenarios. It is mainly the responsibility of the user to maintain balance between these options according to the simulation goal and the availability of data. Figure 4 shows as an example of a good balanced process template for erecting a wall for three construction methods: Precast, In-situ concrete and as bricks walls.

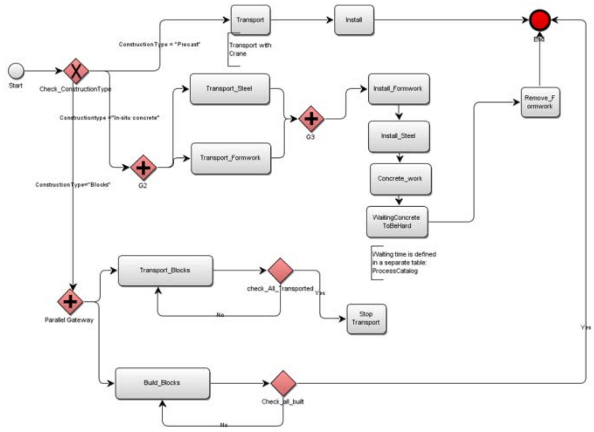


Figure 4: BPMN process model template for wall erecting

By using this template there is no need to use three different templates and link them explicitly to each wall as part of the simulation input, which reduce the time and efforts to prepare the simulation data. The definition of resources and duration value/formula for each single task can be embedded inside the BPMN process template or maintained in separate database.

2.3. Integration of planning data model

Another important domain of information is planning data model, which connects the other basic models together in form of project schedules or what is called "Task list" inside the CST framework.

The start point of planning any project is to establish a list of all the activities which have to carry on in order to deliver the expected result with a certain quality taking into account the limits of time and resources. The most common methods to describe and visualize this list are Gantt charts, activity networks, and PERT diagrams combined with the critical path analysis method (CPM). Using simulation technique as an additional tool to support the planning, offers a lot of benefits to improve the outcome by:

- The duration of any activity will be calculated dynamically during the simulation taking in account the real combination of resources which have been used to do the work and the exact quantities of work from the product mode.
- Integrating the uncertainty factors of resource productivity, quantities and duration calculations.
- Allowing the planer to test the output of many "what if" scenarios rapidly.
- Assist in taking immediate corrective actions in case of unexpected changes during the construction phase.

CST support importing/exporting project schedules from/to Microsoft Project software and offers a handy way to break the project schedules one or more levels down automatically by using special process templates, which serve as a recipe to describe the steps and logic of activities needed to complete the work with each task in the high level schedule.

3. MODELING OF RESOURCES

The modeling of resources inside the simulation model and the interaction between tasks and available resources is very important to ensure the correctness and reliability of simulation results. A simple data model was defined to describe the resource requirements of each activity in a generic way. This model allows defining the minimum set of necessary resources and their (min, max, default). Table 1 shows an example of the required resources definition for the activity "Concrete Work" in the Wall erecting process model.

Table 1: Resource definition for the activity "Wall_ConcreteWork"

Resource From	ResourceTo	Condition	default/min/max amount		
Concrete	Concrete_Sink	Class= "C25"	Volume		
Worker	WorkerBusy		3	2	5
Pump	PumpBusy		1	1	1

Most of traditional planning methods and tools assume that required resources for any activity will be allocated at the start of the activity and released at the end and also assume that the same resources will be used to complete an interrupted. In our approach the dynamic allocation of resources during simulation bypasses these assumptions and resources can be reallocated when they are not necessary any more needed to complete the activities.

CST framework includes the following standard resources:

- Building material: concrete, steel
- Equipment: Formwork panels
- Construction machines: Tower crane, concrete pump, trucks
- Human resources: workers with qualifications
- Construction site facilities: Storage areas, transport roads, entrance and exit gates

The resources inside the simulation model can be defined as abstract objects or as detailed simulation modules. Figure 5 shows the representation of the tower crane object as an abstract resource (top) or as a detailed simulation module with extra functionality to calculate the duration of transport precisely and support 2D and 3D animation.

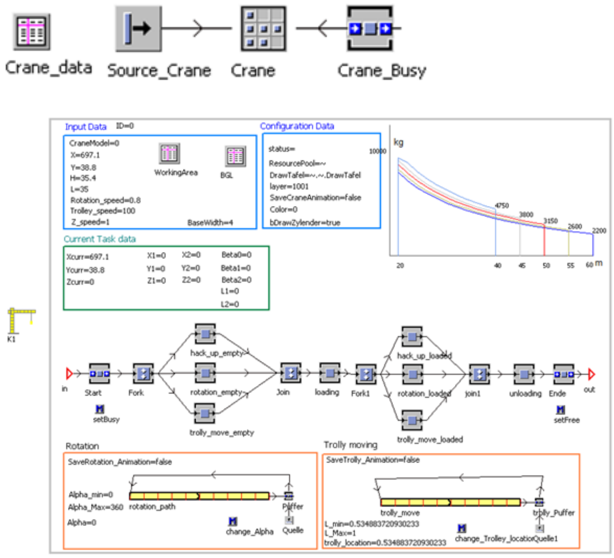


Figure 5: Simulation modules for tower cranes in different level of details

4. CASE STUDIES

4.1. Case Study I: Generic planning model

The first case study is a textbook activity network adopted from (Talbot and Patterson 1979). This study case was presented in (Shih-Ming Chen et al. 2012) using an intelligent scheduling system (IIS) in order to find the near optimum resource distribution for different project objectives. The project objective of this study case is assigned to be the minimum project duration. The activity network consists of twenty activities with equal priorities using six resources with daily availability constraints (Fig. 6).

The shortest project duration was 45 days which is 4 days shorter than the best result obtained from widely used commercial software using CPM method

Activity	Duration (days)	Predecessors	Daily resource requirements R1	R2	R3	R4	R5	R6
A	6	-	5	2	2	2	7	4
B	3	-	3	5	2	3	9	6
C	4	A	2	4	4	2	3	1
D	6	-	5	4	3	5	5	4
E	7	A, B	3	5	2	3	8	0
F	5	C	4	1	4	9	2	5
G	2	D	4	1	4	3	9	8
H	2	A, B	5	5	4	0	9	1
I	2	G, H	3	2	4	3	4	2
J	6	F	1	5	4	6	7	3

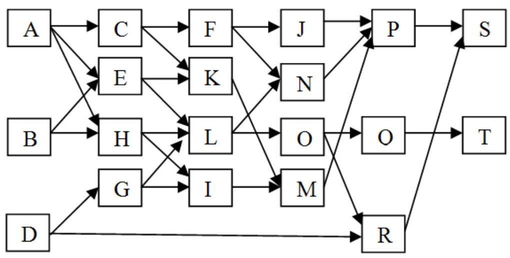


Figure 6: Activity network for the first study case

In this case study we are presenting how to use CST as a generic simulation framework. The definition of activities (duration, resource requirements) and the generic resources have been entered as direct input. Figure 7 shows the simulation model of this case. It consist mainly of 2 inputs components: Task list and Resource pool and three output components: Process pool, Gantt diagram, and a Project monitor to watch the utilization of resources.

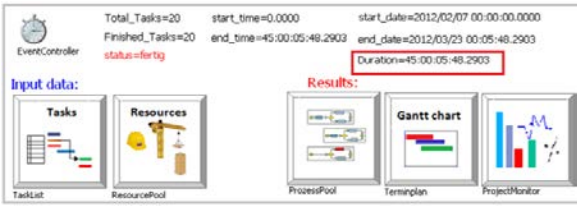


Figure 7: Simulation model for case study I

In this example the start time of all activities has been set to zero and the relationships between tasks and the definition of their resource requirements have been entered directly inside the "Task List" component. The resources R1-R6 defined inside the "Resource Pool" as generic resources and their daily resource limits have been set according to the table in Fig 6. The daily and weekly work time settings have been adjusted to ignore the default weekend days and daily work time. The shortest project duration we got after 100 simulations run was 45 days which is the same compared to the result obtained from IIS. Figure 8 shows the simulation results as Gantt chart and the resource utilization diagrams.

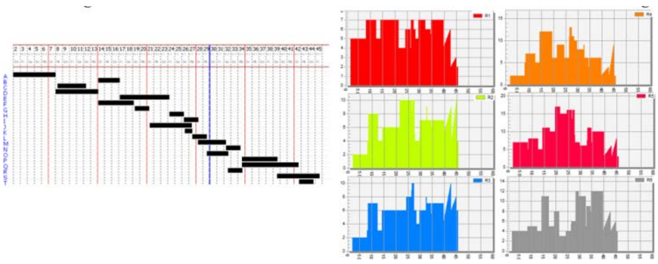


Figure 8: Simulation result of the study case I

4.2. Case Study II: Multi-Story Office Building

The second case study is a simulation model for 18-story office building. This sample project is based on a real project in Germany and was chosen as a reference project in the MEFISTO project. The goal of the simulation study for this sample project is to analyze and compare different planning options and "what if" scenarios. Some of the applied scenarios for this sample project are:

1. Showing the effect of changing the amount of any available key resource on the total duration of the structural work for one story or the whole building. This could be done for example by changing the number of workers, cranes, or formwork panels.
2. Comparing the resource utilization rates and the expected win or lose in the construction duration for two planning options (1) Each story consists of one working section and all workers are qualified to do any kind of structural work. (2) Each story consists of two work sections and the workers are divided to different teams with specific qualifications.
3. Generate detailed construction schedules based on primary schedules and visualize the construction work progress.

In this project a process catalog of all construction activities (formwork, steelwork, concrete work, etc.) for the walls, columns, and slabs was created including the definition of resources.

The graphical representation of the BPMN models made it is easy for all participants to understand how the simulation works. The modular structure and the flexibility of the simulation components allows with a minimum effort to analyze a lot of simulation scenarios using the same simulation model and finally deliver the simulation results in different formats as Gantt charts, 2D animation, 4D Simulation.

Figure 9 shows for example the effect of changing the number of workers (10-40 worker) on the total duration of the structural work for two stories, the utilization rates and the concrete consumption for the case of 12, and 24 workers. The result of this study case showed the advantages of using simulation technique effectively to solve a lot of planning problems which is not possible using traditional methods. Figures 10 and 11 show a part of the simulation results as 2D animation and 4D model.

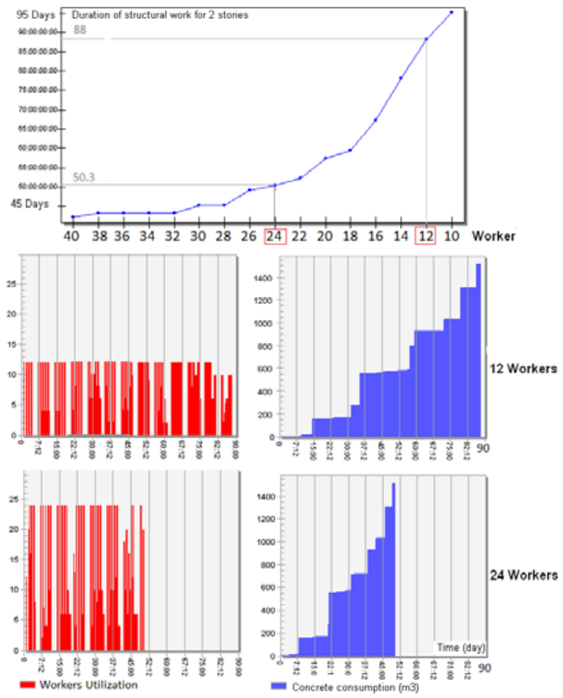


Figure 9: Effect of changing number of workers

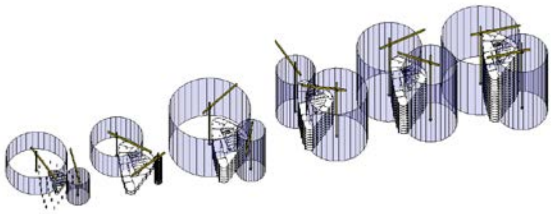


Figure 10: Construction progress as 4D model

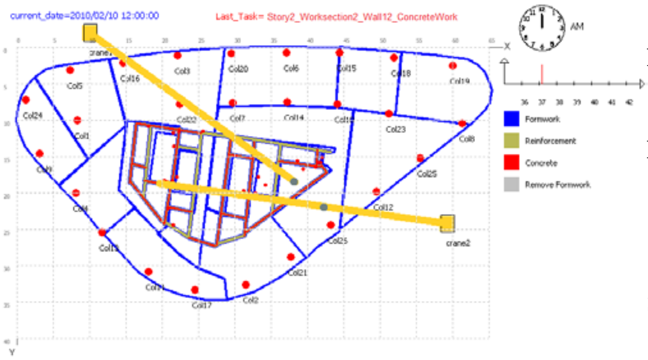


Figure 11: 2D animation during the simulation run

The full supplementary materials related to this case study and other simulation projects can be accessed online at: http://bci52.cib.bau.tu-dresden.de:3000

5. CONCLUSION AND FUTURE RESEARCH

This paper presented the development of a process-based construction simulation toolkit with a focus on the data integration between the basic models and the resource modeling. Two sample cases are presented to validate and explain the concepts behind the development and the possible applications of this toolkit. This research attempts to spread the acceptance and application of simulation technique to support the planning of real projects by allowing the rapid deployment of simulation models. Future work should focus on the integration with other models like the cost and risk models and also on the interaction between the production and logistic operations.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by the German Federal Ministry of Education and Research (BMWF) within the lead research project MEFISTO (http://www.mefisto-bau.de).

REFERENCES

AbouRizk S., (2011) A Construction Synthetic Environment Integrating Visualization and Simulation ConVR 2011- pp 11

Akhavian,R. Behzadan, H. A. 2011. "Dynamic Simulation of Construction Activities Using Real Time Field Data Collection" In European Group for Intelligent Computing in Engineering Workshop.

BPMN 2.0 Specifications http://www.omg.org/spec/BPMN/2.0

Ismail A., IFCWebServer project. 2011 http://www.ifcwebserver.org

Ismail, A. and Scherer R. 2011. "Process-based simulation library for construction project planning" In Proceedings of the 2011 Winter Simulation Conference. S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, eds.

Ismail, A., 2010 "Optimization of Resource Utilization and Project Schedule through an Object Oriented Simulation Library". Forum Bauinformatik Berlin

F. Talbot, J. Patterson. 1979 Optimal methods for scheduling projects under resource constraints. Proj. Manage. Q. pp 26-33

Lucko, G., P. C. Benjamin, and M. G. Madden. Harnessing the power of simulation in the project management / decision support aspects of the construction industry. In Proceedings of the Winter Simulation Conference (2008)

Shih-Ming Chen, F.H. (Bud). Griffis, Po-Han Chen, Luh-Mann Chang. 2012 "Simulation and analytical techniques for construction resource planning and scheduling". Automation in Construction.

Wagner, U., Ismail, A., Kog, F. & Scherer R. 2010. "About the Mapping Problem of Process to Simulation Models". In: Proceedings of ECPPM, Cork, Ireland.

AUTHORS BIOGRAPHY

RAIMAR SCHERER is a Professor and head of the Institute of Construction Informatics at Dresden University of Technology. He obtained his Ph.D. in stochastic earthquake engineering from Technical University of Munich in 1984. His research interest includes product and process modeling, concurrent engineering, virtual organizations, information logistics, Internet-enabled network infrastructures, Semantic Web technology, fuzzy and stochastic methods. He is the founder of the biannual European Conference on Product and Process Modeling (ECPPM) and member of the editorial board of 4 international journals. His email address is Raimar.Scherer@tu-dresden.de.

ALI ISMAIL is a Ph.D. student at the Institute of Construction Informatics. His research interest are in simulation and visualization of construction processes, Building information Modeling, business modeling , and simulation-based decision support for construction projects. He holds a M.S. in Civil Engineering from the Dresden University of Technology. His email address is ali.ismail@tu-dresden.de

AGENT-BASED MODELING TO ANALYZE THE IMPACT OF WORK STATUS ON CHILDBIRTH: A CASE STUDY OF KOREA

Mazhar Sajjad[1, 2], Chang Won Ahn[1, 2]

[1]University of Science and Technology (UST), Korea
[2]Electronics and Telecommunication Research Institute (ETRI), Korea)
[1]sajjad@etri.re.kr, [2]ahn@etri.re.kr

ABSTRACT

To maintain a desirable fertility level for each country population is the top priority. It is very well known that fertility has been steadily declined to a remarkably lower level in Korea. This study explores the relationship between women's work status patterns and fertility in the women's socio-economic context of South Korea. In this paper, our model applying combined micro-simulation and agent-based model (ABM) to investigate the economic influence by using agent's work status i.e. employment and education and its feedback on childbirth decision. This model appears to be a good approximation in describing the childbirth decision based on women's work status. The result shows that women without employment have more probability to go for childbirth than women having part or full time employment. Among employment women having full-time employment are more likely to become mothers than the part-time. Our study suggests that work status play an important role in childbirth transition in Korea.

Keywords: Total fertility rate, work status, trend and agent-based model, micro-simulation, macro-simulation

1. INTRODUCTION

Many developed countries are facing decline in population growth. Fertility has been steadily declined to a remarkably lower level in Korea. Young generations in recent years are intending to have one or two children at most. Korean government have major concerns about what kind of the economic forces should play a role in child birth decisions and what kind of policy could be effective. Korea completed fertility transition under the replacement level in the middle of the 1980s. The TFR (total fertility rate) continued to decline, reaching the lowest level of 1.076 in 2005, but has rebounded to 1.297 in 2012 [1].

The work status and fertility relationship is complex. The direction can be from work status to fertility or the other way around; the direction can be mutual; and there can be some common external factors affecting both behaviors [2]. The purpose of the paper is to analyze the relationship of women's work status patterns and childbirth in the socio-economic and institutional context of South Korea. The issue of the work status is put into perspective in the context of family childbirth decisions and to study the mechanism that create inter-dependencies among various social characteristics. Our model is trying to identify those features that may affect the relationship between work status and childbirth. In our research, we will concentrate on how work status dynamically affects the women childbirth. The model allows us to study demographic dynamics and interpret empirical data in terms of a formal theoretical structure.

Using agent-based modeling of population dynamics, it is possible to design women agent to be both autonomous and heterogeneous in their decision making about when and how many childbirth they would have or not [3]. This study will focus on one direction of the relationship: the role of woman work status in the childbirth decision. In order to head-off future problems, particularly anticipated economic problems, that will arise if the population continues to decline, policy makers and corporations alike have begun making changes to incentivize larger family sizes [8-9]. While important strides have been made, it is apparent that they have been largely unsuccessful. Contrary to the opinion of most policymakers, it is not a logical conclusion that enforcing macro-level policies will have any effect on micro-level behaviors related to fertility choices.

2. SOCIO-ECONOMIC CONTEXT OF SOUTH KOREA

Dating back to the 1950s and the 1960s, Korea was a country with rapid population growth, high population density, high fertility levels and little developed industrial economy [4]. In 1962, the Korean government implemented family planning program to control population growth [5]. Figure 1 shows that Korea's TFRs experienced a sharp decline from above 4 in the 1960s and the 1970s to below the replacement level in 1983. The period 1984-2000 witnessed a stagnated fertility decline, with TFRs lingering between 1.8 and 1.4. In sharp contrast with the fertility decline is the steady increase in women employment rates.

The Equal Employment Act in 1988, prohibited discrimination against women in employment, wages, and working conditions. It also prohibited discrimination against employed women due to marriage, pregnancy or child delivery. Meanwhile, a women public employee target system was established to enhance the recruitment of women in the public sector [6]. The Korean government has also made efforts to promote child care services to help women reduce the burden of childrearing and to facilitate for women to reconcile work and family life [7].

Figure 2, shows Korean women's labor force participation rate by ages and years. M-shaped curves appear for all calendar year periods. For decades, Korean women have followed a similar strategy to arrange their work-family life career: labor market entry – leaving labor market for family life – labor market re-entry when their children need less concern. The patterns indicate that for Korean women, temporarily sacrificing their career at a certain stage of life for the sake of family formation and expansion is entrenched with tradition. If we take a closer look, we can see that labor market exits have gradually shifted from ages 25-29 to ages 30-34. Besides, fewer women drop out of the labor market for family life in most recent years. Women aged 20-44 are more economically active in the latter periods.

3. MODEL DEVELOPMENT

3.1. Agent Based Modeling

Agent-based modeling is a bottom-up approach computational framework for simulating dynamic-processes that involves autonomous and heterogeneous agents having their own decision criteria and growing the population. The ABM offers a new approach to the problem of social phenomena in order to analyze the dynamic mutual-independent relationship among individual behaviors that could lead to emerging evolution. Agent based computational demography includes also micro-simulation that has been used to derive macro-level outcomes from empirical models of micro-level demographic processes, but also formal models of demographic behavior that describe micro-level decision with macro-level outcomes [10].

We are using ABM methodology to explore the economic influence of work status of an agent and its feedback on the childbirth decision. In population dynamics, it is obvious that women play a remarkable role in determining macro-level indexes. In ABM of population dynamics, it is possible to design women agent to be both autonomous and heterogeneous in their decision making about when and how many childbirth they would have or not [4]. Individual micro behavior results in a macro outcome which is could feedback to the individual decisions. Likewise, micro-macro independency could be obtained, which can't be modeled in a traditional micro-simulation [11].

The problem of addressing Korea's declining fertility rate is not simply a matter of predicting that the population is getting 'smaller', but addressing 'why', which inherently rests on individual choices that cause macro-level consequences. Only by addressing the actual ratio of women involved in Korean's population crisis can one begin to experiment with policies that might help turn the problem around.

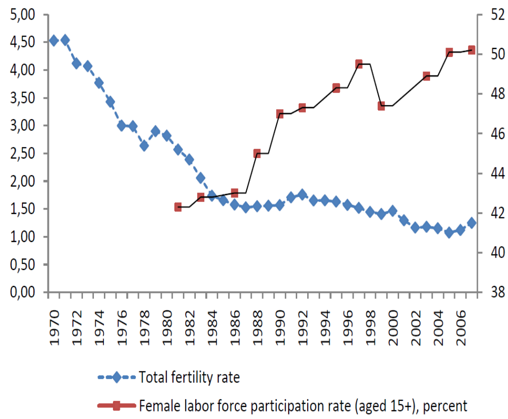


Figure.1. Development of total fertility rate and women labor force participation rate, Korea
Source: Korea Statistical Information System (KOSIS) & LABORSTA Labor Statistics Database, International Labor Organization.

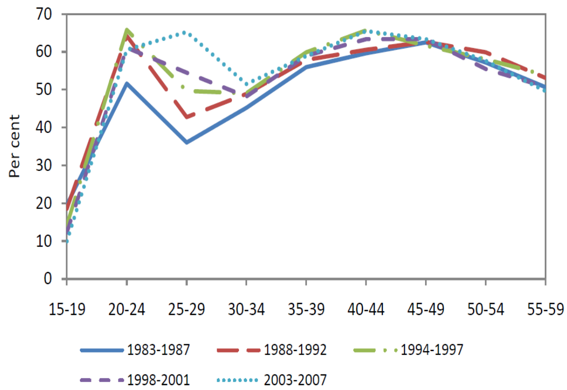


Figure.2. Women labor force participation rate of Korea by age since the early 1980s

Source: LABORSTA Labor Statistics Database, International Labor Organization (data unavailable for 1993 and 2002).

3.2. Model Specification

We setup a single sex model, limited our model to consider only 'woman agent' the most influential partner regarding childbirth decision, although both partners play a major role while deciding about childbirth. In our ABM, we can observe the aggregate effects of decentralized decision making without very strong assumptions to limit agent's behaviors to the scope of the current work. In our model the fertile period of an agent is from 16 to 45 years. In our model, We assume that only married women can give birth to child mean there is no cohabiting state in the context of Korea, every married woman is fertile, all mothers are real, the time spent between the childbirth and actual conception is ignored, the minimum childbirth space is one year, childbirth could be either single or twin and each woman agent has an expecting number of desired childbirth, but may not have as many children as expected due to the family economic outcome or other social factors.

We apply our model to analyze the causes and effects of woman work status on childbirth with the bottom-up approach using Korean census data from 1990 to 2010, since 1990, the Korea TFR (total fertility rate) declined rapidly as the country struggled through the financial crises in 1997, below 1.2 in 2000 and 1.08 in 2005, the lowest in the world [6]. Each individual agent has a household-member-id and three characteristics. The first and foremost time varying dependent variable 'agent work status' with job characteristics like full time, part time and never employed. Factors such as age and education level have also been indicated in many population studies as important in determining fertility rate [12], [13].

Therefore, our two time varying control variables are 'agent age' grouped into 16~20, 21~25, 26~30, 31~35, 36~40 and 40~45 years in order to analyze the relation of work status on various age groups. The 'agent education' status categorized into elementary, high, university in which woman are still enrolled. We are in the process to build agent-based simulation model that incorporates both historical census data on population characteristics that allow us to study the influence of work status on fertility decision by the interaction of various agent characteristics and behavior assumptions at different stages.

Table 1. Summary of agent variable and parameters

Agent Variable	Variable Values
Work Status	• Full Time • Part Time • Unemployed
Childbirth Age	• 16~20 • 21~25 • 26~30 • 31~35 • 36~40 • 41~45
Education Level	• School (elementary, middle, high) • College • University

3.3. Model Implementation

The model is implementing using Anylogic 7.0 professional software, both because of its simple, natural programming style as well as its user-friendly interface [14]. Before using empirical data in in micro-simulation, first we set behavior criteria for childbirth. After marriage, couples determine if they want to have children first based on the woman agent age, in our model the agent age must be within biological fertility range i.e. 15 to 45 years. The agent must be married, because regarding Korean population there is no custom of cohabitating fertility. After marriage the agent decide the desired family size and the minimum fertility gape based on the employment condition. The woman agent age and fertility gape can be tested again in the model after a period of one year in the simulation environment.

To examine the relationship of women employment and childbirth, we apply event history analysis using micro-simulation empirical data. Our model considering those women that follows the above set behavior criteria for childbirth. The observation starts when a woman reaches the age of sixteen till the age of forty-five. We apply Cox proportional hazard model depicted below.

$$h(t|X) = h(t)\exp(X_1\beta_1 + \cdots + X_p\beta_p).$$

Where h(t/X) is the hazard at time 't', the h(t) is the baseline hazard. The $X_{1,\ \ldots\ldots,}$ Xp are the covariates and 'βs' are the hazard ratios (HR) of the covariates. In our model, the covariates $X_{1,\ \ldots\ldots,}$ X_4 are women age, year, work-status and education level of the women respectively and β1,…… β4 are the respective hazard ratio of our four covariates (variables).

Considering the conventional social practice of Korea women leaving the job market for family life, we subtract nine months from the date of any first childbirth to capture the effect of pre-pregnancy work status and job characteristics on first childbirth. In our model, the woman can give maximum two kids at one time as normal fertility outcome. There are several variables that affect the outcome of childbirth, but here we are interesting to analyze the relationship of work status and fertility as shown in table-1. We are in the process to use various empirical models about initial demographic setup and transition from one state to another. Apart from these empirical methods, we are building the interactions model based on these agent characteristics at various stages that affecting the fertility over time.

Table 2, 3 and 4 shows the detail statistics of our sample data regarding the agent's variables. We are analyzing the Korean census data from 1990 to 2010 with five years interval and see the trends of childbirth event under various conditions. Our sample data includes the effect of 1997 Asian financial crises that disturbed the economy of Korea badly and 2005-2010 years data that was the time of economy recovery period after crises. Table-2 indicates the amount of time in months of the number of the women in a specific year and the number of childbirth event occurred in those women.

Table-2 Distributions of women time and childbirth event by main variables

Variables	Variables Categories	Women Time (Months)	Childbirth Event (First Birth)
Year	1990	27672	134
	1995	27290	120
	2000	26625	110
	2005	26245	102
	2010	25890	93
Work Status	Full Time	72052	242
	Part Time	33698	135
	Not Working	27972	182
Education Level	School	63214	268
	College	38243	114
	University	32265	177

Our main variable explanatory variable is the time varying employment status of Korean women. Here we are considering the three main categories: the full-time employment indicates that the women staying in the labor force, the part-time indicates that women is partially participating in the labor force and the un-employment indicates periods when the women have left the labor market.

Another time-varying variable is women education level. It is categorized into three broad groups: School level education combined elementary, middle and high school level. The college and university level indicates the college and university level education respectively.

4. EXPERIMENTAL RESULTS

Our result shows the relative childbirth event by main work status and other control covariates. Using the Cox proportional hazard model, we calculated the hazard ratio of our all four variables. Results shown in table-3 indicate the hazard ratios based on women's age, year, work status and education levels of the women. The hazard ratio-1 include only women's age and colander year to find the calendar year effect on women likelihood to become mother. The hazard ratio-2 and 3 include the work status and education levels to investigate the effects on childbirth.

The hazard ratio-2 indicates that women from age 26 to 30 have more than double intensity to become a mother than other age group. The overall reduction of first childbirth has been event detected. The involvement of work status, education level does not change this decline trend. The hazard ratio-2 includes work status for analysis. Women who have left the employment have more than 100% chances to become a mother than women who are full or part time employed. This discovery indicates that women who leave the work status usually go for childbirth, unemployment extent the family.

The hazard ratio-3 discovered the effects of education levels on childbirth. Women who completed their education or have college level education are most likely to become mothers than the low educated women. It shows that the highly educated women once get married then they are more committed toward childbirth. The agent based simulation interface in Anylogic is shown in figure-3.

Table 3. Childbirth event estimation from the main effect model

Variables	Category	Hazard Ratio-1	Hazard Ratio-2	Hazard Ratio-3
Women's Age	16 ~ 20	0.12 ***	0.27 ***	0.34 ***
	21 ~ 25	0.98 ***	1.00 ***	1.10 ***
	26 ~ 30	2.11 ***	2.06 ***	2.11 ***
	31 ~ 35	0.75 ***	0.67 ***	0.72 ***
	36 ~ 40	0.53 ***	0.46 ***	0.49 ***
	41 ~ 45	0.25 ***	0.18 ***	0.20 ***
Year	1990	1.70 ***	1.74 ***	1.71 ***
	1995	1.53 ***	1.58 ***	1.61 ***
	2000	1.12 ***	1.15 ***	1.14 ***
	2005	0.65 ***	0.68 ***	0.62 ***
	2010	0.58 ***	0.63 ***	0.60 ***
Work Status	Full-Time		1.23 ***	1.25 ***
	Part-Time		0.79 ***	0.92 ***
	Unemployment		1.95 ***	2.10 ***
Education	School		0.89 ***	1.05 ***
	College		1.15 ***	1.18***
	University		1.17 **	1.28 ***

Note: Statistical significance: ***: p<=0.01; **: p<=0.05; *: p<=0.10

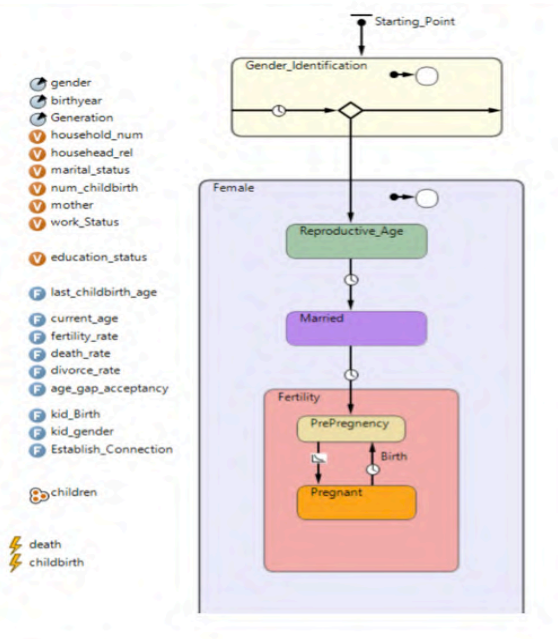


Figure-3 Agent Based Model Interface in Anylogic

4.1 Work Status Characteristics over Time

Our model composed of agent based simulation to find the relationship between work status and fertility. Various work status characteristics of each woman agent changes over time. Figure 3 shows how the hazard ratio by work status changes over time. The agent behavior shows a general decline of childbirth occurs in all types of women since 1990 irrespective of their work status. The result shows that in 1990 the women with part-time employment and with no employment had a high probability of becoming a mother than women in the labor force with full time employment. The Korean economic downturn in 1997 reflects more declines in unemployment and part-time employment than the full time employment. This decline reflects these women intensions to re-adjust their life career under the bad financial condition. When the country economy starts its survival after 2000, again we see positive feedback about childbirth.

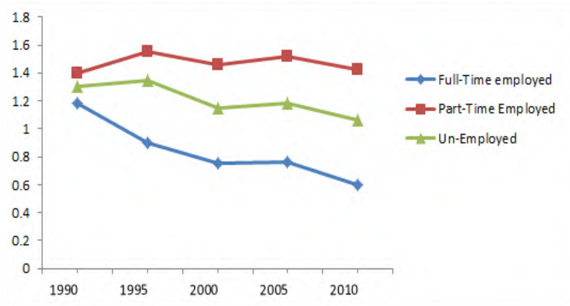


Figure 4 Relation of work status with year during childbirth

5. CONCLUSION

This study investigates the relation between the work status and fertility in the context of South Korea. The model develop in this paper appears to be a good first approximation in describing the fertility based on work status using agent-based simulation technique. We applied agent-based model to study the role of different characteristics interaction for explaining the observed demographic patterns on the transition to motherhood. I have applied event history analysis to the Korean census data to estimate how work status status and job characteristics are related to motherhood entry. Main effects models along with work status characteristics were specified to address our research questions.

The results showed that women who had left the labor market were more likely to become a mother than women continue as full-time or part time in the labor force. Another finding of the study relates to the effect of women education characteristics on motherhood entry. The results reveals that among women who completed their education and women at college level or above are most likely to become mothers. The higher educated women once gets married then those women are more committed to become mother than the lower educated women. Our results reveal that the structure of a society represented by parameters specifying work status patterns has the potential to alter the role of family policies, because the working women who become mothers are faced with the decision on whether to return to the labor force or when to return. Upon return, they have to face challenges with career opportunities for mothers.

References

Korean Statistical Information Service (KOSIS). “Population projections by age and Sex; Number of Incoming Foreigners by Type of Visa and Citizenship; Deaths by cause (103 items) by sex, by administrative Unit.; Population Density by Population Census Accessed”, September 10-22, 2013

Weller, Robert H. “Wife’s employment and cumulative family size in the United States”, 1970 and 1960. Demography. 14/1:43-65, 1970

Frydenlund, Erika. "Dynamic Attitudes about Motherhood: Modeling Family Planning Policies in Japan" [4]. Jones, Gavin and Richard Leete. “Asia’s Family planning programs as low fertility is attained. Studies in Family Planning”. 33/1:114-126, 2002

Rhee, Ock. “Childcare policy in Korea: Current status and major issues, International journal of Child Care and Education Policy”, 1/1:59:72, 2007

Cho, Namhoon. “Policy responses to population ageing and population decline in Korea.Expert group meeting on policy responses to population ageing and population decline. Population Division, Department of Economic and Social affairs”, United Nations Secretariat. Policy responses to population ageing and population decline in Korea. Expert group meeting on policy responses to population ageing and population decline. Population Division, Department of Economic and Social affairs, United Nations Secretariat, October, 2000

Organization for Economic Co-operation and Development (OECD).OECD Economic Surveys: Korea Overview. Paris: OECD, 2012.

Boling, Patricia. “Demography, Culture, and Policy: Understanding Japan’s Low Fertility.” Population and Development Review. Vol. 34. No. 2 : 307-326, 2008

Lewis, Leo. “Go Home and Multiply! Bank Gives Staff an Early Night to Halt Japan’s Falling Birthrate”, The Times. 18 November, 2009

Billari F, “Bridging the gap between micro-demography and macro-demography, in Demography: analysis and synthesis”, Caselli G., Vallin J. & Wunsch G. (Eds.), Academic Press (Elsevier), New York, Vol. 4, pp. 695-707, 2006

Žamac, Jovan, Daniel Hallberg, and Thomas Lindh. "Low fertility and long-run growth in an economy with a large public sector", European Journal of Population/Revue européenne de Démographie, 183-205, 2010

Shirahase, Sawako. “Female’s Increased Higher Education and the Declining Fertility Rate in Japan”, Review of Population and Social Policy. No. 9 : 47-63, 2000

Kaplan, Hillard S. and Jane B. Lancaster. “The Evolutionary Economics and Psychology of Demographic Transition to Low Fertility”, In: Adaptation and Human Behavior: An An thropological Perspective. Lee Cronk. et al (eds). New York: Walter de Gruyter, Inc. 283-322, 2000

http://www.anylogic.com

Cristina Sousa Gomes, Isabel Tiago de Oliveira,Maria Luis Rocha Pinto, Miguel Cabrita,“Fertility, Full-time and Part-time Female Employment in Europe”, CIES e-Working Papers Av. das Forças Armadas, Edifício ISCTE, 1649-026 LISBOA, PORTUGAL

MULTIDISCIPLINARY COORDINATION OF ON-SCENE COMMAND TEAMS IN VIRTUAL EMERGENCY EXERCISES

Theo van Ruijven[(a)], Igor Mayer[(b)], Mark de Bruijne[(c)]

[(a)] Consultant at Capgemini Consulting and Guest Researcher at Delft University of Technology
[(b)] Associate Professor at Delft University of Technology
[(c)] Assistant Professor at Delft University of Technology

[(a)] t.w.j.vanruijven@tudelft.nl, [(b)]i.s.mayer@tudelft.nl, [(c)]m.l.c.debruijne@tudelft.nl

ABSTRACT
The authors present the design and results of a comparative study into multidisciplinary on-scene command teams at work in virtual emergency training exercises. The main question of the study is: How do on-scene command teams coordinate multidisciplinary objectives and tasks, and how does the way this is done determine their performance? The study involves 20 'on-scene command teams' consisting of multiple disciplines, such as police, fire and medical services, municipal officers and infrastructure operators, in a safety region in the Netherlands. Integral video observation by five synchronized cameras was used to capture the coordination processes during the virtual exercises. These integral and synchronized video recordings were then coded. Performance was operationalized by scoring the progress and completion of emergency management tasks for which individual members and/or the team as a whole were responsible. Team coordination was operationalized through network centrality and density measures. Analysis of the data shows that there is wide variation within and among the teams with regard to emergency management performance and coordination patterns. Significant findings are: 1) decentralized coordination is an important factor in emergency management performance; 2) teams that use less coordination during the intermediate phase of emergency management perform significantly better; 3) actors that have a central position in the network achieve their own goals better.

Keywords: emergency response, crisis management, team performance, serious games, game-based training, virtual team training, video observation, network analysis, XVR™

1. INTRODUCTION

Emergency response refers to all operational and procedural tasks conducted individually or jointly by qualified professionals, aimed at normalizing a situation after it has been disrupted by an incident or accident (Haddow, Bullock, and Coppola 2003; Chen et al. 2008). It generally involves tasks such as rescue, medical aid, policing, evacuation, rerouting of traffic, fire-fighting, containment of chemical spills, and many other roles. A wide range of responsibilities are attributed to specific emergency response disciplines such as ambulance and fire services, and the police. *First responders* arriving at the scene of an incident or accident will commonly follow a set of standard operating procedures (SOPs), for which they are trained in professional education. These skills are further developed and maintained in drills and simulations. For safety and efficiency reasons, such training is increasingly 'virtual' and game-based (Harteveld 2012; Benjamins and Rothkrantz 2007; Haferkamp et al. 2011; Mcgrath and McGrath 2005).

Incidents and accidents commonly occur in and around vital infrastructure, such as railways, roads, waterways, power stations and airports. In modern societies, even a small incident or accident may cause considerable disruption of the infrastructure (traffic congestion and power blackout), causing socio-economic loss that may even lead to a political-institutional crisis. Rapid repair and recovery of disrupted infrastructure is therefore an essential part of emergency response. This brings even more actors onto the scene, such as road inspectors and repair men, infrastructure operators, public utility managers and municipal officials, who all have individual objectives that they try to achieve with their own SOPs. Therefore, the objectives and procedures of all actors on the scene need to be prioritized, aligned and sequenced to form an integrated response.

Emergency (response) management refers to all tactical and strategic tasks aimed at the smooth operation of emergency response services, either during the proactive or 'cold' phase, such as planning and training, or the reactive or 'hot' phase. During the 'hot' phase, emergency response management is generally undertaken by a team consisting of representatives of the disciplines involved – a blending of mono-disciplinary professionalism and multidisciplinary teamwork. In the Netherlands, such teams are referred to as COPI, an acronym for *on-scene command team*. In the early phases of an incident or accident, an on-scene command team is faced with a great many, possibly conflicting, objectives and SOPs. Arriving at the scene, a command team needs to prioritize, align and sequence these objectives rapidly. In other words, on-scene command teams must *coordinate* in the most effective manner (Chen et al. 2008). How do they do this? What forms of coordination work well and which do not?

2. STUDY DESIGN

The main research question addressed in this paper is: *How do on-scene command teams coordinate multidisciplinary objectives and procedures, and how does the way this is done determine their emergency management performance?* This question is highly relevant, given the increasing complexity of emergency response management and its consequential gains and losses. Furthermore, there is little empirical understanding about what on-scene command teams actually do and what makes one team more effective than another.

Research into this matter is far from simple. Real-time observation during an emergency faces all kinds of practical complications, scientific limitations and moral objections. Incidents are chaotic and response management is dispersed and lengthy (hours to weeks), thereby requiring a considerable number of observers, or some other way of logging and tracking interactions. The occurrence of incidents and accidents is unpredictable, demanding researchers to 'stand by' over a longer period of time. Such factors make it virtually impossible to collect quantitative or quantifiable data while on-scene command teams are in the midst of a crisis.

Advances in research methods have used the tracking of digital communication during a crisis (data from mobile phone network, etc.) but this type of data is meaningful only for specific purposes in larger scale events (such as determining location and movements) (Landgren and Nulden 2007). Furthermore, it seems highly impractical and unethical to record on camera what first responders do, let alone to distribute questionnaires. The potential suffering of real victims may be confrontational and emotional, not a good context to start 'counting' or 'coding'. Structured interviews are likely to interfere and disturb the performance of an on-scene command team. In short, with respect to emergencies, researchers at best observe in a *stealth mode*. Although this has generally delivered valuable insights, this type of research tends to be evaluative, case-based, qualitative and interpretative if not anecdotal. Other more objective forms of research (i.e., explanatory, comparative and quantitative) seem almost impossible.

There is however a compromise solution, where larger amounts of qualitative data about multiple emergency management events can be turned into quantitative data for comparative analysis. Our study into the coordination and performance of on-scene command teams was carried out using observations of 20 teams of professionals at work in four different scenarios of virtual emergency response training. The four scenarios concerned: 1) Tunnel hazardous materials; 2) Tunnel evacuation; 3) Urban hazardous materials; 4) Port carbon monoxide scenario. The research was conducted between 2011 and 2014 with the support of one of the 25 safety regions in the Netherlands. Written permission was granted by the safety region and participants to make video recordings during all sessions. Results were anonymized and cannot be traced back to individuals or teams. Participants were operational officers – novices to experienced seniors – in one of the relevant disciplines and working in the same safety region, including police, fire services, medical emergency services, the municipality and infrastructure operators.

An on-scene command team is usually composed of officers on duty. As is the case in reality, many of the participants in the exercises did not know each other. The virtual training exercises were part of a mandatory training program for COPI members. One participant in the virtual on-scene command team was assigned a role as leading officer and another as information manager.

The virtual emergency training sessions were prepared and operated in XVR™, based on the Quest3D™ game engine, provided under license by a Dutch company. The XVR™ system allows users to build authentic 3D representations of an emergency situation with building blocks such as cars, victims, hazardous substances, fire and explosions, etc. The Tunnel Scenarios used an accurate 3D model of an actual tunnel in the region. The exercises were managed by an experienced team of professional facilitators who moderated the different aspects of the sessions, from logistics, computer technology and player briefings and debriefings. A technical facilitator operated and supported the player-computer interactions, allowing participants to focus on their task rather than on the controls.

Five synchronized cameras were placed in two adjacent rooms. One camera captured the plenary meetings of the on-scene command team, as well as the briefings and debriefings. Four cameras – roughly one for each actor – were placed in the situational assessment room, also called "the field room," where on-scene command team members engaged individually with the virtual environment while monitoring the virtual emergency situation on a large projection screen. The full on-scene command team interaction for each session was captured through time-synchronization of the five cameras. All data were coded and analyzed afterwards by the first author using the Transana video transcription software package. In addition, a questionnaire (approx. 15 min to fill out) was handed out to collect information about the participants and their experience of the virtual exercise. As the main researcher, the first author also logged notable session information, such as start and end times, facilitators, players, previous virtual training experience, technical or other disturbances. Figure 1 sketches the setting of the exercise and its observation, giving an impression of the players at work.

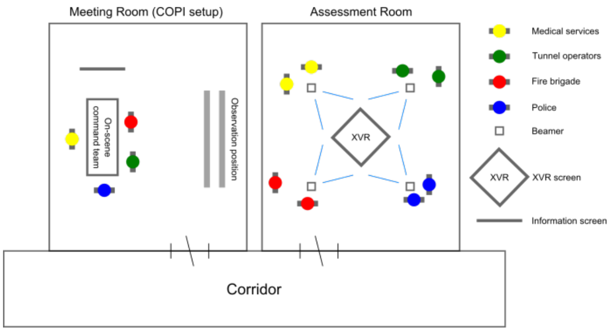


Figure 1: Setting

2.1. Centralized and decentralized coordination

Communication and coordination are crucial factors in explaining team performance (Bettenhausen 1991; Comfort 2007; J. Mathieu et al. 2008). Previous research has found that initial temporal planning contributes to time awareness, coordination and task performance in self-managing teams (Janicik and Bartel 2003). Information processing and the spread of information over emergency response actors has been found to be crucial for explaining crisis management performance (Kapucu 2005). Other studies of communication and coordination networks in emergency management teams have concluded that network analysis is a useful method for studying command and control (Houghton et al. 2006).

To conceptualize and operationalize on-scene command team coordination, we adapted a model presented by Marks et al. (Marks, Mathieu, and Zaccaro 2001). The original model (see Figure 2) breaks down the effort of emergency management teams into episodes called (a) action phases, where the team does the operational work, and (b) transition phases, where the team focuses on evaluation and planning (J. E. Mathieu and Schulze 2006; J. Mathieu et al. 2008; Ilgen et al. 2005). During the phases, Inputs (I) are transformed into Outputs (O) by Procedures (P). Temporally, the input of one phase is the output of another phase.

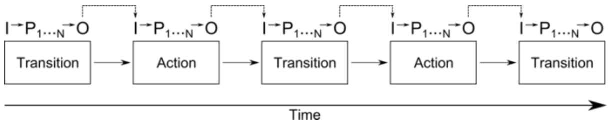


Figure 2: Recurring phase model of team performance (Marks et al. 2001)

Viewed in this way, the various emergency response services conduct their respective operational tasks during the action phases. Since the SOPs may not be aligned and need to be prioritized and sequenced, coordination is highly important. The transition phases consist of meetings of an on-scene command team, where the emergency response is coordinated. The output of the coordination meetings, therefore, can be seen as the input for the operational activities and vice versa.

This straightforward picture of coordination in on-scene command teams, however, is not fully accurate. During the meetings of on-scene command teams, all actors come together at the same location to coordinate the emergency response activities of their respective disciplines. The effectiveness of a single sequence of coordination meetings is disrupted by the chaotic nature of an emergency, most importantly due to time pressure. Many tasks, such as firefighting, rescue or rerouting of traffic, need an immediate response. In other words, they cannot wait until the next coordination meeting, although they do require some immediate form of coordination with other actors.

Another issue is that not all tasks require coordination by all actors. They can be coordinated by a subset of actors, for instance bilaterally or tripartite. Last but not least, on-scene command teams are more like a network of actors than a hierarchy (Leukfeldt et al. 2007). There is no hierarchical leader who takes, oversees or enforces the major decisions, and there are no subordinate actors who simply execute their tasks. Each member of an on-scene command team is responsible for their own decisions with respect to his or her crew, equipment and actions. Since there is no formal hierarchical leader in the form of a single coordinator and person in charge, there is considerable variety in the way coordination in an on-scene command team is organized. To a large extent, the coordination of on-scene command teams depends upon situational circumstances, although it may also depend upon factors related to team composition, team structure and team leadership. We therefore revised Marks' framework, as presented in Figure 3.

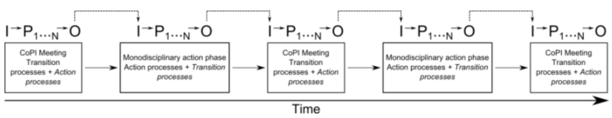


Figure 3: Revised framework of coordination in on-scene command teams

2.2. Hypotheses

Centralized coordination in an on-scene command team meeting is important, but when it is applied too rigidly and is not in sync with decentralized coordination, the team will be less effective and slow down. Decentralized coordination seems necessary to get things done immediately, while centralized coordination seems necessary to maintain sight of the bigger picture. A lack of decentralized coordination may lead to ineffective, slow emergency response. Hence, our first hypothesis (*H1*) is that more decentralized coordination will lead to better emergence response performance by a command team. Correspondingly, a lesser amount of decentralized coordination will contribute to poorer emergency response performance by an on-scene command team.

Decentralized coordination also gives more space to individual actors to pursue their own interests, to achieve their own goals and to coordinate their own preferred standard operating procedures. Hence, our second hypothesis (*H2*) argues that the active involvement of an actor in decentralized coordination is beneficial for the performance of tasks for which this actor is responsible.

2.3. Operationalization of performance

To determine and compare emergency management performance, the observable progress of each individual task during an exercise was coded over time. The continuous nature of time in a training exercise was turned into three periodic intervals – start, middle and end – which correspond with natural steps in the emergency response exercise. Tasks for which no progress was observed at the end of a time interval were coded as '0'. Tasks that had started at the end of an interval but had not yet been finished were coded '1', and completed tasks were coded '2'. A task finished at an earlier interval continued to be coded '2' at later intervals. Thus, a task completed early on in the exercise received a higher score than the same task completed at the end of the exercise.

The end performance of an individual actor could now be calculated as the sum of all scores attributed to the tasks for which this actor was responsible. The end performance of an on-scene command team could then be calculated by summating all scores for all tasks by all actors. Since the number and nature of tasks differed among the four training scenarios, we could only compare the end performances of teams playing the same scenario. Furthermore, some actors were responsible for only one task, while most other actors were responsible for two or more tasks. Thus, the performance scores of actors in the same team and the performance of teams playing different scenarios could not be compared.

In order to achieve a better standard for comparative analysis, all scores were normalized to give a value between 0 (lowest performing actor or team) and 1 (highest performing actor or team), with scores in between based upon distance to highest and lowest team/actor. The normalized scores allow comparison of the performance across all actors and teams, despite some actors having a relatively light task, while others faced severe challenges. The various performance indicators are listed in Table 1.

Level	Indicator	Operationalization
Normalized team performance	Summation of all scores of all tasks for which a team is responsible	Assigning 0, 1, 2 score based on the progress a task has made at each interval, normalized between 0 and 1
Actor performance	Summation of all scores of the tasks for which the actor is responsible	Assigning 0, 1, 2 score based on the progress a task has made at each interval, normalized between 0 and 1

Table 1: Performance indicators

2.4. Operationalization of coordination

In order to measure coordination, we operationalized it into several strong indicators that could be coded from the videos. We decided to take communication – that is, actors talking to each other during the exercise – as a proxy for decentralized coordination. We realize that communication and coordination are not identical, but communication is at least a prerequisite for coordination. Network theory was used to construct the coordination indicators. Networks are webs of ties or links (e.g., communication flows, transportation lines) interconnected by nodes or points (e.g., actors or hubs) (Scott 2000; Kossinets and Watts 2006; Borgatti et al. 2009; Scott 1988).

For our purposes, several indicators can be developed to measure ties (communication) and nodes (actors). A tie can be analyzed in a binary fashion: it exists (1) or not (0). However, ties can also be weighted, for instance on the basis of *intensity*, giving us weaker and stronger ties, commonly pictured by the weight of a line between two nodes. In communication networks, weight may be attributed, for instance, on the basis of the *duration* of the communication or the *amount of data* that is communicated.

Communication between actors during action phases can also be taken as a proxy for decentralized coordination. The amount of decentralized coordination in a team can be measured by examining the communication network's *interconnectedness*, here called *density*. Network density is the number of actual ties (i.e., the number of communication lines between two actors) in a network as a proportion of the maximum number of all possible ties. Network density can also be weighted by taking into account the weight of the ties. In our case, the network density of an on-scene command team can be calculated by looking at the duration of all communication among all actors as a proportion of the maximum number of possible ties and their maximum possible duration.

To determine the importance of an individual node within a network we can use *network centrality* measures, commonly pictured by the size of a node in a network graph (Opsahl, Agneessens, and Skvoretz 2010; Newman 2005; Borgatti 2005; Okamoto, Chen, and Li 2008). *Degree centrality* is derived by calculating the number of ties that one node has with other nodes. Taking into account the direction of communication – sending or receiving – the *In* and/or *Out-degree centrality* of a node can also be calculated. *Betweenness centrality* is an indicator of the intermediate positions of a node in-between other nodes. A high betweenness centrality score implies that an actor is an important 'hub' of information (Scott, 2011). The above indicators can again be weighted by taking into account the duration of the communication. For reasons of simplicity, we decided to develop a weighted indicator for degree centrality only.

Another issue in the analysis was that we needed to decide upon the relative importance of the number of ties (the degree) in relation to the importance of the weight of ties (the strength). In other words, is it more important to be connected to many other team members or to be connected to them for a longer period of time? Opsahl et al. (Opsahl, Agneessens, and Skvoretz 2010) proposed using a *tuning parameter* for weighted degree centrality. Setting the parameter at 0 implies that strength is disregarded, and the indicator thus becomes identical to degree centrality. Setting the parameter at 1 means that the indicator disregards the number of ties. In our analysis below, we set the tuning parameter at 0.5, allowing as much influence to the number of ties as to their weight. Table 2 gives an overview of all of the network density and centrality indicators.

Indicator	Definition	Operationalization
Density	The number of ties as a proportion of the maximum number of ties within a network	The number of communicating emergency response actors as a proportion of the maximum number of communicating emergency response actors
Weighted degree centrality	The number of ties and their weights as a proportion of the maximum number of ties and their maximum weights within a network	The number of communicating emergency response actors and the duration of their communications as a proportion of the maximum number of communicating actors and their maximum duration
Degree centrality	The number of ties between a node and other nodes	The number of emergency response actors that an actor in question is communicating with
In-degree centrality	The number of ties that run from other nodes to the node in question	The number of emergency response actors that communicate to an actor in question
Out-degree centrality	The number of ties that run from the node in question to other nodes	The number of emergency response actors that an actor in question is communicating to
Betweenness centrality	The number of shortest paths between node pairs that pass through the node of interest	The number of times that an actor in question is at the shortest communication path between two other emergency response actors
Closeness centrality	The total graph-theoretic distance of a given node from all other nodes	The distance of an actor in question to all other emergency response actors
Weighted degree centrality	The number of ties between a node and other nodes, adjusted for the weight of the ties (the tuning parameter – alpha – is set at 0.5 in this study)	The number of emergency response actors that an actor in question is communicating with, weighted by the duration of the communication
Weighted in-degree centrality	The number of ties that run from other nodes to the node in question, adjusted for the weight of the ties (alpha = 0.5)	The number of emergency response actors that communicate to an actor in question, weighted by the duration of the communication
Weighted out-degree centrality	The number of ties that run from the node	The number of emergency response actors

	in question to other nodes, adjusted for the weight of the ties (alpha = 0.5)	that an actor in question is communicating to, weighted by the duration of the communication

Table 2: Coordination indicators

3. RESULTS

3.1. Performance

The assessment of the collected data reveals variation in how on-scene command teams perform and how they coordinate. Teams playing the same scenario show marked variety in the sequence of tasks. Furthermore, some teams are significantly better than others; they finish more tasks and are quicker to finish them. Table 3 gives an overview of each team's end score before normalization. Table 4 shows the ranking of the teams after normalization.

Team/ scenario	Tunnel hazmat	Tunnel evacuation	Urban hazmat	Port CO
1	48	41	42	47
2	52	45	45	52
3	50	68	46	43
4	53	52	53	56
5	45		56	
6	46		54	

Table 3 - Team performance (before normalization)

Team/ scenario	Tunnel hazmat	Tunnel evacuation	Urban hazmat	Port CO
Best performing	Team 4 (1.0)	Team 3 (1.0)	Team 5 (1.0)	Team 4 (1.0)
	Team 2 (0.88)	Team 4 (0.41)	Team 6 (0.86)	Team 2 (0.54)
	Team 3 (0.63)	Team 2 (0.15)	Team 4 (0.79)	Team 1 (0.31)
	Team 1 (0.38)	Team 1 (0.0)	Team 3 (0.29)	Team 3 (0.0)
	Team 6 (0.13)		Team 2 (0.21)	
Worst performing	Team 5 (0.0)		Team 1 (0.0)	

Table 4: Team performance (after normalization)

Individual actors also show great variety in their level of performance and coordination. Table 5 shows the normalized performance of the fire services as an example.

Team/ scenario	Tunnel hazmat	Tunnel evacuation	Urban hazmat	Port CO
Best performing	Fire services 2 (1.0)	Fire services 3 (1.0)	Fire services 5 (1.0)	Fire services 4 (1.0)
	Fire services 6 (0.71)	Fire services 1 (0.13)	Fire services 4 (0.86)	Fire services 2 (0.83)
	Fire services 5 (0.57)	Fire services 4 (0.13)	Fire services 2 (0.71)	Fire services 3 (0.67)
	Fire services 4 (0.43)	Fire services 2 (0.0)	Fire services 3 (0.57)	Fire services 1 (0.0)
	Fire services 3 (0.29)		Fire services 1 (0.43)	
Worst performing	Fire services 1 (0.0)		Fire services 6 (0.0)	

Table 5: Performance of the fire services

3.2. Coordination at team level

Decentralized coordination within the teams is measured through *network density*, which varies between 0.16 and 0.34, with a mean of 0.25. This implies that one-third of the members in the high density team spoke with each other during the exercises, while in the low density team this is one sixth. The variation in the *weighted network density* is greater, and varies between 0.05 and 0.29, with a mean of 0.12. In conclusion, in some teams, the actors engage in coordination significantly more than in other teams. Table 6 shows the *normalized weighted densities* of the 20 teams, ranked from high to low density.

Team/ scenario	Tunnel hazmat	Tunnel evacuation	Urban hazmat	Port CO
Highest density (1.0)	Team 6 (1.0)	Team 3 (1.0)	Team 3 (1.0)	Team 3 (1.0)
	Team 2 (0.83)	Team 1 (0.71)	Team 4 (0.86)	Team 2 (0.60)
	Team 1 (0.27)	Team 4 (0.70)	Team 6 (0.85)	Team 1 (0.05)
	Team 3 (0.25)	Team 2 (0.0)	Team 1 (0.71)	Team 4 (0.0)
	Team 5 (0.25)		Team 2 (0.14)	
Lowest density (0.0)	Team 4 (0.0)		Team 5 (0.0)	

Table 6: Team coordination based upon normalized weighted density

3.3. Coordination at actor level

The centrality of individual emergency response actors in the networks varies in different ways. On average, the actors communicate with 3.7 other actors, with a standard deviation of 1.4. The weighted centrality of actors varies more strongly, with an average of 465 and a standard deviation of 307. The standard deviation indicates that the majority of the team members have a weighted degree centrality between 158 and 772, which is a substantial range. In sum, the amount of coordination in which individual emergency response actors are involved varies significantly (see Table 7).

Team/ scenario	Tunnel hazmat	Tunnel evacuation	Urban hazmat	Port CO
Most coordination (1.0)	Fire services 6 (1.0)	Fire services 3 (1.0)	Fire services 6 (1.0)	Fire services 3 (1.0)
	Fire services 5 (0.41)	Fire services 4 (0.24)	Fire services 4 (0.95)	Fire services 2 (0.59)
	Fire services 3 (0.29)	Fire services 1 (0.15)	Fire services 3 (0.76)	Fire services 4 (0.26)
	Fire services 1 (0.29)	Fire services 2 (0.0)	Fire services 5 (0.25)	Fire services 1 (0.0)
	Fire services 4 (0.01)		Fire services 1 (0.21)	
Least coordination (0.0)	Fire services 2 (0.0)		Fire services 2 (0.0)	

Table 7: Coordination of fire services (after normalization)

4. TESTING HYPOTHESES

H1: Teams with a high level of decentralized coordination show better team performance than teams with a low level of decentralized coordination.

Figures 4 and 5 give an overview of each team's weighted network density in relation to team performance. The correlations are weak and statistically not significant. More and longer decentralized coordination – that is, speaking with each other – does not seem to lead to better performance. The hypothesis is rejected.

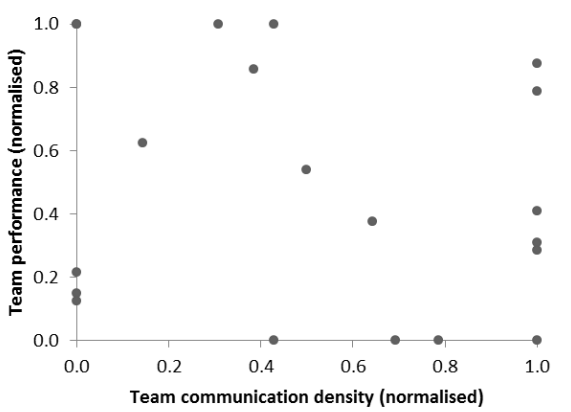


Figure 4: Communication density and performance of on-scene command teams

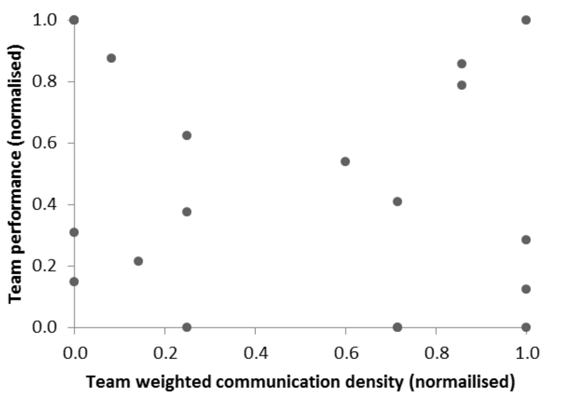


Figure 5: Weighted communication density and performance of on-scene command teams

Following the temporal framework of team coordination (Figures 2 and 3), we decided to analyze how coordination and performance were related 'over time'. We broke up the exercises into three different phases: 1) the initial phase, before the first team meeting; 2) the intermediate phase, between the first and second team meetings; 3) final phase, between the second and last team meetings. We differentiated the density and weighted density for the three episodes and then correlated the latter with overall team performance. This analysis provided interesting results. Again, coordination during the initial and final phases has no significant correlation with performance. Coordination during the intermediate phase of emergency response, however, is significantly negatively correlated with team performance (-.473*). Less coordination during the intermediate phase of emergency response seems to lead to better overall team performance. The normalized outcomes of team performance and weighted communication network density are plotted in Figure 6.

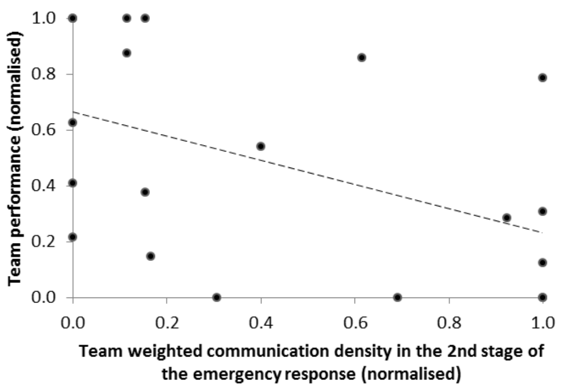


Figure 6: Weighted communication density and team performance (intermediate phase)

H2: Actors who coordinate more with other actors have a better actor performance.

Figure 7 shows the normalized weighted degree centrality – the indicator for the occurrence and duration of coordination – for all actors in relation to their performance. We found no statistically significant correlation. Being the centre point of the network does not yield better individual performance. Breaking up the exercise into three phases does not lead to refined conclusions. The hypothesis is rejected.

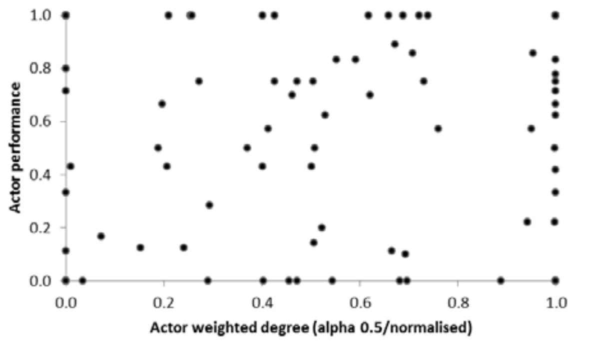


Figure 7 - Weigthed communication density of all actors in relation to their performance

Further analysis using advanced centrality measures yielded two significant results. The *in-degree* of emergency management actors during the *intermediate phase* of emergency response is positively and significantly correlated (.29**) to actor performance. The network graphs for the fire services (in Black) in the Tunnel evacuation scenario are presented in table 8. Network graphs in Figures 8 to 10 are based upon the weighted centrality of the actors, which means that the size of the nodes indicates the amount of actor communication. Network graphs in Figures 11 to 13 are based upon in-degree centrality. The larger the node, the more communication received by the actor and the higher the actor's status. The differences in results between weighted centrality and in-degree are striking. The fire services in Team One performed much better than those in the other teams.

There is also a significant correlation between the *weighted betweenness centrality* of actors and their performance (.27**). The network graphs for the medical services (in Black) based upon weighted degree centrality and weighted betweenness centrality in the Tunnel hazardous materials scenario are presented in table 9, with network graphs 14-16 indicating *weighted degree centrality* and Figure 17-19 indicating *weighted betweenness centrality*. Actors with a central position in the network have a higher betweenness centrality, indicated by a large node. Eccentric actors have no betweenness centrality and a small node.

Figure 8 - WST evacuation scenario, Team One: medium weighted degree centrality	Figure 9 - WST evacuation scenario, Team Two: high weighted degree centrality	Figure 10 - WST evacuation scenario, Team Three: low weighted degree centrality

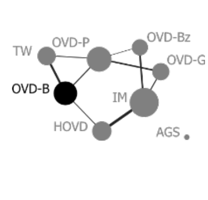
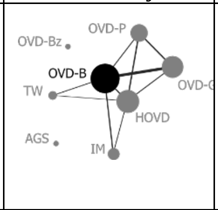
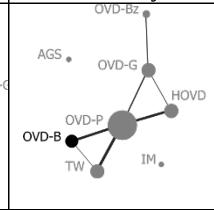

		

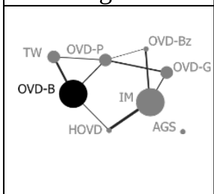
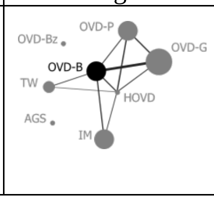
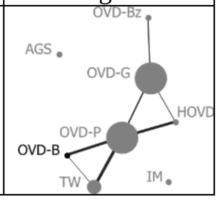
Figure 11 - WST evacuation scenario, Team One: high in-degree	Figure 12 - WST Evacuation scenario, Team Two: medium in-degree	Figure 13 - WST Evacuation scenario, Team Three: low in-degree
		

Table 8 - Network graphs of the fire services in the Tunnel Evacuation Scenario
(*Legend: AGS = advisor on hazardous materials, IM = information manager, HOVD = chief officer, OVD-B = fire services, OVD-Bz = municipality, OVD-G = medical emergency services, OVD-P = police, TW = tunnel operator*)

Visual comparison of the two types of centrality-based networks suggests a correlation between the two measures. This is confirmed statistically (.34**). The relation between *weighted betweenness* and non-weighted degree is even stronger (.54**). Taking into consideration that the weighted betweenness centrality was set at .5 (see above), the increase in correlation suggests that it is primarily the relation with an actor, and not the duration of the communication, that matters. The fact that *betweenness centrality* is positively and moderately correlated with actor performance indicates that it is the position of an actor in the network that is most important for this actor's performance.

Figure 14 - WST hazardous materials scenario, Team One: weighted degree centrality	Figure 15 - WST hazardous materials scenario, Team Two: weighted degree centrality	Figure 16 - WST hazardous materials scenario, Team Six: weighted degree centrality

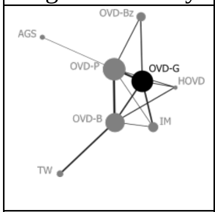
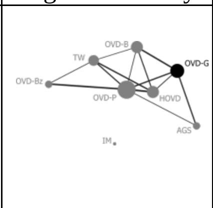
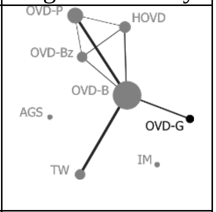

		

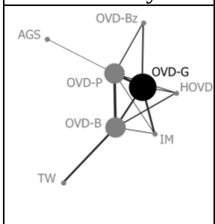
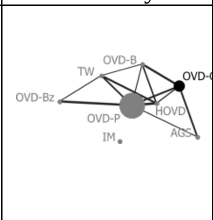
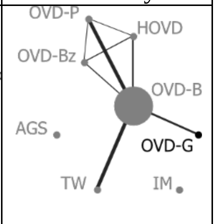
Figure 17 - WST hazardous materials scenario, Team One: weighted betweenness centrality	Figure 18 - WST hazardous materials scenario, Team Two: weighted betweenness centrality	Figure 19 - WST hazardous materials scenario, Team Six: weighted betweenness centrality
		

Table 9 - Network graphs for the medical services in the Tunnel Hazardous Materials Scenario
(*Legend: AGS = advisor on hazardous materials, IM = information manager, HOVD = chief officer, OVD-B = fire services, OVD-Bz = municipality, OVD-G = medical emergency services, OVD-P = police, TW = tunnel operator*)

5. DISCUSSION

Although the two main hypotheses are not substantiated, more in-depth analysis of the data yielded some interesting findings. The level of coordination in the intermediate phase of emergency response management seems to be determinant for the performance of an on-scene command team. This should be noted, since theory often suggests that the initial phase of an emergency is the most crucial. In contrast with expectations, less coordination, not more, during the intermediate phase, leads to better performance.

Another relevant finding is that advanced centrality measures (in-degree and weighted betweenness degree) are more useful in explaining actor performance than the more comprehensive degree indicators. It is not the overall amount of coordination in a team, but a few more complex traits of coordination that help us understand the differentiations in performance. Qualitative interpretation of the data seems necessary to understand why this is the case. However, for reasons of space, this challenge cannot be taken up in this publication.

The framework developed by Marks et al. (Marks, Mathieu, and Zaccaro 2001), which we adjusted for emergency management teams, proved useful as a starting point for understanding emergency management performance. Temporal differentiation seems to be relevant when trying to understand the relationship between coordination and emergency management performance. Decentralized coordination (network density, the advanced centrality indicators) is relevant to understanding performance. At the same time, we believe that the model needs to be further revised.

6. CONCLUSION

The core question of this study was: *How do on-scene command teams coordinate multidisciplinary objectives and procedures, and how does the way this is done determine their emergency management performance?* Our analysis led to four answers:

1. Emergency management performance and coordination patterns within and among on-scene command teams show wide variation.
2. Decentralized coordination is an important factor in emergency management performance.
3. Teams that use less coordination during the intermediate phase of emergency management perform significantly better.
4. Actors that have a central position in the network achieve their own goals better.

ACKNOWLEDGMENTS

This publication and the PhD research on which it is based was made possible through the support of the Next Generation Infrastructures Foundation http://www.nextgenerationinfrastructures.eu/. The authors wish to thank E-semble and all staff and participants from the anonymized safety region in the Netherlands for their support.

REFERENCES

Benjamins, T, and L J M Rothkrantz. 2007. "Interactive Simulation in Crisis Management." *Proceedings of ISCRAM 2007*, 978–90.

Bettenhausen, K L. 1991. "Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed." *Journal of Management* 17 (2): 345–81.

Borgatti, Stephen P, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. 2009. "Network Analysis in the Social Sciences." *Science (New York, N.Y.)* 323 (5916): 892–95. doi:10.1126/science.1165821.

Borgatti, Stephen P. 2005. "Centrality and Network Flow." *Social Networks* 27 (1): 55–71. doi:10.1016/j.socnet.2004.11.008.

Chen, Rui, Raj Sharman, H. Raghav Rao, and Shambhu J. Upadhyaya. 2008. "Coordination in Emergency

Response Management.” *Communications of the ACM*. doi:10.1145/1342327.1342340.

Comfort, L K. 2007. “Crisis Management in Hindsight: Cognition, Communication, Coordination, and Control.” *Public Administration Review* 67 (1): 189–97.

Haddow, George D, Jane A Bullock, and Damon P Coppola. 2003. *Introduction to Emergency Management*. *Introduction to Emergency Management*. Butterworth-Heinemann. doi:10.1111/1540-6210.00244.

Haferkamp, Nina, Nicole C. Kraemer, Conor Linehan, and Massimiliano Schembri. 2011. “Training Disaster Communication by Means of Serious Games in Virtual Environments.” *Entertainment Computing* 2 (2). International Federation for Information Processing: 81–88. doi:10.1016/j.entcom.2010.12.009.

Harteveld, Casper. 2012. “Making Sense of Virtual Risks; A Quasi Experimental Investigation Into Game-Based Training.” Deltares Select Series 11/2012. Delft, The Netherlands: IOS Press. doi:10.4233/uuid:c3d38ad9-88cb-403c-b69e-ba22b005a26f.

Houghton, Robert J, Chris Baber, Richard McMaster, Neville A Stanton, Paul Salmon, Rebecca Stewart, and Guy Walker. 2006. “Command and Control in Emergency Services Operations: A Social Network Analysis.” *Ergonomics* 49 (12-13). Taylor & Francis: 1204–25. doi:10.1080/00140130600619528. http://dx.doi.org/10.1080/00140130600619528.

Ilgen, Daniel R, John R Hollenbeck, Michael Johnson, and Dustin Jundt. 2005. “Teams in Organizations: From Input-Process-Output Models to IMOI Models.” *Annual Review of Psychology* 56: 517–43.

Janicik, G A, and C A Bartel. 2003. “Talking about Time: Effects of Temporal Planning and Time Awareness Norms on Group Coordination and Performance.” *Group Dynamics: Theory, Research, and Practice* 7 (2): 122–34.

Kapucu, N. 2005. “Interorganizational Coordination in Dynamic Context: Networks in Emergency Response Management.” *Connections* 26 (2): 33–48.

Kossinets, Gueorgi, and Duncan J Watts. 2006. “Empirical Analysis of an Evolving Social Network.” *Science (New York, N.Y.)* 311 (5757): 88–90. doi:10.1126/science.1116869.

Landgren, Jonas, and Urban Nulden. 2007. “A Study of Emergency Response Work: Patterns of Mobile Phone Interaction.” *Organization* 42 (2). ACM Press: 1323–32. doi:10.1145/1240624.1240824.

Leukfeldt, E R, K W C van der Straten, M P Kruis, and W Ph Stol. 2007. *Ter Plaatse: Alledaagse Samenwerking Tussen de Primarie Hulpdiensten*. *Veiligheidsstudies*. Den Haag: Boom Juridische uitgevers.

Marks, M A, J E Mathieu, and S J Zaccaro. 2001. “A Temporally Based Framework and Taxonomy of Team Processes.” *The Academy of Management Review* 26 (3): 356–76.

Mathieu, J E, and W Schulze. 2006. “The Influence of Team Knowledge and Formal Plans on Episodic Team Process-Performance Relationships.” *The Academy of Management Journal* 49 (3): 605–19.

Mathieu, J, M T Maynard, T Rapp, and L Gilson. 2008. “Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse into the Future.” *Journal of Management* 34 (3): 410–76.

Mcgrath, Dennis, and Susan P McGrath. 2005. “Simulation and Network-Centric Emergency Response.” In *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Vol. 2005. Orlando, FL: NTSA.

Newman, M. E. J. 2005. “A Measure of Betweenness Centrality Based on Random Walks.” *Social Networks* 27 (1): 39–54. doi:10.1016/j.socnet.2004.11.009.

Okamoto, K, W Chen, and XY Li. 2008. “Ranking of Closeness Centrality for Large-Scale Social Networks.” In *Frontiers in Algorithmics*, 186–95. doi:10.1007/978-3-540-69311-6_21.

Opsahl, Tore, Filip Agneessens, and John Skvoretz. 2010. “Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths.” *Social Networks* 32 (3): 245–51. doi:10.1016/j.socnet.2010.03.006.

Scott, John. 1988. “Social Network Analysis.” *Sociology* 22 (1): 109–27. doi:10.1177/0038038588022001007.

———. 2000. *Social Network Analysis: A Handbook*. 幌. Sage. doi:citeulike-article-id:244827.

APPROXIMATE AND EXACT CORRECTIONS OF THE BIAS IN CROSTON'S METHOD WHEN FORECASTING LUMPY DEMAND: EMPIRICAL EVALUATION

Adriano O. Solis[(a)], Francesco Longo[(b)], Somnath Mukhopadhyay[(c)], Letizia Nicoletti[(d)], Vittoria Brasacchio[(e)]

[(a)] School of Administrative Studies, Faculty of Liberal Arts & Professional Studies, York University, Toronto, Canada
[(b) (d) (e)] Department of Mechanical, Energy and Management Engineering, University of Calabria, Rende, Italy
[(c)] College of Business Administration, University of Texas at El Paso, El Paso, Texas, USA

[(a)] asolis@yorku.ca, [(b)] francesco.longo@unical.it, [(c)] smukhopadhyhyay@utep.edu, [(d)] letizia.nicoletti@unical.it,
[(e)] vittoria.brasacchio@gmail.com

ABSTRACT
Croston's method was developed to forecast intermittent demand, employing separate exponential smoothing estimates of the average demand size and the average interval between demand occurrences. Syntetos and Boylan reported an error in Croston's mathematical derivation of expected demand, and proposed an approximate correction now usually referred to as the Syntetos-Boylan approximation. Subsequently, Shale, Boylan, and Johnston derived the expected bias in Croston's method and proposed an 'exact' correction factor. Both approximate and exact corrections have been derived analytically. In the current study, we empirically investigate whether or not there are actually significant improvements in terms of statistical forecast accuracy as well as inventory control performance obtained by applying the approximate or exact correction.

Keywords: intermittent/lumpy demand forecasting, forecast accuracy, bias correction, inventory control, modeling and simulation

1. INTRODUCTION

Demand for an item of inventory is *intermittent* when there are time intervals in which there are no demand occurrences. Intermittent demand is said to be *lumpy* when there are large variations in the sizes of actual demand occurrences.

Croston (1972) noted that under simple exponential smoothing (SES), which has frequently been used for forecasting demand, a biased estimate arises since forecasts are based on an average of the recent demand occurrences. This bias is greatest immediately following a demand occurrence. Because inventory replenishment decisions are usually taken after a reduction in stock, there can be serious consequences of an upward bias in the demand forecast. To address this upward bias, Croston proposed a method of forecasting intermittent demand using separate exponential smoothing estimates of the average demand size and the average interval between demand occurrences, and combining these to obtain a demand forecast. Leading statistical forecasting software packages include Croston's method (Syntetos and Boylan 2005; Boylan and Syntetos 2007).

While Croston applied a single smoothing constant α, Schultz (1987) proposed the use of separate smoothing constants, α_i and α_s, in updating inter-demand intervals and nonzero demand sizes, respectively. However, Mukhopadhyay, Solis, and Gutierrez (2012) investigated separate smoothing constants, α_i and α_s, in forecasting lumpy demand and reported no substantial improvement in forecast accuracy.

Syntetos and Boylan (2001) established the presence of a positive bias in Croston's method, called an 'inversion bias', arising from an error in Croston's mathematical derivation of expected demand. Syntetos and Boylan (2005) proposed a correction factor of $[1-(\alpha/2)]$ applied to Croston's original estimator of mean demand, where α is the smoothing constant in use for updating the inter-demand intervals. The revised estimator yields an *approximately* unbiased estimator, and is now usually referred to as the Syntetos-Boylan approximation (SBA) in the literature on intermittent demand forecasting (e.g., Gutierrez, Solis, and Mukhopadhyay 2008; Boylan, Syntetos, and Karakostas 2008; Babai, Syntetos, and Teunter 2010; Mukhopadhyay, Solis, and Gutierrez 2012).

Levén and Segerstedt (2004) proposed a modification to Croston's method, which they called a 'modified Croston procedure', involving a new method for estimating the mean and variance of the forecasted demand rate. Boylan and Syntetos (2007), however, found that the smoothing method for estimating the variance is based on an invalid forecast accuracy measure, and that the new method of estimating mean demand produces biased forecasts.

Shale, Boylan, and Johnston (2006) derived the expected bias when the arrival of orders follows a Poisson process, and extended their work to other inter-arrival distributions. They specified $[1-(\alpha/(2-\alpha))]$ as an 'exact' correction factor (hereafter referred to in this paper as SBJ) to remove the inversion bias in Croston's method. However, the SBJ method has not been cited or applied in recent intermittent demand forecasting literature. For instance, Syntetos and Boylan (2010) analyze the "most well-cited intermittent demand estimation procedures" with respect to the variance of their estimates, but do not even mention SBJ. Only the variances of demand estimates using SES, Croston's method, SBA, and an "exactly unbiased modification" of Croston's method (Syntetos 2001) are reported.

A critical issue in forecasting intermittent/lumpy demand is the assumption of a distribution of demand occurrence. Boylan (1997) proposed three criteria for assessing the suitability of demand distributions:

1. *A priori* grounds for modelling demand.
2. The flexibility of the distribution to represent different types of demand.
3. Empirical evidence.

Syntetos and Boylan (2006) argued that compound distributions can represent demand incidence and demand size by separate distributions. Noting that the negative binomial distribution (NBD) is a compound distribution with variance greater than the mean, with "empirical evidence in its support (Kwan 1991)," Syntetos and Boylan declared the NBD to meet all the above three criteria. They accordingly selected the NBD to represent intermittent demand over lead time (plus review period) in their stock control simulation model. Among others, Boylan, Syntetos, and Karakostas (2008) and Syntetos, Babai, Dallery, and Teunter (2009) have also conducted empirical investigations of stock control using the NBD to characterize intermittent demand over the lead time (plus review period). These latter studies have cited Syntetos and Boylan's (2006) declaration that the NBD "satisfies both theoretical and empirical criteria."

Use of the NBD to characterize demand may indeed have been found by previous researchers to apply to intermittent (but not very erratic) demand. Our investigations, however, show that the NBD may not hold for much of lumpy demand distributions. In the current study, we use a two-stage simulation approach to characterize lumpy demand, as earlier discussed and applied, for instance, by Solis, Longo, Nicoletti, and Yemialyanava (2013). The first stage uses a uniform distribution, with probability z_1 of zero demand, to determine whether or not a demand occurs in the given period. If the first stage leads to a demand actually occurring in the period, the second stage estimates the demand size using an NBD. In the process, Pr(X=0) in the NBD is applied to adjust the probability z_1 of zero demand in the first stage.

It should be emphasized that this two-stage approach was not intended to "accurately" capture the actual lumpy demand distribution. The objective is to simulate demand distributions that mimic as closely as possible the lumpy demand distributions. We, therefore, address in this study the apparent inadequacy of the NBD, in spite of its "satisfying both theoretical and empirical criteria," for characterizing intermittent demand. As in Solis, Longo, Nicoletti, and Yemialyanava (2013), we illustrate how the two-stage simulation process better characterizes a greater proportion of lumpy demand distributions we investigate.

In this paper, we report on the preliminary results of our empirical investigation of both statistical accuracy and inventory control performance of four intermittent demand forecasting methods which employ exponential smoothing: SES, Croston's, SBA, and SBJ. We evaluate the improvements associated with applying approximate and exact corrections (SBA and SBJ, respectively) which address the positive bias in Croston's method. We first seek to characterize lumpy (i.e., both intermittent *and* erratic) demand by applying the NBD or two-stage approximations. We proceed to evaluate forecast accuracy using a number of error statistics, and then consider inventory control performance.

This paper is organized as follows. In section 2, we discuss the industrial dataset and how data partitioning is conducted for purposes of empirical evaluation, as well our application of the NBD and two-stage approximations to characterize demand. In the next section, we first discuss the statistical measures of forecast accuracy that we use, and proceed to report on our empirical investigation of forecasting performance on the performance blocks of actual data as well as on the simulated demand distributions. In section 4, we report on our empirical investigation of inventory control performance. We present our conclusions in the final section.

2. INDUSTRIAL DATASET AND DEMAND CHARACTERIZATION

2.1. Industrial Dataset and Partitioning

In this paper, we apply the SES, Croston's, SBA, and SBJ methods to stock-keeping units (SKUs) in a regional warehouse of a firm operating in the professional electronics sector. The SKUs represent end items, sub-assemblies, components, and spare parts that are used for building projects, retail sales, or servicing of professional electronic products. The raw data consist of actual withdrawals from stock as reported in the company's enterprise resource planning system over a period of 61 months. The transactional data are aggregated into usage quantities per month, which we treat as a surrogate measure of monthly demand while recognizing that the inventory on hand when a demand occurs may not meet the required quantity. The 61 months of "demand" data are divided into initialization, calibration, and performance measurement blocks (as in Boylan, Syntetos, and Karakostas 2008) with our blocks consisting of 20, 20, and 21 months, respectively.

Syntetos, Boylan, and Croston (2005) proposed a scheme to classify demand patterns into four categories (smooth, erratic, intermittent, and lumpy) for the purpose of establishing 'regions' of superior forecasting performance between Croston's method and SBA. The scheme (hereafter referred to as SBC) is based on the use of two statistics: CV^2 and *ADI*, the squared coefficient of variation of demand and average inter-demand interval, respectively. The four categories are delimited by cutoff values for CV^2 and *ADI* as follows: (i) *smooth*, with $ADI < 1.32$ and $CV^2 < 0.49$; (ii) *erratic*, with $ADI < 1.32$ and $CV^2 > 0.49$; (iii) *intermittent*, with $ADI > 1.32$ and $CV^2 < 0.49$; and (iv) *lumpy*, with $ADI > 1.32$ and $CV^2 > 0.49$. Recently, Heinecke, Syntetos, and Wang (2013) empirically evaluated the SBC cutoff values for CV^2 and *ADI* in comparison with alternative approaches proposed by Kostenko and Hyndman (2006), and found that SBC results in inferior forecasting performance overall in comparison with the latter alternatives. For purposes of the current study, we will nonetheless continue to apply the relatively simple SBC fixed cutoff values to classify demand patterns. 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; Altay, Rudisill, and Litteral 2008; Boylan, Syntetos, and Karakostas 2008; Mukhopadhyay, Solis, and Gutierrez 2012).

In this paper, we report findings on a set of ten SKUs which have thus far been evaluated for purposes of the current study. Demand statistics are presented in Table 1. All 10 SKUs exhibit lumpy demand ($ADI > 1.32$ and $CV^2 > 0.49$) according to the SBC categorization scheme. We have yet to find a SKU with intermittent demand ($ADI > 1.32$ and $CV^2 < 0.49$) according to the scheme. [Solis, Longo, Nicoletti, and Yemialyanava (2013) earlier investigated a separate set of SKUs (nine with lumpy demand, and six with erratic demand) coming from the firm's central warehouse. That earlier study also evaluated the simple moving average method, but did not consider SBJ. Hence, no comparison was made between the approximate and exact corrections of the bias in Croston's method, which is the focus of the current study.]

Table 1: 10 SKUs with Lumpy Demand

SKU #	**1**	**2**	**3**	**4**	**5**
Mean	2.0492	1.0656	2.4918	1.9672	2.7541
Std Dev	2.8427	1.1954	3.2641	3.0549	4.2531
CV^2	1.9244	1.2585	1.7159	2.4115	2.3848
ADI	1.4186	1.6486	1.3864	1.7429	1.5641
z (% of Zero Demand)	31.15%	40.98%	29.51%	44.26%	37.70%
Mean Nonzero Demand	2.9762	1.8056	3.5349	3.5294	4.4211
Std Dev of Nonzero Demand	2.9999	1.0370	3.3831	3.3596	4.6652
SKU #	**6**	**7**	**8**	**9**	**10**
Mean	6.5410	2.5082	6.9016	3.4426	2.1639
Std Dev	9.1462	3.8454	10.1648	4.3609	2.8412
CV^2	1.9552	2.3506	2.1692	1.6046	1.7240
ADI	1.4878	1.5641	1.6053	1.6053	1.6486
z (% of Zero Demand)	34.43%	37.70%	39.34%	39.34%	40.98%
Mean Nonzero Demand	9.9750	4.0263	11.3784	5.6757	3.6667
Std Dev of Nonzero Demand	9.6728	4.2074	10.9477	4.3208	2.8586

2.2. Demand Characterization Using a Negative Binomial Distribution

As previously stated, Syntetos and Boylan (2006) declared that the NBD "satisfies both theoretical and empirical criteria" to characterize demand. It is a discrete probability distribution with density function

$$f(x;r,p) = \binom{r+x-1}{x} p^r (1-p)^x I_{\{0,1,2,\ldots\}}(x). \tag{1}$$

with two parameters r and p. The parameter r is a positive integer. The parameter p is a real number satisfying $0 < p \leq 1$, and is a probability of "success" in a Bernoulli trial, while r is a target number of successes (e.g., Mood, Graybill, and Boes 1974). The random variable X represents the number of failures, in a succession of the Bernoulli trials, preceding the rth success. The NBD has mean

$$\mu = E[X] = \frac{r(1-p)}{p} \tag{2}$$

and variance

$$\sigma^2 = V[X] = \frac{r(1-p)}{p^2}. \tag{3}$$

Clearly, the variance of the NBD is greater than its mean. The NBD, when $r = 1$, reduces to a geometric (or Pascal) distribution with density function

$$f(x;p) = p(1-p)^x I_{\{0,1,2,\ldots\}}(x). \tag{4}$$

Solving (2) and (3) simultaneously, we 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 parameters of an NBD with mean μ and variance σ^2. We use the mean $\overline{x}$ and the variance s^2 of the 61-month demand time series in place of μ and σ^2, respectively. However, while r is supposed to be integer-valued, the expression (6) is real-valued. Thus, in attempting to characterize the actual demand distribution using an NBD approximation, we investigate rounded up and rounded down values of $\hat{r}$ while at the same time adjusting $\hat{p}$ to obtain acceptable NBD parameters r and p.

We used AnyLogic as our simulation platform. To address mathematical modeling not doable within the AnyLogic standard library, some code was written in Java. For the NBD parameters tested, we performed 100 runs each consisting of 100 months (for a total of 10,000 months) in each experiment.

As a rule of thumb, we operationalize 'reasonably acceptable' approximation in terms of mean, standard deviation, CV^2, and *ADI* of the simulated distribution all being within ±20% of those of the actual demand distribution and with a fairly small difference between simulated and actual proportion z of zero demand.

In each of SKUs 1, 2 and 3, we found the suggested NBD approximation to yield a simulated distribution which fairly closely resembles the actual demand distribution. The simulation results are reported in Table 2. For SKUs 1 and 3, the adjusted r value is 1 (i.e., the NBD reduces to a geometric distribution).

Table 2: Three SKUs with Reasonably Acceptable NBD Approximations

SKU #	1	2	3
Mean	2.0492	1.0656	2.4918
Std Dev	2.8427	1.1954	3.2641
CV^2	1.9244	1.2585	1.7159
ADI	1.4186	1.6486	1.3864
z (% of Zero Demand)	31.15%	40.98%	29.51%
*r^*	0.6962	3.1246	0.7607
*p^*	0.2536	0.7457	0.2339
SIMULATION	**NBD**	**NBD**	**NBD**
r	1	4	1
p	0.3458	0.7897	0.2904
Mean	1.9053	1.0567	2.4137
Std Dev	2.3690	1.1648	2.8935
CV^2	1.5460	1.2151	1.4371
ADI	1.5249	1.6464	1.4100
z (% of Zero Demand)	34.43%	39.27%	29.09%
Simulated vs Actual Mean	93.0%	99.2%	96.9%
Simulated vs Actual Std Dev	83.3%	97.4%	88.6%
Simulated vs Actual CV^2	80.3%	96.6%	83.7%
Simulated vs Actual ADI	107.5%	99.9%	101.7%
Difference in Simulated vs Actual z	3.28%	-1.71%	-0.42%

2.3. Demand Characterization Using a Two-Stage Distribution

For the seven other SKUs (4-10), we were unable to find NBD approximations that are reasonably acceptable (as set forth in the previous sub-section).

We briefly summarize the alternative, two-stage distribution as applied by Solis, Longo, Nicoletti, and Yemialyanava (2013). In each time period, the first stage is based on applying the continuous uniform distribution defined over the real number interval (0,1). Stage 1 can be viewed as a Bernoulli process, which has a fixed probability of "success" or "failure". It determines whether or not a demand occurs. If the random number generated in stage 1 is less than or equal to z_1 (a "failure"), demand for the period is set equal to zero and the demand generation process moves on to the next period. If the random number generated in stage 1 is greater than z_1 (a "success"), the demand generation process moves to stage 2 in which an NBD is used to simulate the demand size. It follows that there will still be some probability of zero demand in stage 2. At the conclusion of stage 2, the demand generation process moves to the next time period, again starting with stage 1 of the two-stage process.

To estimate the parameters of the NBD in stage 2 of the demand simulation process, the mean $\bar{x}_{nz}$ and variance s_{nz}^2 of the nonzero demands are calculated and used to obtain first approximations of the parameters $\hat{p}_{nz}$ and $\hat{r}_{nz}$ in line with (5) and (6). We round up or down $\hat{r}_{nz}$ to some integer value and adjust $\hat{p}_{nz}$ accordingly. The corresponding negative binomial probability $P_0 = \Pr(X = 0) > 0$ is then used to find z_1 (as applied in the first stage), as follows:

$$z_1 = \frac{z - P_0}{1 - P_0}, \tag{7}$$

provided $z > P_0$. The resulting proportion of zero demand periods arising from the two-stage distribution is then closer to z. We refine the parameter estimate $\hat{p}_{nz}$ while the mean, standard deviation, CV^2, *ADI*, and z of the actual and simulated distributions are compared.

We note that this two-stage approach did not yield acceptable characterizations of demand for SKUs 1 and 2; only the NBD approximations for these two SKUs (as presented in Table 2) were reasonably close to actual demand statistics. In the case of SKU 3, on the other hand, both NBD and two-stage approximations are both fairly close to the actual demand distribution (refer to Table 3), but with the two-stage approximation appearing to yield a somewhat better characterization.

Table 3: SKU 3 – Comparison of NBD and Two-Stage Approximations

SKU #	3	
Mean	2.4918	
Std Dev	3.2641	
CV^2	1.7159	
ADI	1.3864	
z (% of Zero Demand)	29.51%	
*r^*	0.7607	
*p^*	0.2339	
SIMULATION	**NBD**	**Two-Stage**
r	1	1
p	0.2904	0.2602
Mean	2.4137	2.6693
Std Dev	2.8935	3.2648
CV^2	1.4371	1.4960
ADI	1.4100	1.4186
z (% of Zero Demand)	29.09%	29.52%
Simulated vs Actual Mean	96.9%	107.1%
Simulated vs Actual Std Dev	88.6%	100.0%
Simulated vs Actual CV^2	83.7%	87.2%
Simulated vs Actual ADI	101.7%	102.3%
Difference in Simulated vs Actual z	-0.42%	0.01%

We present in Table 4 the simulation results for the demand distribution approximations, using the NBD approximation for SKUs 1 and 2 and the two-stage approach for SKUs 3-10.

We have yet to evaluate a SKU in the current dataset for which neither approximation method leads to an acceptable characterization, albeit with only 10 SKUs evaluated thus far. [We must quickly point out, however, that Solis, Longo, Nicoletti, and Yemialyanava (2013) reported that both the NBD and two-stage approximations fail in the case of SKUs with demand distributions that are lumpier.]

3. EMPIRICAL INVESTIGATION OF FORECASTING PERFORMANCE

3.1. Smoothing Constants and Forecast Accuracy Measures

In the context of intermittent demand, low values of the smoothing constant α have been recommended, and values in the range 0.05-0.20 are considered realistic (Croston 1972; Willemain, Smart, Shockor, and DeSautels 1994; Johnston and Boylan 1996). We test four α values: 0.05, 0.10, 0.15, and 0.20 (as in Syntetos and Boylan 2005, 2006; Gutierrez, Solis, and Mukhopadhyay 2008; Mukhopadhyay, Solis, and Gutierrez 2012).

Table 4: Lumpy Demand Approximations

SKU #	**1**	**2**	**3**	**4**	**5**
Mean	2.0492	1.0656	2.4918	1.9672	2.7541
Std Dev	2.8427	1.1954	3.2641	3.0549	4.2531
CV^2	1.9244	1.2585	1.7159	2.4115	2.3848
ADI	1.4186	1.6486	1.3864	1.7429	1.5641
z (% of Zero Demand)	31.15%	40.98%	29.51%	44.26%	37.70%
*r^*	0.6962	3.1246	0.7607	0.5254	0.4946
*p^*	0.2536	0.7457	0.2339	0.2108	0.1523
Mean of nonzero demand	2.9762	1.8056	3.5349	3.5294	4.4211
Std Dev of nonzero demand	2.9999	1.0370	3.3831	3.3596	4.6652
r^ nonzero	-	-	1.5796	1.6058	1.1270
p^ nonzero	-	-	0.3089	0.3127	0.2031
SIMULATION	**NBD**	**NBD**	**Two-Stage**	**Two-Stage**	**Two-Stage**
r	1	4	1	1	1
p	0.3458	0.7897	0.2602	0.2726	0.1987
Pr(*X* = 0)	0.3458	0.3889	0.2602	0.2726	0.1987
Final zero proportion in stage 1	-	-	4.72%	23.37%	22.26%
Mean	1.9053	1.0534	2.6693	1.9950	3.1393
Std Dev	2.3690	1.1959	3.2648	2.9137	4.4296
CV^2	1.5460	1.2888	1.4960	2.1331	1.9910
ADI	1.5249	1.7074	1.4186	1.7989	1.6124
z (% of Zero Demand)	34.43%	41.44%	29.52%	44.42%	37.99%
Simulated vs Actual Mean	93.0%	98.9%	107.1%	101.4%	114.0%
Simulated vs Actual Std Dev	83.3%	100.0%	100.0%	95.4%	104.2%
Simulated vs Actual CV^2	80.3%	102.4%	87.2%	88.5%	83.5%
Simulated vs Actual ADI	107.5%	103.6%	102.3%	103.2%	103.1%
D in Simulated vs Actual z	3.28%	0.46%	0.01%	0.16%	0.29%
SKU #	**6**	**7**	**8**	**9**	**10**
Mean	6.5410	2.5082	6.9016	3.4426	2.1639
Std Dev	9.1462	3.8454	10.1648	4.3609	2.8412
CV^2	1.9552	2.3506	2.1692	1.6046	1.7240
ADI	1.4878	1.5641	1.6053	1.6053	1.6486
z (% of Zero Demand)	34.43%	37.70%	39.34%	39.34%	40.98%
*r^*	0.5548	0.5123	0.4940	0.7609	0.7925
*p^*	0.0782	0.1696	0.0668	0.1810	0.2681
Mean of nonzero demand	9.9750	4.0263	11.3784	5.6757	3.6667
Std Dev of nonzero demand	9.6728	4.2074	10.9477	4.3208	2.8586
r^ nonzero	1.1904	1.1854	1.1935	2.4791	2.9845
p^ nonzero	0.1066	0.2274	0.0949	0.3040	0.4487
SIMULATION	**Two-Stage**	**Two-Stage**	**Two-Stage**	**Two-Stage**	**Two-Stage**
r	1	1	1	2	2
p	0.0973	0.2073	0.0877	0.2754	0.3868
Pr(*X* = 0)	0.0973	0.2073	0.0877	0.0758	0.1496
Final zero proportion in stage 1	27.36%	21.41%	33.51%	34.37%	30.60%
Mean	6.6577	3.0033	6.8982	3.4690	2.1988
Std Dev	9.1379	4.1380	10.2014	4.3599	2.7861
CV^2	1.8838	1.8984	2.1870	1.5796	1.6056
ADI	1.5108	1.6028	1.6491	1.6548	1.6998
z (% of Zero Demand)	33.82%	37.62%	39.37%	39.58%	41.18%
Simulated vs Actual Mean	101.8%	119.7%	100.0%	100.8%	101.6%
Simulated vs Actual Std Dev	99.9%	107.6%	100.4%	100.0%	98.1%
Simulated vs Actual CV^2	96.4%	80.8%	100.8%	98.4%	93.1%
Simulated vs Actual ADI	101.5%	102.5%	102.7%	103.1%	103.1%
D in Simulated vs Actual z	-0.61%	-0.08%	0.03%	0.24%	0.20%

To compare intermittent demand forecasting methods, Eaves and Kingsman (2004) used three traditional measures of accuracy: *mean absolute deviation* (MAD), *root mean squared error* (RMSE), and *mean absolute percentage error* (MAPE). We apply these three, as well as two additional error statistics.

MAPE is the most widely used scale-free forecast accuracy measure. The traditional MAPE definition fails when demand is intermittent, due to division by zero. We use an alternative specification of MAPE as a ratio estimate (e.g., Gilliland 2002):

$$\text{MAPE} = \left(\sum_{t=1}^{n} |E_t| \Big/ \sum_{t=1}^{n} A_t \right) \times 100 \cdot \tag{8}$$

We use as a second scale-free error statistic the *mean absolute scaled error* (MASE), which was proposed by Hyndman and Koehler (2006). It uses the in-sample mean absolute error from the naïve forecast as a benchmark. The scaled error for period *t* is

$$q_t = \frac{e_t}{\left(\sum_{i=2}^{n} |Y_i - Y_{i-1}| \right) \Big/ (n-1)} \cdot \tag{9}$$

MASE is calculated as follows:

$$\text{MASE} = \text{mean}\left(|q_t| \right). \tag{10}$$

If MASE is less than one, the forecasting method being considered performs better than the in-sample naïve forecasts. In comparing different forecasting methods, a smaller MASE indicates better performance.

A third scale-free error statistic that we used is *percentage best* (PB), which refers to the percentage of time periods in which one particular method outperforms all of the other methods with respect to a specified criterion. We applied smallest absolute error as performance criterion. PB has been used in previous intermittent demand forecasting studies (e.g., Syntetos and Boylan 2005, 2006; Gutierrez, Solis, and Mukhopadhyay 2008; Mukhopadhyay, Solis, and Gutierrez 2012).

3.2. Forecast Accuracy: Performance Block

The exponential smoothing constant α was selected from among the candidate values (0.05, 0.10, 0.15, or 0.20) for each of the SES, Croston's, SBA and SBJ methods, taking into consideration four forecast accuracy measures (RMSE, MAD, MAPE, and MASE). We did not consider PB, as the 20 months in the calibration block are apparently insufficient to apply PB as an appropriate accuracy measure. For all ten SKUs thus far evaluated, minimum values of MAD, MAPE, and MASE were consistently associated with the same α values. We accordingly selected α values (as reported in Table 5) based upon minimum MAD, MAPE, and MASE in the calibration block. (In the case of five SKUs, the minimum RMSE actually yielded α values consistent with the other three error measures.)

We proceeded to calculate the resulting error statistics (MAD, MAPE, and MASE) when applying SES, Croston's, SBA and SBJ methods to actual demand data in the performance block (the final 21 months). These error statistics are summarized in Table 5. SES resulted in the best forecast accuracy for SKU 10. For the other nine SKUs, SBJ and SBA resulted in the best error statistics. As expected, SBJ yielded "slightly better" accuracy measures, though generally only at the fourth or fifth significant digit. The improvement arising from SBJ's 'exact' correction over SBA's approximate correction factor is, therefore, hardly significant.

3.3. Forecast Accuracy: Simulated Demand

In evaluating forecast accuracy over the 10,000 months of simulated demand (100 runs of 100 months each), we found a more pronounced overall superiority of SBJ and SBA over SES and Croston's methods using the three scale-free error statistics (MAPE, MASE, and PB). These error statistics are summarized in Table 6. Once again, SBJ yielded "better" accuracy measures only generally at the fourth or fifth significant digit. The improvement arising from SBJ's 'exact' correction over SBA's approximate correction factor is, therefore, hardly significant.

Table 5: Error Statistics when Applying Forecasting Methods to Actual Demand in the Performance Block

SKU #	1	2	3	4	5
Smoothing Constants Selected in Calibration Block					
SES	0.05	0.05	0.05	0.05	0.05
Croston	0.05	0.05	0.05	0.05	0.05
SBA	0.05	0.05	0.05	0.05	0.05
SBJ	0.05	0.05	0.05	0.05	0.05
MAD					
SES	1.674	0.670	2.223	2.261	1.909
Croston	1.637	0.649	2.179	2.211	2.029
SBA	1.625	0.652	2.158	2.202	1.982
SBJ	1.625	0.652	2.157	2.202	1.980
Best MAD	SBJ/SBA	SBJ/SBA	SBJ	SBJ/SBA	SBJ
MAPE					
SES	95.0%	87.9%	101.5%	101.0%	174.3%
Croston	92.9%	85.2%	99.5%	98.8%	185.2%
SBA	92.2%	85.5%	98.5%	98.4%	180.9%
SBJ	92.2%	85.5%	98.5%	98.4%	180.8%
Best MAPE	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ
MASE					
SES	0.676	0.670	0.765	0.669	1.336
Croston	0.661	0.649	0.750	0.654	1.420
SBA	0.656	0.652	0.743	0.651	1.387
SBJ	0.656	0.652	0.743	0.651	1.386
Best MASE	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ
SKU #	**6**	**7**	**8**	**9**	**10**
Smoothing Constants Selected in Calibration Block					
SES	0.05	0.05	0.05	0.2	0.05
Croston	0.05	0.05	0.05	0.05	0.2
SBA	0.05	0.05	0.05	0.05	0.2
SBJ	0.05	0.05	0.05	0.05	0.2
MAD					
SES	8.152	2.079	7.538	2.268	2.600
Croston	7.980	2.032	7.164	2.287	2.787
SBA	7.936	1.991	7.064	2.263	2.718
SBJ	7.935	1.990	7.062	2.263	2.710
Best MAD	SBJ	SBJ	SBJ	SBJ/SBA	SES
MAPE					
SES	113.4%	121.3%	135.3%	103.5%	101.1%
Croston	111.0%	118.5%	128.6%	104.4%	108.4%
SBA	110.4%	116.1%	126.8%	103.3%	105.7%
SBJ	110.4%	116.1%	126.8%	103.3%	105.4%
Best MAPE	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ/SBA	SES
MASE					
SES	0.751	1.149	1.319	0.744	0.853
Croston	0.735	1.123	1.254	0.751	0.915
SBA	0.731	1.100	1.236	0.743	0.892
SBJ	0.731	1.100	1.236	0.742	0.889
Best MASE	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ	SES

4. EMPIRICAL INVESTIGATION OF INVENTORY CONTROL PERFORMANCE

Recent studies that have looked into both forecast accuracy and inventory control performance of intermittent demand forecasting studies have applied a (T,S) periodic review system, where T and S denote the review period and the base stock (or 'order-up-to' level), respectively. These include 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), and Teunter, Syntetos, and Babai (2010).

We simulate the performance of a (T,S) inventory control system over the 10,000 months of simulated demand (100 runs of 100 months each) generated using the NBD or two-stage approximations. We assume full backordering, with inventory reviewed on a monthly basis (T = 1). The reorder lead time for most SKUs is about one month (L = 1).

Table 6: Error Statistics when Applying Forecasting Methods to the Simulated Demand Distributions

SKU #	1	2	3	4	5
Smoothing Constants Selected in Calibration Block					
SES	0.05	0.05	0.05	0.05	0.05
Croston	0.05	0.05	0.05	0.05	0.05
SBA	0.05	0.05	0.05	0.05	0.05
SBJ	0.05	0.05	0.05	0.05	0.05
MAPE					
SES	81.6%	84.5%	90.4%	106.5%	102.7%
Croston	81.8%	83.5%	89.6%	105.0%	102.6%
SBA	81.1%	83.3%	89.0%	104.3%	101.8%
SBJ	81.1%	83.3%	88.9%	104.3%	101.7%
Best MAPE	SBJ/SBA	SBJ/SBA	SBJ	SBJ/SBA	SBJ
MASE					
SES	0.777	0.759	0.761	0.788	0.778
Croston	0.779	0.750	0.755	0.777	0.777
SBA	0.773	0.748	0.750	0.772	0.771
SBJ	0.772	0.748	0.749	0.772	0.771
Best MASE	SBJ	SBJ/SBA	SBJ	SBJ/SBA	SBJ/SBA
PB					
SES	37.5%	34.6%	35.1%	38.5%	39.7%
Croston	20.5%	26.4%	15.4%	13.5%	20.6%
SBA	0.1%	1.2%	0.3%	0.4%	0.3%
SBJ	42.0%	37.8%	49.2%	47.7%	39.5%
Best PB	SBJ	SBJ	SBJ	SBJ	SES
SKU #	**6**	**7**	**8**	**9**	**10**
Smoothing Constants Selected in Calibration Block					
SES	0.05	0.05	0.05	0.2	0.05
Croston	0.05	0.05	0.05	0.05	0.2
SBA	0.05	0.05	0.05	0.05	0.2
SBJ	0.05	0.05	0.05	0.05	0.2
MAPE					
SES	100.9%	100.8%	109.8%	101.4%	98.5%
Croston	99.9%	99.9%	108.4%	97.9%	101.7%
SBA	99.2%	99.2%	107.5%	97.3%	99.1%
SBJ	99.2%	99.2%	107.5%	97.3%	98.8%
Best MAPE	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ/SBA	SES
MASE					
SES	0.771	0.778	0.788	0.810	0.783
Croston	0.763	0.771	0.778	0.783	0.808
SBA	0.758	0.766	0.772	0.779	0.788
SBJ	0.758	0.766	0.772	0.779	0.786
Best MASE	SBJ/SBA	SBJ/SBA	SBJ/SBA	SBJ/SBA	SES
PB					
SES	12.4%	13.4%	12.7%	31.4%	29.4%
Croston	11.4%	12.8%	12.2%	18.3%	14.8%
SBA	0.2%	0.3%	0.2%	0.3%	0.7%
SBJ	49.3%	47.4%	47.4%	31.5%	38.4%
Best PB	SBJ	SBJ	SBJ	SBJ	SBJ

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 ‘training sample’ (corresponding to the combined initialization and calibration blocks). We seek a safety stock level of $k \cdot s_{tr}$, with ‘safety factor’ k. This approach is different from that suggested under an assumption that daily demand is identically and independently normally distributed during the protection interval (e.g., Silver, Pyke, and Peterson 1998). The replenishment quantity Q_t at the time of review is

$$Q_t = (T + L) \cdot F_t + k \cdot s_{tr} - I_t + B_t, \quad (11)$$

where 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.

4.1. Service Levels

According to Silver, Pyke, and Peterson (1998), the two most commonly specified service level criteria in inventory systems are:

- Probability of no stockout (PNS) per review period, and
- Fill rate (FR), the average percentage of demand to be satisfied from on-hand inventory.

FR is considered to have considerably more appeal for practitioners.

We used two values of the target PNS (90% and 95%) and two values of the target FR (95% and 98%) in simulating inventory control performance. These target PNS and FR values are comparable with ‘reasonable’ levels tested in inventory systems studies – for instance, 80%, 90%, 95%, or 97.5% for PNS and 95%, 98%, 99% or 99.9% for FR (Solis, Longo, Nicoletti, Caruso, and Fazzari 2014). We performed simulation searches to find the safety factor k that would yield the target PNS or FR.

4.2. Average Inventory on Hand

For a 95% target FR, resulting averages of inventory on hand are reported in Table 7. We proceeded to index the average inventory on hand using Croston’s method as base (Croston index = 100). These indices, reported in Table 8, are all very close to 100. In fact, the indices for SBA and SBJ differ by at most 0.1. Moreover, all SBA and SBJ indices are between 98.9 and 100.3, which indicate that average levels of inventory on hand do not differ much from those arising using Croston’s method. Indices for a 98% target FR are all even closer to 100.

In the case of a 90% target PNS, average inventory on hand levels are reported in Table 9. In applying an index of 100 to average inventory on hand under Croston’s method (see Table 10), we find the resulting indices for SBA and SBJ to be roughly equal for each of the 10 SKUs. These SBA and SBJ indices are all very close to 100 (between 99.8 and 100.4). With a 95% target PNS, the SBA and SBJ indices all fall between 99.9 and 100.7, and are again roughly equal for each SKU.

Table 7: Average Inventory on Hand for a 95% Target Fill Rate

SKU #	1	2	3	4	5
SES	5.3849	2.1224	7.4255	7.6331	11.3890
Croston	5.3636	2.1561	7.4242	7.6112	11.3918
SBA	5.3618	2.1335	7.4268	7.6104	11.4288
SBJ	5.3625	2.1322	7.4273	7.6101	11.4271
SKU #	**6**	**7**	**8**	**9**	**10**
SES	22.5350	10.2828	27.3020	9.3131	5.9441
Croston	22.5247	10.2460	27.2997	9.2100	6.0214
SBA	22.5171	10.2382	27.2846	9.2028	5.9952
SBJ	22.5207	10.2404	27.2797	9.2028	5.9916

Table 8: Indices of Average Inventory on Hand for a 95% Target Fill Rate

SKU #	1	2	3	4	5
SES	100.4	98.4	100.0	100.3	100.0
Croston	100.0	100.0	100.0	100.0	100.0
SBA	100.0	99.0	100.0	100.0	100.3
SBJ	100.0	98.9	100.0	100.0	100.3
SKU #	**6**	**7**	**8**	**9**	**10**
SES	100.0	100.4	100.0	101.1	98.7
Croston	100.0	100.0	100.0	100.0	100.0
SBA	100.0	99.9	99.9	99.9	99.6
SBJ	100.0	99.9	99.9	99.9	99.5

Table 9: Average Inventory on Hand for a 90% Target Probability of No Stockout

SKU #	1	2	3	4	5
SES	2.9487	1.2462	4.0906	3.5699	5.7023
Croston	2.9623	1.2558	4.0598	3.5567	5.6738
SBA	2.9616	1.2557	4.0614	3.5520	5.6752
SBJ	2.9616	1.2559	4.0614	3.5530	5.6723
SKU #	**6**	**7**	**8**	**9**	**10**
SES	12.3393	5.3399	13.7299	6.1841	3.6508
Croston	12.2887	5.3147	13.6642	6.0096	3.6979
SBA	12.3222	5.3185	13.6631	5.9979	3.7144
SBJ	12.3241	5.3185	13.6692	5.9985	3.7122

Table 10: Indices of Average Inventory on Hand for a 90% Target Probability of No Stockout

SKU #	1	2	3	4	5
SES	99.5	99.2	100.8	100.4	100.5
Croston	100.0	100.0	100.0	100.0	100.0
SBA	100.0	100.0	100.0	99.9	100.0
SBJ	100.0	100.0	100.0	99.9	100.0
SKU #	**6**	**7**	**8**	**9**	**10**
SES	100.4	100.5	100.5	102.9	98.7
Croston	100.0	100.0	100.0	100.0	100.0
SBA	100.3	100.1	100.0	99.8	100.4
SBJ	100.3	100.1	100.0	99.8	100.4

4.3. Cumulative Backlogs

For reasonable target service levels, the occurrence of backlogs is minimized with the provision of safety stock levels. Therefore, reporting on average backlog per period will lead to averages of well under one unit. We accordingly record the cumulative backlogs over an entire 100-month simulation run.

Table 11 shows the average (across 100 replications) of the cumulative backlogs over 100-month intervals when the target FR is 98%. The absolute differences (with respect to results arising from the use of Croston's method) are all less than 0.1 unit, 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 11: Mean 100-Month Backlogs for a 98% Target Fill Rate

SKU #	1	2	3	4	5
SES	3.86	2.14	5.41	3.98	6.45
Croston	3.87	2.14	5.43	3.98	6.47
SBA	3.88	2.14	5.44	3.99	6.48
SBJ	3.87	2.14	5.44	3.99	6.48
SKU #	**6**	**7**	**8**	**9**	**10**
SES	13.98	6.26	13.84	7.04	4.51
Croston	14.03	6.24	13.87	7.10	4.51
SBA	13.99	6.25	13.87	7.11	4.52
SBJ	14.01	6.26	13.89	7.09	4.52

Table 12 provides analogous results under a 95% target PNS. The absolute differences in average cumulative backlogs over 100-month intervals of SBA or SBJ with respect to those arising from the use of Croston's method are all less than 0.2 unit. The absolute differences between SES and Croston's method cumulative backlogs are all well under 1 unit. Essentially the same observations apply for a target PNS of 90%

Table 12: Mean 100-Month Backlogs for a 95% Target Probability of No Stockout

SKU #	1	2	3	4	5
SES	14.88	7.50	19.39	19.07	27.92
Croston	14.83	7.68	19.68	19.04	28.03
SBA	14.84	7.68	19.69	19.03	27.88
SBJ	14.84	7.68	19.69	19.03	27.91
SKU #	**6**	**7**	**8**	**9**	**10**
SES	50.97	24.60	59.48	20.00	14.64
Croston	51.43	24.52	58.85	20.56	14.69
SBA	51.35	24.46	58.93	20.44	14.56
SBJ	51.35	24.46	58.96	20.45	14.54

5. CONCLUSION

Croston's method (1972) was developed to forecast intermittent demand, employing separate exponential smoothing estimates of the average demand size and the average interval between demand occurrences. Syntetos and Boylan (2001) reported an error in Croston's mathematical derivation of expected demand, leading to a positive bias. Syntetos and Boylan (2005) then proposed an approximate correction, SBA. Subsequently, Shale, Boylan, and Johnston (2006) derived the expected bias in Croston's method and proposed an 'exact' correction factor, SBJ.

Both the approximate correction (SBA) and the exact correction (SBJ) have been derived analytically. In the current study, we empirically investigate, using an industrial dataset involving SKUs exhibiting lumpy demand, whether or not there are actually significant improvements in terms of statistical forecast accuracy as well as inventory control performance obtained by applying the approximate or exact correction. We evaluate SES (the original exponential smoothing method), Croston's method, SBA, and SBJ by way of modeling and simulation. This paper constitutes a very preliminary report, limited to ten SKUs, all exhibiting lumpy demand, that have thus far been subjected to extensive simulation experiments.

We first attempt to characterize lumpy demand using a suggested NBD approximation (Syntetos and Boylan 2006). Failing to find a reasonably acceptable NBD approximation, we try characterizing the demand distribution using an alternative two-stage simulation approach involving the continuous uniform distribution (stage 1) and the NBD (stage 2). The two-stage alternative allows us to better characterize demand for most of the 10 SKUs. Having characterized demand, we then simulate forecasting performance. Based on the 10 SKUs evaluated, we have, as expected, found overall superior forecast accuracy of the bias corrections (SBA and SBJ) over both Croston's method and SES. However, we have not found significant differences in forecast accuracy between the SBA (approximate) and SBJ (exact) corrections.

Moreover, in terms of inventory control performance, we have observed very minute differences in average inventory on hand and average cumulative backlogs.

We reiterate, nonetheless, that this is a very preliminary report based upon an investigation of 10 SKUs. In particular, Table 1 shows that the mean monthly demands range between 1.07 and 6.54 (with mean nonzero demands ranging between 1.81 and 9.98). We anticipate being able to report, by the time of the conference, on a larger set of SKUs which will also include some with higher mean monthly demands.

REFERENCES

Altay, N., Rudisill, F., Litteral, L.A., 2008. Adapting Wright's modification of Holt's method to forecasting intermittent demand. *International Journal of Production Economics*, 111 (2), 389-408.

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., 1997. The centralization of inventory and the modelling of demand. Unpublished Ph.D. Thesis, University of Warwick, UK.

Boylan, J.E. and Syntetos, A.A., 2007. The accuracy of a modified Croston procedure. *International Journal of Production Economics*, 107 (2), 511-517.

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.
Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23 (3), 289-304.
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.
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.
Heinecke, G., Syntetos, A.A., and Wang, W., 2013. Forecasting-based SKU classification. *International Journal of Production Economics*, 143 (2), 455-462.
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.
Kostenko, A.V. and Hyndman, R.J., 2006. A note on the categorization of demand patterns. *Journal of the Operational Research Society*, 57 (10), 1256-1257.
Kwan, H.W., 1991. On the demand distributions of slow moving items. Unpublished Ph.D. Thesis, Lancaster University, UK.
Levén, E. and Segerstedt, A., 2004. Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics*, 90 (3), 361-367.
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.
Schultz CR., 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society*, 38 (5), 453-458.
Shale, E.A., Boylan, J.E., and Johnston, F.R., 2006. Forecasting for intermittent demand: the estimation of an unbiased estimate. *Journal of the Operational Research Society*, 57 (5), 588-592.
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., Longo, F., Nicoletti, L., and Yemialyanava, A., 2013. Applying modeling and simulation to evaluate statistical accuracy and inventory control performance of lumpy demand forecasting methods. *Proceedings of the 12th International Conference on Modeling & Applied Simulation*, 212-220, September 25-27, Athens, Greece.
Solis, A.O., Longo, F., Nicoletti, L., Caruso, P., and Fazzari, E., 2014. A modelling and simulation approach to assessment of a negative binomial approximation in a multi-echelon inventory system. *International Journal of Simulation and Process Modelling*, forthcoming.
Syntetos, A.A., 2001. Forecasting of intermittent demand. Unpublished Ph.D. Thesis, Brunel University of Buckinghamshire New University, UK.
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. 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. and Boylan, J.E., 2010. On the variance of intermittent demand estimates. *International Journal of Production Economics*, 128 (2), 546-555.
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., 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., 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.
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.
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.

Willemain T.R., Smart C.N., Shockor J.H., DeSautels P.A., 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10 (4), 529-538.

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 Ph.D. 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. He has served as Program Chair of the 2013 Summer Computer Simulation Conference of the Society for Modeling & Simulation International.

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 also been Guest Editor of two issues of *International Journal of Simulation and Process Modelling*. 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, and/or member of the International Program Committee.

Somnath Mukhopadhyay is an Associate Professor of Operations and Supply Chain Management in the College of Business Administration at the University of Texas at El Paso. He has published in numerous well-respected journals like *Decision Sciences*, *INFORMS Journal on Computing*, *Journal of Forecasting*, *Communications of the AIS*, *IEEE Transactions on Neural Networks*, *Neural Networks*, *Neural Computation*, *International Journal of Production Economics*, and *Journal of World Business*. He received MS and Ph.D. degrees in Management Science from Arizona State University. He was a visiting research assistant in the parallel distributed processing group of Stanford University. He has over 15 years of industry experience in building and implementing mathematical models.

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.

Vittoria Brasacchio 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 2012. She was recently an international visiting graduate student scholar at the School of Administrative Studies, York University, Canada.

MANAGING CYBER SECURITY RESOURCES VIA SIMULATION-BASED OPTIMIZATION

Pasquale Legato[(a)], Rina Mary Mazza[(a)]

[(a)]Department of Informatics, Modeling, Electronics and System Engineering
University of Calabria
Via P. Bucci 42C
87036 Rende (CS), ITALY

[(a)]legato@dimes.unical.it, [(a)]rmazza@dimes.unical.it

ABSTRACT
Simulation-based optimization (SO) has been applied in many different application areas with the objective of searching for the settings of controllable decision variables that yield the minimum (maximum) expected performance of a stochastic system. Here we propose an SO method to deal with computer/network security related to systems for conditional access. The basic idea consists in designing and developing a simulation-based optimization tool to evaluate cyber attack tolerance along with the related performance degradation. In particular, we optimize training-based recovery actions aimed at restoring the target quality of service level for the services under attack while enhancing the knowledge of the human resources (i.e. analysts) engaged in defending cyber security assets. An illustrative example is presented to show how system performance varies according to whether the analysts in a cyber defense team (i.e. the controllable decision variables) are called to work alone or in consultation with other analysts.

Keywords: simulation optimization, cyber security, team formation and collaboration

1. INTRODUCTION
Simulation-based optimization (SO) is the practice of searching for the settings of controllable decision variables that yield the minimum (maximum) expected performance of a stochastic system that is represented by a simulation model (Fu and Nelson 2003). In an SO procedure, a structured iterative approach calls an optimization algorithm to decide how to change the values for the set of input parameters (i.e. configuration) and then uses the responses generated by the simulation runs to guide the selection of the next set. The logic of this approach is shown in Figure 1.

SO methods have been applied to applications with a single objective, applications that require the optimization of multiple criteria, and applications with non-parametric objectives. (Carson and Maria 1997) review the area of simulation optimization by providing a critical review of the methods employed and presenting applications developed in the area. A similar work is somewhat proposed in (Wang and Shi 2013), but without major differences in content. Fields of SO application include, but are not limited to, energy, environment, economics, health, manufacturing, high tech, education, government, and defense.

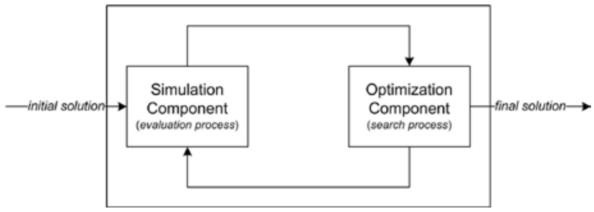


Figure 1: Logic of an SO Procedure

Whatever be the application field of interest, the appeal of SO is that it allows one to work with essentially arbitrarily complex simulation models, freeing the modeler from the tyranny of restricting model complexity to tractable forms (Pasupathy and Henderson 2011). In (cit. op.), the authors develop and promote a library (www.simopt.org) of over 50 simulation optimization problems intended to spur development and comparison of simulation-optimization methods and algorithms with respect to practical guarantees of performance.

To our knowledge, SO applied to cyber security has not received great attention in the past literature. (Fischer et al. 2010) present an effective simulation methodology called optimal splitting technique for rare events (OSTRE) with applications to cyber security. On one side, a splitting methodology is used to create separate copies of the simulation whenever it gets close to the rare event in order to multiply the promising runs that are "near" the rare event and, thus, improve the efficiency of the simulation. On the other, the notion of optimal computing budget allocation is applied to determine a good allocation of simulation runs at the intermediate levels (i.e. levels measure proximity to the rare event). The overall methodology is applied to simulate the link performance of an Internet Protocol (IP) network under a worm attack where the worm propagation creates a denial of service in many parts of the Internet and, thus, changes the traffic loading on the network. (Masi et al. 2011) extends (Fischer et al. 2010) by studying the sensitivity of the benefit of splitting to the number and location of the levels and also examining equal-allocation splitting.

(Zhang et al. 2012) present a learning environment to provide users with a unique way of improving their understanding of cyber intelligence with respect to the identification, tracking, analysis and countering of security threats in the cyberspace. In their system, a simulation engine drives the environment dynamics and changes; an optimization engine based on a multi-objective genetic algorithm implements decision making mechanisms; and data mining techniques provide for adaptation.

(Kiesling et al. 2013) introduce an approach that is based on an adversary-centric view and combines modeling and simulation-optimization techniques to detect ongoing attacks and prevent their successful execution. On the simulation side, human adversaries are represented as agents that make active decisions in attacking a system by means of attack patterns (e.g. brute force, SQL injection, social attack, spearfish attack, keylogger, and backdoor installation), deliberately exploiting dynamic interactions of vulnerabilities. On the optimization side, a Multi-objective genetic algorithm metaheuristic is introduced to optimize information systems and enable decision-makers to study how its security may be improved (e.g., by adding physical, technical, operational, and organizational security controls) while trading off multiple objectives (e.g., effectiveness against different types of adversaries, cost, risk, awareness of attacks). The overall model returns non-dominated efficient portfolios, i.e., there is no other portfolio with equal or better values for all objectives and a strictly better value in one of the objectives.

In our work we focus on computer/network security related to *systems for conditional access* by which we refer to digital systems that administer certain rights of their users pertaining to the access of documents, confidential data or, even more importantly, digital payment systems. Our ongoing research activity currently consists in designing and developing a qualitative and quantitative simulation-based optimization tool to evaluate attack tolerance, along with the related performance degradation. In particular, we optimize training-based recovery actions aimed at restoring the target quality of service level for the digital services under attack while enhancing the knowledge of the human resources (i.e. analysts) engaged in defending, alone or in cooperation with others, the cyber security assets to which they are assigned.

The rest of the paper is organized as follows: the statement of the problem is presented in section 2. In section 3, we present the meta-heuristic technique for the optimization of the controllable decision variables sets. Section 4 illustrates the practical usefulness of the tool by means of optimizations for a sample scenario. Conclusions and directions for further research investigations are presented in Section 5.

2. PROBLEM STATEMENT

Digital service systems are a fast-growing IT market area in which data exchange, transactions and payments are increasingly implemented by using advanced technologies, devices and network architectures (e.g. cloud computing, mobile devices, etc.). Within this context, cyber security has become an "enabling factor" for the use of digital systems, due to the fact that the comprehension and control of risk scenarios is assuming a particularly critical role, especially because of the latest emerging risk characteristics. These can be summarized as large-scale network attacks, associated with either fraudulent or denial-of-services activities, that exploit the increased vulnerability associated with new technologies (such as smartphones), growing availability of cloud services and infrastructure virtualization.

In the above market scenario, a major response can certainly come from deploying cyber defense security analysts. The main job of a cyber defense security analyst entails auditing computer networks and information systems for vulnerabilities, developing solutions for security issues and investigating security breaches. These analysts are also often responsible for training both (junior) co-workers and clients with respect to the best computer security practices.

In fulfilling the major of the above purposes, besides taking appropriate off-the-shelf preventive measures by installing firewalls and anti-virus software, a security analyst monitors server logs and network traffic for unusual or suspicious activity or data flow, often with the support of automation software and applications designed to detect and filter intrusion. Obviously, in a corporate-based perspective the contribution to providing an efficient protection of the cyber assets cannot be delivered regardless of the types of skills (e.g. the ability to carry out triage analysis, escalation analysis, correlations analysis, forensic analysis as described in (D'Amico and Whitley (2007)) and levels of skills (e.g. expert, average and novice) possessed by the analysts. As a result, both company activity scheduling and knowledge-sharing policies among analysts must undergo a systematic approach. To begin with, these activities require the formalization of skill types and levels. If different types of skills are defined with letters (e.g. 5 different types of skills are labeled from "A" to "E") and different levels of skills are defined with numbers (e.g. expert level is 3, average level is 2 and novice level is 1), then the cyber defense skill map can be formalized by means of a 2-entry table such as the one depicted by Table 1 in which the cyber defense security staff is supposed to consists of 10 units.

Table 1: Example of Skill Types and Levels of the Cyber Defense Security Staff

Analyst/Skill	A	B	C	D	E
analyst 1		3		3	3
analyst 2	1	2	2		
analyst 3				1	1
analyst 4	3	2			
analyst 5		2	1		3
analyst 6		2	2	3	
analyst 7		1	3	2	
analyst 8			3		1
analyst 9	3		1	2	
analyst 10	3	2	1		

Besides “who knows how to do what”, from the above table one may derive additional information. For instance, since analyst 1 is a senior worker in three skills (i.e. B, D and E), while analyst 3 is a junior worker bearing two of the same skills held by analyst 1 (i.e. D and E), if analyst 3 is meant to improve on skills D and E or start training on skill B, then he/she is likely to be engaged by senior management in teamwork with analyst 1. This and similar options lead to considering different knowledge-sharing policies and practices that may be addressed by the company.

In general, depending on their roles and/or tasks performed within a specific workflow, cyber defense analysts in a company may be called to *i*) work alone or *ii*) in consultation with other analysts who are committed to a common mission and are willing to share the knowledge that is necessary to fulfill that mission (Kvan and Candy 2000). The former case may apply in small companies that operate a non-collaborative policy because analyst staff is limited in number and each unit is dedicated to monitoring a specific cyber asset (see left-hand side of Figure 2). The latter case may apply in more complex organizations in which a team of analysts, each bearing specific knowledge and behavioral characteristics, exploit a set of rules to study macro-level patterns emerging from micro-level interactions among team members (see right-hand side of Figure 2).

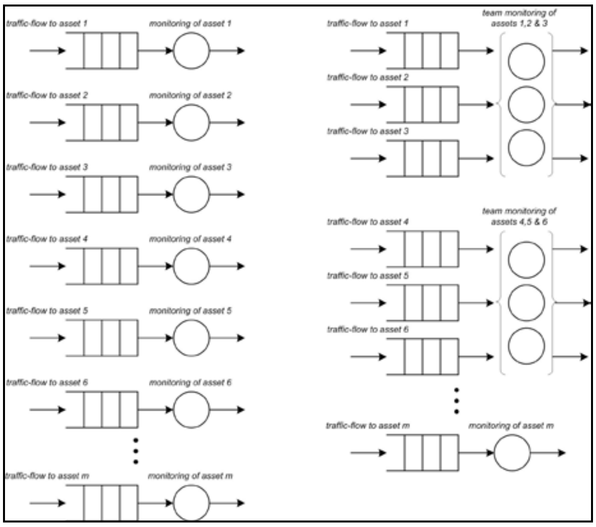


Figure 2: No Knowledge-sharing vs Knowledge-sharing Working Procedure

Whatever be the working procedure, let us consider the case in which a company operates an evaluation program according to which an analyst generates a certain number of credits for every attack mitigated. A credit is a measure of security performance ranking from 1 to 4 according to the type of attack. If the analyst works alone, every type of threat for which he/she is skilled will be detected and mitigated according to the proper service time - the service time depends on the type of attack and the analyst’s (expert, average or novice) level of skill. As a result of attack mitigation, the analyst will be “rewarded” with the entire credit. If no such expertise is held by the analyst, the lack of ability to mitigate the malicious attack will have a negative impact on the entire system and likely cause a “loss” of overall performance. On the other hand, if the analyst works in consultation with other analysts, two situations may occur depending on whether or not the analyst holds the appropriate skill to detect and manage an attack. If he/she is properly skilled, then the attack will be dealt with according to the expertise of the analyst who, in turn, will be rewarded the entire credit at the end of the attack management process. Vice versa, if the analyst is obliged to consult with his/her team members in order to acquire the necessary know-how to manage the attack, then as a result of the ongoing interaction process: the service times of all the interacting team members will be inflated; the status of the skill level of the “enquiring” analyst will change thanks to the learning process he/she is undergoing; and the final credit will be shared among the team members that took part in the knowledge-sharing process. Of course, if none of the team members hold the appropriate skill to manage the attack, in the same way as the work alone *modus operandi*, this lack will have a negative impact on the entire system and cause a loss of overall performance.

In the present study we consider both of the above policies, but individually and propose for each configuration a qualitative/quantitative simulation-based optimization methodology to evaluate attack tolerance, along with the related performance degradation. Specifically, under a given attack scenario by which we mean different types and rates of attacks targeting a predefined set of cyber assets, the resulting model is aimed to estimate the following (average) performance metrics:

- percentage of attacks mitigated;
- resource (analyst) utilization;
- number of credits gained

and in addition for the knowledge-sharing policy among team members:

- number of cyber defense security analysts per team;

- cyber defense security team composition in terms of skill types and levels held by every single analyst assigned to every single team;
- knowledge gain.

3. SO METHODOLOGY

It should be clear at this point that the problem at hand is a problem of cost-performance-security evaluation of security services, where both (scarce and costly) human resources have to be allocated to important assets in a rational way in order to face the execution of activities and transactions by several actors under different policies. In this section we show how simulation-based optimization takes over whenever optimal resource allocation is more properly modeled in stochastic environment, due to the exclusive capability of discrete-event simulation to reproduce attacker activities against vulnerable assets and defender responses under security controls and policies. We expect to achieve a rational cost-effective organization of security analysts devoted to activity and resource monitoring, along with a rational cooperation and training of skilled personnel.

To show that this is the case here, we present an integer programming based mathematical formulation of the simplified decision problem of allocating analysts from several sources to a given asset to face multiple types of possible threats. It is inspired from the classical multi-choice multi-knapsack problem (Hifi, Michrafy and Sbihi 2004). Let's assume that:

- $i = 1,..., I$ analysts are available, each with a certain number of skills measured in terms of capability of covering to some extent a fixed set of threats;
- $j = 1,..., J$ assets are given as subjected to $k = 1,..., K$ threats;
- r_{ij}^k is the expected reward achieved (amount of risk covered) with respect to threat k by allocating analyst i to asset j;
- analyst i could be allocated to asset j (through the binary x_{ij} variable) at an expected cost c_{ij}.

We want to minimize the total allocation cost under the constraint of guaranteeing a fixed level of coverage (R_k) against each threat.

The formulation follows.

$$\min \sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij} x_{ij} \tag{1}$$

$$\sum_{i=1}^{I} \sum_{j=1}^{J} r_{ij}^k x_{ij} \geq R_k \qquad k = 1,..., K \tag{2}$$

$$\sum_{j=1}^{J} x_{ij} = 1 \qquad i = 1,..., I \tag{3}$$

$$x_{ij} \in \{0,1\} \qquad i = 1,..., I; \quad j = 1,..., J \tag{4}$$

In principle, a mathematical programming formulation as the one just presented can be embodied in a simulation-based optimization algorithm (see, for example, the context-specific chart illustrated in Figure 3). Once both the objective and the set of constraints have been formulated, i.e. formalized under some linear or non linear functions and inequalities, and an initial feasible solution is available, then statistical analysis of the simulation output data gathered from one sufficiently-long run or multiple simulation runs allows to estimate the expected values of the predefined security rewards and costs corresponding to the current feasible solution. An iterative search process is then applied with the aim of exploiting the neighborhood of the current solution and sometimes suitably escaping from local minima to explore the whole feasible set of solutions.

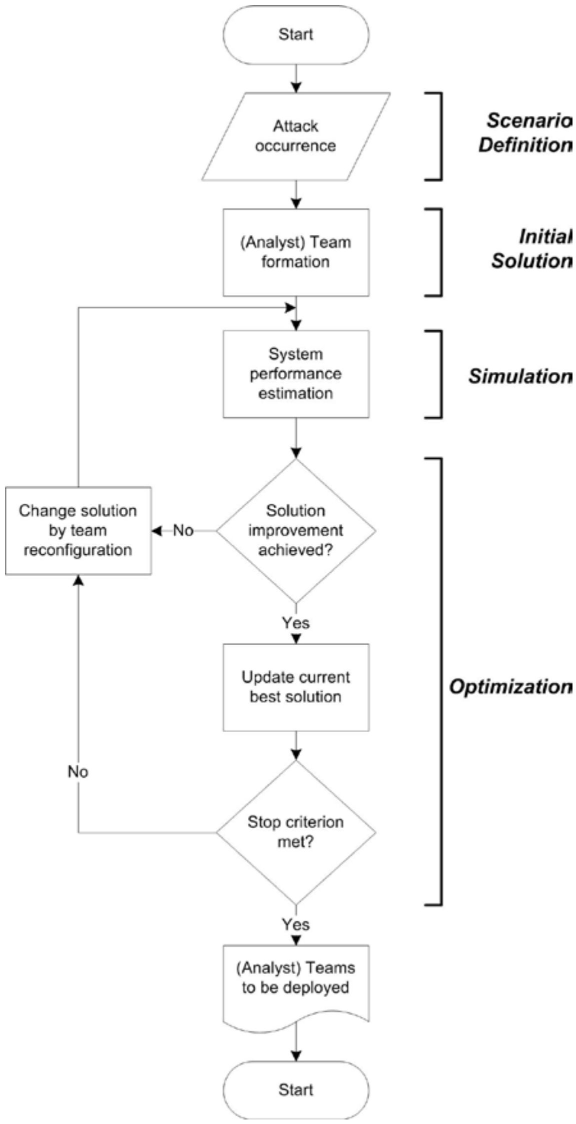


Figure 3: The Context-Specific SO Procedure

In a "complete" cyber environment, one could accept the deterministic measure of the c_{ij} parameter as a pure (monetary) cost. However, the inadequacy of such a formulation becomes evident by recalling the more complex policies and objectives underlying the complex cooperation and training settings under the randomly occurring events described in the previous section. So, we focus on a simulation-based optimization approach and simply put aside the IP formulation. As for the search process, here two different simulation optimization algorithms are tailored to the problem: *simulated annealing* and *ant colony optimization*.

3.1. Heuristic Methods

Heuristic methods represent the latest developments in the field of direct search methods that are frequently used for simulation optimization. Many of these techniques balance the global search for promising solutions within the entire feasible region (*exploration*) with the local search of promising sub-regions (*exploitation*), thereby resulting in efficient global search strategies. In the following we tailor two different types of meta-heuristics to our team formation problem: *simulated annealing* and *ant colony optimization*. The reason for choosing to illustrate these two meta-heuristics is due to the fact that, given a fixed time budget, they represent two different philosophies in the implementation of a computationally expensive activity such as simulation optimization. In the former, more computational time is devoted to estimate via simulation (*evaluation process*) the solutions found because only one neighbor solution is sampled and more time is left for the simulation process. In the latter, more time is devoted to find improved solutions on the optimization side (*search process*) because of the wider sampling of the neighborhood. In general, for any problem it is impossible to determine *a priori* which option carries a better pay-off.

3.1.1. Simulated Annealing

Originally introduced by (Kirkpatrick, Gelatt and Vecchi 1983), simulated annealing (SA) was developed on the similarities between combinatorial optimization problems and statistical mechanics. In the field of metal sciences, the annealing process is used to eliminate the reticular defects from crystals by heating and then gradually cooling the metal. In our case, a reticular defect could be seen as grouping analysts in teams that are not able to "properly" cover cyber assets and, thus, guarantee a given quality of service level when the above assets undergo an attack.

Technically speaking, the annealing process is aimed to generate feasible teams of analysts, explore them in a more or less restricted amount and, finally, stop at a satisfactory solution. To avoid getting caught in local minima, during the exploration process a transition to a worse feasible solution (higher-energy state) can occur with probability

$$p = e^{\Delta/T} \qquad (5)$$

where Δ is the difference between the values of the objective function (energy) of the current solution (state) θ and the candidate solution θ_t and T is the process temperature. A prefixed value of T determines the stop of the entire process and it usually decreases according to a so-called *cooling schema*. Unfortunately, in the literature there is no algorithm that can determine "correct" values for the initial temperature and cooling schema, but, as suggested by empirical knowledge simple cooling schemas seem to work well (Ingber 1993).

In the following, some pseudo-code is given for the original SA algorithm for a minimization problem.

Algorithm 1: Simulated Annealing

$\theta \leftarrow$ *initial solution*
for *time = 1* **to** *time-budget* **do**
    $T \leftarrow$ *cooling-schema[time]*
    if $T = 0$ **then**
        Present *current solution* as the estimate of the optimal solution and **stop**
    Generate a random neighbor θ_t of the current solution θ by performing a *move*.
    $\Delta = f(\theta) - f(\theta_t)$
    if $\Delta > 0$ **then**
        $\theta \leftarrow \theta_t$
    else
        $\theta \leftarrow \theta_t$ *(with probability* $p = e^{\Delta/T}$ *)*
end for

When customizing the SA algorithm to our problem, some choices need to be made.

To begin with, choosing the proper cooling schema has great impact on reaching a global minimum. In particular, it affects the number and which analysts are assigned to a team (solutions) that will be evaluated by running the SA algorithm. To this end, the so-called simple mathematical cooling schema $T_{i+1} = \alpha \cdot T_i$ has been tested, and the best results are returned for an initial temperature $T_0 = 100$ and a decreasing rate $\alpha \approx 0.9$.

The "move" definition for neighborhood generation is very context-sensitive. For our problem, a move must be defined with respect to the feasibility (or lack thereof) of a team by taking into account the analysts' skills, as well as the constraint that limits the number of analysts that can communicate and, thus, be assigned to the same team. Some examples of moves are:

- move analyst l from team i to team j ($i \neq j$);
- swap analyst l and analyst k ($l \neq k$), originally assigned to team i and team j ($i \neq j$), respectively.

As far as the stopping criteria are concerned, designers can choose among the following possibilities:

- stop when the algorithm has reached a fixed number of iterations n or an upper bound on the available time-budget;
- stop when the current solution has not been updated in the last m iterations;
- stop when the cooling schema has reached a *lower bound* on the temperature.

3.1.2. Ant Colony Optimization

Ant colony optimization (ACO) is a population-based metaheuristic for combinatorial optimization problems which was inspired by the capability of real ants to find the shortest path between their nest and a food source. Dorigo (1992) developed the first ACO algorithm called ant system and applied it to solve the traveling salesman problem (TSP). In this problem, an ant builds a tour by moving from one city to another until all cities have been visited and the objective is to find the tour that minimizes the distance traveled in visiting all cities.

The probability that ant *k* in city *i* chooses to go to city *j* is given by the following rule

$$p_k(i,j)=\begin{cases}\dfrac{[\tau(i,j)]\cdot[\eta(i,j)]^{\beta}}{\sum_{g\in J_k(i)}[\tau(i,g)]\cdot[\eta(i,g)]^{\beta}} & \text{if } j\in J_k(i)\\ 0 & \text{otherwise}\end{cases}\tag{6}$$

where $\tau(i,j)$ is the pheromone associated to the connection from city *i* to city *j*, $\eta(i,j)$ is a simple heuristic guiding the ant, for example $\eta(i,j)=1/d(i,j)$ where $d(i,j)$ is the distance between the two cities, and β is used to define the importance of the heuristic information as opposed to the pheromone information.

Once ant *k* has built a tour, the pheromone trail is updated according to

$$\tau(i,j)=\rho\cdot\tau(i,j)+\sum_{k=1}^{m}\Delta\tau_k(i,j)\tag{7}$$

where ρ is the evaporation rate and

$$\Delta\tau_k(i,j)=\begin{cases}\dfrac{1}{L_k} & \text{if } (\mathrm{i},\mathrm{j})\in \text{tour of ant } k\\ 0 & \text{otherwise}\end{cases}\tag{8}$$

is the pheromone increase on all the edges visited by the all the ants (the more the ants visit an edge, the greater the pheromone they leave).

In the following, some pseudo-code is given for a basic ACO algorithm for a minimization problem.

Algorithm 2: Ant Colony Optimization

P $\leftarrow$ pheromone initialization
$\theta_{gb} \leftarrow$ global best solution is null
for *time = 1* **to** *time-budget* **do**
  $\Theta_{iteration} \leftarrow \{\}$
  for *j=1* **to** *n° of ants*
    $\theta \leftarrow$ build a solution based on P
    if θ is feasible **then**
      if $(f(\theta)< f(\theta_{gb}))$ or θ_{gb} is null **then**
        $\theta_{gb} \leftarrow \theta$
      $\Theta_{iteration} \leftarrow \Theta_{iteration} \cup \{\theta\}$
  end for
  P $\leftarrow$ pheromone update
end for
return θ_{gb}

Some variants to the original algorithm have been proposed such as:

- update the pheromone trail by allowing only the best ant to place pheromone after an iteration of the algorithm $\tau(i,j)=\rho\cdot\tau(i,j)+\Delta\tau_{ij}(\text{best iteration ant})$;
- update the pheromone trail every γ iterations by allowing only the best global ant to place pheromone $\tau(i,j)=\rho\cdot\tau(i,j)+\Delta\tau_{ij}(\text{best global ant})$.

Once again we must think of customization: the TSP is an ordering problem, while in our case we face a grouping problem. AS has been applied to solve other grouping problems such as bin packing, cutting stock (see, for example, Levine and Lucatelle 2004) and graph coloring (see, for example, Costa and Hertz 1997).

Rather than visiting cities, in our problem an ant moves to connect analysts with different skills and, thus, form teams to defend a given set of cyber assets. In doing so, the ant leaves a pheromone trail between analysts *i* and *j* which may be seen as the global goodness of teaming *i* and *j*. So the probability of an ant *k* connecting analyst *i* with *j* is still given by (6), where $\eta(i,j)$ is the number of different skills obtained when teaming the analysts. Once ant *k* has teamed all the analysts the pheromone trail is updated according to (7) where ρ is the evaporation rate and

$$\Delta\tau_k(i,j)=\begin{cases}\dfrac{\sum_{j\in K_k(i)}\tau(i,j)}{\eta(i,j)} & \text{if } I\neq\{\ \}\\ 1 & \text{otherwise}\end{cases}\tag{9}$$

is the pheromone value given by the sum of all the pheromone values between analyst *i* and all the analysts connected to *i* (including *j* obviously)

As for the stopping criteria, designers can chose among the same options given for the customization of the SA algorithm.

4. ILLUSTRATIVE EXAMPLE

4.1. Preliminary Verification

The SO model has been implemented in Microsoft Visual Basic 6.0 Professional and experiments have all been run on a personal computer equipped with a 2.26Hz Inter Core™2 duo processor and 3 GB of RAM. Input data and SO parameters are specified in the proper sections of a simple GUI panel, as the one illustrated in Figure 4. In particular, one must first define the attack scenario by specifying attack interarrivals (in time units) and composition (in percentage) with respect to different types of attacks. Skills are then defined by specifying for every analyst which skills he/she features and the level of competence for each skill (0=no skill, 1=novice, 2=average, 3=expert). After inserting the number of teammates (here ranging between 1 and 10), the input stage is then completed by providing the settings for the simulated annealing-based SO scheme and the simulation settings which are, respectively, the initial temperature along with the cooling rate of the SA procedure, and the time horizon, the number of simulation runs to be performed and the simulation seed.

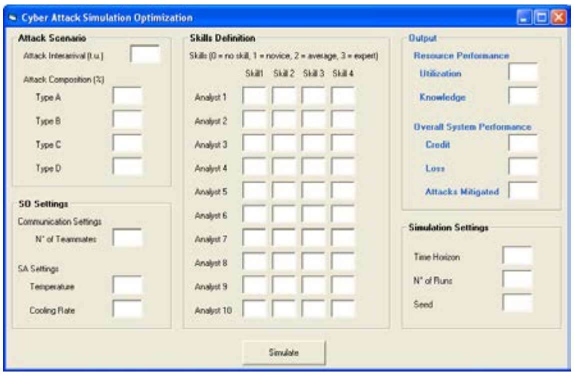


Figure 4: Snapshot of the SO Tool

The design and implementation of the SO tool depicted in Figure 4 has been carried-out in compliance with all the conventional steps used to guide a thorough and sound simulation study (Banks et al. 2001). For the time being, due to the unavailability of real-world input data as driving force for the SO model, here we will focus on illustrating the "predictive" capability of the model by explicitly reporting on the verification, rather than validation step of the study.

Verification has been performed with respect to the input parameters and logical structure of the SO model by combining three classes of techniques: common-sense techniques, thorough documentation and traces. A set of *ad hoc* instances have been used for verification by running the model in boundary cognitive conditions which allow expecting a predetermined system behavior. For instance, lets us consider the case in which every analyst *i* has an expert competence in every skill *j* (i.e. *SkillMap*(*i*,*j*)=3 for every (*i*,*j*)). Now, since every single analyst is skilled at the highest level for any type of attack, whatever be the number of teammates in the given scenario, the SO model is likely to return a fairly stable *utilization* of analysts (U) along with high accomplishments in both *credit* (C) and *number of attacks mitigated* (AM), while no *loss* (L) or whatsoever gain in *knowledge* (K) should be recorded. As a result, as one may see from Table 1, in which all of the above indices are averaged over (suitable) multiple runs and plotted as the number of teammates grows from 1 to 10, there are no significant changes in system performance whatever be the level of collaboration between analysts inserted by the user.

Table 2: Model Verification in a "Boundary" Scenario

Teammates	U	K	C	L	AM
1	0.66	0	14349	0	5736
2	0.66	0	14066	0	5708
3	0.67	0	14389	0	5751
4	0.66	0	14254	0	5705
5	0.66	0	14358	0	5736
6	0.67	0	14397	0	5756
7	0.67	0	14388	0	5758
8	0.66	0	14294	0	5720
9	0.66	0	14314	0	5702
10	0.67	0	14443	0	5768

The results of this and other similar experiments allow us to be confident in the correctness of the SO model.

4.2. Problem Set-up

In this subsection we present the problem set-up involved in the second set of experiments designed to estimate performance under a given attack scenario.

As far as system features are concerned, we consider 10 cyber assets which are attacked, according to an exponential renewal process, on average every 200 time units (λ, the average interarrival rate, is thus equal to 1/200) from a combination of 4 different types of attacks (i.e. 60% type A, 25% type B, 10% type C and 5% type D). Defense is provided by a team of 10 cyber analysts, where each analyst is skilled according to the data reported in Table 3. In the given scenario, analysts respond to attacks by working alone (n° of teammates=1) or in cooperation with other analysts (n° of teammates>1). The rate (μ) of the attack mitigation activity depends on the type of attack, the skill level held by the analyst and if mitigation occurs alone or in cooperation with other analysts. In the later case, mitigation times are inflated by 25%.

Table 3: Skill Types and Levels of the Cyber Defense Security Staff

Analyst/Skill	A	B	C	D
analyst 1	1	0	1	0
analyst 2	3	0	0	0
analyst 3	0	2	0	0
analyst 4	1	0	3	0
analyst 5	2	0	0	0
analyst 6	1	3	0	0
analyst 7	0	0	1	3

analyst 8	2	0	0	0
analyst 9	0	3	0	0
analyst 10	3	0	1	0

The initial temperature and the cooling rate of the SA scheme are set equal to 100 and at least 0.948, respectively, so that at least 100 different team-formation and assignment configurations are considered for the given scenario. The time horizon is fixed at 14400 time units (i.e. one four-hour labor shift) and, based on system variance, from 10 to 30 runs are performed for each experiment in order to obtain point estimates and/or construct 95% confidence intervals for estimating resource (i.e. analyst) *utilization* and *knowledge* gain and system *credit*, *loss* and *number of attacks mitigated*. Here, for clarity of illustration, we prefer using (stable) point estimates to show the numerical results in the next subsection.

4.3. Numerical Experiments

In this subsection, the illustration and related discussion of the numerical results returned by the previously defined scenario of the SO model is quite clear. In summary, except for the case in which an analyst is called to work alone, cooperation in small teams seems to return better performances.

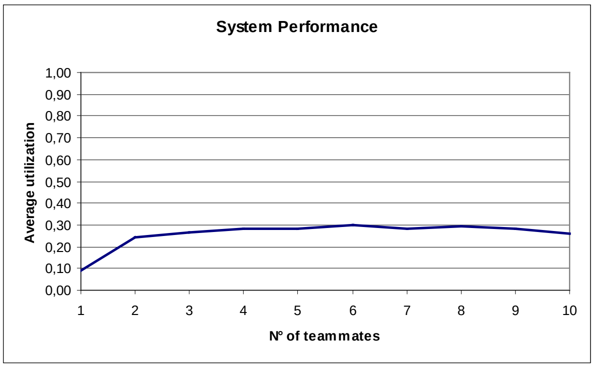


Figure 5: Trend of Average Analyst Utilization

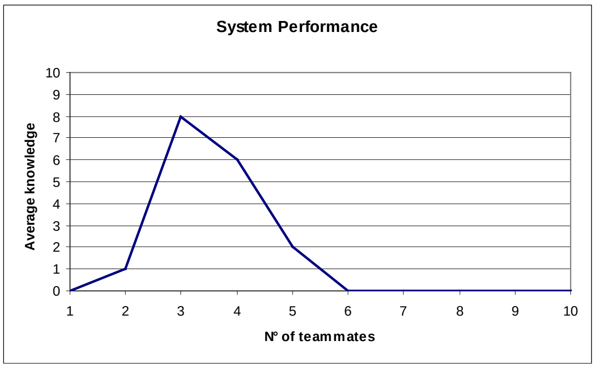


Figure 6: Trend of Average Knowledge Gain

To begin with, let us consider the (average) values of what we believe to be two measures of resource performance: analyst utilization and knowledge gain. Figures 5 and 6, respectively, show how utilization reaches its peak when the number of *teammates* is equal to 3, while utilization stays rather stable for $teammates > 3$. As one may see from Table 4, this happens in conjunction with a specific asset-analyst assignment and subsequent team formation in which analysts with complementary skills have been teamed together.

Table 4: Details of Best Asset-Analyst Assignment and Team Formation when Number of Teammates=3

Analyst	Asset	Teammates	Skills Covered
1	10	6 & 9	1, 2 & 3
2	5	4 & 10	1 & 3
3	8	5 & 7	1, 2, 3 & 4
4	3	2 & 8	1 & 3
5	6	3 & 6	1 & 2
6	7	1 & 5	1, 2 & 3
7	9	3 & 10	1, 3 & 4
8	1	4 & 9	1, 2 & 3
9	2	1 & 8	1, 2 & 3
10	4	2 & 7	1, 3 & 4

As for the so-called system performance measures, system loss is totally overcome when $teammates \geq 4$, while the credit acquired and the number of attacks mitigated do not seem to experience significant changes beyond the above value of *teammates* (see Figures 7, 8 and 9, respectively). This is even more evident if one plots the point-by-point difference between system credit and loss over the number of teammates: for $teammates \geq 4$ this trend is stable on a value approximately equal to 1200 units of credit.

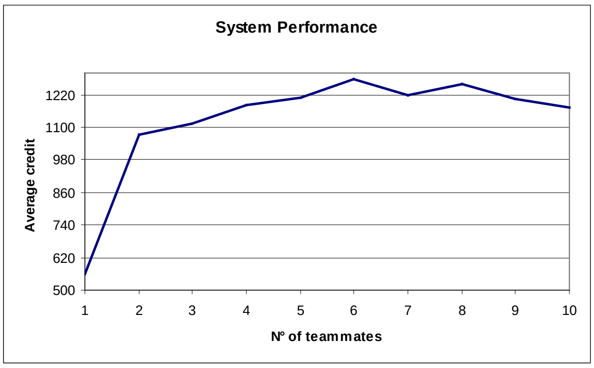


Figure 7: Trend of Average System Credit

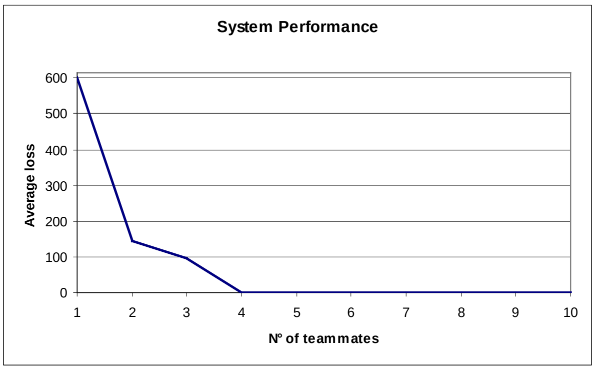


Figure 8: Trend of Average System Loss

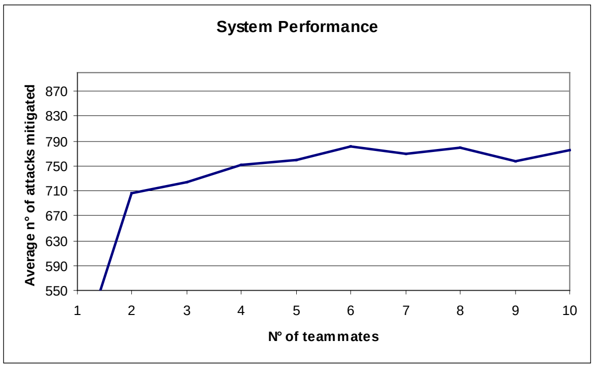


Figure 9: Trend of Average Number of Attacks Mitigated

So now the final question should be whether or not there are changes in system behavior when the average interarrival rate (λ) of the attacks grows significantly, for example by a factor of four (e.g. average interarrival times go from 200 to 50 time units), while clearly keeping satisfied the classical stability condition $\lambda/\mu < 1$ where we recall μ being the service rate (i.e. attack mitigation rate). Regardless of the growth of the interarrival rate, numerical experiments not reported here for lack of space state that only minor changes occur in system behavior such as system loss that goes to zero at a lower speed and specifically when $teammates = 10$, rather than $teammates \geq 4$.

5. CONCLUSIONS

In this paper we presented a simulation-based optimization model to assess attack tolerance in cyber systems when man-monitored assets are targeted by different types of attacks and different attack rates. The opportunity of teaming cyber defense analysts to work together in attack mitigation, rather than implementing a non-cooperative working policy seems to suitably fit the twofold purpose of cyber security defense unit: protection and learning. Numerical evidence shows that resource (analyst) utilization and knowledge, as well as system loss, credit and number of attacks mitigated benefit from the formation of small, yet well-assorted teams as far as the skills held by analysts belonging to the same team are concerned. In the proposed set of sample experiments, the best overall performance is recorded for a number of teammates ranging between 3 and 4.

Future research effort will focus on input modeling pertaining to both data and organizational matters. In the former case, a finer representation of the stochastic cyber attacks process will be provided via phase-type distribution-based models to match, for instance, the dynamics of attack graphs. In the latter, new policies will be considered. In particular, the current version of the SO model may be seen as the in vitro lab to assess the proficiency of proactive policies within the wider underlying attacker-defender game logic in cyber security.

ACKNOWLEDGMENTS

This work was partly supported by the Cyber Security Technological District funded within the Italian National Operational Programme for Research and Competitiveness 2007-2013 under grant number PON03PE_00032_2.

REFERENCES

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2000. *Discrete-Event System Simulation*. 3rd ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Carson, Y. and Maria, A., 1997. Simulation optimization: methods and applications. In *Proceedings of the 1997 Winter Simulation Conference*, Andradóttir, S., Healy, K.J., Withers, D.H., and Nelson, B.L. (eds.), 118-126. December 7-10, 1997, Atlanta (Georgia, USA).

Costa, D. and Hertz, A., 1997. Ants can colour graphs. *Journal of the Operational Research Society*, 48:295–305, 1997.

D'Amico A. and Whitley, K., 2007. The Real Work of Computer Network Defense Analysts. In *VizSEC 2007. Proceedings of the Workshop on Visualization for Computer Security*. J.R. Goodall, G. Conti, K.-L. Ma (eds.), Springer-Verlag Berlin Heidelberg, 19-37.

Dorigo, M., 1992. Ottimizzazione, apprendimento automatico, ed algoritmi basati su metafora naturale. Thesis (PhD), Politecnico di Milano, Italy.

Fischer, M.J., Masi, D.M.B., Chen, C.-H., and Shortle, J.F., 2010. Simulating non-stationary congestion systems using splitting with applications to cyber security. In In *Proceedings of the 2010 Winter Simulation Conference*, Johansson, B., Jain, S., Montoya-Torres, J., Hugan, J., and Yücesan, E. (eds.), 2865-2875, December 5-8, 2010, Baltimore (Maryland, USA).

Fu, M. and Nelson, B., 2003. Guest Editorial. *ACM Transactions on Modeling and Computer Simulation* 13(2), 105–107.

Hifi, M., Michrafy, M. and Sbihi, A., 2004. Heuristic algorithms for the multiple-choice multidimensional knapsack problem. *The Journal of the Operational Research Society*, 55(12):1323-1332.

Ingber, L., 1993. Simulated annealing: Practice versus theory. *Mathematical Computer Modelling*, 18(11):29-57.

Kiesling, E., Ekelhart, A., Grill, B., Strauss, C. and Stummer, C., 2013. Simulation-based optimization of information security controls: an adversary-centric approach. In *Proceedings of the 2013 Winter Simulation Conference*, Pasupathy, R., Kim, S.-H., Tolk, A., Hill, R., and Kuhl, M.E. (eds.), 2054-2065, December 8-11, 2013, Washington DC (USA).

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M. P., 1983. Optimization by Simulated Annealing. *Science*, New Series, 220(4598):671-680.
Kvan, T. and Candy, L., 2000. Designing Collaborative Environments for Strategic Knowledge in Design. *Knowledge-Based Systems* 13(6), 429-438.
Levine, J. and Lucatelle, D., 2004. Ant colony optimisation and local search for bin packing and cutting stock problems *Journal of the Operational Research Society*, 55(7):705–716.
Masi, D.M.B., Fischer, M.J., Shortle, J.F., and Chen, C.-H., 2010. Simulating network cyber attacks using splitting techniques. In In *Proceedings of the 2011 Winter Simulation Conference*, Jain, S., Creasey, R.R., Himmelspach, J., White, K.P., and Fu, M. (eds.), 3217-3228, December 5-8, 2010, Baltimore (Maryland, USA).
Pasupathy, R. and Henderson, S.G., 2011. SimOpt: a Library of Simulation Optimization Problems. In *Proceedings of the 2011 Winter Simulation Conference*. Jain, S., Creasy, R.R., Himmelspach, J., Whote, K.P. and Fu, M. (eds.), 4080-4090, December 11-14, 2011, Phoenix (Arizona, USA).
Wang, L.-F. and Shi, L.-Y., 2013. Simulation optimization: a review on theory and applications. *Acta Automatica Sinica*, **39**(11): 1957−1968.
Zhang, B., Shafi, K., and Abbass, H.A., 2012. Robo-teacher: a computational simulation based educational system to improve cyber security. *Robot Intelligence Technology and Applications*, Kim, J.-H., Matson, E.T., Myung, H., and Xu, P. (eds.) AISC 208 179-186.

AUTHORS BIOGRAPHY

Pasquale LEGATO is an Associate Professor of Operations Research in the Department of Informatics, Modeling, Electronics and System Engineering 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. His email is legato@deis.unical.it and his web page is www.deis.unical.it/legato.

Rina Mary MAZZA is the Research Manager in the Department of Informatics, Modeling, Electronics and System Engineering at the University of Calabria, Rende (CS, Italy). She received a Ph.D. degree in Operations Research from 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 terminal containers. Her current research interests include discrete-event simulation and optimum-seeking by simulation in complex systems. Her e-mail address is rmazza@deis.unical.it.

MODEL TRANSFORMATION FROM BPMN FORMAT TO OPENERP WORKFLOW

Zhe Liu [(a)], Gregory Zacharewicz[(b)]

University of Bordeaux, Lab. IMS, UMR 5218, 33405 Talence, France.

[(a)]liuzhe.ch@gmail.com, [(b)]gregory.zacharewicz@ims-bordeaux.fr

ABSTRACT

This article presents a model transformation from a conceptual modeling language to a platform-specific model. Business people do not have IT background when creating views of their process nevertheless at the end they look at implementing their design. The idea is to support the model transformation from concept to implementation. The source model is BPMN (Business Process Model and Notation) that is a graphical representation used for conceptual modeling in business process management. The target is OpenERP that is an open-source enterprise resource planning (ERP) and customer relationship management (CRM) software which consist in two functional modules, end-user processes and workflow engine. In this research, using model-driven architecture, a mapping relation is built up to bridge the two different models. And with the help of XSLT, the transformation can be achieved, tested and validated.

Keywords: model transformation, BPMN, workflow, OpenERP

1. INTRODUCTION

Business process management and workflow management are both common and important approaches in business process. They help people intuitively understand what to do in business process.

Among the many business process management methods, BPMN is a symbolic language used for specifying business processes using a process models.

In the meantime, OpenERP is the most popular ERP software for European small and medium enterprises. Today it has a number of collaborators and participants all over the world (Pinckaers et al. 2013).

As mentioned above, OpenERP is widely used in European SMEs, which means it has a large amount of users. However, considering about business process modeling, most people are still accustomed to use BPMN. Converting BPMN files to OpenERP workflow will extend a new function in business process management of OpenERP and help people have a better understanding of their business process, have a better communication and cooperation, improve efficiency, and reduce mistakes and omission.

2. BACKGROUND

2.1. Workflow

There are many definitions of workflow given by different persons or organizations. To define standards for the interoperability of workflow management systems, an association, the Workflow Management Coalition (WfMC), founded in 1993.

According to the definition given by WfMC (1996):

Workflow is "*the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant (also called resource, which means a human or machine) to another for action, according to a set of procedural rules*."

Its synonyms are workflow management, workflow computing and case management.

2.2. Process

As for process, it has different meanings in different domains. To find an accepted definition, we can see the one defined by WfMC (1996):

A process is "*the representation of a business process in a form which supports automated manipulation, such as modelling, or enactment by a workflow management system*."

And as for business process, defined by WfMC, it is "*a set of one or more linked procedures or activities which collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships*."

2.3. BPMN

BPMN is flow-chart based notation for defining business processes. BPMN is an agreement between multiple modeling tools vendors, who had their own notations, to use a single notation for the benefit of end-user understand and training. (Stephen 2004)

BPMN consists of a set of graphical elements. The four basic kinds of elements are flow objects (events, activities and gateways), connecting objects (sequence flow, message flow and association), swimlanes (pool and lane) and artifacts (data object, group and annotation).

2.4. Signavio
With the widespread use of BPMN, many BPMN drawing software and plugins are released. Apart from Visio as we all know, EdrawMax, intalio designer, bonitaBPM, Signavio are all BPM drawing tools.

In this project, we choose Signavio web as the source tool from the BPMN drawing tools mentioned above. The Signavio Process Editor is a network-based business process modeling tool.

User can export drawing in different formats, Signavio archive (SGX), BPMN 2.0 XML, PNG, SVG and PDF in Signavio, which provides many choices. Considering the OpenERP workflow format, the output format of Signavio is more suitable and easy to use.

2.5. OpenERP
OpenERP system provides a flexible modular architecture, including financial management, purchasing/sales management, inventory management, MPR, CRM, human resource management, point-of-sale management, project management, etc. (Xiao 2010).

Among all the functional modules of OpenERP, business process management has two parts, one is end-user processes, and another is workflow engine. In this topic, we also consider these two aspects.

2.5.1. OpenERP workflow
Workflows represent the company's different document flows. They are used to define the behavior of a given file. Developers and system implementers use workflows to determine which object perform the actions, which actions should be performed and at which moments the actions performed.

A workflow is expressed by a directed graph, in which the nodes are called "activities" and the arrowed lines are called "transitions".

Workflow activities represent the nodes of workflows. These nodes are the actions to be executed. They state the work that needs to be performed in the OpenERP server, such as modifying the content of some records, or sending emails.

Workflow transitions are the conditions to be satisfied to go from one activity to the next one; they control how the workflow progresses from activity to activity. Transitions are represented by one-way arrows joining two activities.

In OpenERP workflow format, three main elements exists which are workflow, workflow activity and workflow transition.

2.5.2. OpenERP process
In OpenERP, enterprise process is a view, that is to say, it is a display interface, rather than a real flow. The actual flow is workflow. So, enterprise process is a display view in a different form, like the form view. While workflow is a function that changes the object's state, not a view which is used to display. However, the difference between enterprise process and other views is that each node in enterprise process can be associated with different objects. As a consequence, it can present action menu of multiple objects at the same time in enterprise process, unlike the form view which only can show one object and its related functions (report, wizard, etc.) (Xiao 2010).

Processes form a structure for all activities that enable the company to function effectively. Processes represent workflows across all of a company and the documents associated with the workflows. For this reason, user processes are associated to workflows. Processes are used by end-users to locate an action for more complete handling, and to help them understand the problems which have not been handled in OpenERP.

Corresponding to workflow activity and workflow transition of workflow, process has process node and process transition to present activities and transitions.

3. GENERAL DESIGN

3.1. MDA
The Model-Driven Architecture (MDA) is a framework based on a set of OMG standards that uses models to describe their transformation into other models or complete systems (Mellor et al. 2003).

The OMG's standards include UML (Unified Modeling Language) modeling notation, MOF (Meta Object Facility), XMI (XML Metadata Interchange), CWM (Common Warehouse Metamodel), CORBA (Common Object Request Broker Architecture) middleware, etc. These standards define the key concepts of the MDA.

3.1.1. Models in MDA
According to the different levels of abstraction, MDA identifies three model types, computation independent model (CIM), platform independent model (PIM) and platform specific model (PSM), which are also mentioned above. To make it more clearly, next, each level will be introduced them one by one.

A computation independent model doesn't show any details of the system, since it is defined by the business requirements. CIM is also called "domain model". It focuses on the requirements of the systems.

A platform-independent model is defined by Mellor et al. (2003) as *a model that contains no reference to the platforms on which it depends*. PIM is used to present the aspects of system characters that are unlikely to change with the change of the platform which the model depends on (Robert and Bernhard 2007).

The OMG defines a platform as "*a set of subsystems and technologies that provide a coherent set of functionality through interfaces and specified usage patterns*".

A platform-specific model is defined by Mellor et al. (2003) as *the result of weaving a PIM with the platforms on which it depends*. PSM describes a system in which platform specific details are integrated with the elements in a PIM.

In general, when developing an MDA-based application, the first step is to build a platform independent model (PIM), conveyed by UML on the basis of the proper core model. Platform architects then convert this common application model into one target platform which is a specific platform such as CCM, EJB, or MTS. Standard mappings allow tools to perform automatically some of the transformation.

The work in the following stage is to generate application code. The system needs to produce several types of code and configuration files. (Soley 2000).

The essence of MDA is to distinguish platform-independent model and platform-specific model.

3.1.2. Models in this case

In this topic, we focus on the process from PIM to PSM without considering CIM.

Generally, the model mapping process goes from PIM to PSM, finally to coding.

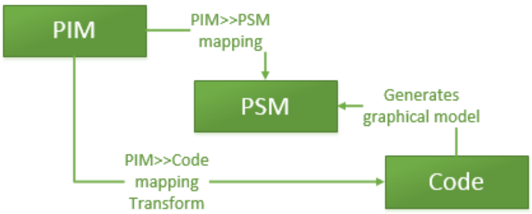


Figure 1 Model mapping in MDA of this project

However, in this transformation, it begins from PIM, after building the mapping relationship between PIM (BPMN in this case) and PSM (OpenERP workflow and business process in this case), it firstly goes to code (workflow and business process XML files) according to the mapping and then generates PSM from XML code with the help of OpenERP platform. See figure 1.

3.2. Model mapping

In the process of the development based on MDA, the transformation between the models is a very important part.

Before transforming PIM, it is necessary to identify the target platform. On the basis of the platform, then designers identify the metamodels and the mapping techniques which will be used with the metamodels (Mellor et al. 2004).

The following part explains the related concepts and design in the mapping process.

3.2.1. Model and metamodel

Defined by Mellor et al. (2003), a model is *a coherent set of formal elements describing something (for example, a system, bank, phone, or train) built for some purpose that is amenable to a particular form of analysis*, such as communication of ideas between people and machines. In other words, a model is an abstraction of the real objects or processes.

Each model, both the source model and the target model, is expressed in a sort of language. The target model's language may define different objects or processes from the source model's language. To describe the models better, we need to define the two languages in one way or another by building a model of the modeling language — a so-called metamodel. A metamodel defines the structure and well-formed rules of the model confirms to it (Tom and Pieter 2006). Which means, as a model is an abstraction of the real world, a metamodel is an abstraction of the model.

Figure 2 Model, metamodel, and platform.

See figure 2. The OMG' s Meta Object Facility (MOF) defines a model as an instance of a metamodel. A metamodel can describe a particular platform.

3.2.2. Mapping and model transformation

A model transformation *defines a relation between two sets of models*. (Robert and Bernhard 2007) If one set of models is called as a source model and the remaining one as a target model, then an automated process will take the source models as input and produce the target models as output, according to the transformation rules. We call this process as model transformation (Kleppe et al. 2003; Sendall and Kozaczynski 2003).

To realize the transformation, we build a "bridge" from source models to target models. All model transformations performed are based on to the generic transformation architecture (see figure 3) (Jouault et al. 2008). The "bridge" is a mapping technique (Bazoun et al. 2013). A mapping between models takes one or more models as its input and produces one output model. The mapping technique describes rules for the transformation. These rules are described at the metamodel level.

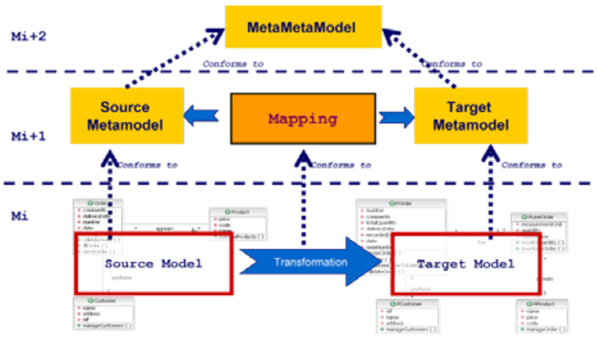


Figure 3 Transformation architecture

In this project, we make a new transformation from BPMN to OpenERP workflow.

In consideration of this case, figure 3 also shows instances of PIM and PSM models and metamodels, and how these instances relate to one another. The source model here refers to the PIM model instance, BPMN. The target models refer to the PSM model instances, OpenERP workflow and process, which contains the semantic information of the original PIM model instance, as well as the information added as a result of the mapping technique. Because the PSM metamodel describes a platform (here is OpenERP), the PSM also includes the elements that workflow and process depend on in OpenERP platform.

As mentioned above, the model mapping in this case is a little different from the general model mapping in MDA. See figure 4.

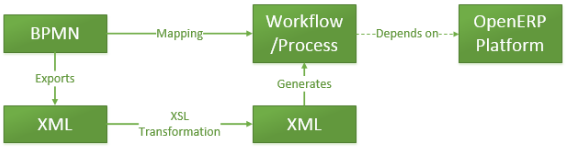


Figure 4 Model transformation in this project

To realize the transformation, first, their mapping relationship needs to be established. In accordance with the mapping, with the help of coding, the original BPMN files can be transformed to the target model files. Then, depending on the OpenERP platform, workflow graphs and process graphs can be generated.

3.2.3. Mapping relation in this case

Core elements of BPMN associated with workflow are flow objects (event, activity and gateway) and sequence flow.

To describe workflow, OpenERP workflow has three elements which are workflow, workflow activity and workflow transition. In general, pool matches workflow. Event and activity (task and sub-process) correspond to workflow activity. Sequence flow and gateway equals to workflow transition.

To describe process, OpenERP process also has three elements which are process, process node and process transition. Process is similar with workflow; the paper doesn't focus too much on describing it.

Even though BPMN and workflow are both used to model the business process, they are still two different modeling language and have totally different file formats. It is not closely coincident between their elements and attributes. The node activities and the transmission method to the next node of the flow that the two modeling methods describe in the business process are not exactly in the same way. Many concepts of them are not overlapping each other.

Figure 5 describes the corresponding relation between BPMN and OpenERP workflow and process in concept, and builds the mapping between source metamodel and target metamodel.

From the workflow perspective, it only has two major elements (namely, workflow activity and workflow transition) to describe business process, there are yet many sub-elements or attributes to make further explanation or give limiting conditions.

For workflow, in more details, "id" or "name" attributes of elements in BPMN format correspond to "id" or "name" attributes of related elements in workflow format. Start event means the value of workflow activity's "flow_start" attribute is "True", and end event means the value of workflow activity's "flow_stop" attribute is "True".

Figure 5 Mapping in concept

Gateway's attribute "gatewayDirection" is also used to describe workflow activity's attributes, which are split_mode (gatewayDirection= "Diverging") and join_mode (gatewayDirection="Converging"). Sequence flow's "sourceRef" attribute corresponds to "act_from" attribute of workflow transition, and sequence flow's "targetRef" attribute corresponds to "act_to" attribute of workflow transition.

For some elements in BPMN, such as data object and text annotation, there is no corresponding content in workflow. And for some other elements in BPMN, we can see some contents which describe the same concepts in workflow. However, the corresponding contents in workflow are not node elements just like in BPMN, but some sub-elements or attributes describe the node elements. For instance, about how to describe the trend of divergence and convergence in the business process flow. In BPMN format, we use "gatewayDirection" attribute of gateway element to describe the trend of divergence and convergence. But in workflow, we use "join_mode" and "split_mode" attributes of workflow activity to indicate whether the activity node is going to diverge or converge the flow. This makes the node describes flow divergence and convergence become a child node of the activity node from a sibling node.

Some of these attributes are hard, even impossible to map. For example, "kind", an attribute of workflow activity which shows the action type the system should execute when flow gets to this node, has four values in total, "dummy", "function", "subflow" and "stopall". Among the four values, "function" is not able to be simply mapped from BPMN. If the value of "kind" equals to "function", it indicates to execute Python code defined in "action", and execute "server action" defined in "action_id". Since the two attributes, "action" and "action_id", are being used only when the value of "kind" equals to "function", they are also difficult to map. The common case is, define a "write" method in "action", and modify the state of the related object. For the "function" type nodes, Python code defined in "action" will return "False" or a client action id.

3.3. XSLT

Since Signavio web has been choosen as the source tool, only the output format of Signavio web is considered as source model.

In OpenERP, the workflow input and the process input are both in a XML format, in this case, we choose BPMN 2.0 XML as Signavio exported format for simplicity.

As the source format and the target format are both XML formats, XSLT is used to achieve the transformation.

XSL (Extensible Stylesheet Language) is a language which is used to present XML data in a readable format. XSLT is a language for transforming XML documents into other forms, such as XML documents in another format, text documents, HTML documents, XHTML, or into XSL Formatting Objects which can then be converted to PDF, PostScript and PNG. The original document is not changed; rather, a new document is created based on the content of an existing one (Clark 1999).

To perform the XSL transformation, the programmer firstly needs to provide one or more XSLT stylesheets, which are also XML documents and describe the transformation rules that the original XML documents should follow, to translate the original XML document into an XML document conforming to the syntax in another format. The XSLT processor then reads in the original XML documents, performs transformations using XSLT stylesheets and produces the output documents needed. See figure 6.

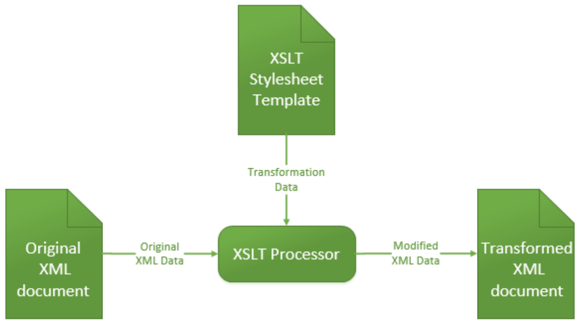


Figure 6 XSLT

As figure 6 illustrates, the original XML data is input into the processor as one input, and an XSLT stylesheet is input as a another input. The output will then be saved directly in the local folder (or a web browser in other situation) as an output XML document. The XSLT stylesheet provides XML formatting instructions, while the XML provides raw data (Burke 2001).

To transform the imported BPMN file to workflow format XML file and process format XML file, XSL code has been designed to process input files.

4. IMPLEMENTATION

In this project, to perform the mapping transformation from BPMN to coding, Java is used. There are four ways to operate XML in Java, DOM (Document Object Model), SAX (Simple API for XML), DOM4J, and JDOM (Java Document Object Model). DOM and SAX are basic means to parse XML, DOM parses XML based on the tree structure of the document, while SAX parses XML based on events stream.

With the help of these methods, there exist different XSLT processors. Among them, Saxon was chosen to be used for Java and .NET.

After developing the program, the transformation can be performed. See the realization process in figure 7

As figure 7 shows, with the business process diagram created in Signavio web, firstly export the diagram in BPMN 2.0 XML format from Signavio, then the original XML document is obtained as one input of the Saxon processor.

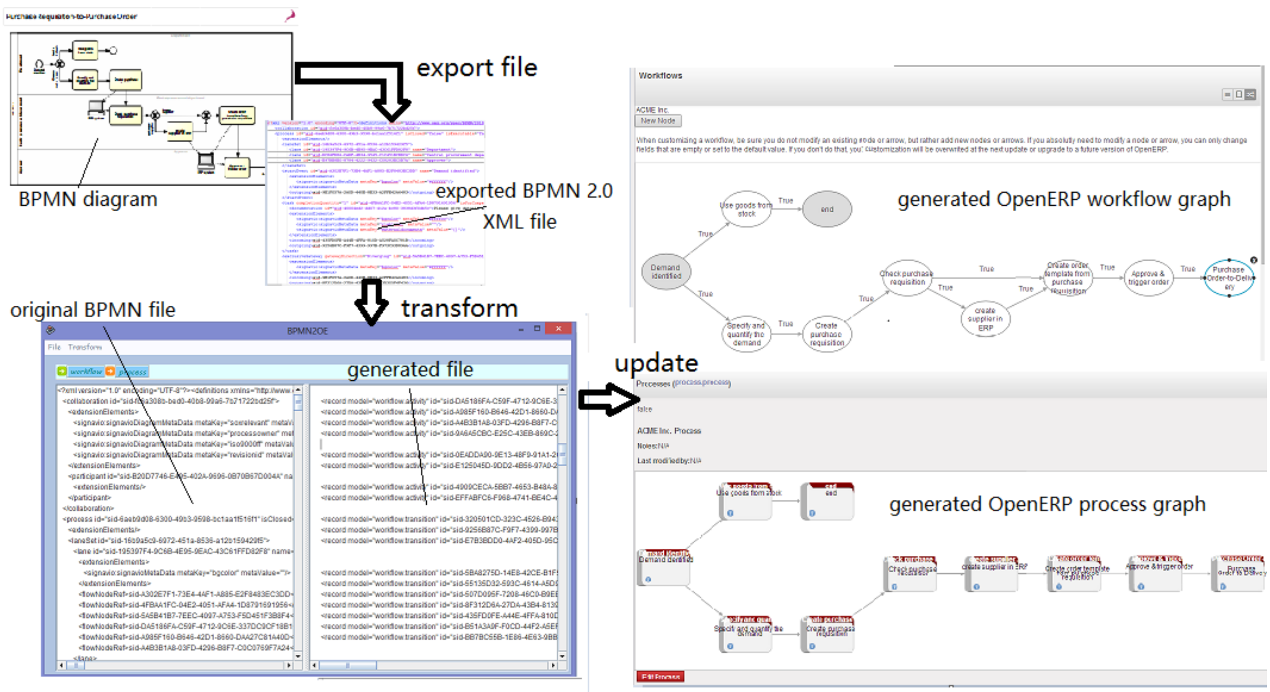


Figure 7 Implementation steps

Open the application and choose the source XML document to be transformed, the source document will be shown in the left pane. The XSLT stylesheets are designed before and provided in the application as the second input of the Saxon processor. Clicking on the button "workflow" or "process", the application will automatically generate the transformed XML document confirms to OpenERP file format (workflow or process), which will be shown on the right pane. Save the file in the proper folder and update the related module in OpenERP, the workflow or process graphs can be generated easily.

5. CONCLUSION

The paper presented a means of converting PIM BPMN to OpenERP workflow and process at PSM, and it gave an original mapping relation between BPMN and OpenERP workflow and process. This work has permitted to perform the XML files' transformation with the help of the code, thus it achieves the implemented conversion from BPMN to OpenERP workflow and process.

The next step of the work will consist in the reverse transformation from OpenERP workflow and process to BPMN.

REFERENCES

Bazoun H., Zacharewicz G., Ducq Y., Boye H.: Transformation of extended actigram star to BPMN2.0 and simulation model in the frame of model driven service engineering architecture. SpringSim (TMS-DEVS) 2013: 20

Burke E. M., 2001. Java and XSLT [M]. O'Reilly Media, Inc.

Clark J, 1999. XSL transformations (XSLT) version 1.0. W3C recommendation [J].

France R., Rumpe B., 2007. Model-driven development of complex software: A research roadmap[C]. 2007 Future of Software Engineering. IEEE Computer Society: 37-54.

Jouault F., Allilaire F., Bézivin J., Kurtev I., 2008. ATL: A model transformation tool [J]. Science of computer programming, 72(1): 31-39.

Kleppe A, Warmer J., Bast W., et al., 2003. The model driven architecture: practice and promise [J].

Mellor S. J., Clark A. N., Futagami T., 2003. odel-driven development: guest editors' introduction [J]. IEEE software, 20(5): 14-18.

Mellor S. J., Scott K., Uhl A., Weise D., 2004. Model-driven architecture [J]. Computing Reviews, 45(10): 631.

Mens T., Van Gorp P., 2006. A taxonomy of model transformation [J]. Electronic Notes in Theoretical Computer Science, 152: 125-142.

Pinckaers F., Gardiner G., Van Vossel E., 2013. Open ERP, a modern approach to integrated business management (Release 7.0.0).

Sendall S., Kozaczynski W., 2003. Model transformation: The heart and soul of model-driven software development [J]. IEEE software, 20(5): 42-45.

Soley R., 2000. Model driven architecture [J]. OMG white paper, 308: 308.

WfMC, 1996. Workflow Management Coalition Glossary and Terminology. Available from: http://www.aiai.ed.ac.uk/project/wfmc/ARCHIVE/DOCS/glossary/glossary.html

White S. A. 2004. Business Process Modeling Notation v1.0. For the Business Process Management Initiative (BPMI).

Xiangfu X, 2010. OpenERP applications and development foundation.

ADDRESSING STRATEGIC CHALLENGES ON MEGA CITIES THROUGH MS2G

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, Simonluca Poggi
Simulation Team
Via Cadorna 2, 17100 Savona, Italy
Email: {matteo.agresta, simonluca.poggi}@simulationteam.com
URL: www.simulationteam.com

Fabio Camponeschi, Marco Camponeschi
Dark Energy srl
Via Europa 46, 19020 Follo, Italy
Email: {fabio.camponeschi, marco.camponeschi}@darkenergysoftware.it
URL: www.darkenergysoftware.it

ABSTRACT

This paper proposes the combination of modeling and simulation and serious games in order to create an effective framework to address crowdsourcing. Indeed interoperable simulation, intelligent agents and serious games have great potential and a new architecture is proposed in this paper. Such architecture that is based on modern technologies and standards allow today to develop an effective cloud solution that could deliver simulation as a service (SaaS). This approach is defined as MS2G (Modeling, Simulation and Serious Games) and could be used to address also other areas such as capability development, training, etc.
In this paper MS2G is proposed to adopt it in order to face the strategic challenges related to future scenarios involving megacities; hence, the large metropolitan areas are expected to require simulation as support tool to address complex aspects such as energy, logistics, food, safety, security. Crowdsourcing in this framework provide a great chance for both policy makers and citizens to work together evaluating new solutions.

Key Words: Crowdsourcing, Megacities, Serious game, Modeling & Simulation, Collective Intelligence, E-government

1. INTRODUCTION

During last year the evolution of interoperability standards for modeling and simulation enabled development of new generation of models; for instance the case of intelligent agents devoted to reproduce complex behaviors have been effectively applied to several domains such as health care, country reconstruction, disaster relief, defense (Bruzzone, Tremori, Massei 2011, Diaz et al. 2013).; the idea is to get benefits from the engagement procedures of serious games and from simulation to guarantee interoperability.
The diffusion of new technological solutions supporting IoT (Internet of Things) are enabling cloud applications; these phenomena are contributing to enhance simulation potential being able to be distributed over a wide community of users (Weber 2010); to succeed in this task it is necessary to guarantee that these innovative generation models should adopt these new technological standards and solutions (Bruzzone et al.2013b).
Therefore it is also fundamental to guarantee that these models should be effective in being distributed to many users; so usability, engagement, friendliness, immersive and intuitive interface become crucial aspects to be included (Connolly et al. 2012).
From this point of view the methodologies used within Serious Games (SG) for guarantee user engagement as well as the relative immersive technologies have a great potential supporting these aspects.
Another crucial element is represented by the necessity to automate many objects in the models by attributing intelligent behaviors; indeed most of the complex systems that could benefit from the innovative simulators are characterized by being composed by many different entities dynamically interacting in a pretty sophisticated way (Bossomaier et al. 2009); in this sense the adoption of intelligent agents (IA) is a very smart idea and the availability of interoperable agents ready to be federated with simulators represents a strategic advantage (Bruzzone, Tremori, Massei 2011).
Due to these reasons, the authors propose to integrate Modeling and Simulation (M&S), Intelligent Agents and Serious Games (SG) within a combined approach defined as MS2G (Modeling, Simulation and Serious Games).

The idea is to get benefits from the engagement procedures of serious games and from simulation to guarantee interoperability.
The current paper address this problem in relation to create a crowdsourcing framework for Mega Cities of the future; by MS2G all stakeholders will be able to “wear the heat of the city major” and to play the game trying to solve complex real problems such as power supply systems, transportations & households.
The MS2G will allow considering sustainability (i.e. social, economic and environmental) under existing constraints (i.e. available space, technology capabilities, renewable energies etc.).

2. MS2G AS COMBINATION OF M&S AND SERIOUS GAMES

The authors proposes MS2G (Modeling, Simulation and Serious Games) as combined approach to address crowdsourcing and to further diffuse M&S.
It is evident that simulation offers many opportunities to be applied with demonstrable benefits as technology able to investigate new ideas, to test assumptions, to acquire knowledge, to predict future scenarios and to prepare solutions (Bruzzone, Kerckhoffs 1996). Indeed simulation in the last decades has been used in a wide range of field, such as: engineering, physics, medicine, logistic and military operation, astronomy, etc. with many different scopes (Massei 2006; Merkuriev et al. 2009).
Another interesting sector, characterized by different characteristics, it is represented by serious games, where the focus is usually shifting from entertainment to education and training keeping benefits of user engagement (Crookall 2010). For instance, serious games have educational values that are based on learning concepts advocated by constructivist psycho-cognitive theories; they guarantee intrinsic motivation, generate cognitive conflicts and provide situated learning (Mouaheb et al. 2012).
Mike Zyda provided an update and a logical approach to the term in his 2005 article and his definition begins with "game" and proceeds from there:

- Game: "a physical or mental contest, played according to specific rules, with the goal of amusing or rewarding the participant."
- Video Game: "a mental contest, played with a computer according to certain rules for amusement, recreation, or winning a stake."
- Serious Game: "a mental contest, played with a computer in accordance with specific rules that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives."

A Serious Game is not simply a simulation; it has the look and feel of a game, a chance to win, but correspond to non-game events or processes from the real world, including for example military operations (even though many popular entertainment games depicted business and military operations). Serious Games with special emphasis on Homeland Security have been developed in many application, for instance as GNU-licensed Open Source engine (Darken et al. 2007) as well as web based multiplayer game (Bruzzone et al. 2014).
Hence, games are focusing on creating an engaging, self-reinforcing context in which to motivate, educate and train the players. This could be applied to very different sectors, for instance an application of monitoring obese people through a cloud-based Serious Game was developed in order to motivate them for physical exercising in a playful manner; the monitoring process focuses on obtaining various health and exercise-related parameters of obese during game-play in order to contribute to their weight loss (Alamri et al. 2014); in similar way serious games was used for measuring sense of presence affecting training effectiveness by online virtual simulation (De Leo et al. 2014).
Hence, there are games designed for supporting crowdsourcing using more realistic knowledge and tools (Eickhoff et al. 2012); so there is a strong potential in combining realistic simulation with computer games.
The innovative concept proposed by the authors is to develop a new generation of solutions devoted to support crowdsourcing that combines interoperable simulation and serious games; the main focus of serious game component is on the user engagement, while the interoperable simulation guarantees to be able to create scalable, reliable and modular scenarios where experts could develop experience and construct trustiness on the models.
In addition, the idea to deploy the simulator over a cloud and to be accessible over a server function allows getting benefits of SaaS (Simulation as a Service) paradigm. Indeed, the proposed architecture is scalable and could be easily distributed in a network, while users could access these services from the web without specific re-configuration needs; in this way the community of users and the knowledge repository generated by these new MS2G solutions is further empowered.
In the proposed example all the info exchanged with the user are presented within a plug in for standard web browsers.
From this point of view it is crucial to introduce also Intelligent Agents (IA) able to support agent driven simulation; this allows reproducing complex mission environments where multi entities should adopt “intelligent behaviors” reacting dynamically to the scenario evolution as well as to the different actor actions.
From these points of view the authors are integrating the IA-CGF (Intelligent Agent Computer Generated Forces), an innovative set of HLA interoperable IAs developed by Simulation Team and tested in multiple applications (Bruzzone, Tremori, Massei 2011).

Due to these reasons it emerges evident the opportunity to combine M&S and SG into MS2G as framework to create new engaging web games able to operate over a wide spectrum of platforms including mobile devices; MS2G approach guarantees interoperability with multiple models and simulators reinforcing aspects such as modularity, composability, reusability and capability to interact with real equipment and systems.
The authors propose to adopt the interoperable simulation standards to keep this solution open to be federated also with other models by using HLA, High Level Architecture (Kuhl et al. 1999, Longo et al., 2013-a); so this new approach, defined as MS2G, strongly relies on these capabilities. In addition MS2G could be adapted to address also other additional service respect crowdsourcing (e.g. training, dissemination, education) reinforcing the potential of this approach; indeed the users could finalize and test hypotheses, parameter changes, solutions and decisions within the simulator environment; the simulator results able to evaluate the impact of these changes and of the different assumptions on the final results.
The authors develop an architecture based on the integration of distributed interoperable M&S with a web serious games; the solution enable to create a web service by deploying simulation over a cloud and guaranteeing access from distributed users.
The paper focus on the development of an application addressing crowdsourcing in relation to future megacity challenges. Indeed MS2G allows to create virtual worlds where stakeholders could test different assumptions and consequence interacting dynamically with the simulator; by this approach crowd sourcing and data mining, become available to a large user community. The users will be able to take decisions and to evaluate different solutions related to food, logistics, energy, environment respect future urban scenarios. Indeed in these areas, simulation based approaches have already proved to be able to provide effective solutions and support the decision making process (specific examples can be found in Longo et al. 2013-b, Longo 2012). The interoperable stochastic simulator (part of the proposed architecture) could evaluate also the risk related to factors affected by uncertainty such as fluctuations in prices, potential natural disasters, population demand forecasts (Bruzzone et al.2008a).

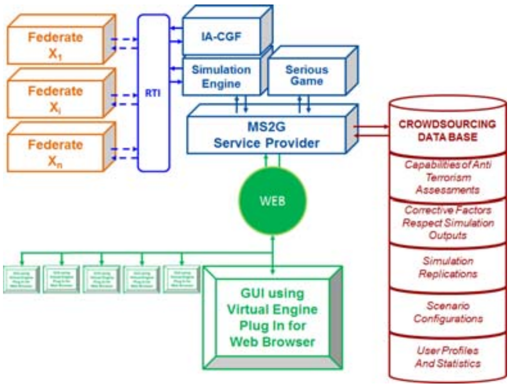


Figure 1: MS2G Architecture

The proposed architecture is represented in figure 1.
The main concept is based on the combination of Discrete Event Simulation, Web Based Simulation, Virtual Simulation and Intelligent Agents with the aim of introducing Decision Makers, Subject Matter Experts and other Stakeholders in an intuitive and interactive experience.

3. CROWDSOURCING AND MS2G

MS2G is a powerful instrument to analyze complex system such as a megacity, in particular for reproducing the evolution and interaction of several entities (Bossomaier et al.2009; Bruzzone et al. 2004, 2009, 2013a). Complex systems require to model different sub-components and entities as well as their interactions; this could be effectively done by modeling the different elements and to let them to interoperate (Kuhl et al. 1999).
The complex systems are pretty challenging to be studied due to the high influence of entities interactions and stochastic elements generating not linear components; these aspects lead to the generation of emergent behaviors that are not intuitive and often not predictable; due to these reasons simulation appears as the main investigation tool; therefore in order to identify solutions to control or direct the evolution of complex systems it is evident the necessity to develop some analysis and experimental approach; from this point of view the use of crowdsourcing results pretty promising.
Indeed “collective” intelligence is a business model that is already used by several companies by means of virtual communities; this approach is considered such as an open enterprise producing ideas, solutions and analysis from the whole community.
It was defined the collective intelligence phenomenon as "the capacity of human communities to evolve towards higher order complexity and harmony, through such innovation mechanisms as differentiation and integration, competition and collaboration” (George Pór, 1995).
Collective intelligence was already available potentially also in the groups, families and communities, but in order to produce effective results it is necessary to access a large, effective, efficient, quick and capillary network; indeed this potential is guaranteed today by new communications based on internet, social networks, IoT (e.g. connected device such as smartphone and computer). Today it is also possible to create and access large data quantities that could be used in order to characterize a particular process or a group of processes.
Indeed, crowdsourcing, or the outsourcing of tasks to the general internet public, has been made possible by technological innovations related to "Web 2.0" (Kleemann et al. 2008),
Therefore Crowd-sourcing is a new work methodology that unifies the experience, the knowledge as well as the different points of view of

many experts towards a collective intelligence for addressing a specific problem (Bruzzone et al. 2012).
The basic principle behind crowdsourcing is that more heads are better than one; it is evident that this could be strongly supported by innovative internet services and resources available on the web 24/7 for a large community through dedicated web platforms (Bruzzone et al. 2012). In this way users access to many information and are entitled to:

- increase transparency of available information
- get access to the available data
- share trusted information
- guarantee the storage of the information

The idea is to maximize the collective intelligence in order to find solutions or reach consensus about political decisions, such a new power plant, or a new transport infrastructure among the citizens (Bruzzone et al. 2013c).
The innovative concept related to use MS2G into crowdsourcing is the possibility to create an interactive virtual world to each stakeholder; in this way he could conduct his own experiments and test his hypotheses and ideas; in addition the MS2G world becomes available for sharing the experimental results in a large community

4. THE MEGACITIES CONTEXT

The global urban population is growing by 65 million annually and the number of people that lives in town and city is already more than 50% of the total world population generating 80% of total GDP (United Nation, 2014; PWC 2010). The cities represent only 2% of earth's surface, but they use 75% of global resources. (Bugliarello 1999).
Today the situation of Cities, Metropolitan and Urban Areas overpassing 10 million inhabitants are distributed especially over emerging countries (see figure 2); it is interesting to note that many new locations in Far East, Latin America and Africa are approaching such threshold.
Urbanization will be one of the major drivers for the next century and represents a major challenge. McKinsey Global Institute, in its forecasting, is reporting that in 2025 2.0 billion of people (25% of the total world population) will live in only 600 urban centers ("city 600") that will generate 60% of global GDP.
Continuing population growth and urbanization are projected to add 2.5 billion people to the world's urban population by 2050, with nearly 90 per cent of the increase concentrated in Asia and Africa (United Nation, 2014); in that period the size of household is declining, the biggest 600 cities are likely to account for 250 million new households.
In this context, the big challenge for the future is to preserve a harmonized growth of mega urban centers, preserving a good balance between urbanization, development, and quality of life for the inhabitants (Fujita et al.1999).

Sustainability should be considered as a major element by a comprehensive approach, where obviously simulation could provide significant benefits (Dupont, 2013; Fiala 2008; Cabezas et al. 2002)
A megacity is usually defined as "a metropolitan area with a total population in excess of ten million people".(How Big Can Cities Get?" New Scientist Magazine, 2006). Therefore a megacity could be represented by a single metropolitan area or two or more metropolitan aggregated areas.

Figure 2: Current Mega Cities around the World

In order to understand and analyze this problem more effectively it is necessary to consider several layers:

- the social layer (cultural diversity and variety, education, art, living conditions, transport, security, health care, innovation, etc.)
- the economical layer (work & mass unemployment, improvement of infrastructure, new technologies, decentralization, repartition of wealth, capital equipment, etc.)
- the ecological layer (energy sources, sustainable development, air and water pollution, noise pollution, traffic jam, water supply, urban sprawl, urban environment protection, public transportation, waste management, etc)

In GlobeScan research (2007) three city archetypes are defined in relation to different parameters:

- Emerging cities: Emerging megacities are characterized by high grow rates driven both by migration and natural growth with a rate of 3-6% per year. Emerging cities have a younger population profile
- Transitional cities: Tansitional megacities have a lower natural population growth but an high migration ; they usually grow by 2-3% per year. Several of these cities are seeing the first signs of an ageing population.
- Mature Cities: Mature megacities have much slower growth rates than both Emerging and Transitional megacities, at around 1% on average. Mature megacities also have older population profiles.

It is important to outline that megacities attract the most talented people and attract more investments, stimulating local and global economic growth with the city network In this context the competition among the megacities is extremely high; every city tries to keep the most talented people and to capture as much investments as possible from worldwide (Shannon et al. 2012).

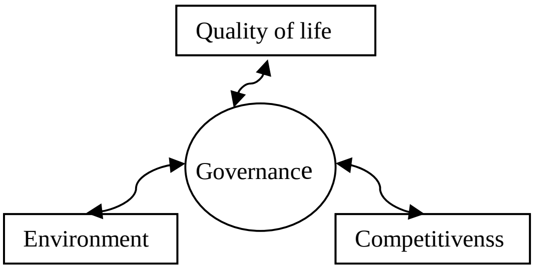


Figure 3: Governance Basic Model

As much as a city is competitive, as much results attractive enabling to enroll a higher number of new citizen. Therefore the Authorities and City Governance need to find a balance among competitiveness, quality of life and Environment as proposed in figure 3 (Fu-Liu et al. 2006).
This results in a quite hard task for the city majors who have to take decision within this really complex problems; indeed a megacity is a really complex system with a huge number of interactions within people, different systems & layers including environmental issues. These systems are continuously changing and growing, expanding in time and space and asking for more space and more energy while it is evident that they are affected by constraints such as limited economic, environmental and space resources.
To evaluate these aspects it could be useful to adopt effective metrics; from this point of view an interesting study covering 112 Chinese cities respect sustainability proposed 18 individual indicators grouped into five key categories: basic needs, resource efficiency, environmental impact, the built environment, and commitment to future sustainability (Jonathan et al 2012). Indeed, in the megacity context there are many challenges that policy maker need to face, such as:

- ***Transport***: Congestion costs are often extremely huge both for economy and environment; improving public and private mobility is absolutely needed (Rodrigue et. al 2006)
- ***Food Logistic***: Food distribution in a megacity is a big challenge, and an efficient urban logistic could reduce traffic and pollution, increasing competitiveness. In addition traceability reduces and mitigates events like contaminations that are extremely since they can affect the health status of a large number of people.
- ***Reverse Logistic & Waste Management***
- It is important to consider both the flows of the supply chain, because in a megacity tons of waste are produced every day and need to be managed. An efficient reverse logistic can resize packaging reducing waste production (Pfohl et al. 2010)
- S***afety and Security:*** these are big challenges; usually in these contexts the organized crime is the biggest security challenge for megacities as well as terrorism. Furthermore protection against natural disaster such as earthquakes, flooding, pandemic events, could be required considering the high density of population.
- ***Power supply***: Energy and power supply is fundamental, and the use of renewable energy can reduce the environmental impact
- ***Environment***: Air, ground and water pollution should be considered; indeed water quality is one of the major issues in particular for Emerging cities
- ***Quality of Life and Healthcare***: Healthcare infrastructure and quality of life are key parameters for making a city more attractive for future developments.

Considering these aspects improving governance is usually the first step towards better and more competitive cities, by the way holistic solutions are desired, but obviously difficult to achieve; indeed, the main obstacle to strategic management and policy makers is often the poor coordination between the different levels of municipal government.
Indeed many megacities have many overlapping in the administrative bodies with limited responsibilities, which limit the efficiency in strategic planning.
Another important aspect is to involve not only the decision makers in the decision process, but also the citizens generating consensus.
This result could be achieved thanks to the internet revolution due to the large number of population that have an internet access: for instance today researches have been carried out on using new internet-connected devices for introducing new forms of active citizenship and e-government (Zhu Yi et al, 2011; Shuchu Xiong &Yihui Luo, 2010; Aljebly, R. & Aboalsamh, H. 2011; Jooho Lee &Soonhee Kim, 2014)
These considerations make evident the potential provided by MS2G into developing new solutions by creating an effective problem solving engine getting benefits from crowdsourcing concept and simulation capabilities.

5. MEGACITY GAME ASPECT

In this paper we propose a serious game developed for simulating decision in a megacity; by this approach the stakeholders are able to play role of city managers and operate over the web by trying to solve problems in the megacity, taking high level decisions and evaluating through the MS2G their impact.

The mega city is proposed through a dynamic evolving virtual 3D representation as presented in the following figure 4; therefore in the town there are Hot-Spots, these elements (see figure 5) represent the opportunity to change the parameters and to select the different alternatives, for instance the power generation configuration balancing different sources among nuclear, fossil, wind and solar facilities.
In the simulated scenario is possible to consider the population growth as well as the need for new resources such as new households, additional electric power and food. Population increase is an opportunity for developing the megacity, but is also a challenge in terms of resource demand, available space, jobs, energy etc.

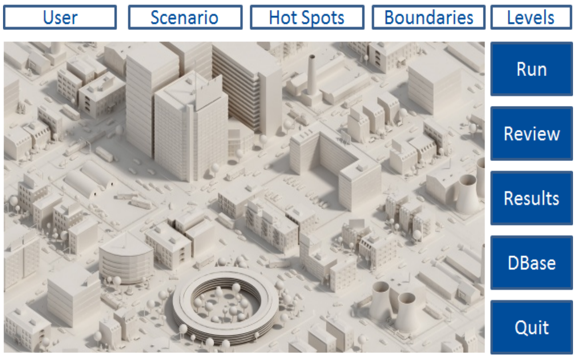


Figure 4: Example of Megacity Crowdsource MS2G

In this framework the decision making is a complex problem since there isn't only one correct solution and due to the multivariable parameters that could be used to compare alternatives; in addition the high number of elements dynamically interacting in the megacity further increase the system complexity.
For example, building a new coal power plant in order to satisfy the growing demand of electricity, increases the level of air pollution, but provides also new job opportunities; on the other side, a huge investment in wind farms results very space consuming, so available space for green areas or new households is strongly affected. Other solutions like offshore wind farms, that have a better energy density and less interference with the local systems, could result in unprofitable investments or can lead to an increase of the cost of energy for the final consumers.
MS2G makes use of the idea of gaming, simulation and crowdsourcing as powerful instrument to share information among the stakeholders by testing alternative solutions, saving and sharing previous results among the stakeholder community; in addition the experimental results provided by the simulator could be presented in the virtual world to the players in a very intuitive way.
Indeed players could try to solve problem by themselves, while they understand the effects of their decisions and assumptions; obviously each choice and solution is evaluated in terms of metrics (i.e. money, space, environmental impacts, etc.)
MS2G in this context should take in account a high number of variables obtained through simulation over the whole city by modeling different subsystems such as human subsystem, environmental layer.
The human system takes in account the variables related to the human presence such as the population, the size of the city area, the industrial area, the number of jobs, the annual income, the quality of life etc. The environmental system, takes in account the green area in the city, the emissions and the quality of air, water and land. The results of the simulation consist of both Global effects and local effects among different timescale, considering both short terms and long terms effects and consequences.
For this reason the two dimensions that are considered are to map the results are related to:

- Time (short and long term)
- Space (local and global effects)

In order to proper model megacity, the human behavior need to be simulated considering aspects such as political, cultural, economic, ethnic, religious elements as well as unemployment, taxes, etc.
From this point of view the MS2G approach in this case benefits of reusing models developed in IA-CGF to address these elements and to simulate urban tensions, interest groups and disorders (Bruzzone et al. 2008b).
In general MS2G could include different models addressing many diverse phenomena:

- Finance: profits, losses, cost of energy, etc.
- Political aspects: political acceptation or disfavor
- Pressure into ecosystem
- Water, materials, energy flows
- Ecosystem changes: evolutions of some parameters as disaster frequency, sea level, etc.

This confirms the advantage of adopting interoperable simulation able to federate different models into a common federation for reproducing mutual interactions among these phenomena.
Single results are available on the virtual world within the Hot-Stop in order to compare the different values (e.g. cost, GW, GWh, Area, CO2 emitted, water consumption, energy consumption, etc.).
Also Global Results are proposed in terms of:

- Water footprint as lake equivalent
- Carbon foot print as a black/grey cloud equivalent on the city
- Energy final cost
- Food cost
- Water cost
- Green residual area
- Building area
- Results in term of cost are stochastic according to scenarios.
- Other outputs can be the population status in term of Welfare
- General satisfaction (peak, blackout,…)
- Psychological status
- Urban Disorder Probability

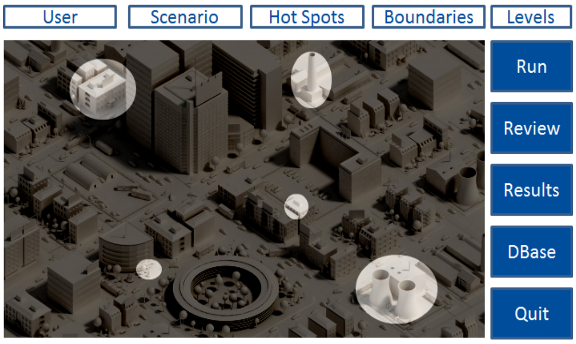


Figure 5: Hot Spots in the MS2G

Examples of Human Input Variables	**Units of measure**
Population	#
Pro capita house water consumption	liters/day
Pro capita food consumption	kcal/day
Total new productive area	m^2
Total new building area	m^2
Development rate (different for mature, transitional and emerging city)	%

Examples of Environmental Input Variables	**Units of measure**
Altitude	m
Area of the city	km^2
Wind data (important both for wind farm and for pollution),	m/s
Total surface of the green areas	km^2
Industrial areas and households areas	km^2
Wind data	m/s
Water availability	kg/inhab
Rain data	cl/m^2day

6. AN EXAMPLE OF APPLYING MS2G: PLANNING ELECTRIC POWER SUPPLY

One of the problems that the user may have to face is planning the electric power supply in the megacity; due to these reason, it is proposed a basic example related to this aspect. The megacity is already configured including general data defining its own characteristics and including among the other electric average consumptions (e.g. daily average consumption per capita) related to its population and services; all these parameters could be modified by the players. The user is entitled to choose among different options in order to satisfy electric power demand; the energy available sources in the library are:

- Thermal Power Stations
  - With CCS (carbon capture and storage)
  - Without CCS
- Wind Farms
  - On shore
  - Off shore
- Solar Panel Farms
- Nuclear Power Plants

For example for what concerns wind turbines, several different choices are available, with different power productions, sizes, costs and impact as proposed in the following figure.

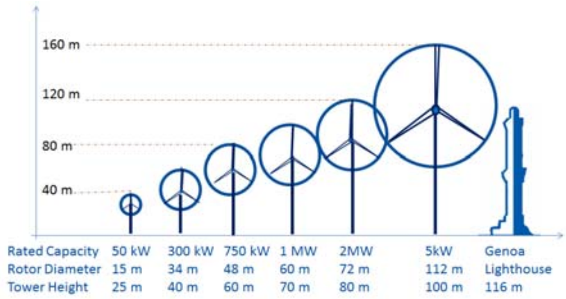


Figure 6: Models used for Wind Turbine in MS2G


Figure 7: GUI on Wind Turbine Parameters

Each alternative is characterized by different parameters including:

- Space occupancy
- Emissions
- Energy costs
- Installation costs
- Managing costs
- Power curves

The scenario has one Hot-Spot for each kind of power plant located in different parts of the megacity virtual world; for instance there are Hot-Spots corresponding to:

- Breeding area
- Agriculture area
- Desalination area
- Port Area

By clicking on the Hot-Spots user is entitled to set access to the setup of the related parameters while in City Hall Hot Spot it is possible to define:

- % of energy production by Wind
- % of energy production by Fossil
- % of energy production by Solar
- % of energy production by Nuclear
- % of energy Imported from outside

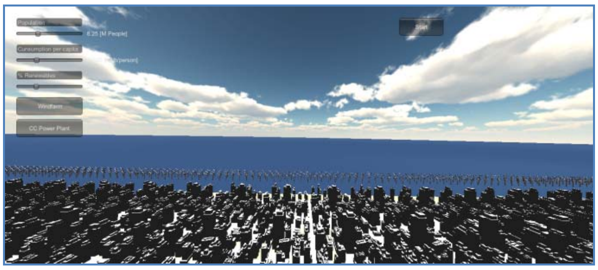


Figure 8: MS2G presenting Wind Farms immersed in the Megacity

In reference to the proposed scenario it was carried out a simulation scenario addressing wind farms as proposed in figure 7; indeed the model presents the results immersed in the virtual world as presented in the following figure 8.

Examples of Megacity Parameters		
Population	20'000'000	inhab.
City Area	3000	km^2
Wind Behavior	Statistical Distributed	m/s
Density	6'666	inhab. /km^2
Per capita Consumption	17'445	kWh/year
Population Growth	3	%
Taxation Rate	50	%

Examples of Wind Farm Parameters		
Tower height	50	m
Base size	100	m^2
Power curve slope	100	(kW/turbine)/(m/s)
Maximum power	1000	kW/turbine
Min Wind threshold	5	m/s
Max Wind threshold	27	m/s
Installation cost	180'000	€/turbine
Managing cost	3'000	€/turbine per year
Rotor diameter	42	M
Nominal power	600	kW/turbine
Power Curve	0,05	%
Life Cycle	20	year
Investment horizon\	15	year
BEP	0,05	%
Percentage of Power supply by Wind Farm	60	%
Slope limit	15	m/s

Examples of Simulation Output		
Wind efficiency	85,47	%
Number of Wind Turbines	55'913	
Energy cost pro capita	2'563	€/year
Active hours	7'488	h/year
Unit Power	156'000	kW
Energy generated	3'744'000	kWh/year
UAC	863'501'680	€/year
Annual revenue required	1'025'534'539	€/year
Energy required	2,0934E+11	kWh/year
Energy cost	0,1469	€/kWh

The development of offshore wind farms guarnatees better energy density and lesser interference with other local systems; this leads to an interesting solution for realization of the wind potential.
The simulator results include the number of wind tower, UAC, annual revenue and the energy cost.

The MS2G has been implemented in Java, C# and by using Unity as virtual engine and experimented over a virtual expansion of a large Mediterranean town currently counting over 4 million inhabitants and expected to growth in the future years; ANOVA was used in order to validate experimental results (Montgomery 2010).

7. CONCLUSIONS

The authors present an innovative approach to support crowdsourcing in relation to strategic problem solving; the combined use of M&S, SG and IA is proposed in terms of architecture.
The system has been developed and implemented as a web service that could be easily accessed through web browser and even deployed on mobile platforms; currently the authors are proceeding in extending the models integrated in the scenario and in carrying out experimental analysis to finalize the validation and verification process along the dynamic development phase. The MS2G approach is currently under development also in relation to other application fields such as logistics and emergency management.

REFERENCES

Alamri, A., Hassan, M.M., Hossain, M.A., Al-Qurishi M.; Aldukhayyil Y.; Shamim Hossain M. (2014) "Evaluating the impact of a cloud-based serious game on obese people", Computers in Human Behavior, January 2014, Vol.30, pp.468-475

Aljebly, R.; Aboalsamh, H., (2011) "E-government — A survey on achieving advanced electronic government services," Current Trends in Information Technology (CTIT), 2011 International Conference and Workshop on , vol., no., pp.143,149, 26-27

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., Massei, M., Tremori, A., Poggi, S., Nicoletti, L., Baisini, C. (2014) "Simulation as enabling technologies for agile thinking: Training and education aids for decision makers" , International Journal of Simulation and Process Modelling, Vol. 9, Issue 1-2, 2014, Pages 113-127

Bruzzone A.G., Massei M., Agresta M., Ferrando A. (2013a) Modeling Fresh goods supply chain contamination I3M, Athens, Greece

Bruzzone, A.G., Fontaine J.G., Gough E. (2013b) "White Paper on Modelling and Simulation Within Maritime Research and Experimentation", CMRE Techical Report, SP

Bruzzone, A.G., Massei, M., Poggi, S., Dallorto, M., Franzinetti, G., Barbarino, A. (2013c) "Quantitative simulation of comprehensive sustainability models as game based experience for education in decision making", Proc.of SESDE, Athens, September 25-27, pp.119-126

Bruzzone A.G., Marques H., Cantice G., Turi M. (2012) "Innovative C2 and Simulation for Crowdsourcing as Force Multiplier", Proceedings of European Modeling and Simulation Symposium, Wien, Austria, September

Bruzzone, A.G., Tremori, A., Massei, M., (2011) “Adding Smart to the Mix,” Modeling, Simulation & Training: the International Defence Training Journal, 3, 25-27

Bruzzone A.G., M. Massei, E. Bocca (2009) Fresh Food Supply Chain Simulation Based Case Studies in Logistic, pp. 127-146

Bruzzone, A.G., Poggi, S., Bocca, E., Longo, F., Madeo, F., Rondinelli, S. (2008a) "Workload forecast algorithm optimization for re-organizing retail network", Proceedings of MAS 2008, Campora S.G. Italy; September 17-19 September, pp.77-82

Bruzzone, A., Scavotti, A., Massei, M., Tremori, A.(2008b) "Metamodelling for analyzing scenarios of urban crisis and area stabilization by applying intelligent agents", Proceedings of 20th European Modeling and Simulation Symposium, Amantea, Italy, September 17-19, pp.521-526

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.

Bruzzone A.G., Kerckhoffs (1996) "Emergency Management" in the Proceedings of European Simulation Symposium, Simulation in Industry Genoa, Italy, October, Vol. I & II, ISBN 1-56555-099-4

Bugliarello, G., (1999) “Megacities and the developing world,” The Bridge, Volume 29,

Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M.(2012) "A systematic literature review of empirical evidence on computer games and serious games", Computers and Education, Volume 59, Issue 2, September, Pages 661-686

Crookall, D. (2010) "Serious games, debriefing, and simulation/gaming as a discipline", Simulation and Gaming, Volume 41, Issue 6, December, Pages 898-920

Darken, CJ, Anderegg, BG, McDowell, PL, (2007) "Game AI in Delta3D", Proceedings of IEEE Symposium on Computational Intelligence and Games, Honolulu, USA, April

De Leo, G., Diggs, L.A., Radici, E., Mastaglio, T.W. (2014) "Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment", Simulation in Healthcare, Vol.9, Issue 1, February, Pages 1-6

Diaz R., Behr J., Toba A.L., Giles, B., Manwoo N., Longo F., Nicoletti L. (2013). Humanitarian /emergency logistics models: A state of the art overview. Simulation Series, 45 (11), pp. 261-268.

DuPont (2013) “Global Reporting Initiative Report”, Technical Report, Wilmington, USA

Eickhoff, C., Harris, C.G., De Vries, A.P., Srinivasan, P. (2012) "Quality through flow and immersion: Gamifying crowdsourced relevance assessments", Proceedings of SIGIR, Portland, OR, August 12-16, pp. 871-880

Kleemann F., Günter G., Rieder K. (2008) Un(der)paid Innovators: The Commercial Utiliza-tion of Consumer Work through Crowdsourcing Science, Technology & Innovation Studies Vol. 4, No. 1, July 2008 ISSN: 1861-3675

Fujita, Masahisa, Paul R. Krugman, and Anthony J. Venables, (1999) “The Spatial Economy: Cities, Regions and International Trade”, MIT Press, Cambridge MA, USA

Fu-Liu Xu, Shan-Shan Zhao, Richard W. Dawson, Jun-Yi Hao, Ying Zhang, Shu Tao (2006) A triangle model for evaluating the sustainability status and trends of economic development”, Ecological Modelling, Volume 195, Issues 3–4.

Cabezas H. and B. Fath, (2002). Towards a theory of sustainable systems,” Fluid Phase Equilibria, vol. 194–197, pp. 3–14.

Jonathan W., G. Xiao, and Lan Xue, (2010) The Urban Sustainability Index: A new tool formeasuring China’s cities, Urban China Initiative.

Jooho Lee; Soonhee Kim, (2014) "Active Citizen E-Participation in Local Governance: Do Individual Social Capital and E-Participation Management Matter?," System Sciences (HICSS), 2014 47th Hawaii International Conference on , vol., no., pp.2044,2053.

Kuhl, F., Weatherly, R., Dahmann, J., (1999) “Creating Computer Simulation Systems: An Introduction to the High Level Architecture”, Prentice Hall, Upper Saddle River, USA

Longo F., 2012. Sustainable supply chain design: An application example in local business retail. Simulation, 88(12), pp. 1484-1498.

Longo F., Nicoletti L., Chiurco A., 2013-a. Cooperative training for ships and tugboats pilots based on interoperable simulation. Proceedings of the 25th European Modeling and Simulation Symposium, EMSS 2013; Athens; Greece; 25 September 2013 through 27 September 2013, pp. 695-703

Longo, F., Huerta, A., Nicoletti, L., 2013-b. Performance analysis of a Southern Mediterranean seaport via discrete-event simulation. Journal of Mechanical Engineering, 59(9), pp. 517-525.

Massei M. (2006) Logistics and process solution for supply chain of fresh food in retail, Proc. of MAS2006, Bergeggi October

McKinsey Global Institute (2011) “Urban world: Mapping the economic power of cities”, McKinsey Global Institute Technical Report

GlobeScan and MRC McLean Hazel (2007) “Megacity Challenges, A stakeholder perspective A research project conducted by Sponsored by Siemens”, Siemens Technical Report

Merkuryev Y., Galina Merkuryeva, Miquel Angel Piera, Antoni Guasch (2009) Simulation-based Case Studies in Logistics: Education and Applied Research. Springer-Verlag, London. 232

Montgomery D.C. (2000) "Design and Analysis of Experiments", John Wiley & Sons, New York

Fiala, N. (2008) “Measuring sustainability: Why the ecological footprint is bad economics and bad environmental science,” Ecolog. Econom., vol. 67, pp. 519–525

Mouaheb, Houda ;Fahli, Ahmed, Moussetad, Mohammed ;Eljamali Said (2012) “The Serious Game: What Educational Benefits?”, Procedia - Social and Behavioral Sciences, 2012, Vol.46, pp.5502-5508

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

Por, George (1995). "The Quest for Collective intelligence". In K. Gozdz. Community Building: Renewing Spirit and Learning in Business. San Francisco: New Leaders Press.

PWC (2010) “Cities of Opportunity”, Pricewaterhousecoopers Technical Report

Rodrigue J.P., C. Comitos, B. Slack (2006) The Geography of Transport Systems, Routledge

Shannon B., M. Lindsay & J. Woetzel (2012) “New models for sustainable growth in emerging-market cities. A new tool, the urban sustainability index, highlights five themes of sustainable development for cities in emerging economies” McKinsey Global Institute.

Shuchu Xiong; Yihui Luo, (2010)"Research of E-Government Collaboration Service Model," Internet Technology and Applications, 2010 International Conference on , vol., no., pp.1,5, 20-22

United Nation, Department of Economic and Social Affairs, (2014) World Urbanization Prospects.

Weber, R.H. (2010) "Internet of Things - New security and privacy challenges", Computer Law and Security Review, Vol 26, Issue 1, January, Pages 23-30

Zhu Yi; Sun Li-Jian; Zhang Yu, (2011)"The Data Exchange Research of E-Government Application System Basing on Spatial Information," Management and Service Science (MASS), 2011 International Conference on , vol., no., pp.1,5, 12-14

Zyda, M. (2005), "From visual simulation to virtual reality to games". IEEE Computer, September

Author's Index