

THE 28TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 26 - 28 2016
CYPRUS



EDITED BY
AGOSTINO G. BRUZZONE
EMILIO JIMÉNEZ
FRANCESCO LONGO
LOUCAS S. LOUCA
LIN ZHANG

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2016

ISBN 978-88-97999-68-3 (Paperback)
ISBN 978-88-97999-76-8 (PDF)

© 2016 DIME UNIVERSITÀ DI GENOVA

RESPONSIBILITY FOR THE ACCURACY OF ALL STATEMENTS IN EACH PAPER RESTS SOLELY WITH THE AUTHOR(S). STATEMENTS ARE NOT NECESSARILY REPRESENTATIVE OF NOR ENDORSED BY THE DIME, UNIVERSITY OF GENOA. PERMISSION IS GRANTED TO PHOTOCOPY PORTIONS OF THE PUBLICATION FOR PERSONAL USE AND FOR THE USE OF STUDENTS PROVIDING CREDIT IS GIVEN TO THE CONFERENCES AND PUBLICATION. PERMISSION DOES NOT EXTEND TO OTHER TYPES OF REPRODUCTION NOR TO COPYING FOR INCORPORATION INTO COMMERCIAL ADVERTISING NOR FOR ANY OTHER PROFIT - MAKING PURPOSE. OTHER PUBLICATIONS ARE ENCOURAGED TO INCLUDE 300 TO 500 WORD ABSTRACTS OR EXCERPTS FROM ANY PAPER CONTAINED IN THIS BOOK, PROVIDED CREDITS ARE GIVEN TO THE AUTHOR(S) AND THE CONFERENCE.

FOR PERMISSION TO PUBLISH A COMPLETE PAPER WRITE TO: DIME UNIVERSITY OF GENOA, PROF. AGOSTINO BRUZZONE, VIA OPERA PIA 15, 16145 GENOVA, ITALY. ADDITIONAL COPIES OF THE PROCEEDINGS OF THE EMSS ARE AVAILABLE FROM DIME UNIVERSITY OF GENOA, PROF. AGOSTINO BRUZZONE, VIA OPERA PIA 15, 16145 GENOVA, ITALY.

ISBN 978-88-97999-68-3 (Paperback)
ISBN 978-88-97999-76-8 (PDF)

THE 28TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 26-28 2016, CYPRUS

ORGANIZED BY



DIME - UNIVERSITY OF GENOA



LIOPHANT SIMULATION



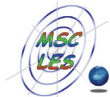
SIMULATION TEAM



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. INEGNERIA Y AGRIMENSURA



UNIVERSITY OF LA LAGUNA



CIFASIS: CONICET-UNR-UPCAM



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



NATIONAL RUSSIAN SIMULATION SOCIETY



CEA - IFAC



UNIVERSITY OF BORDEAUX



UNIVERSITY OF CYPRUS

I3M 2016 INDUSTRIAL SPONSORS



CAL-TEK SRL



LIOTECH LTD



MAST SRL



SIM-4-FUTURE

I3M 2016 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 Group



HALLDALE MEDIA GROUP: MILITARY SIMULATION AND TRAINING MAGAZINE



HALLDALE MEDIA GROUP: THE JOURNAL FOR HEALTHCARE EDUCATION,
SIMULATION AND TRAINING



SAGE
SIMULATION TRANSACTION OF SCS



DE GRUYTER
INTERNATIONAL JOURNAL OF FOOD ENGINEERING

EDITORS

AGOSTINO BRUZZONE

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

EMILIO JIMÉNEZ

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

FRANCESCO LONGO

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

LOUCAS S. LOUCA

UNIVERSITY OF CYPRUS, CYPRUS
lslouca@ucy.ac.cy

LIN ZHANG

BEIHANG UNIVERSITY, CHINA
zhanglin@buaa.edu.cn

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2016

GENERAL CO-CHAIRS

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

PROGRAM CHAIR

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

THE 28TH EUROPEAN MODELING & SIMULATION SYMPOSIUM, EMSS 2016

GENERAL CO-CHAIRS

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

PROGRAM CO-CHAIRS

LOUCAS S.LOUCA, *UNIVERSITY OF CYPRUS, CYPRUS*
LIN ZHANG, *BEIHANG UNIVERSITY, CHINA*

EMSS 2016 INTERNATIONAL PROGRAM COMMITTEE

MICHAEL AFFENZELLER, *UPPER AUSTRIA UAS, AUSTRIA*
JONI A. AMORIM, *UNIVERSITY OF SKÖVDE, SWEDEN*
MAJA ATANASIJEVIC-KUNC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
FELIX BREITENECKER, *TECHNICAL UNIVERSITY OF WIEN, AUSTRIA*
AGOSTINO BRUZZONE, *UNIVERSITY OF GENOA, ITALY*
ALESSANDRO CHIURCO, *UNIVERSITY OF CALABRIA, ITALY*
MARIA PIA FANTI, *POLYTECHNIC UNIVERSITY OF BARI, ITALY*
IDALIA FLORES, *UNIVERSITY OF MEXICO, MEXICO*
SERGIO GALLO, *UNIV. OF MODENA AND REGGIO EMILIA, ITALY*
RICARDO GONÇALVES, *UNIVERSIDADE NOVA DE LISBOA, PORTUGAL*
WITOLD JACAK, *UPPER AUSTRIA UAS, AUSTRIA*
EMILIO JIMÉNEZ, *UNIVERSITY OF LA RIOJA, SPAIN*
ANDREAS KÖRNER, *VIENNA UNIVERSITY OF TECHNOLOGY, AUSTRIA*
JUAN IGNACIO LATORRE BIEL, *UNIV. PÚBLICA DE NAVARRA, SPAIN*
FRANCESCO LONGO, *MSC-LES, UNIVERSITY OF CALABRIA, ITALY*
EDGAR ALONSO LOPEZ-ROJAS, *BLEKINGE INSTITUTE OF TECHNOLOGY, SWEDEN*
LOUCA LOUCA, *UNIVERSITY OF CYPRUS, CYPRUS*
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*
RICCARDO MELLONI, *UNIV. OF MODENA AND REGGIO EMILIA, ITALY*
GALINA MERKURYEVA, *RIGA TECHNICAL UNIVERSITY, LATVIA*
SAURABH MITTAL, *DUNIP TECHNOLOGIES, USA*
MIGUEL MÚJICA MOTA, *AMSTERDAM UNIVERSITY OF APPLIED SCIENCES, NETHERLANDS*
GASPER MUSIC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
TUNCER ÖREN, *UNIVERSITY OF OTTAWA, CANADA*
ANTONIO PADOVANO, *UNIVERSITY OF CALABRIA, ITALY*
ELEONORA PANTANO, *MIDDLESEX UNIVERSITY, UK*
CHUMMING RONG, *UNIVERSITY OF STAVANGER, NORWAY*
PEER OLAF SIEBERS, *UNIVERSITY OF NOTTINGHAM, UNITED KINGDOM*
CHRYSOSTOMOS STYLIOS, *TECHNOLOGICAL EDUCATIONAL INSTITUTE OF EPIRUS, GREECE*
FEI TAO, *BEIHANG UNIVERSITY, CHINA*
WALTER UKOVICH, *UNIVERSITY OF TRIESTE, ITALY*
ANN WELLENS, *UNIVERSITY OF MEXICO, MEXICO*
THOMAS WIEDEMANN, *UNIVERSITY OF APPLIED SCIENCES AT DRESDEN, GERMANY*
LEVENT YILMAZ, *AUBURN UNIVERSITY, USA*
RICHARD ZALUSKI, *CENTER FOR STRATEGIC CYBERSPACE + SECURITY SCIENCE, UNITED KINGDOM*
BERNARD P. ZEIGLER, *RTSYNC CORP, USA*
LIN ZHANG, *BEIHANG UNIVERSITY, CHINA*

TRACKS AND WORKSHOP CHAIRS

DISCRETE AND COMBINED SIMULATION

CHAIRS: GASPER MUSIC, UNIVERSITY OF LJUBLJANA, SLOVENIA;
THOMAS WIEDEMANN, HTW DRESDEN FB INFORMATIK, GERMANY

INDUSTRIAL PROCESSES MODELING & SIMULATION

CHAIR: AGOSTINO BRUZZONE, DIME, UNIVERSITY OF GENOA, ITALY

INDUSTRIAL ENGINEERING

CHAIR: FRANCESCO LONGO, MSC-LES, UNIVERSITY OF CALABRIA, ITALY

AGENT DIRECTED SIMULATION

CHAIRS: TUNCER ÖREN, UNIVERSITY OF OTTAWA, CANADA; LEVENT YILMAZ, AUBURN UNIVERSITY, USA

PETRI NETS BASED MODELLING & SIMULATION

CHAIRS: EMILIO JIMÉNEZ, UNIVERSITY OF LA RIOJA, SPAIN; JUAN IGNACIO LATORRE, PUBLIC UNIVERSITY OF NAVARRA, SPAIN

WORKSHOP ON CLOUD MANUFACTURING

CHAIRS: PROF. LIN ZHANG, BEIHANG UNIVERSITY, CHINA; PROF. FEI TAO, BEIHANG UNIVERSITY, CHINA;

SIMULATION OPTIMIZATION APPROACHES IN INDUSTRY, AVIATION, SERVICES AND TRANSPORT

CHAIRS: IDALIA FLORES, UNIVERSITY OF MEXICO; MIGUEL MÚJICA MOTA, AMSTERDAM UNIVERSITY OF APPLIED SCIENCES

HUMAN-CENTRED AND HUMAN-FOCUSED MODELLING AND SIMULATION

CHAIRS: AGOSTINO BRUZZONE, DIME, UNIVERSITY OF GENOA, ITALY; PEER OLAF SIEBERS, UNIVERSITY OF NOTTINGHAM, UK

WORKSHOP ON SOFT COMPUTING AND MODELLING & SIMULATION

CHAIRS: MICHAEL AFFENZELLER, UPPER AUSTRIAN UNIVERSITY OF APPLIED SCIENCES, AUSTRIA; WITOLD JACAK, UPPER AUSTRIAN UNIVERSITY OF APPLIED SCIENCES, AUSTRIA

WORKSHOP ON CLOUD COMPUTING

CHAIR: CHUMMING RONG, UNIVERSITY OF STAVANGER, NORWAY

ADVANCED SIMULATION FOR LOGISTICS SYSTEMS

CHAIRS: MARIA PIA FANTI, POLYTECHNIC OF BARI, ITALY; CHRYSOSTOMOS STYLIOS, TECHNOLOGICAL EDUCATIONAL INSTITUTE OF EPIRUS, GREECE; WALTER UKOVICH, UNIVERSITY OF TRIESTE, ITALY

MODELLING AND SIMULATION APPROACHES IN AND FOR EDUCATION

CHAIRS: MAJA ATANASIJEVIC-KUNC, UNIV. LJUBLJANA, SLOVENIA; ANDREAS KÖRNER, VIENNA UNIV. OF TECHNOLOGY, AUSTRIA

MODELLING AND SIMULATION IN PHYSIOLOGY AND MEDICINE (COMMON TRACK EMSS-IWISH)

CHAIRS: MAJA ATANASIJEVIC-KUNC, UNIV. LJUBLJANA, SLOVENIA; FELIX BREITENECKER, VIENNA UNIV. OF TECHNOLOGY, AUSTRIA

SIMULATION AND MODELLING FOR OCCUPATIONAL HEALTH AND SAFETY

CHAIRS: RICCARDO MELLONI, UNIV. OF MODENA AND REGGIO EMILIA, ITALY; SERGIO GALLO, UNIV. OF MODENA AND REGGIO EMILIA, ITALY

RETAIL-ORIENTED SIMULATION

CHAIR: DR. ELEONORA PANTANO, MIDDLESEX UNIVERSITY, UK

COMPUTER SIMULATION FOR SECURITY

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

WORKSHOP ON MODEL ENGINEERING FOR SYSTEM OF SYSTEMS

CHAIRS: PROF. LIN ZHANG, BEIHANG UNIVERSITY, CHINA; PROF. BERNARD P. ZEIGLER, RTSYNC CORP, USA; DR. SAURABH MITTAL, DUNIP TECHNOLOGIES, USA

MODELLING, COMPUTER SIMULATION AND VISUALISATION OF HERITAGE

CHAIRS: GALINA MERKURYEVA, RIGA TECHNICAL UNIVERSITY; RICARDO GONÇALVES, UNIVERSIDADE NOVA DE LISBOA (UNINOVA)

GENERAL CO-CHAIRS' MESSAGE

WELCOME TO EMSS 2016!

After 27 successful editions, the 28th European Modeling and Simulation Symposium faces the challenge to preserve and increase the prestige of the event. After so many years, indeed, EMSS builds on a rich heritage of knowledge and experiences. Hence, it is without exaggeration that we deem the Symposium an acknowledged and firm reference appointment for M&S theorists and practitioners.

Year by year EMSS raises new expectations that, until now, it has always been able to meet or even outperform. Therefore, in this edition, EMSS carries on with the long tradition started in 1988 bringing new scientific endeavors and cutting-edge knowledge.

In terms of coverage, EMSS collects many and various contributions encompassing advanced M&S approaches and paradigms, latest technologies as well as innovative industrial paradigms. In the general framework of the 13th International Multidisciplinary Modelling & Simulation Multiconference, EMSS is co-located with, all the attendees will have the possibility to get in touch with the latest advances of M&S methodologies, techniques and applications in Industrial Engineering, Business, Economy, Finance and Commerce.

It goes without saying that the EMSS value is tied to all those people involved in it: IPC members, reviewers and authors to whom we are deeply grateful. Therefore we take this opportunity to thank warmly IPC members, reviewers and, above all, authors that are the real pillars of the EMSS success. Moreover, year by year, EMSS takes place in a wonderful but different location so that work activities could be surrounded by a superb frame.

This year we are proud to welcome all of you in Larnaca, Cyprus with its exclusive and memorable landmarks.



Francesco Longo
DIMEG, University of
Calabria, Italy



Emilio Jiménez
University of La Rioja,
Spain

ACKNOWLEDGEMENTS

The EMSS 2016 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The EMSS 2016 IPC would like to thank all the authors as well as the reviewers for their invaluable work.

A special thank goes to Prof. Loucas S. Louca from University of Cyprus as Local Organizer and to all the organizations, institutions and societies that have supported and technically sponsored the event.

I3M 2016 INTERNAL STAFF

MATTEO AGRESTA, *SIMULATION TEAM, ITALY*
LUIGI BRUNO, *DIMEG, UNIVERSITY OF CALABRIA*
AGOSTINO G. BRUZZONE, *DIME, UNIVERSITY OF GENOA, ITALY*
ALESSANDRO CHIURCO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
RICCARDO DI MATTEO, *SIMULATION TEAM, ITALY*
CATERINA FUSTO, *CAL-TEK SRL, ITALY*
FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
GIANLUCA MAGLIONE, *SIMULATION TEAM, ITALY*
MARINA MASSEI, *DIME, UNIVERSITY OF GENOA, ITALY*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
ANTONIO PADOVANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
LUCA ROTUNDO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
CATALDO RUSSO, *CAL-TEK SRL, ITALY*
GIUSEPPE SPEZZANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
MARCO VETRANO, *CAL-TEK SRL, ITALY*



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



Index

Using pipeline and binary-tree as CPANS for solving the sorting problem. A comparative study	1
M. Rossainz-López, M. I. Capel, O. Carrasco-Limón, B. Sánchez-Rinza	
Dual use bicycle path network designing and exploitation environment - velorouter	10
E. Ginters, M. Baltruks, I. Sakne, Y. Merkurjev	
Development and simulation-based evaluation of an algorithm for the retrieval-in-sequence for shuttle systems	15
T. Lienert, W. A. Günthner	
Scalable simulation models of railway traffic	25
R. Novotny, A. Kavicka	
HLA listener – A GUI driven mapping of HLA standard services	32
M. Ali, Y. Mohamed	
Forecasting construction duration of watermain upgrade project by using discrete event simulation	39
D. Kim, Y. Zong, Y. Mohamed, S. AbouRizk	
A flexible and graphical operational effectiveness evaluation tool	48
W. Li, S. Lin, Z. Liu, P. Ma, M. Yang	
A tree-search based heuristic for a complex stacking problem with continuous production and retrieval	56
S. Raggl, B. Andreas, F. Tricoire, M. Affenzeller	
Optimization of paint shop department using discrete event simulation	62
P. Raska, Z. Ulrych	
No more conflicts: the development of a generic airport model in a sequence-optimization framework	67
P. Scala, M. Mujica Mota, D. Delahaye	
High-performance web simulation	73
J. Voracek	
Interactive distributed simulation realised as a web application	78
S. Kartak	
Process simulation as support in learning theoretical-practical within the lean manufacturing laboratory of UTFV	87
C. Q. Aviles, W. Guerrero Coronel, O. A. González Manzanares	
Radio frequency safety device applied to chair children in cars	94
R. da Cunha Santos, K. Barros de Oliveira, M. A. Cruz Moreira, S. J. Correia Neto, G. Gomes Cunha	
The method of nested simulations supporting decision-making process within a mesoscopic railway simulator	100
R. Diviš, A. Kavička	

Analysis of uncertainty in engineering design optimization problems	107
P. Fleck, M. Kommenda, M. Affenzeller, T. Prante	
Generation of dispatching rules for job sequencing in single-machine environments	117
J. Karder, A. Scheibenpflug, S. Wagner, M. Affenzeller	
Simulating the assorting planning process to support the decision-making of grocery wholesalers: the mexican market case	122
S. Flores Monroy	
Use of web services for the localization of rolling stock with utilization technology oracle spatial	128
J. Fikejz, J. Merta, J. Oborník	
Emulation as part of an integrated logistics learning environment	134
S. Langer, W. Hofmann, S. Lang, T. Reggelin, D. Weigert	
An advanced approach to simulation of super-wide bandwidth information and communication systems combining microwave and photonic industrial technologies	141
M.E. Belkin, V. Iakovlev, A. Sigov, Y. Tyschuk, V. Golovin	
A simulation model of human decision making in emergency conditions	148
F. De Felice, D. Falcone, A. Petrillo, A. G. Bruzzone, F. Longo	
Performance of FDS versions 5 and 6 in passenger car fire computer simulation	155
P. Weisenpacher, J. Glasa, V. Sipkova	
Two time steps predictive control application to a bioprocess simulator	162
S. Altuntaş, H. Hapoğlu	
Simulation and optimization of the manufacturing process of e- bicycles: the case of innovaunam	168
P. González-Hernández, A. Huerta-Barrientos	
Simulation optimization of pedestrian evacuation in mexican subway: the case of Pino Suarez station	175
A. Vera-Morales, A. Huerta-Barrientos	
Preserving and managing heritage buildings in a sustainable way with integration of BIM – a review	183
L. Seduikyte, A. Kylili, J. Sadauskiene, P. Fokaides	
Discrete event simulation to improve cutting-welding workshop in a shipyard	191
M. Rouco-Couzo, D. Crespo-Pereira, A. Garcia-del-Valle, M. Cebra-Fernandez, R. Rios-Prado	
Improving planning and resource utilization of a shipbuilding process based on simulation	197
M. Cebra-Fernandez, D. Crespo-Pereira, A. Garcia-del-Valle, M. Rouco-Couzo	
A simulation-based tool to support decision-making in logistics design of a cans packaging line	204
V. G. Achkar, V. Bär, F. Cornú, C. A. Méndez	

Simulation of longwall coal mining technologies	212
V. V. Okolnishnikov, S. V. Rudometov, S. S. Zhuravlev, V. V. Sinoviev	
Simulation-optimization of schedules with sequence dependent setup times based on Petri net models	216
G. Mušič	
Minimizing the impact of escalator failures in metro tacubaya subway station on user's mobility	224
H. O. I. Bautista-Martínez, A. Huerta-Barrientos, G. de Jesús Portillo-Villasana	
Automatic simulation models generation of warehouses with milkruns and pickers	231
A. Vieira, L. S. Dias, G. B. Pereira, J. A. Oliveira, M. Sameiro Carvalho, P. Martins	
A simulation model for assigning secure waiting areas on subway platforms to minimize accidents	242
G. de Jesús Portillo-Villasana, A. Huerta-Barrientos	
Paysim: a financial mobile money simulator for fraud detection	249
E. A. Lopez-Rojas, A. Elmir, S. Axelsson	
Analysis and evaluation of performance issues of parallel software on multi-core processors	256
F. Wiesinger, M. Tunca, M. Bogner	
Balancing of a manufacturing workstation, using a simulation and optimization model	263
F.D. Chaves Almanza, A. Wellens	
Service level definition from forecast error cost evaluation	270
A. Sosa, A. Wellens	
Supplier selection: single sourcing vs. dual sourcing with sustainability considerations	277
M. Accorsi, R. Montanari, M. Rinaldi, P. Centobelli, E. Bottani	
A proposal for optimizing urban mobility in Mexico City based on the public transport network	282
I. Flores De La Mota, A. Huerta Barrientos	
Simulation optimization for an age-based vaccines distribution strategy against the spread of influenza epidemic	288
J. Lara de Paz, I. Flores	
Modelling a biomass boiler using an artificial neural network	296
P. Ruiz, G. Sorrosal, C. E. Borges, A. M. Macarulla	
Comparison of PSO and DE in the task of optimal control of chaotic Lozi map	303
R. Senkerik, M. Pluhacek, I. Zelinka, A. Viktorin	
Mesosopic simulation models for the automotive industry	309
A. Hennig, T. Reggelin, D. Wolff	
Modeling and simulation of the effects of thermal expansion mismatch and temperature on thermal induced stresses in modern nano pMOS	314
A. El Boukili	

Simulation system of mining works	323
A. N. Starodubov, V. V. Sinoviev, V. V. Okolnishnikov	
Intelligent simulation for tourism and cultural heritage management	328
G. Merkuryeva, V. Bolshakov, J. Martins, R. Gonçalves, G. Bueno Benito	
Using NodeJS for discrete simulation in a web-based environment	334
T. Wiedemann	
A comparative study of hyperelastic constitutive models for an automotive component material	338
R. Tobajas, D. Elduque, C. Javierre, E. Ibarz, L. Gracia	
An intelligent, interactive and interoperable platform for the 21st century museums	348
F. Longo, L. Nicoletti, A. Padovano, A. Solis, A. Monzon	
Augmented reality and mobile technologies for maintenance, security and operations industrial facilities	355
A. G. Bruzzone, F. Longo, L. Nicoletti, M. Vetrano, L. Bruno, A. Chiurco, C. Fusto, G. Vignali	
Virtual and augmented reality as enablers for improving the service on distributed assets	361
A. G. Bruzzone, M. Massei, G. Maglione, M. Agresta, G. Franzinetti, A. Padovano	
Author's Index	367

USING PIPELINE AND BINARY-TREE AS CPANS FOR SOLVING THE SORTING PROBLEM. A COMPARATIVE STUDY

M. Rossainz-López^(a), Manuel I. Capel^(b), O. Carrasco-Limón^(a), B. Sánchez-Rinza^(a)

^(a) Faculty of Computer Science, Autonomous University of Puebla, San Claudio Avenue and South 14th Street, San Manuel, Puebla, Puebla, 72000, México

^(b) Software Engineering Department, College of Informatics and Telecommunications ETSIT, University of Granada, Daniel Saucedo Aranda s/n, Granada 18071, Spain

^(a) rossainz@cs.buap.mx, ^(b) manuelcapel@ugr.es, ^(a) odondavidcarrasco95@gmail.com, ^(a) brinza@hotmail.com

ABSTRACT

This paper proposes the model of the High Level Parallel Compositions or CPANS (Acronym in Spanish) to communication patterns/interaction Pipeline and Binary Tree for implementing a sorting algorithm by Structured Parallel Programming approach based on the concept of Parallel Objects. The CPANS Pipeline and TreeDV are displayed using the paradigm of object orientation and sorting problem is solved using two different algorithms; it is using a pipeline process to sort a dataset in disordered (CPAN Pipe) and one that by quick sort uses a binary tree for the ordering of the same dataset disordered by divide and conquer technique (CPAN TreeDVQS). Each proposal of CPAN contains a predefined set of restrictions of synchronization between processes (maximum parallelism, mutual exclusion and synchronization of producer-consumer type), and the use of synchronous, asynchronous and asynchronous future communication modes. Sorting algorithms, their design and implementation as CPANS and comparative performance metrics on a parallel machine 64 processors are shown.

Keywords: CPAN Pipeline, CPAN Binary Tree, Structured Parallel Programming, Communication Patterns.

1. INTRODUCTION

At moment the construction of concurrent and parallel systems has less restraints than ever, since the existence of parallel computation systems, more and more affordable, of high performance, or HPC (High Performance Computing) has brought to reality the possibility of obtaining a great efficiency in data processing without a great rise in prices. Even though, open problems that motivate research in this area still exist, efficient affordable parallel computing is a reality today. We are interested, in particular, to do research work that has to do with parallel applications that use predetermined communication patterns, among other component--software. At least, the following ones have currently been identified as important open problems: The lack of acceptance structured parallel programming environments of use to develop applications (Bacci and et-al 1999), The necessity to have patterns or High Level Parallel Compositions, the Determination of a

complete set of patterns as well as of their semantics (Corradi, and Zambonelli, 1995), the adoption of an object-oriented approach (Corradi and Leonardi, 1991, Darlington 1993). The High Level Parallel Compositions or CPANS are parallel patterns defined and logically structured that, once identified in terms of their components and of their communication, can be adopted in the practice and be available as high level abstractions in user applications within an OO-programming environment (Rossainz 2005, Rossainz and Capel 2008). The process interconnection structures of most common parallel execution patterns, such as pipelines, farms and trees can be built using CPANS, within the work environment of Parallel Objects that is the one used to detail the structure of a CPAN implementation and to solve the problem of sorting.

2. HIGH LEVEL PARALLEL COMPOSITIONS (CPAN)

A CPAN comes from the composition of a set three object types: An object manager (Figure 1) that represents the CPAN itself and makes an encapsulated abstraction out of it that hides the internal structure. The object manager controls a set of objects references, which address the object Collector and several Stage objects and represent the CPAN components whose parallel execution is coordinated by the object manager.

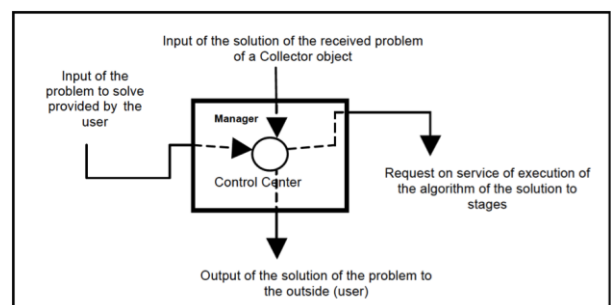


Figure 1: Component MANAGER of model CPAN

The objects Stage (Figure 2) are objects of a specific purpose, in charge of encapsulating an client-server type interface that settles down between the manager and the slave-objects. These objects do not actively participate in the composition of the CPAN, but are considered external entities that contain the sequential algorithm

that constitutes the solution of a given problem. Additionally, they provide the necessary inter-connection to implement the semantics of the communication pattern which definition is sought. In other words, each stage should act a node of the graph representing the pattern that operates in parallel with the other nodes. Depending on the particular pattern that the implemented CPAN follows, any stage of it can be directly connected to the manager and/or to the other component stages.

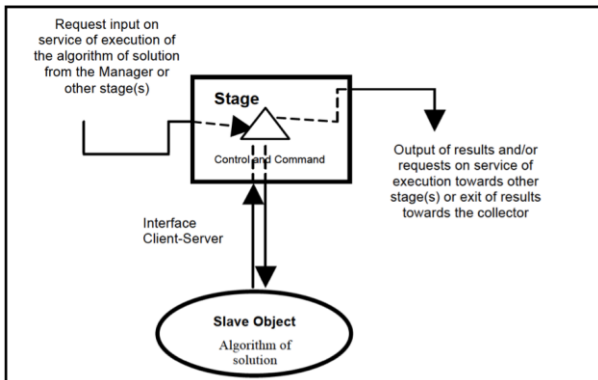


Figure 2: Component Stage of model CPAN and its associated slave object

The Collector object (Figure 3) we can see an object in charge of storing the results received from the stage objects to which is connected, in parallel with other objects of CPAN composition. That is to say, during a service request the control flow within the stages of a CPAN depends on the implemented communication pattern. When the composition finishes its execution, the result does not return to the manager directly, but rather to an instance of the Collector class that is in charge of storing these results and sending them to the manager, which will finally send the results to the environment, which in its turn sends them to a collector object as soon as they arrive, without being necessary to wait for all the results that are being obtained.

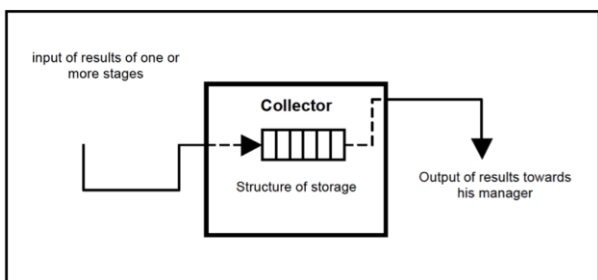


Figure 3: Component Collector of model CPAN

In summary, a CPAN is composed of an object manager that represents the CPAN itself, some stage objects and an object of the class Collector, for each petition that should be managed within the CPAN. Also, for each stage, a slave object will be in charge of implementing the necessary functionalities to solve the sequential version of the problem being solved (Figure 4). For details CPAN model, see (Rossainz and Capel 2014).

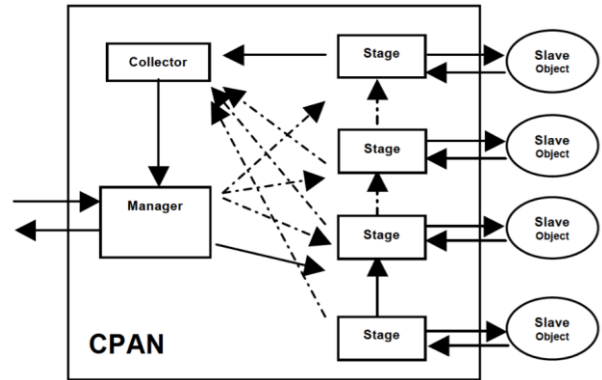


Figure 4: Internal structure of CPAN. Composition of its components

The Figure 4 shows the pattern CPAN in general, without defining any explicit parallel communication pattern. The box that includes the components, represents the encapsulated CPAN, internal boxes represent compound objects (collector, manager and objects stages), as long as the circles are the objects slaves associated to the stages. The continuous lines within the CPAN suppose that at least a connection should exist between the manager and some of the component stages. Same thing happens between the stages and the collector. The dotted lines mean more than one connection among components of the CPAN.

2.1. The CPAN seen as composition of parallel objects

Manager, collector and stages are included in the definition of a Parallel Object (PO), (Corradi 1991). Parallel Objects are active objects, which is equivalent to say that these objects have intrinsic execution capability (Corradi 1991). Applications that deploy the PO pattern can exploit the inter-object parallelism as much as the internal or intra-object parallelism. A PO-instance object has a similar structure to that of an object in Smalltalk, and additionally defines a scheduling politics, previously determined that specifies the way in which one or more operations carried out by the instance synchronize (Danelutto and Orlando 1995, Corradi 1991). Synchronization policies are expressed in terms of restrictions; for instance, mutual exclusion in reader/writer processes or the maximum parallelism allowed for writer processes. Thus, all the parallel objects derive from the classic definition of a class plus the synchronization restrictions (mutual exclusion and maximum parallelism), which are now included in that definition (Birrel 1989). Objects of the same class share the specification contained in the class of which are instances. The inheritance allows objects to derive a new specification from the one that already exists in the super-class. Parallel objects support multiple inheritance in the CPAN model.

2.2. Communication types in the parallel objects of CPAN

Parallel objects define 3 communication modes: synchronous, asynchronous communication and synchronous future communication.

1. The synchronous communication mode stops the client activity until it receives the answer of its request from the active server object (Andrews 2000).
2. The asynchronous communication does not delay the client activity. The client simply sends the request to the active object server and its execution continues afterwards. Its use in application programming is also easy, because it is only necessary to create a thread and start it to carry out the communication independently from the client (Andrews 2000).
3. The asynchronous future will delay client activity when the method's result is reached in the client's code to evaluate an expression. The asynchronous futures also have a simple use, though its implementation requires of a special care to get a syntactical construct with the correct required semantics. For details see (Lavander and Kafura 1995).

The asynchronous and asynchronous future communication modes carry out the inter-objects parallelism by executing the client and server objects at the same time.

2.3. The base classes of any CPAN

As it has already been described, a CPAN comes from the composition of a set of objects of three types. In particular, each CPAN is made up of several objects: an object manager, some stage objects and a collector object for each request sent by client objects of the CPAN. Also, for each stage of the CPAN, a slave object will be in charge of implementing the sequential part of the computation that is sought and carried out in the application or in the distributed and parallel algorithm. In PO the necessary base classes to define the manager, collector, stages objects that compose a CPAN - the implementation details are in (Rossainz, Pineda and Domínguez 2014) - are the next ones:

1. Abstract class `ComponentManager`: It defines the generic structure of the component manager of a CPAN, from which will be derived all the manager instances depending on the parallel behavior that is assumed in the CPAN creation. All specific instances of a manager accept a list of n-associations as input. An association is a pair of elements, that is, an object slave and the name of the method that has to be executed by this object. The objects slaves are external entities that contain a sequential algorithm that have to be executed by one of their methods. Once the manager has obtained the list of n-associations, it will generate the concrete stages, one for each association and then each stage becomes responsible for an object slave together with its execution method. In turn,

each stage is connected to each other, in accordance with the parallel pattern that has been implemented in the CPAN. Finally, the manager carries out a computation by the execution of one of its methods. To achieve the computation phase, it is necessary to pass on the input data that it requires to start to the method. The manager then generates a component collector and sends its reference to the stages, as well as the input data. The stages start processing the data according to the connection configuration that they keep to each other, results will be passed on as they become available. At the end the collector will gather the results sent by the stages to return them to the manager, which finally will transfer these results to the CPAN environment or to the code that uses them.

2. Abstract class `ComponentStage`: It defines the generic structure of the component stage of a CPAN, as well as their interconnections, from which will be derived all the concrete stages depending on the parallel behavior that is assumed in the creation of the CPAN. All specific instances of a stage accepts a list of associations slave-object/method as input to work with them, whether they are connected or not with the following stage of the list of associations and depending on the parallel pattern they are willing to implement. When the manager send in parallel a command to the stages, each one of them makes the object-slave to carry out the execution of its method, then the stage captures the results and sends them to the following stage or to the collector, depending on the implemented structure.
3. Concrete class `ComponentCollector`: It defines the concrete structure of the component collector of any CPAN. This component fundamentally implements a multi-item buffer, where it will store the results of stages that have the reference of this collector. This way one can obtain the result of the calculation initiated by the manager.

2.4. The Synchronization restrictions of a CPAN

It is necessary to have synchronization mechanisms available when parallel request of service take place in a CPAN, so that the objects that conform it can negotiate several execution flows concurrently and, at the same time, guarantee the consistency in the data that being processed. Within any CPAN the restrictions MAXPAR, MUTEX and SYNC can be used for correct programming of their methods.

1. MAXPAR: The maximum parallelism or MaxPar is the maximum number of processes that can be executed at the same time. That is to say the MAXPAR applied to a function represents the maximum number of processes that can execute that function concurrently. In the case of CPAN, the maximum parallelism is applied to the functions of the `ComponentManager` class and to the functions of the `ComponentStage` class.

- MUTEX: The restriction of synchronization *mutex* carries out a mutual exclusion among processes that want to access to a shared object. The mutex preserves critical sections of code and obtains exclusive access to the resources. In the case of the CPANs, the restriction mutex applied to a function represents the use of that function on the part of a process every time. In other words, the mutex allows that only one of the processes executes the function, blocking all the other processes trying to make use of the service until one of the ones that execute it finishes. The mutex within the CPAN is applied to the functions of an object collector.
- SYNC: The restriction SYNC is not more than a producer/consumer type of synchronization; it is of use, for instance, for programming the methods of the componentCollector class. SYNC helps to synchronize these methods when accessing the shared resource at the same time, which in this case is a multi-item list.

The details of the algorithms and their implementation can be seen in (Rossainz, Pineda and Domínguez 2014).

3. CONSTRUCTION OF A CPAN

With the base-classes of the PO model of programming, it is now possible to build concrete CPANs. To build a CPAN, first it should have made clear the parallel behavior that the user application needs to implement, so that the CPAN becomes this pattern itself. Several parallel patterns of interaction have long been identified in Parallel Programming, such as farms, pipes, trees, cubes, meshes, a matrix of processes, etc. Once identified the parallel behavior, the second step consists of elaborating a graph of its representation, as an informal design of the objective system. This practice is also good for illustrating the general characteristics of the desired system and will allow us to define its representation with CPANs later on, by following the pattern proposed in the previous section. When the model of a CPAN has already been made clear, it defines a specific parallel pattern; let's say, for example, a tree, or some other mentioned pattern, and then the following step will be to do its syntactic definition and specify its semantics. Finally, the syntactic definition prior to any programmed CPAN is transformed into the most appropriate programming environment, with the objective of producing its parallel implementation. It must be verified that the resulting semantics is the correct one. To attain this, we use several different examples to demonstrate the generality and flexibility of the application CPAN-based design and the expected performance and quality as a software component. Some support from an integrated development environment (IDE) for Parallel Programming should be provided in order to validate the component satisfactorily. The parallel patterns worked in the present investigation have been the pipeline and the binary-tree to solve the sorting problem using two different algorithms.

4. THE CPAN PIPELINE

It is presented the technique of the parallel processing of the pipeline as a High Level Parallel Composition or CPAN, applicable to a wide range of problems that you/they are partially sequential in their nature. The CPAN Pipe guarantees the parallelization of sequential code using the pattern PipeLine.

4.1. The technique of the Pipeline

Using the technique of the Pipeline, the idea is to divide the problem in a series of tasks that have to be completed, one after another, see figure 5. In a pipeline each task can be executed by a process, thread or processor for separate (De Simone 1997, Robbins and Robbins 1999).

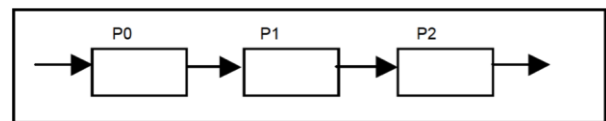


Figure 5: Pipeline

The processes of the pipeline are sometimes called stages of the pipeline (Roosta 1999). Each stage can contribute to the solution of the total problem and it can pass the information that is necessary to the following stage of the pipeline. This type of parallelism is seen many times as a form of functional decomposition. The problem is divided in separate functions that can be executed individually, but with this technique, the functions are executed in succession.

The technique of parallel processing pipeline is then presented as a High Level Parallel Composition applicable to solving a range of problems that are partially sequential in nature, so that the Pipe CPAN guarantees code parallelization of sequential algorithm using the pattern Pipeline.

4.2. Representation of the Pipeline as a CPAN

The Figure 6 represents the parallel pattern of communication Pipeline as a CPAN.

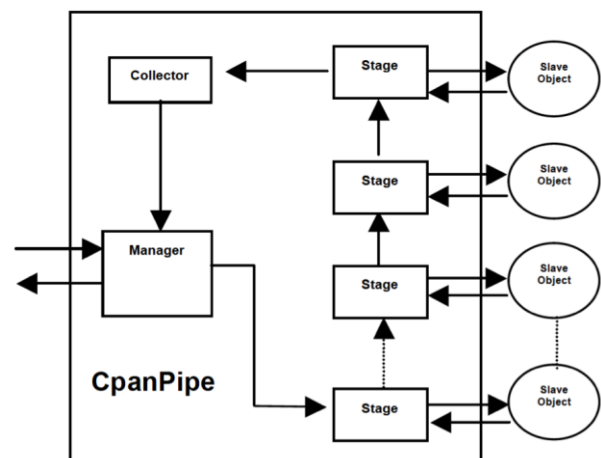


Figure 6: The CPAN of a Pipeline

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

5. PARALLEL ALGORITHM SORTING WITH PIPELINE

Using a PipeLine is useful to introduce a scheme of parallelization of a sorting algorithm, so that to solve the problem have to perform a series of operations on a data set. Each of these transactions is considered a stage in the data processing and each is executed by a separate process that synchronizes with the above processes and form respective next stage. The complete data processing ends when they have passed through every stage (Wilkinson and Allen M., 1999).

Pipeline processing a serial data sequence they pass through the pipeline stages. Each stage is associated with a process that performs a specific operation when a fact comes through its associated slave object. Completed this operation, passes the result to the next stage. In a parallel sorting algorithm with a pipeline 3 phases are distinguished (Barry and Allen 1999; Bletloch 1996; Roosta 1999):

- The initial charge: data is allocated to all processes associated with the stages of the pipeline. In this phase the processes are running the same code in the second phase, the difference is that you must initialize properly to receive the first data, they will come from the previous stage or initial program load.
- The processing of the data stream with maximum efficiency: Processes behave cyclically in execution. Data support the previous stage, process and send the result to the next stage. Each process has to be synchronized with that of the previous stage to not send new data when it has not yet finished processing the data streams; but also to the next step, to not send the result to the process of this stage is not ready to receive it. The final process has a special behavior with respect to the processes associated with the above steps as you have to run a routine or exit code and presentation of results of the program. Its operation is to obtain the data sent by the process of the last stage of the pipeline and send them to an output device or send a termination condition the main program. The series of results it produces the last process must match the expected result of the algorithm has been parallelized, if the pipeline has been successfully parallelized.
- Download: In this last stage the processes send the result of the last processed data and themselves detect termination situation, as they will no longer receive more data from the input stream and should not pose any global control in the program tells them when they have finished. Processes for transmitting the data stored in its stages before completion, is usually

introduced a special value at the end of the input sequence used to unload the pipeline.

To implement the parallel sorting algorithm, a pipeline process is used, which receives an unordered set of integers by a routine or entry code. It is obtained as a result the ordered sequence of integers ascending. The number of values in the input sequence cannot be greater than the number of pipeline stages. Each pipeline processes can store an integer, which will be the largest that has been received so far from the previous step. In each iteration, a process receives a integer, compared to the one that had stored and sends the smaller of the next stage of the pipeline, while the highest is stored (Barry and Allen 1999; Bletloch 1996). For more details see (Rossainz, Capel and Domínguez 2015).

6. THE CPAN TREEDV

The programming technique is presented it Divide and Conquer as a CPAN, applicable to a wide range of problems that can be parallelizable within this scheme, in particular to solve sorting problems in parallel (Rossainz and Capel 2012).

6.1. The technique of the Divide and Conquer

The technique of it Divide and Conquer it is characterized by the division of a problem in sub-problems that have the same form that the complete problem (Brassard and Bartley 1997). The division of the problem in smaller sub-problems is carried out using the recursion. The method recursive continues dividing the problem until the parts divided can no longer follow dividing itself, and then they combine the partial results of each sub-problem to obtain at the end the solution to the initial problem (Brassard and Bartley 1997). In this technique the division of the problem is always made in two parts, therefore a formulation recursive of the method Divide and Conquer form a binary tree whose nodes will be processors, processes or threads.

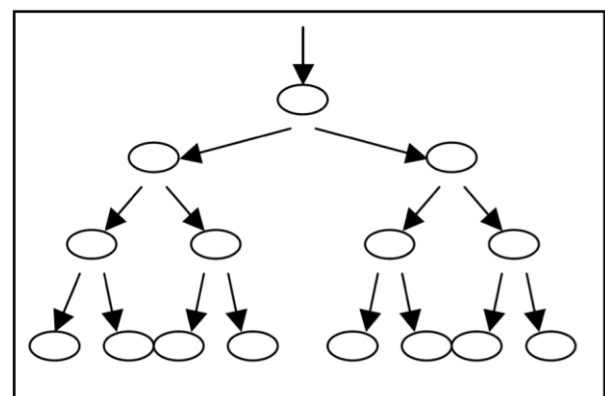


Figure 7: Binary Tree

The node root of the tree receives as input a complete problem that is divided in two parts. It is sent to the node left son, while the other is sent to the node that represents the right son (figure 7). This division process

is repeated of recursive form until the lowest levels in the tree. Lapsed a certain time, all the nodes leaf receives as input a problem given by its node father; they solve it and the solutions (that are the exit of the node leaf) are again correspondents to its progenitor. Any node father in the tree will obtain his children's two partial solutions and it will combine them to provide an only solution that will be the node father's exit. Finally the node root will give as exit the complete solution of the problem, (Brinch Hansen 1993). This way, while in a sequential implementation a single node of the tree can be executed or visited at the same time, in a parallel implementation, more than a node it can be executed at the same time in the different levels, it is, when dividing the problem in two sub-problems, both can be processed in a simultaneous way.

6.2. Representation of the TreeDV as a CPAN

The representation of the patron tree that defines the technique of it Divide and Conquer as CPAN has their model represented in figure 8.

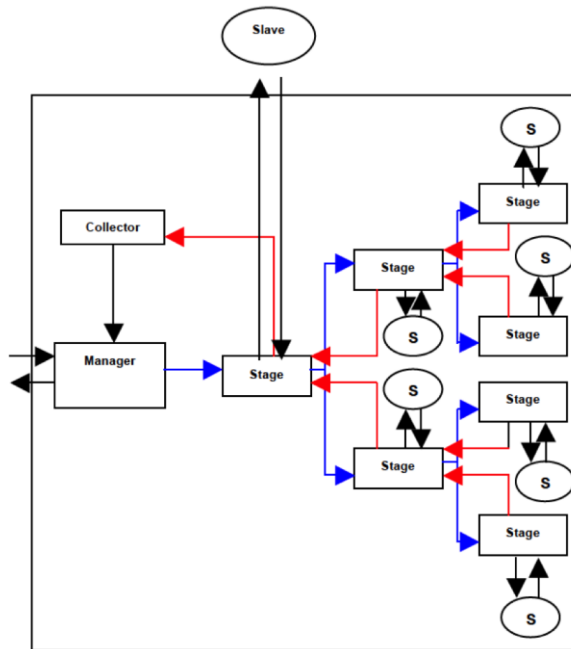


Figure 8: The Cpan of a TreeDV

Contrary to the previous model, where the objects slaves were predetermined outside of the pattern CPAN, in this model an object slave is only predefined statically and associated to the first stage of the tree. The following objects slaves will be created internally by the own stages in a dynamic way, because the levels of the tree depend from the problem to solve and a priori the number of nodes that can have the tree is not known, neither its level of depth.

6.3. Quicksort sorting algorithm using the CPAN TreeDV

The Quicksort sorting was created by Hoare and is based on the paradigm of divide and conquer. As a first step the algorithm selects as a pivot one of the elements

of the data set you have to order. The array is then partitioned on either side of the pivot: elements are moved so that those greater than the pivot are to its right, whereas the others are to its left. If now the sections of the array on either side of the pivot are sorted independently by recursive and parallel calls of the algorithm (Brassard and Bratley 1997), in this case through the stage TreeDV CPAN objects, the final result is a completely sorted array, no subsequent merge step being necessary.

Algorithm QuickSort(T[i..j])

```

{
  var l;
  if ( j-i is sufficiently small then insert(T[i..j]) )
  else {
    l= pivot(T[i..j]);
    QuickSort(T[i..l-1]);
    QuickSort(T[l+1..j]);
  }
}

```

To balance the sizes of the two subinstances to be sorted, we would like to use the median element as the pivot. Unfortunately, finding the median takes more time it is worth. For this reason we simply use an arbitrary element of the array as the pivot, hoping for the best.

Algorithm pivot(T[i..j])

```

{
  var l;
  p=T[i]; k=i; l=j+1;
  repeat { k=k+1; } until ( T[k]>p ) or ( k>=j );
  repeat { l=l-1; } until ( T[l]<=p );
  while ( k<l )
  {
    swap(T[k],T[l]);
    repeat { k=k+1; } until ( T[k]>p );
    repeat { l=l-1; } until ( T[l]<=p );
  }
  swap(T[i],T[l]);
  return l;
}

```

Suppose subarray $T[i..j]$ is to be pivoted around $p=T[i]$. One good way of pivoting consists of scanning the subarray just once, but starting at both ends. Pointers k and l are initialized to i and $j+1$, respectively. Pointer k is then incremented until $T[k]>p$, and pointer l is decremented until $T[l]<=p$. Now $T[k]$ and $T[l]$ are interchanged. This process continues as long as $k<l$. Finally, $T[i]$ and $T[l]$ are interchanged to put the pivot in its correct position (Brassard and Bratley 1997).

7. PERFORMANCE

Performance analysis of CPANS Pipeline and TreeDV solving sorting problems are shown. The aim is to show that, at least for these problems, the performances obtained are "good" based on the model of the CPAN.

CpanPipe and CpanTreeDV performance to implement a parallel sorting algorithm was carried out on a parallel computer with 64 processors, 8 GB of main memory, high-speed buses and distributed shared memory architecture. Performance measures obtained in implementing the CpanPipe and CpanTreeDV that solves the problem of sorting using an Pipeline and Binary Tree respectively, is carried out with the following restrictions execution:

- Parallel implementation of sequential sorting algorithm based on a pipeline in the case of CPAN Pipe and parallel-sequential implementation is Quicksort sorting algorithm based on a binary tree using the technique of Divide and Conquer for the case of CpanTreeDV.
- In both cases, both the CPAN Pipe as CPAN TreeDV, it is implemented the same sequential algorithm of comparing values in each of the slave objects associated with the stages of CPANs.
- 50000 a set of whole numbers randomly obtained in the range of 0-50000 ordered, allowing make a sufficient charge for processors and thereby observe the performance improvement CpanPipe and TreeDV,
- CpanPipe and CpanTreeDV execution for 2, 4, 8, 16 and 32 full-time processors.

The methodology that has been followed for the analysis of performance CPANs is:

1. The CPANs Pipeline and TreeDV are compiled in their sequential and parallel versions and run on the corresponding cpuset,
2. The following parameters of the execution performance of CPANs are measured. They show their behavior.
 - 2.1. Runtime of each CPAN, including its sequential version and measurement of page faults caused in the system during its execution.
 - 2.2. Cycles per instruction (CPI) for each CPAN, including sequential versions.
 - 2.3. Page faults caused during the execution of the CPANs.
 - 2.4. Magnitude speedup for each execution of the CPANs in Cpuset about their sequential versions.
 - 2.5. Upper bound of the magnitude speedup for each CPAN using Amdahl's law.

Tables 1, 2 and figures 9, 10, show the series of measurements obtained including their corresponding sequential versions for Cpan Pipe and TreeDV, execution time in seconds, cycles per instruction executed, magnitude speedup found and the upper bound on the magnitude of speedup using for that Amdahl's law.

Table 1: Cpan Performance Pipe Parallel to the Management of 50000 Sorting integers

CPAN PIPE	Sequential Pipe	CPU SET 2	CPU SET 4	CPU SET 8	CPU SET 16	CPU SET 32
Runtime in seconds	238.5	127.27	123.83	116.97	115.63	111.03
CPU time in seconds	231.82	224.79	217.8	203.98	198.3	191.89
CPI	1.862	0.909	0.904	0.901	0.886	0.871
Speedup	1	1.87	1.93	2.04	2.06	2.15
Amdahl	1	1.89	3.39	5.63	8.42	11.19

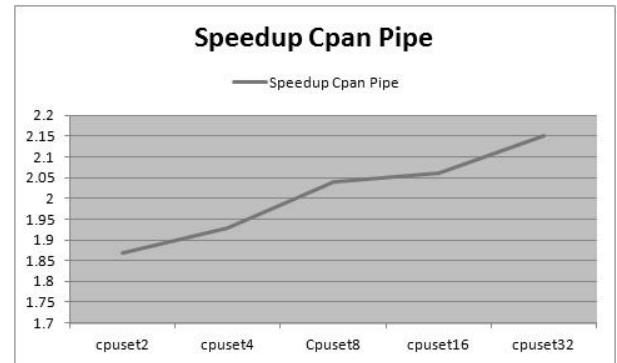


Figure 9: Scaling the magnitude of CpanPipe Speedup for 2, 4, 8, 16 and 32 exclusive processors

Table 2: Cpan Performance Binary Tree Parallel to the Management of 50000 Sorting integers

CPAN TREETDVQS	Sequential TREETDVQS	CPU SET 2	CPU SET 4	CPU SET 8	CPU SET 16	CPU SET 32
Runtime in seconds	20.55	11.48	7.21	4.59	3.68	3.40
CPU time in seconds	7.60	7.18	6.31	6.88	6.45	6.35
CPI	1.261	0.900	0.892	0.858	0.849	0.836
Speedup	1	1.79	2.85	4.48	5.58	6.04
Amdahl	1	1.82	3.08	4.71	6.40	7.80

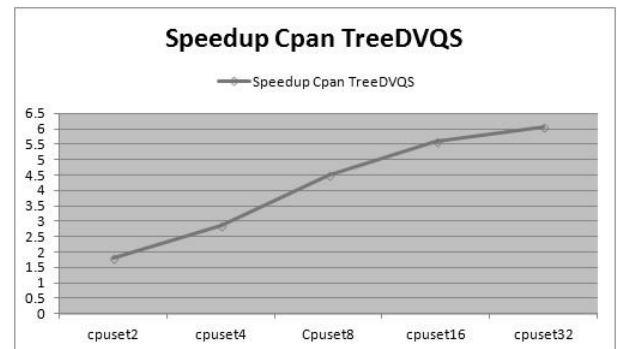


Figure 10: Scaling the magnitude of CpanTreeDV Speedup for 2, 4, 8, 16 and 32 exclusive processors

The measurements obtained from page faults caused by the implementation of Pipeline and TreeDV CPANs shown in Table 3 and Figure 11.

A measure page fault is useful to see if the CPANs cause excessive paging, especially if they use a lot of memory for execution, which has not been.

Table 3: Page faults in the execution of CPANS

CPANS	SEQ		Cpuset2		cpuset4		cpuset8		cpuset16		cpuset32	
	ma-	mi-	ma-	mi-	ma-	mi-	ma-	mi-	ma-	mi-	ma-	mi-
	yor	nor	yor	nor	yor	nor	yor	nor	yor	nor	yor	nor
Pipe	0	25	0	37	0	37	0	37	0	37	0	37
TreeDVQS	0	25	0	38	1	38	0	38	0	38	0	38

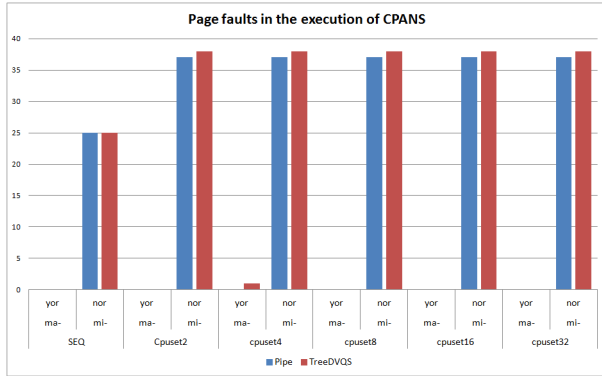


Figure 11: Equivalence of page faults in the execution of the CPANS

Parallel executions of CPANS have a time shorter than the time used by their corresponding sequential versions, as expected. The execution times of their parallel versions CPANS improve as the number of processors is increased, ie, as is increasing the number of processors with which CPANS are executed, their execution times are decreasing. A value of the magnitude called speedup is appreciated ever upward on improving execution times of parallel CPANS respect to its sequential counterpart, but always below the levels of Amdahl's Law calculated, obtaining "good" yields. Improved CPI is obtained by increasing the number of processors, that is, a larger number of processors used in performing the CPANS, the lower the value of the ratio of cycles per instruction. This indicates that while the number of instructions in the execution of the application within cpusetX remains more or less constant, the number of cycles per instruction decreases, resulting in a gain in the final value of the CPI.

8. CONCLUSIONS

We have implemented communication patterns Pipeline and Binary Tree as CPANS and with them has solved the problem of sorting parallelizing two sequential algorithms different, it is using a pipeline process to sort a dataset in disordered and one that by quick sort uses a binary tree for the ordering of the same dataset disordered by divide and conquer technique.

The implemented CPANS can be exploited, thanks to the adoption of the approach oriented to objects. Well-known algorithms that solve sequential problems in algorithms parallelizable have transformed and with them the utility of CPANS has been proven.

It has become a sequential sorting algorithm on a parallelizable algorithm using a Pipeline and a Binary Tree as CPAN. It has been proven performance Cpan

Pipe and Cpan TreeDV by metrics Speedup, Amdahl's Law and efficiency to demonstrate that parallel behavior CpanPipe is better than its sequential counterpart. Furthermore the speedup TreeDVQS CPAN is much closer to the upper bound (Amdahl law) that the speedup obtained in CPAN Pipe. The runtime in seconds of CPAN TreeDV to solve the problem of sorting is much less than the CPAN Pipe that solves the same problem with the same input size, according increase the number of processors. The same applies to the CPU time in the execution of those CPANS (see Table 1, 2 and Table 3).

REFERENCES

- Andrews G.R., 2000. Foundations of Multithreaded, Parallel, and Distributed Programming, Addison-Wesley
- Brassard G., Bratley P., 1997. Fundamentos de Algoritmia, Prentice-Hall. 1997.
- Bacci, Danelutto, Pelagatti, Vaneschi, 1999. SKIE: A Heterogeneous Environment for HPC Applications. *Parallel Computing* 25.
- Birrell, Andrew, 1989. An Introduction to programming with threads. *Digital Equipment Corporation*, Systems Research Center.
- Blelloch, Guy E., 1996. Programming Parallel Algorithms. *Communications of the ACM*. Volume 39, Number 3.
- Brinch Hansen, 1993. Model Programs for Computational Science: A programming methodology for multicomputers, *Concurrency: Practice and Experience*, Volume 5, Number 5.
- Barry W., Allen M., 1999. Parallel Programming. Techniques and Applications Using Networked Workstations and Parallel Computers. *Prentice Hall*. ISBN 0-13-671710-1.
- Corradi A., Leonardi L., 1991. PO Constraints as tools to synchronize active objects. *Journal Object Oriented Programming* 10, pp. 42-53.
- Corradi A, Leonardo L, Zambonelli F., 1995. Experiences toward an Object-Oriented Approach to Structured Parallel Programming. *DEIS technical report no. DEIS-LIA-95-007*.
- Danelutto, M.; Orlando, S; et al., 1995. Parallel Programming Models Based on Restricted Computation Structure *Approach. Technical Report-Dpt. Informatica. Università de Pisa*.
- Darlington et al., 1993, Parallel Programming Using Skeleton Functions. *Proceedings PARLE'93, Munich (D)*.
- De Simone, et al. 1997. Designs Patterns for Parallel Programming. *PDPTA International Conference*.
- Lavander G.R., Kafura D.G. 1995. A Polimorphic Future and First-class Function Type for Concurrent Object-Oriented Programming. *Journal of Object-Oriented Systems*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.7088&rep=rep1&type=pdf>

- Robbins, K. A., Robbins S. 1999. “UNIX Programación Práctica. Guía para la concurrencia, la comunicación y los multihilos”. *Prentice Hall*.
- Roosta, Séller, 1999. Parallel Processing and Parallel Algorithms. *Theory and Computation. Springer*.
- Rossainz, M., 2005. Una Metodología de Programación Basada en Composiciones Paralelas de Alto Nivel (CPANs). *Universidad de Granada, PhD dissertation, 02/25/2005*.
- Rossainz, M., Capel M., 2008. A Parallel Programming Methodology using Communication Patterns named CPANS or Composition of Parallel Object. *20TH European Modeling & Simulation Symposium*. Campora S. Giovanni. Italy.
- Rossainz, M., Capel M., 2012. Compositions of Parallel Objects to Implement Communication Patterns. *XXIII Jornadas de Paralelismo. SARTECO 2012*. Septiembre de 2012. Elche, España.
- Rossainz M., Capel M., 2014. Approach class library of high level parallel compositions to implements communication patterns using structured parallel programming. *26TH European Modeling & Simulation Symposium*. Campora Bordeaux, France.
- Rossainz M., Pineda I., Dominguez P., Análisis y Definición del Modelo de las Composiciones Paralelas de Alto Nivel llamadas CPANs. *Modelos Matemáticos y TIC: Teoría y Aplicaciones 2014*. Dirección de Fomento Editorial. ISBN 987-607-487-834-9. Pp. 1-19. México.
- Rossainz M, Capel M., Domínguez P., 2015. Pipeline as high level parallel composition for the implementation of a sorting algorithm. *27TH European Modeling & Simulation Symposium*. Campora Bergeggi, Italy.
- Wilkinson B., Allen M., 1999. Parallel Programming Techniques and Applications Using Networked Workstations and Parallel Computers. *Prentice-Hall*. USA.

DUAL USE BICYCLE PATH NETWORK DESIGNING AND EXPLOITATION ENVIRONMENT - VELOROUTER

Egils Ginters^(a), Mikelis Baltruks^(b), Inita Sakne^(b), Yuri Merkuryev^(a)

^(a) Faculty of Computer Science and Information Technology, Riga Technical University, 1 Kalku Street, Riga LV-1658, Latvia

^(b) Sociotechnical Systems Engineering Institute, 4 Cesu Street, Valmiera LV-4201, Latvia

^(a)egils.ginters@ieec.org, ^(b)mikelis.baltruks@va.lv, ^(b)inita.sakne@va.lv, ^(a)merkur@itl.rtu.lv

ABSTRACT

One of the types of clean transport is cycling. In order to ensure the successful functioning of an integrated and intermodal urban transport system (Merkuryev, Zenina and Romanovs 2015) cycling has to be one of its natural components. Due to limited funding it is important to understand which bike path network would be more efficient and ensure the greatest possible population transfer from other types of transport to bicycles, because it is impossible to construct bike paths, racks and lease points wherever it would be desirable.

Keywords: bike path network design, occupancy simulation, multi-agent based models, VeloRouter

1. INTRODUCTION

Previous FP7-ICT-2011-7 FUPOL project No. 287119 (2011-2015) "Future Policy Modelling" (Ginters et al 2014; Aizstraus et al 2014; Buil et al 2015) gave the opportunity to perform market research and understand the needs of potential users, interested in sustainable transport schemes development. The market analysis involves significant amount of objects. Among them would be mentioned Bikemap (<http://www.bikemap.net>), veloroutes.org (<http://veloroutes.org>), routebuilder.org (<http://www.routebuilder.org>) and EuroVelo (<http://www.eurovelo.com/en>). Similar products were Pro-Plan, Topo-Plan, Bike Node Plan (http://www.routeyou.com/route/routeplanner/overview_all/choose-the-route-planner-that-fits-your-needs), MapMyRide (<http://www.mapmyride.com>), CycleStreets (<http://www.cyclestreets.net/about/>), plotaroute.com (<http://www.plotaroute.com>), Flattest Route (<http://www.flattestroute.com>), Velomap.org (<https://www.velomap.org>), RouteLoops (<http://www.routeloops.com>) and several specialised route planners, for example for mountain biking BikeRouteToaster.com (<http://www.bikeroutetoaster.com>).

The bicycle route planner in the city of Bern Pro Velo Bern (<http://www.veloroutenplaner.ch>) is based on OpenStreetMap data. The proposed routes are generated automatically based on this data. The routing can be

used all over Switzerland. Nevertheless, it may happen that the suggested route can't be driven all the way or that cycling is banned on certain sections. Based on the analysis implemented (Ginters et al 2013) and research mentioned above, it was concluded that existing cycling routes design and planning products mainly offer the capabilities of publishing and planning routes, but do not provide functionality necessary for municipalities to build justified bicycle path network, which is critical for sustainability of transport schema.

2. VELOROUTER CONCEPT

The multi agent-based bicycle path network and exploitation simulator (VeloRouter) is designed in the ABM/MAS Repast Symphony and uses OpenStreetMap spatial data. The product has dual applicability as it is adapted to both of audiences: the needs of municipalities and cyclists (see Figure 1).

The municipality is interested in some basic question: Is the offered cycling route map satisfactory? This is recognized by summarizing potential comments coming from "Cyclists" during the project deliberations "Requests". However the second question is: which potential cycling route sections should be built first? VeloRouter provides municipalities with bicycle path discussion and geofencing opportunities by receiving feedback from cyclists directly and by using semantic search tools on social networks. Municipality can publish ("Route Visualization") bicycle path network map intended giving possibilities to the citizens for discussing and voting "for" or "against" the project. Cyclists also can design and publish their own routes. Statistical analysis of the data ensures summarizing of viewpoints to select most suitable solution for policy crafting in the current conditions. Naturally that accuracy of the forecasting will depend on the quantity and distribution of the audience, and changes in bicycle path network already introduced.

The cyclists want to know what the occupancy of a route will be in certain meteorological conditions on a specific date, as well as if the route is suitable for the group i.e. terrain etc.? In route occupancy simulation model each agent is a cyclist or a group of cyclists that move on a chosen route considering traffic restrictions,

the quality of the route and influence of other cyclists on the occupancy. The occupancy of the route will depend also of particular time (specific hours),

working days or vacations and season as well: summer or winter.

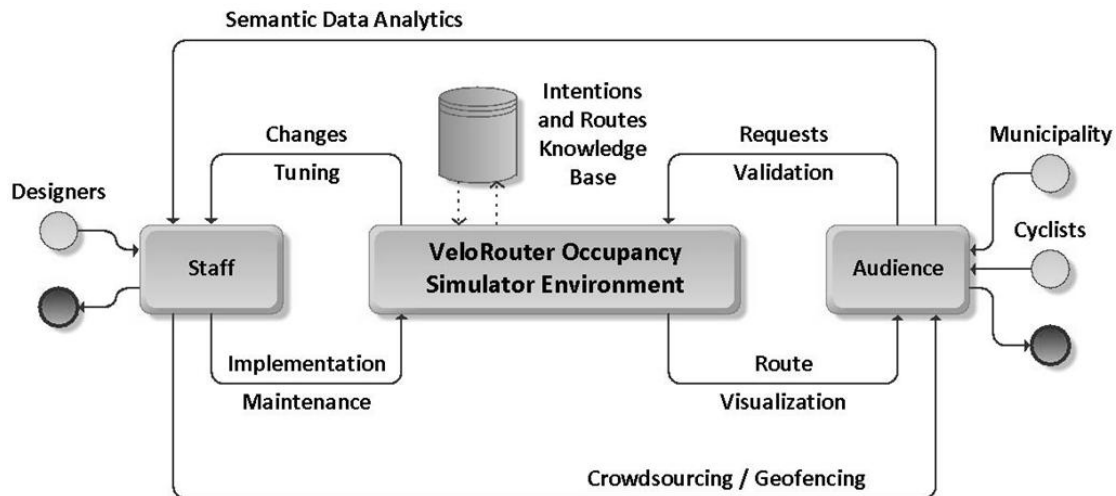


Figure 1: VeloRouter Conceptual Model

During VeloRouter usage Intensions and Routes Knowledge Base is continually and automatically updated. With each usage this improves the quality of simulated occupancy predictions. The database contains data entered by the municipality, which was gathered during interviews, voting data, information added by cyclists, as well as the planned routes of cyclists. Content is complemented by crowdsourcing data from social networks Facebook, LinkedIn etc. Semantic search methods are used also for occupancy simulation model and crafting scenarios validation.

In the conceptual model of VeloRouter there are two active informative subjects "Staff" and "Audience". "Staff" represents developers and maintainers of the VeloRouter information system. Developers design the VeloRouter Occupancy Simulator Environment, as well as are responsible for model tuning and changes management. "Staff" also supports search operations in semantic webs and processes obtained data for simulator validation. "Audience" are employees of the municipality who publish bicycle path projects ("Routes Visualization") and communicate with "Cyclists", who vote and make recommendations.

Figure 2 shows a possible Larnaca project screen. There are different types of projects. "My project" is a user created bicycle route network on the map in the allowed planning region. Simultaneously there can be multiple allowed planning regions. Project has to be published before others can view it and cast votes. It gets status "Published project". Although there is a project that everybody uses and it is called "Exploitation project" and is set by administration. In Exploitation project users can view existing road network, plan individual routes and view their terrain, view traffic restrictions,

simulate occupancy as well as vote, comment and share edits of the project.

It is also possible to analyze project's statistics - see most used road sections, most commented areas and road sections of the region, vote distribution etc. Statistics is municipality's object of interest, as it offers recognition of the offered cycling scheme's acceptance as well as shows which sections of the cycling network scheme are the most important ones.

Occupancy simulation results are important also for municipality as they give an idea about possible outcome, if number of cyclists increased in specific areas or if some traffic restrictions are made. Another one of important concepts is planning region. It is geographical area on the map within which manager of VeloRouter licence has rights to plan cycling routes. Area (region) is set during the talks of VeloRouter owners and manager of licence. Abstracting from financial factors, there is a requirement set for manager to provide full functioning. Functioning options are determined by the number of possible cyclists and size of the road network. If licence manager does not possess powerful enough resources for deployment of VeloRouter, then licence purchase can be refused or allowed planning region reduced.

A particularly important factor is provision of full occupancy simulation as it will be inconvenient if manager does not use HPC solutions. There is no problem with simulations with small number of cyclists, although analyzing real transport schemes where cyclists are over a thousand then the real-time conditions are forgettable. It isn't a problem for municipality as their simulation can be done in even a day, but user's waiting time shouldn't exceed seconds.

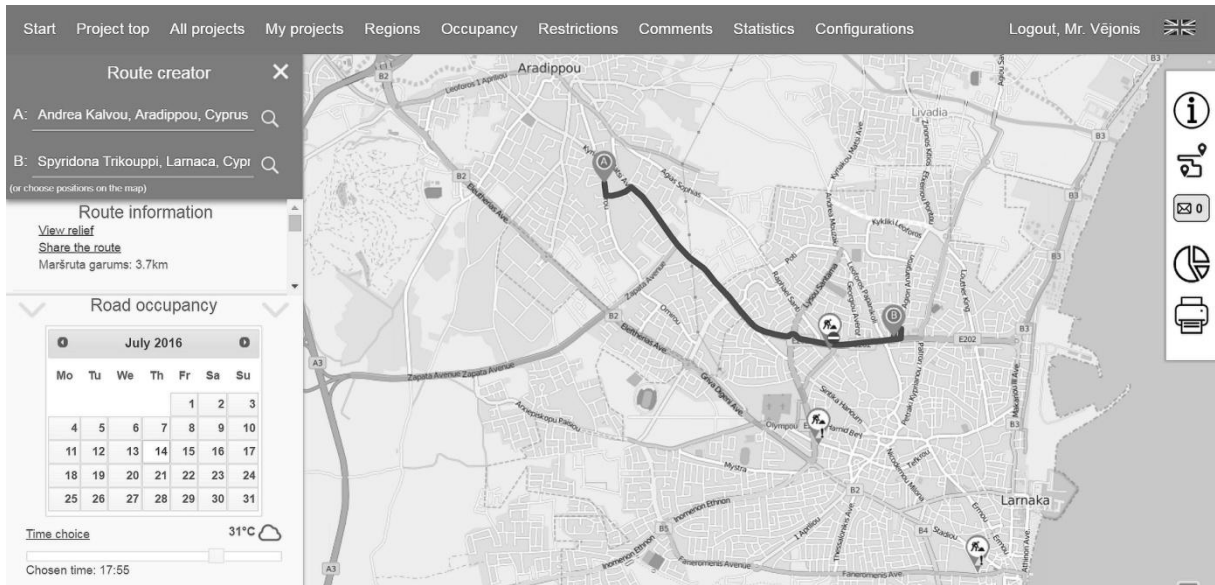


Figure 2: Introduction to Larnaca VeloRouter Project

Above described situation and related problems were examined while developing Skopje Bicycle Intermodality Simulator (Ginters et al 2014), where it was found that the municipality has two fundamental questions - where to build and how to build?!

Although cyclist would not be ready to wait longer amount of time for the construction of his route. Collaboration with user has to be done in real time.

There are different types of simulation, starting with different DEVS solutions up to use of system dynamics modeling, however in situations where accurate enough situation analysis is needed, micro-simulation is unavoidable which determined the use of ABM/MAS.

3. AGENT-BASED OCCUPANCY SIMULATION MODEL

Each cyclist is an agent. For modeling purposes historical data is used, that is stored in "Intentions and Routes Knowledge Base". Municipality employees enter the data, that is obtained in interviews with potential bicycle path users in different places of the city - malls, mass events etc., as well as semantic search results in social networks and Internet environment. Database also contains cyclists' previously calculated routes and intended plans. Occupancy modeling is done for route chosen by user for travelling in specified time interval and meteorological conditions. Riding season is also taken into account (see Figure 3).

The occupancy on the route is specified as:

$$O_{\Delta t}^{AB} = (L^{AB}, Q_{\Delta t}^{AB}, T^{AB}, N_{\Delta t}^{AB}, N_{oth\Delta t}^{AB}, t_0^{AB}, \Delta t^{AB}, M_{\Delta t}^{AB}) \quad (1)$$

where,

- $O_{\Delta t}^{AB}$ - occupancy on the route from the point A to B depending on the location of the agent on the route;
- L^{AB} - length of the route;
- $Q_{\Delta t}^{AB}$ - quality of the route with possible traffic limitations;
- T^{AB} - terrain of the route;
- $N_{\Delta t}^{AB}$ - number of the cyclists in the group;
- $N_{oth\Delta t}^{AB}$ - other simulated cyclists on the route;
- t_0^{AB} - travel start time;
- Δt^{AB} - travel time;
- $M_{\Delta t}^{AB}$ - meteorological conditions/season during travel time Δt^{AB} .

Since database stores all the historical data, then it is possible to determine where (in any given time) any agent will be. That means that by taking average riding speed it is possible to calculate possible number of agents in any time interval in any given route section. The smaller given time interval, wider road network and larger number of cyclists, the larger is load of computing resources. Therefore to ensure VeloRouter performance capabilities in real-time mode, there is a simulation-off option.

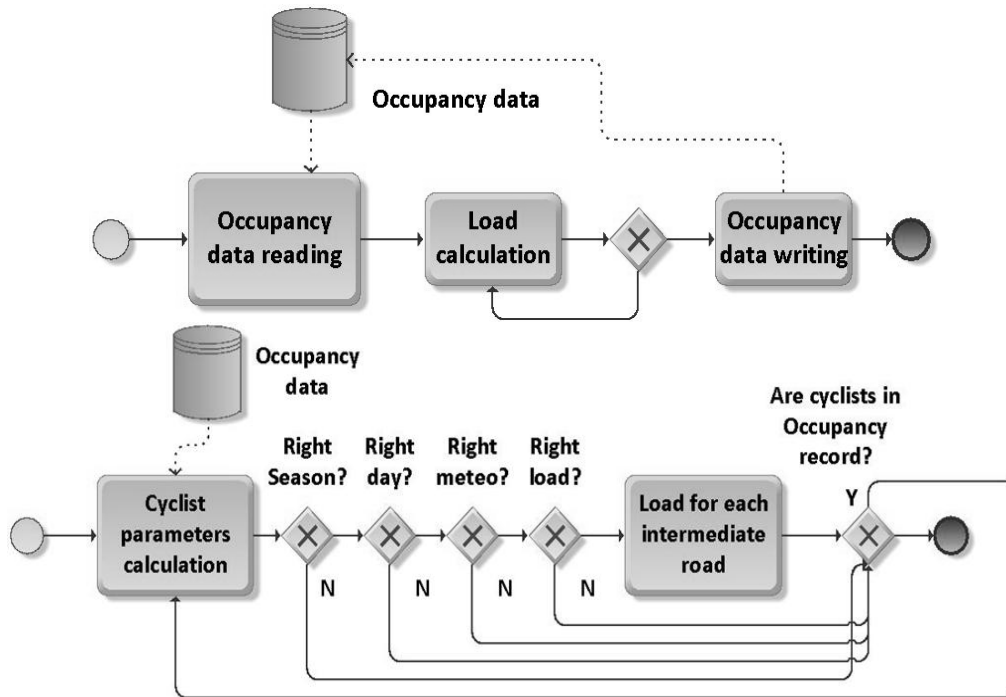


Figure 3: Occupancy Simulation BPMN2

Simulation results are visualized on the route map in form of a movie by changing the marking color indications in time of occupied route links.

4. CONCLUSIONS

VeloRouter (Ginters et al 2016) is one of the first simulators that takes into account municipality needs, performs a social function and the same time is cyclist-friendly. Software environment is designed in conformity with open source code principles. It is relatively easy adaptable for the needs of big cities and smaller ones.

Basically all cycling route designers offer route planning capabilities, however there are very few that propose route load predictions for a route on a specific date taking into account meteorological conditions and quality of the route.

VeloRouter complies with above mentioned requirements as well as provides options of publishing user designed routes, that created alternatives to municipality offered solutions. In fact, it is a municipality and user collaboration platform, that allows to reach a compromise solution during the construction of cycling scheme.

Semantic search has a significant role as it speeds up the database filling process with reliable data. In this way the quality of route occupancy modeling is being improved.

VeloRouter provides news website specific options of comments and votes that enables municipalities to make sure that their proposed transport scheme is correct.

Lack of feedback is obvious road to misguided decisions.

Information statistics analysis section prepares information for region development planners as well as is useful to cyclists by showing the most popular of possible route sections.

Open-source solution gives a chance to other developers to edit the existing software and contributes to the applied model verification.

Open question is about visualization of modeling results. Currently existing cycling route planning tool methods are used, which do not allow to fully benefit from the advantages of semantic solutions. The user sees the existing situation as the result of visualization, but it is not possible to simply view similarities and influencing factors. Here the usage of semantic analytics and visualization would be preferable, offering advanced supplemented visualization solutions. It is possible to use ready-made software tools, for example SemaVis (Nazemi et al 2014; Ginters and Aizstrauts et al 2014), however the majority are commercial products that would reduce VeloRouter audience. So one of the following VeloRouter directions of development could be development of new semantic data analysis and visualization solutions.

VeloRouter route section occupancy modeling is based on use of ABM/MAS, that is useful to municipality as it shows route sections which could be reconstructed in order to increase their permeability, however, there is a series of technological problems.

Agent-based simulation (ABM/MAS) supports relatively precise load prediction so unpleasant incidents during a trip can be limited. However main ABM/MAS application bottleneck is necessity to know location of each agent on the map during simulated travel session asking for significant calculation resources if agent amount or the set of possible routes are growing. Thus make limitation for planning region size. Using Cloud and HPC solutions it is possible more or less to remove bottleneck although it is reasonable to continue the simulation algorithms enhancing.

REFERENCES

- Merkuryev Y., Zenina N., Romanovs A., 2015. Intelligent Transport Measures as a Component of Cyber-Physical Systems: case Study for Adazi City. In: Bruzzone A.-G., Del Rio Villas D., Longo F., Merkurjev Y., Piera M.-A., eds. Proceedings of 17th International Conference on Harbor, Maritime & Multimodal Logistics Modelling and Simulation (HMS 2015). ISBN 978-88-97999-58-4. pp. 57-65. September 21-23, 2015, Bergaggi (Italy).
- Ginters E., Aizstrauts A., Dreija G., Ablazevica M., Stepucevs S., Sakne I., Baltruks M., Piera Eroles M.-A., Buil R., Gusev M., Velkoski G., 2014. Skopje Bicycle Inter-modality Simulator – e-involvement through simulation and ticketing. In: Proceedings of 26th European Modelling & Simulation Symposium (EMSS 2014). ISBN 978-88-97999-38-6 (paperback), ISBN 978-88-97999-32-4. pp. 557-563. 10-12 September, Bordeaux (France).
- Sonntagbauer S., Ginters E., Aizstrauta D., Lapans A., Valtenbergs V., Sakne I. et al., 2013. D4.1 - FUPOL Simulator Software Requirements Report. 231.
- Aizstrauts A., Ginters E., Baltruks M., Gusev M., 2014. Architecture for Distributed Simulation Environment. *Procedia Computer Science*, Elsevier, ISSN 1877-0509, Volume 43, doi:10.1016/S1877-0509(15)00231-8, 18-26.
- Buil R., Piera M.-A., Gusev M., Ginters E., 2015. MAS Simulation for Decision Making in Urban Policy Design: Bicycle Infrastructure. In: Bruzzone G.-A., Del Rio Vilas D., Longo F., Merkurjev Y., Piera M.-A., eds. Proceedings of 17th International Conference of Harbour, Maritime & Multimodal Logistics Modelling and Simulation (HMS 2015). ISBN 978-88-97999-49-2. pp. 95-103. 21-23 September, Bergaggi (Italy).
- Ginters E., Aizstrauts A., Baltruks M., Merkurjev Y., Novickis L., Grundspenkis J., Grabis J., 2016. VeloRouter - Technology for Urban Transport Intermodal Sustainability. City Planning and Urban Design Conference, 07-09 April 2016, DAKAM, Istanbul (Turkey).
- Nazemi K., Breyer M., Burkhardt D., Stab C., Kohlhammer J., 2014. SemaVis - A New Approach for Visualizing Semantic Information. In: *Towards the Internet of Services: The Theseus Program*, Springer.
- Ginters E., Aizstrauts A., Baltruks M., Nazemi K., Burkhardt D., Sonntagbauer P., Sonntagbauer S., Gutierrez JM., 2014. FUPOL Simulators and Advanced Visualization Framework Integration. In: Proceedings of 26th European Modelling & Simulation Symposium (EMSS 2014). ISBN 978-88-97999-38-6 / EMSS 2014 (paperback) ISBN 978-88-97999-32-4. Bordeaux, France, 523-530.

AUTHORS BIOGRAPHY

Egils Ginters is full time Professor at the Faculty of Computer Science and Information Technology of Riga Technical University. He is Senior member of IEEE, member of European Social Simulation Association (ESSA) and Latvian Simulation Society. He participated and/or coordinated EC funded research and academic projects: FP7 FUPOL project No. 287119 (2011-2015), FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2014) and other. He has more than 165 scientific articles related with the research fields.

Mikelis Baltruks has master degree on Sociotechnical Systems Modelling and holds a bachelors degree in Information technologies. He is developer at Sociotechnical Systems Engineering institute with interests about simulation modeling. participated in FP7 FUPOL project No. 287119 (2011-2015). He has also more than one year experience at international student organization AIESEC.

Inita Sakne is technical assistant at Sociotechnical Systems Engineering institute working in different ERDF and FP7 projects.

Yuri Merkurjev is Professor, Habilitated Doctor of Engineering, Head of the Department of Modelling and Simulation at Riga Technical University. His professional interests include modeling and simulation of complex systems, discrete-event systems simulation methodology, supply chain simulation and management, as well as education in the areas of simulation and logistics management. Prof. Merkurjev is well experienced in performing research projects at both international and national levels. In particular, he has served as RTU coordinator of the 6th FWP Specific targeted research project NMP-032378 ECLIPS „Extended Collaborative Integrated Life Cycle Supply Chain Planning System”, as well as overall coordinator of the research project 2.1/ELRI-184/2011/14 "Integrated Intelligent Platform for Monitoring the Cross-Border Natural-Technological Systems" (under the Estonia-Latvia-Russia cross-border cooperation Programme within European Neighborhood and Partnership instrument 2007-2013). He authored more than 330 scientific publications, including 7 books and 6 textbooks.

DEVELOPMENT AND SIMULATION-BASED EVALUATION OF AN ALGORITHM FOR THE RETRIEVAL-IN-SEQUENCE FOR SHUTTLE SYSTEMS

T. Lienert^(a), W. A. Günthner^(b)

^{(a),(b)}Institute for Materials Handling, Material Flow, Logistics – Technical University of Munich

^(a)lienert@fml.mw.tum.de, ^(b)kontakt@fml.mw.tum.de

ABSTRACT

Autonomous vehicle-based storage and retrieval systems are used in order to supply picking or production areas based on the goods-to-person principle. In these contexts, it is often required that the point-of-use has to be supplied in a specific sequence of the retrieval loads. This sequence is normally established in the pre-storage area. Depending on the system configuration of an autonomous vehicle-based storage and retrieval system, it is possible to establish the required sequence at every input/output location directly out of the system. In order to enable this retrieval-in-sequence, we present an efficient algorithm that is based on the time-window routing method. For the purpose of evaluation of the performance, we implemented the algorithm in a simulation environment and conducted a series of simulation experiments.

Keywords: autonomous vehicle-based storage and retrieval systems, shuttle systems, time-window routing, sequencing

1. INTRODUCTION

In addition to ordinary automated storage and retrieval systems (AS/RS), a new technology has been developed in the past few years which is based on autonomous vehicles. Autonomous vehicle-based storage and retrieval systems (AVS/RS), also known as shuttle systems, are used for storing small unit loads, as well as pallets, in order to supply picking or production areas based on the goods-to-person principle and to store articles – both for dynamic buffers and for low-access applications (VDI- Richtlinie 2692).

Shuttle systems are characterised by horizontally-operating vehicles. These vehicles travel on each tier along a rail system within the storage rack. For vertical movements, storage and retrieval transactions, the vehicles may use lifts that are positioned at fixed locations along the periphery of the storage rack system (Malmberg 2002).

Typical features of shuttle systems, as compared to conventional stacker-crane-based AS/RS, are a higher

performance and a better scalability. The main disadvantages are higher investment and an increasingly complex storage management (Kartnig et al. 2012). Depending on specific scenarios, storage units might have to be provided in a certain sequence at the point-of-use. This requirement may, for example, arise in the following settings:

- Supply of picking areas in sequence of customer orders
- Supply of gates in sequence of the delivering order of the trucks
- Supply of production areas in the production sequence

Geinitz considered a stacker-crane-based AS/RS and described the effect that retrievals in a required sequence lead to a loss of throughput (Geinitz 1998). Due to this loss of throughput, the required sequence is normally established in the pre-storage area; this consumes space and requires additional material handling systems.

Concerning shuttle systems, neither an algorithm for the retrieval-in-sequence, nor the effect of sequencing, have been described until now. In this paper, we will fill this gap. We first specify the shuttle system configuration that allows sequencing within the storage system. Subsequently, we present an efficient routing-based sequencing algorithm and, finally, we investigate the loss of performance, measured by the throughput, by performing a simulation study.

2. SCOPE OF THE PAPER

This chapter provides an overview of different shuttle system configurations and specifies the configuration that we consider in this paper. Furthermore, we give a literature review about the research done so far which deals with shuttle systems.

2.1. Shuttle System Configurations

In the course of recent developments, different system configurations have evolved. In order to categorise the system configurations, we introduce movement axes that describe the movement space of the vehicles. The x-axis corresponds to the storage aisles. In every system configuration, the vehicles move along the x-axis in order to execute the storage and retrieval requests.

Based on the format of vehicle assignment to storage tiers, Heragu et al. distinguish two different configurations: Shuttle systems that use tier-to-tier vehicles and systems that use tier-captive vehicles. In the tier-to-tier system, the vehicles may move from one tier to another tier using a lift for the vertical movement along the y-axis. In the tier-captive system, each vehicle is dedicated to a single tier and therefore cannot move to another tier. Lifts are used only to move the unit loads to the destination tier (Heragu et al. 2011). We extend that distinction by considering the aisles to which shuttle vehicles are dedicated as well.

In the aisle-to-aisle configuration, there are cross-aisles integrated into the storage rack. These aisles are orthogonally-positioned to the storage racks and correspond to the z-axis. As a consequence, a vehicle can travel from one aisle to another aisle on the same tier. It can reach every position on that tier. In contrast, in the aisle-captive configuration, vehicles are firmly assigned to aisles and movements along the z-axis are not allowed. Figure 1 provides an overview of the four different configurations that result from the different movement spaces of the vehicles.

Autonomous vehicle-based storage and retrieval system				
Aisles-related movements	vehicles are dedicated to a single storage aisle	vehicles can use cross-aisles to change the storage aisle		
Tiers-related movements	vehicles operate on single tier	vehicles can use lifts for horizontal movements and change the tier		
Configurations	aisle - captive tier-captive	aisle-captive tier-to-tier	aisle-to-aisle tier-captive	aisle-to-aisle tier-to-tier
Movement space	x-axis	x-axis, y-axis	a-axis, z-axis	x-axis, y-axis, z-axis

Figure 1: Shuttle System Configurations

A retrieval-in-sequence directly out of the system is possible, only if every storage unit can be provided at every input/output-location (I/O-location). Given the assumption that the I/O-locations are connected to the storage systems by the lifts, the precondition for the retrieval-in-sequence holds for both aisle-to-aisle and tier-to-tier configuration as it is the more generic one and developed algorithms can be simplified to match the aisle-to-aisle and tier-captive configuration.

Further advantages of the considered configuration are a simple scalability and a good redundancy. As every shuttle can reach every single position within the system, it is possible to run the whole system with a single shuttle. If needed, more and more shuttles can be added to achieve a higher performance. Should a single shuttle fail, depending on the specific layout, it might

nevertheless still be possible to reach every single position within the system.

In more detail, the system configuration, which we are considering, can be described as follows.

- Shuttles travel along the x-axis in order to perform storage and retrieval transactions.
- Shuttles are able to change storage aisles by using cross-aisles along the z-axis.
- Shuttles can operate only on the tier they are currently moving on, as they do not have a lifting unit at their disposal.
- Shuttles have unit load capacity.
- The storage rack is single-deep.
- Shuttles are able to change the tier by using a shuttle lift for vertical movements along the y-axis.
- Lifts have single capacity.
- Shuttles do not leave the storage system. On the input/output-level, shuttles remain in the lift while the handover of storage units takes place.
- Every point of use is supplied by a single lift.

The developed algorithm is generic and does not depend on a specific layout. The layout of the storage system may vary, e.g. the number of tiers, the number of aisles and cross-aisles per tier and the number and positions of the lifts. Figure 2 shows an example of the considered configuration.

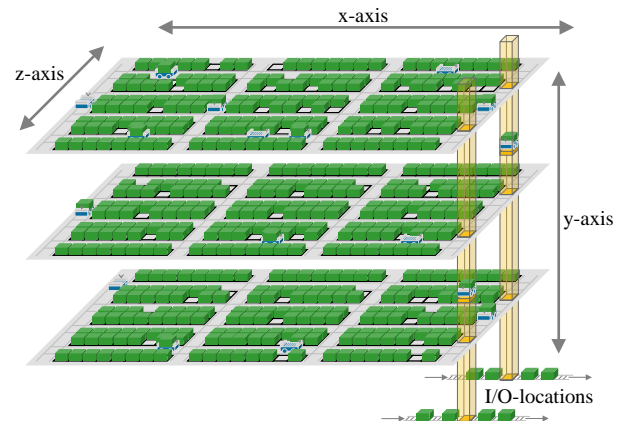


Figure 2: System Example

The most important disadvantage of the considered configuration lies in the complex control strategies that are necessary to run the system in a robust and efficient way. As every single shuttle can reach every single storage position within the system, the question has to be answered as to which shuttle executes which storage or retrieval request. On the other hand, different routing options might exist for reaching a desired position. It therefore has to be decided which path should be taken. Finally, the shuttles share the same rail system, so collisions need to be avoided, as do deadlocks among the shuttles – situations where the shuttles block each other – have to be dealt with.

2.2. Literature Review

Research concerning AVS/RS can be divided into two categories: performance analysis and development of control algorithms. Most of the research papers investigate the performance of the considered system against different parameters, like the storage capacity, rack configuration and numbers of vehicles and lifts. In some papers, various controlling strategies have been developed and tested against each other. Basically, in the literature, two different approaches to performing the system evaluation can be found. Analytical models have been developed, generally based on queuing networks, to investigate the system behaviour. The second approach is based on simulation. In order to identify the impact of different system parameters and controlling strategies, simulation studies were conducted. Malmborg first proposed the idea of an analytical conceptualising tool for AVS/RS. He compared AVS/RS and AS/RS technologies by varying the system configuration (Malmborg 2002). Marchet et al. developed a framework for the design of AVS/RS. The authors considered a tier-captive and aisles-captive configuration and included costs in their model (Marchet et al. 2013). Ekren et al. investigated the effect of several design factors on the performance of a tier-to-tier and aisle-to-aisle AVS/RS. The authors vary the dwell point location, the I/O-location and use basic scheduling rules for both single and dual command scheduling (Ekren et al. 2010). Recent work is more focused on aisle-captive shuttle-systems.

The VDI-Guideline 2962 presents a framework for performance calculation, both for tier-to-tier and tier-captive configurations, in order to estimate possible throughputs achievable with an AVS/RS. The performance calculation is based on the separate determination of the mean cycle times of the lift and of the shuttle. Waiting times of the shuttles for the lift and different controlling strategies are not considered (VDI-Richtlinie 2692). Eder and Kartnig provided an analytical model in order to identify the ideal rack geometry depending on the storage capacity (Eder and Kartnig 2015). Lerher developed an analytical travel-time model for the computation of cycle times for the double-deep storage rack system (Lerher 2015). Carlo and Vis considered an AVS/RS system in which two non-passing lifts share a single mast to transport loads from the horizontally-operating shuttles to the I/O location and vice-versa. The study deals with the scheduling problem of these two lifts, i.e. which lift is going to handle which request and in which order (Carlo and Vis 2012).

Research can be found within the area of automated guided vehicle systems concerning the issue of deadlock-handling that arises in aisle-to-aisle configurations. Kim et al. define a system deadlock as a situation where one or more concurrent processes in a system are blocked forever because the requests for resources by the processes can never be satisfied (Kim et al. 1997). Three different approaches can be distinguished in deadlock handling. Deadlock-

prevention is a static approach. A set of generic rules ensures that a deadlock could never occur. This leads to poor resource utilisation and a low performance. In the detection and recovery approach, deadlocks are allowed to occur. They have to be detected and the system uses mechanisms for recovery. As some deadlocks might be hard to discover and, furthermore, deadlocks can overlap each other, detection and recovery might end in an inefficient way. The most frequently used approach is deadlock-avoidance. By using an online control policy, the resources will be dynamically allocated so that a deadlock will never occur (Liu and Hung 2001). Penners considered a simplified isolated tier of an aisle-to-aisle system (Penners 2015). He adapted two deadlock-avoiding routing-algorithms that have been developed for automated guided vehicles and compared the performance by conducting a simulation study. He came to the conclusion that the time-window routing method, presented by ter Mors et al. (ter Mors et al. 2007), achieves a considerably higher throughput than the modified Banker's routing, which was described by Kalinovic et al. (Kalinovic et al. 2011).

In the literature on AVS/RS, it has not been possible to find a paper that deals with sequencing the retrieval loads within a tier-to-tier and aisle-to-aisle configuration. We will therefore present an algorithm that is based on the time-window-routing method and make use of it for the retrieval-in-sequence. As an analytic evaluation of the algorithm is hard due to the complexity of the sequencing problem, we will follow the simulation-based approach to evaluate the performance, which is widespread within the area of warehouse design and management (Roodbergen et al. 2015), (Curico and Longo 2009).

3. SEQUENCING ALGORITHM

In this section, we briefly describe the system requirements and the options we have for establishing the sequence. We briefly introduce the underlying routing algorithm and finally describe the routing-based sequencing algorithm.

3.1. System Requirements

We follow the assumption that the required sequence of the storage units has to be established at the I/O-location; this means that the final sequence must be made by the lifts. As every point-of-use is supplied by a single lift, we maintain a separate sequence for every lift. As a consequence, we have different sequences within the system that are independent from each other.

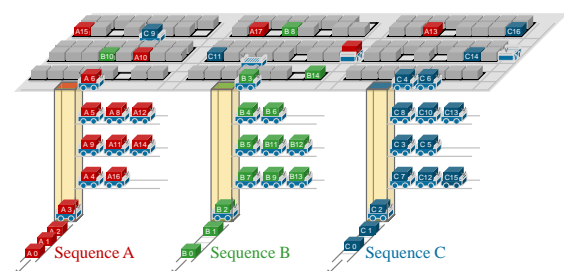


Figure 3: Example with three Independent Sequences

Figure 3 shows a system with three lifts and therefore three independent sequences A, B and C. Within a sequence, the unit loads are marked by increasing sequence numbers. These sequence numbers are unique within a sequence. As the final sequence is established by the lift, the control strategy for the lifts is obvious and simple. For every lift, we will maintain a counter which stores the current sequence number. After having finished a transportation task, the lift will search for the next following sequence number, travel to the corresponding tier, transfer the waiting shuttle to the I/O-location and update the counter. In order to enable this procedure, we have to ensure that on every tier, the shuttles are waiting in an increasing sequence number for the lift; this is the more sophisticated task.

3.2. Sequencing Options

In order to establish the sequence on a single tier, there are different options to be considered:

- Sequencing by dispatching
- Sequencing by clearance
- Sequencing by routing

Dispatching refers to a rule used to select a shuttle to execute a storage or a retrieval request. Egbelu and Tanchoco characterise two different dispatching concepts: tasks-initiated dispatching rules and shuttle-initiated dispatching rules (Egbelu and Tanchoco 1984). In our system, we will use a vehicle-initiated dispatching rule. Whenever a vehicle completes a retrieval request, it will choose the next retrieval request from the set of available requests. The shuttle will choose the retrieval request according to the FIFO rule, this means it will choose a request with the lowest sequence number of one of the independent sequences. By so doing, we pre-sequence the retrieval requests (sequencing by dispatching). Nevertheless, we cannot ensure that the shuttles will arrive in ascending sequence number at the lift. A shuttle to which a retrieval request was assigned with a higher sequence number might be faster in executing the retrieval.

In order to ensure the correct sequence at the lift, we can sequence the shuttles by clearance. In this case, a shuttle waits at its position, after having picked up the retrieval load, until it is cleared by the predecessor on that tier. In figure 4, the shuttle with the sequence number 1 clears the shuttle with the sequence number 2 as soon as it arrives at the lift. We ensure the sequence at the lift, but, depending on the size of a tier, there might be a huge loss of performance, because the second shuttle might take a certain time to travel to the lift. Desirable would be for the shuttle with the sequence number 2 to arrive just after the shuttle with the sequence number 1 had arrived at the lift. This can be achieved by the sequencing by routing, upon which we will focus from now on. The basic idea could be summarised as follows: instead of routing the shuttles from the retrieval location to the lift, we route the

shuttles backwards from the lift to the retrieval locations, ensuring the desired sequence at the lift.

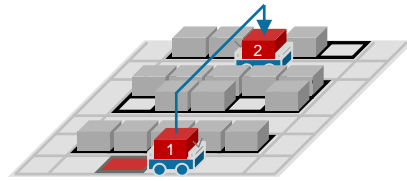


Figure 4: Sequencing by Clearance

3.3. Underlying Routing Algorithm

In order to avoid deadlocks among shuttles moving on the same tier, we make use of the time-window routing method. The basic idea of the algorithm was invented by Kim and Tanchoco, who developed a conflict-free and shortest-time algorithm for routing automated guided vehicles in a bidirectional path network that is based on Dijkstra's shortest path algorithm. The idea of their algorithm consists of modelling the flow path as a graph. Every layout segment corresponds to a single node within the graph (see Figure 5). For each node, the algorithm maintains a list of time-windows reserved by routed vehicles and a list of free time-windows available in which vehicles could be routed. Every free time-window corresponds to a node in the so-called time-window graph. The arcs between these nodes represents the reachability among the free time-windows. The algorithm then routes vehicles through the nodes of the time-window graph instead of the physical nodes of the path network (Kim and Tanchoco 1991). Ter Mors et al. presented an improved version of the time-window routing which is based on the A*-algorithm; their version provides a better worst-case performance and calculates a solution in real-time (ter Mors et al. 2007).

We make use of their time-window routing method in a slightly different way. Instead of constructing the whole time-window graph, we investigate, in every iteration of the routing process, every free time-window on every neighbour node, if that free time-window is reachable from the current time-window. Some conditions must hold for this, e.g. a minimal length or a minimal overlapping of the free time-windows.

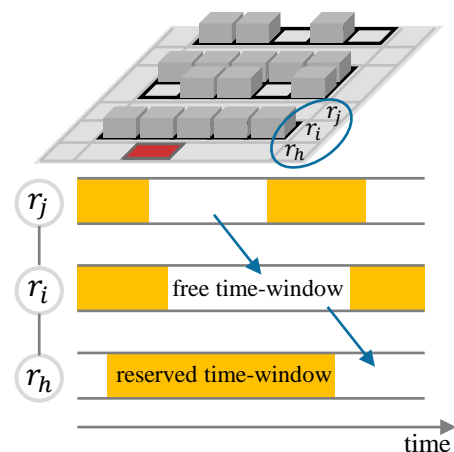


Figure 5: Concept of the Time-Window Routing

The algorithm finds the fastest path for the vehicle from the start node to the destination node at the specified start time under the given reserved time-windows for other vehicles, if such a path exists. The output of the algorithm consists of the sequence of nodes along the path which have to be visited in order to reach the destination and the time intervals, the shuttle will occupy these nodes. The corresponding time-windows will be reserved and the vehicle can travel deadlock-free through the layout.

One might object in that, should a shuttle be delayed, the routing is then no longer robust and deadlocks might occur. However, as Maza and Castagna proved, if the node's crossing order of the shuttles, based on the conflict-free scheduled dates, is fulfilled, then the absence of conflict is guaranteed even if the arrival times are not (Maza and Castagna 2005). In summary, using the time-window method, we do know when a shuttle is likely to arrive at the lift but, furthermore, we also know the sequence in which shuttles arrive at the lift. If the sequence is not correct, we could intervene.

3.4. Routing-Based Sequence-Algorithm

Whenever a shuttle selects the next retrieval request after having finished the current retrieval request, it will travel to the corresponding tier by lift. As soon as the shuttle arrives at that tier, we will run the routing-based sequence algorithm, shown in figure 6.

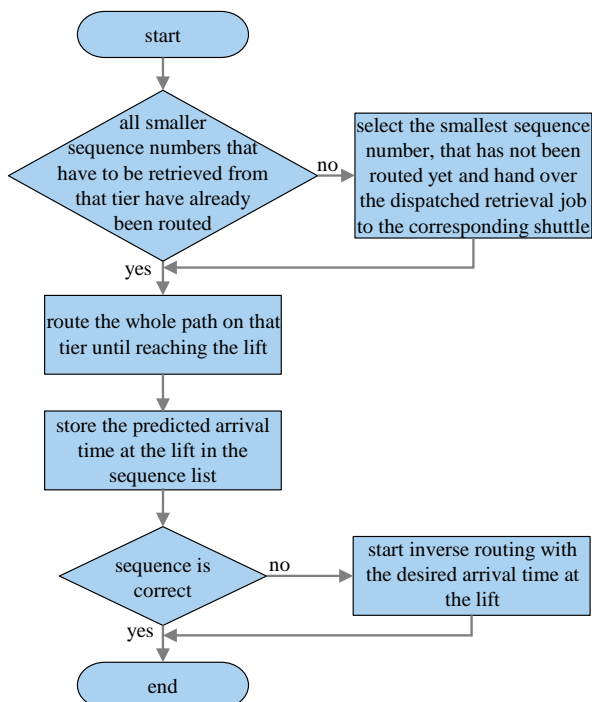


Figure 6: Routing-Based Sequencing Algorithm

Firstly, we will check if all the smaller sequence numbers that have to be retrieved from that tier have already been routed; this would mean that the assigned shuttles are already travelling on that tier and performing the actions. Note that, due to the described dispatching rule, all the smaller sequence numbers must

have already been assigned to a shuttle. If all the smaller sequence numbers have not yet been routed, the shuttle would switch its assigned retrieval task and would retrieve the load with the smallest sequence number which has not yet been routed, but is already assigned to another shuttle. As a consequence, this shuttle would retrieve the released sequence number.

Subsequently, the whole path will be routed on that tier; this means, in case of a double-cycle, the path from the entering lift to the storage location, the path from the storage location to the retrieval location and, finally, the path from the retrieval location to the outgoing lift. After having finished the routing, the computed arrival time will be stored in the sequence list and the sequence can be checked. Should the sequence be incorrect, we start the inverse routing from the lift to the retrieval location. Instead of the start time, the inverse routing algorithm requires the desired arrival time at the destination node. The arrival time has to be chosen in such a way that the shuttle might arrive right after the predecessor at the lift.

In the example of figure 7, there are currently three sequence numbers that have to be retrieved from the considered tier. A shuttle travels on that tier and stores the blue box at the designated location. Afterwards, it moves to the storage location of the sequence number 1, picks it up and travels to the lift. The estimated arrival time at the lift is stored in the corresponding sequence list.

Table 1: Sequence List

Sequence No.	Arrival Time
1	38
2	
3	

Now that a second shuttle has arrived at the tier that has been assigned to retrieve the sequence number 2, we run the algorithm. Firstly, it can be seen that all the smaller sequence numbers have already been routed by searching for the arrival time in the sequence list. Then, the path to the storage location of the green box, the path from the storage location to the retrieval location of sequence number 2 and, finally, the path to the lift will be completely routed. According to the routing, the shuttle will arrive at the lift at $T_{arrival(lift)} = 30s$. This information will be stored in the sequence list.

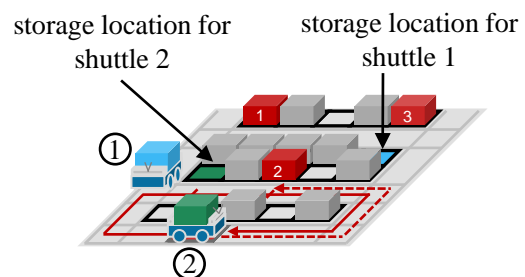


Figure 7: Inverse Routing from the Lift to the Retrieval Location

Table 2: Updated Sequence List

Sequence No.	Arrival Time
1	38
2	30
3	

As can be seen, sequence number 2 will arrive at the lift before sequence number 1. As a consequence, we start the inverse routing with the desired arrival time $T_{arrival(lift)} = 42s$. Assuming a shuttle needs four seconds for entering an intersection, changing the moving direction by 90 degrees, and leaving the intersection entirely, this is the earliest possible arrival time at the lift for the second shuttle after the arrival of the first shuttle.

Only the last segment from the retrieval location to the lift will be rerouted, marked by the dotted line in figure 7. As the shuttle will now arrive later than originally planned, it will wait for a certain time at the retrieval location after having picked up the load.

We follow the idea of the time-window routing, but instead of routing the shuttle ahead, we let the shuttle route backwards. We therefore have to modify the achievable conditions to decide whether a free time-window on neighbour node is reachable from the current free time-window or not. A free time-window is defined by its start and its end time.

- $Start_{i,k}$: Start time of the k-th free time-window on the node r_i
- $End_{i,k}$: End time of the k-th free time-window on the node r_i

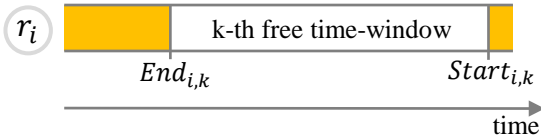


Figure 8: Free Time-Window

Note, that as we are routing backwards, it counts as $Start_{i,k} > End_{i,k}$.

We define the following time-stamps and time durations which describe the movement of a shuttle through the nodes of the layout-graph.

- $T_{entry(r_i)}$: Time-stamp, when the shuttle enters the node r_i
- $T_{arrival(r_i)}$: Time-stamp, when the shuttle resides completely on the node r_i and has completely left the previous node
- $T_{exit(r_i)}$: Time-stamp, when the shuttle has left the node r_i completely

- t_{trans} : Time a shuttle needs to enter or exit a node. More precisely, for the time t_{trans} the shuttle occupies two subsequent nodes
- $t_{cross(r_i)}$: Time a shuttle needs to cross the node r_i , should the node's length exceed the length of a shuttle
- $t_{wait(r_i)}$: Waiting time on the node r_i

The following figure illustrates the relations between the defined variables:

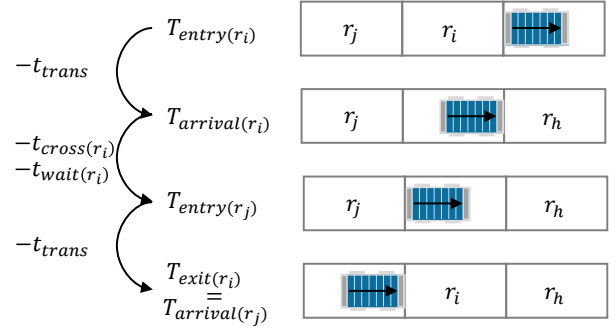


Figure 9: Time-Stamped and Time Durations Describing the Shuttle Movement.

To decide whether a free time-window on a neighbour node is reachable from the current time-window, there are some conditions that must hold:

$$Start_{j,l} - End_{j,l} \geq t_{trans} + t_{cross(r_j)} + t_{trans} \quad (1)$$

$$Start_{j,l} > End_{i,k} \quad (2)$$

$$Start_{i,k} > End_{j,l} \quad (3)$$

The first condition states that the free time-window on the neighbour node has the minimal size; this means that the shuttle might have time to enter the node, cross it and leave it within the free time-window. Condition (2) and (3) ensure that the two time-windows overlap. If, until now, all the conditions have held, the entry time into the free time-window and thereby into the node is calculated:

$$T_{entry(r_j)} = \min\{Start_{j,l}, T_{arrival(r_i)} - t_{cross(r_i)}\} \quad (4)$$

The entry into the free time-window cannot occur before the free time-window starts and, of course, not before the shuttle has crossed the current node in its entirety. If the entry time corresponds to the start time of the free time-window, the shuttle has to wait on the current node. After having calculated the entry time, the last conditions can be verified:

$$T_{entry(r_j)} - t_{trans}(r_i) = T_{exit(r_i)} \geq End_{i,k} \quad (5)$$

$$T_{entry(r_j)} - End_{j,l} \geq t_{trans} + t_{cross(r_j)} + t_{trans} \quad (6)$$

Condition (5) ensures that the shuttle might leave the current node before the time-window ends, and condition (6) ensures that the remaining size of the free time-window on the neighbour node is sufficient to enter the node, cross and leave it again. If all these conditions also hold, the free time-window can be reached from the current time-window.

In the example shown in figure 10, the free time-windows on both nodes overlap. The entry time corresponds to the arrival time on the node r_i minus the time it takes to cross this node. Furthermore, the shuttle can leave the node r_i before the free-time window ends as $T_{exit}(r_i) \geq End_{i,k}$. Finally, the remaining size of the free time-window on node r_j is sufficient to cross it and leave it, which is marked by the hatched elements. As a result, the free time-window on node r_j is reachable from the current time-window on node r_i with the given entry time $T_{entry}(r_i)$ into that node.

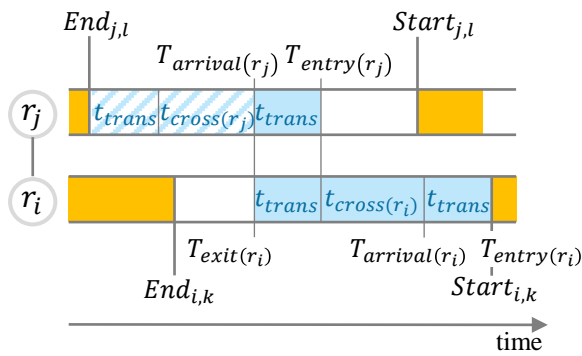


Figure 10: From the Current Time-Window on Node r_i the Free Time-Window on Node r_j is Reachable.

There are two cases which cause difficulties when applying the inverse routing. Firstly, it might happen that there is no route found which ends at the desired arrival time. Applying the time-window routing algorithm, the shuttle can delay its departure on the start node until a route is found. However, as we are routing backwards, we have to delay the arrival time manually and start the inverse routing again.

The second problem might arise under certain traffic conditions within the system. The sequence at the lift is incorrect and an inverse routing is required, given the desired arrival time at the lift. The inverse routing leads to an earlier departure time at the retrieval location than was originally predicted, by routing the path from the storage to the retrieval location on the tier. This is, of course, not valid as the shuttle could not start travelling to the lift before it had finished loading the retrieval load. We have to delay the desired arrival time for the inverse routing again until a valid departure time is found. Figure 11 clarifies the described problem.

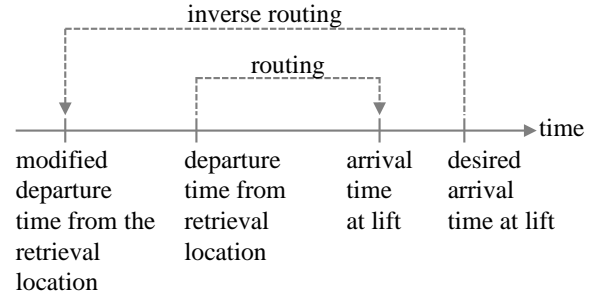


Figure 11: Invalid Inverse Routing

In the subsequently-described simulation, we delayed the arrival time in both cases in steps of one second. Once the route is found, the time-windows will be reserved. The shuttle can leave the lift and start travelling on the tier. So, the requested sequence at the lift will be guaranteed.

4. SIMULATION-BASED EVALUATION

In order to test the sequencing algorithm and to quantify the loss of throughput by the sequencing, we implemented the algorithm in a simulation environment and conducted a series of simulation experiments. The modelling and implementation is briefly described before the experiment settings are clarified and the results presented and discussed.

4.1. Modelling and Implementation

We implemented the simulation model using the discrete-event simulation software, Plant Simulation. This software offers various customisable modules, e.g. roads, warehouses, working stations and vehicles. The control of information and material flows is realised by procedures programmed by the user and assigned to a specific event.

In order to model the storage system, we divided the layout of a single tier into various resources. A resource is a layout segment with capacity for exactly one shuttle. Every resource corresponds to a node within the graph that is used for the time-window-based routing. Resources are storage elements, cross-aisles segments and intersections. In order to avoid having to control the driving direction of the vehicles, we allowed the vehicles to only travel in one direction. We modelled the different directions by two opponent path segments which are part of the same resource.

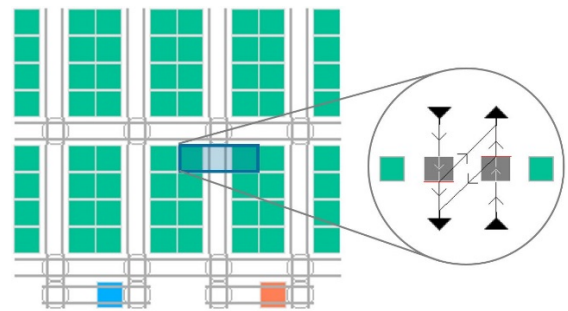


Figure 12: Modelling Concept

Figure 12 shows a single tier and a storage element resource. As time-windows are maintained for every resource, it is ensured that only one shuttle can occupy a resource at a time, even though it consists of several path segments. The routing and sequencing algorithm is implemented in a central controlling module.

The model allows for the varying of a number of parameters, such as the number of shuttles or the used dispatching strategy. The layout of the considered storage system can be generated automatically or individually constructed. As the shuttles can only reside on a resource they have reserved in advance, the question arises as to how to initialise a simulation run. We solved this problem by introducing a virtual buffer for every lift within the system. At the beginning, the shuttles were distributed equally to these virtual buffers. From these buffers, the shuttles requested the lift one after another and were inserted into the storage system.

4.2. Simulation Experiments

In order to test the sequencing algorithm, we built a model consisting of four tiers with 320 storage locations each. We integrated four lifts, which are supplied by the shuttles, to obtain more than one independent sequence. Basically, the sequencing algorithm runs on a single tier, but in order to also respect the influences of the lifts, we extended the system to three more tiers. Variations of the layout were not part of the simulation study and should be considered in further research.

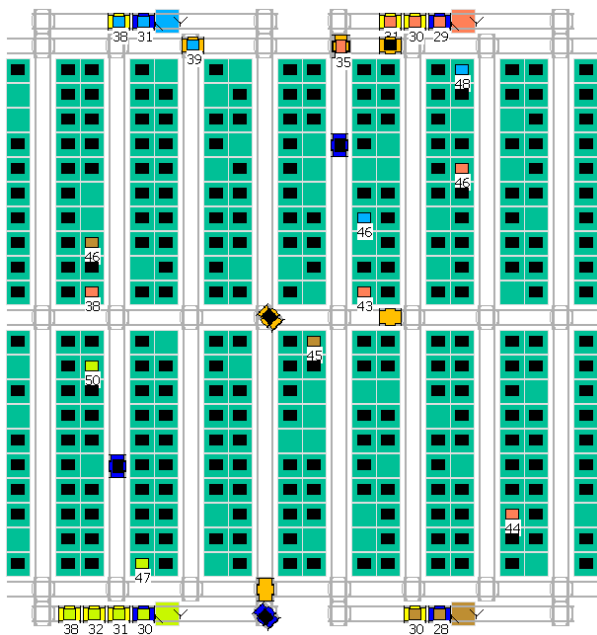


Figure 13: A Single Tier of the Simulation Model

The shuttles work constantly in double-cycles. Every time a retrieval request is fulfilled, the shuttle selects the retrieval job with the oldest time-stamp (and with the smallest sequence number) from the available retrieval requests. Furthermore, a storage load is handed over to the shuttle. The storage location is chosen randomly on the same tier upon which the assigned retrieval load is

situated. By so doing, additional travel with the lifts is avoided and the storage ratio on every tier is kept constant. We generated a new retrieval task for one of the sequences every time a retrieval of that sequence was completed. The retrieval location was chosen randomly.

Figure 13 shows the layout of a single tier from the simulated system. Shuttles are moving on that tier and performing retrieval and storage transactions. As soon as the storage loads are requested, they are coloured according to the lift that connects the storage system with the corresponding I/O-location. Furthermore, the sequence numbers are assigned. As can be seen, the shuttles wait for the lift in increasing sequence number of the retrieval loads. Numbers that do not appear have to be retrieved from a different tier than the one that is shown. We varied the numbers of shuttles working in the system from 5 to 65 in steps of 5. We ran the simulation twice with every number of shuttles. Firstly, the retrieval-in-sequence was required and then it was not. We conducted 5000 double-cycles and measured the time needed to fulfil these cycles. As the storage and retrieval locations were randomly determined, we did three replications per experiment and calculated the mean time for the evaluation.

4.3. Results and Discussion

The developed algorithm for the retrieval-in-sequence was validated by conducting the simulation experiments. At every I/O-location, the sequence numbers of the retrieval units were stored and the correctness of the sequence was verified.

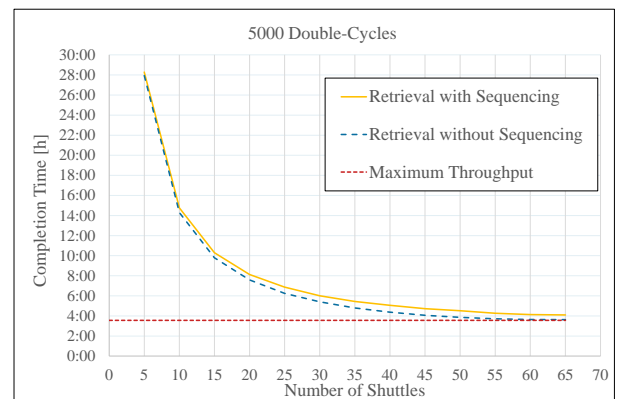


Figure 14: Results of the Simulation Experiments

Figure 14 shows the evolution of the completion time for 5000 double-cycles as a function of the numbers of shuttles working in the system, both for the retrieval with sequencing and retrieval without sequencing. As expected, the completion time lessens with the number of shuttles working in the system, and the retrieval with sequencing requires a higher completion time than the retrieval without.

The loss of performance caused by the retrieval with sequencing for the different numbers of shuttles is presented by figure 15. As can be seen, the loss of throughput is small when there are only a few shuttles

in the system. With only a few shuttles, the interference among these is relatively small. In other words, the probability that two shuttles execute a retrieval request on the same tier for the same sequence is small.

The loss of throughput rises to 14.60% with 50 shuttles working in the system. With a higher number of shuttles, the loss of throughput lowers slightly again.

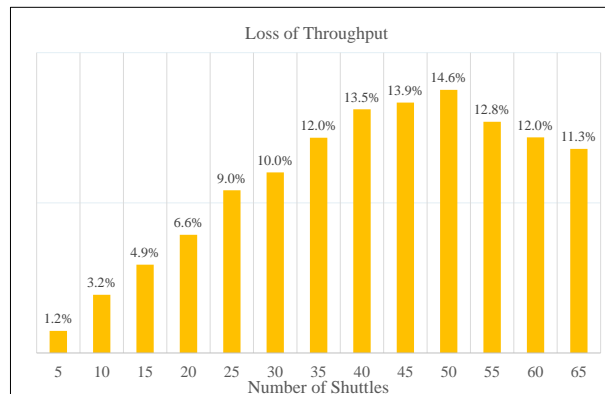


Figure 15: Loss of Throughput caused by the Retrieval in Sequence

The maximum system throughput is limited by the lifts. If the lifts work continuously, the capacity cannot be augmented by adding more shuttles. Using a chaotic storage assignment strategy, every tier is piloted by the shuttles with the same frequency. The mean time of a double-cycle of a lift can be calculated using a combinatorial approach, similar to the one presented in the VDI-Guideline (VDI- Richtlinie 2692). The result of this maximum throughput calculation is presented by the dotted line in figure 14. At a certain number of shuttles, the throughput cannot be further significantly augmented. This number is reached earlier if the retrieval in sequence is not required which explains the decreasing loss of throughput for a high number of shuttles.

From the chart in figure 14, information can also be obtained as to how many more shuttles are needed to compensate for the loss of throughput caused by the retrieval-in-sequence. For instance, if there are currently 30 shuttles in the system, approximately five more shuttles are needed to reach the same throughput with the retrieval-in-sequence.

5. SUMMARY

In this paper, we considered an autonomous vehicle-based storage and retrieval system, where the vehicles move along the x-axis and the z-axis on a tier and use lifts to change the tier along the y-axis. We described an algorithm for the retrieval-in-sequence that is based on the time-window routing method and which ensures the absence of deadlocks. The basic idea of the sequencing algorithm consists of an inverse routing that is applied if the required sequence is not correct. We modelled and implemented the storage system in a simulation environment in order to test the algorithm and to

quantify the loss of throughput by the retrieval-in-sequence.

An interesting topic for future work would be the determination of the number of shuttles needed to achieve the maximum system throughput. This number clearly depends on the number of lifts, the layout and the specific parameters of the lifts and vehicles. Furthermore, different dispatching strategies, such as task-initiated dispatching rules and different layout options, should be investigated and analysed.

REFERENCES

- Carlo H. J., Vis I. F. A., 2012. Sequencing dynamic storage systems with multiple lifts and shuttles. *International Journal of Production Economics* 140: pp. 844-853.
- Curico, D., Longo, F., 2009. Inventory and internal logistics management as critical factors affecting the Supply Chain performances. *International Journal of Simulation and Process Modelling. International Journal of Simulation and Process Modelling* 5: pp. 278-288
- Eder M., Kartnig G., 2015. Throughput analysis of S/R shuttle systems and ideal geometry for high performance. *Proceedings of the XXI International Conference MHCL*, pp. 193-198. September 23-25, Vienna (Austria).
- Egbelu P. J., Tanchoco J. M. A., 1984. Characterization of automatic guided vehicle dispatching rules. *International Journal of Production Research* 22: pp. 359-374.
- Ekren B. Y., Heragu S. S., Krishnamurthy A., Malmberg C. J., 2010. Simulation based experimental design to identify factors affecting performance of AVS/RS. *Computers & Industrial Engineering* 58: pp. 175-185.
- Geinitz, J., 1998. Unerkannte Abhängigkeiten mindern die Leistungsfähigkeit automatisierter Lager. *Marktbild Lager* 18: pp. 16-18.
- Heragu S. S., Cai X., Krishnamurthy A., Malmberg C. J., 2011. Analytical models for analysis of automated warehouse material handling systems. *International Journal of Production Research* 49: pp. 6833-6861.
- Kalinovic L., Petrovic T., Bogdan S., Bobanac V., 2011. Modified banker's algorithm for scheduling in multi-agv systems. *Automation Science and Engineering (CASE)*, pp. 351-356. August 24-27, Trieste (Italy).
- Kartnig G., Grösel B., Zrníc N., 2012. Past, State-of-the-Art and Future of Intralogistics in Relation to Megatrends. *FME Transactions* 40: pp. 193-200.
- Kim C. W., Tanchoco J. M. A., 1991. Conflict-free shortest-time bi-directional AGV routing. *International Journal of Production Research* 29: pp. 2377-2391.
- Kim C. W., Tanchoco J. M. A., Koo P., 1997. Deadlock Prevention in Manufacturing Systems with AGV Systems: Banker's Algorithm Approach. *Journal*

- of Manufacturing Science and Engineering 119: pp. 849-854.
- Lerher T., 2015. Travel-time model for double-deep shuttle-based storage and retrieval systems. *International Journal of Production Research*, pp. 1-22.
- Liu F., Hung P., 2010. Real-time deadlock-free control strategy for single multi-load automated guided vehicle on a job shop manufacturing system. *International Journal of Production Research* 39: pp. 1323-1342.
- Malmberg C. J., 2002. Conceptualizing tools for autonomous vehicle storage and retrieval systems. *International Journal of Production Research* 40: pp. 1807-1822.
- Marchet G., Melacini M., Perotti S., Tappia E., 2013. Development of a framework for the design of autonomous vehicle storage and retrieval systems. *International Journal of Production Research* 51 pp. 4365-4387.
- Maza S., Castagna P., 2005. A performance-based structural policy for conflict-free routing of bi-directional automated guided vehicles. *Computers in Industry* 56: pp. 719-733.
- Penners L. T. M. E., 2015. Investigating the effect of layout and routing strategy on the performance of the Adapto system. Master's Thesis. Eindhoven University of Technology.
- Roodbergen, K. J., Vis I. F. A., Taylor, G.D., 2015. Simultaneous determination of warehouse layout and control policies. *International Journal of Production Research* 53: pp. 3306-3326.
- ter Mors A. W., Zutt J., Witteveen C., 2007. Context-Aware Logistic Routing and Scheduling. *Proceedings of the Seventeenth International Conference on Automated Planning and Scheduling*, pp. 328-335. September 22-26, Providence (USA).
- VDI-Richtlinie 2692 Blatt 1, 2015. Automated vehicle storage and retrieval systems for small unit loads. Berlin: Beuth.

member of several associations, federations and companies.

AUTHORS' BIOGRAPHY

Thomas Lienert, M. Sc. was born in Hamburg, Germany, and studied industrial engineering and management at the Karlsruhe Institute of Technology and the University of Seville with the focus on operations research and logistics systems. He has been working as a research assistant at the Institute for Materials Handling, Material Flow and Logistics, Technical University of Munich since 2014. His research mainly deals with the development of control strategies for autonomous vehicle-based storage and retrieval systems.

Prof. Dr.-Ing. Dipl.-Wi.-Ing. Willibald A. Günthner is head of the Institute for Materials Handling, Material Flow and Logistics, Technical University of Munich. He is co-founder of the Wissenschaftliche Gesellschaft für Technische Logistic e.V. and advisory board

SCALABLE SIMULATION MODELS OF RAILWAY TRAFFIC

Radek Novotny^(a), Antonin Kavicka^(b)

^{(a),(b)}Faculty of Electrical Engineering and Informatics, University of Pardubice

^(a)radek.novotny3@student.upce.cz, ^(b)antonin.kavicka@upce.cz

ABSTRACT

This contribution deals with the methodology of building scalable simulation models reflecting railway traffic. Those models apply different level of abstraction (granularity) to diverse parts of a simulating system. The areas of a railway system that are supposed to be studied in detail are typically modelled on a microscopic level. The sections of infrastructure, which require just rough traffic evaluations, are investigated on a macroscopic level. The resulting methodology contains a variety of functionalities supporting the construction of variant configurations related to a scalable simulator of railway traffic for the needs of different simulation scenarios.

Keywords: scalable simulation model, hybrid model, railway traffic

1. INTRODUCTION

Modelling a railway infrastructure and corresponding railway traffic represents an important part of the research focused on railway system optimizations. For such purposes the researchers use the experimental method of computer simulation within the frame of which the applied level of *abstraction/granularity* plays an important role. The mentioned granularity defines the level of details which are observed in the simulating system. On the basis of the required granularity level different simulating systems are supposed to be utilized. Traffic simulations can be classified according to the applied level of details.

Generally, concerning transport simulations we can classify models according to their applied level of detail as *microscopic*, *mesoscopic*, and *macroscopic*. Microscopic simulation models are characterized by examination with great level of resolution. Individual mobile entities (rolling stock) and their interaction are observed. For these reasons it is necessary to include detailed information about particular sub-systems (*fixed subsystem* - typically consisting of infrastructure, *mobile subsystem* - consisting of mobile objects, and *management subsystem* - specifies technological processes) in the simulation model. Utilizing microscopic simulation models is suitable for examining small segments of railway (e.g. for an operational check of a railway station after its infrastructure or its graphical timetable of rail transport have been changed, or after the handling technology has been changed, etc.) mainly because of their high demands for detail and computational demand.

While the listed types of investigation are typical for their detailed examination of interactions of individual mobile entities, for evaluating, for example, track or station capacities, it is sufficient to examine only rough characteristics of transport flows (individual rolling stock is not examined). For the purposes of examination we advantageously use macroscopic simulation models, which enable to simplify individual subsystems and thus conduct simulation experiments even above extensive segments of a railway network, but all of that at the expense of the level of detail. However, there is a compromise in the form of mesoscopic simulation models which are trying to balance the level of detail and computational demand.

Considering the concrete aim of the examination (e.g. investigating the throughput or the influence of delay), it is important to select the level of detail for each individual subsystem of the simulating system. (Krivý and Kindler 2003; Burghout 2004).

Traditional approaches apply the same granularity level for the entire simulator. Such a homogenous approach does not enable to combine for example microscopic and macroscopic levels of details within one simulator. Designers of traffic simulations are motivated to use methodologies for building scalable traffic simulators in order to be able to combine and interconnect various submodels of infrastructure constructed on different levels of details (Hansen and Pachl 2008). Then, different traffic models are applied to the mentioned infrastructural submodels – certainly relevant transformations of traffic flows are supposed to be carried out on the boundary between corresponding submodels. The mentioned approach has to be supported by appropriate software tools (e.g. infrastructure editors and integrated simulation environments etc.).

2. CURRENT STATE

Typical contemporary solutions use the same granularity within the whole railway simulating system. Certain scalability can be enabled by applying different level of details for diverse parts of input data. For example, in the simulation tools *Villon* (Simcon 2014) and *OpenTrack* (Huerlimann and Nash 2010) a larger area of a railway network can be modelled. Border railway stations of that area are not supposed to be investigated on a detailed level because the primary attention is paid to central parts of the mentioned area. Thus, those border stations can be specified by means of

simplified processes and skeletonized layouts of tracks. However, the trains and their driving dynamics are modelled on a unified level of abstraction within all parts of the studied network.

There are various approaches of building scalable simulation models focused on railway traffic. The published concept of simulation model (Burghout 2004; Cui, and Martin 2011) features the option to change the level of detail dynamically during the simulation experiment aimed on the operation of the whole railway network (*scalable simulation model with continuous changeover*). Changing the level of detail from microscopic to macroscopic (*bottom-up*) or vice versa (*top-down*) can be applied by, for example, mere executed change of an appropriate graphic zoom. In the mentioned concept of the scalable model, we can use a uniform graphic visualization of selected rolling objects for various levels of zoom within the infrastructure – *Figure 1*). Concerning the level of detail, the interconnection of different infrastructure submodels can be achieved by model aggregation, in which each node or edge on a higher level can be considered a set of nodes or edges of a lower level. However, there is no uniform approach in the field of traffic simulations which would deal with the consistency of (railway) traffic dynamics in relation to various submodels within the entire simulator.

An approach based on parallelly conducting both, microscopic and macroscopic simulation (of the entire investigated system) can be used for the change of level of detail in traffic simulation - the person conducting the experiment will only see the switch between levels. However, computational demand is a rather inconvenient disadvantage. We are trying to reduce such high demand by introducing scalable simulation models.

Another approach is the combination of submodels with different level of detail within one simulator, which is created statically before starting the simulation (*hybrid model*). Within the simulation experiment conducted above the hybrid model it is possible to, for example, to focus on detailed examination only of narrow operational areas and to manage with only rough operational characteristics in adjacent areas (Magne, Rabut, and Gabard 2000; Gille, Klemen, and Siefer 2008).

Current solutions of a hybrid model are characterized by the utilization of cooperation of several simulation tools. A possible approach is to use a microscopic simulator for the area of a railway network of high importance and to use a macroscopic simulator for all other areas. The railway network is then divided into areas in which an investigation of traffic is investigated using different simulation tools (Casas, Perarnau, and Torday 2011).

A different approach, which is considered a borderline hybrid model, is using the cooperation of simulation tools in a scenario in which each tool solves its own constituent task above the whole area of a railway network. The base of such approach is a microscopic simulator which includes detailed information of individual subsystems and which conducts detailed simulation of train movement dynamics and with it connected investigation, results of which (e.g. aggregated infrastructure model) are subsequently used for, for example, evaluating station throughput or for finding optimal (the shortest) train routes, etc., in a macroscopic simulator (Montero, Codina, and Barcelo 2001). RailSys microscopic tool (Radtke and Bendfeldt 2011) and the macroscopic tool NEMO (Sewcyk and Kettner 2001; Kettner, Sewcyk, and Eickmann, 2003) are both examples of simulators utilizing such approach.

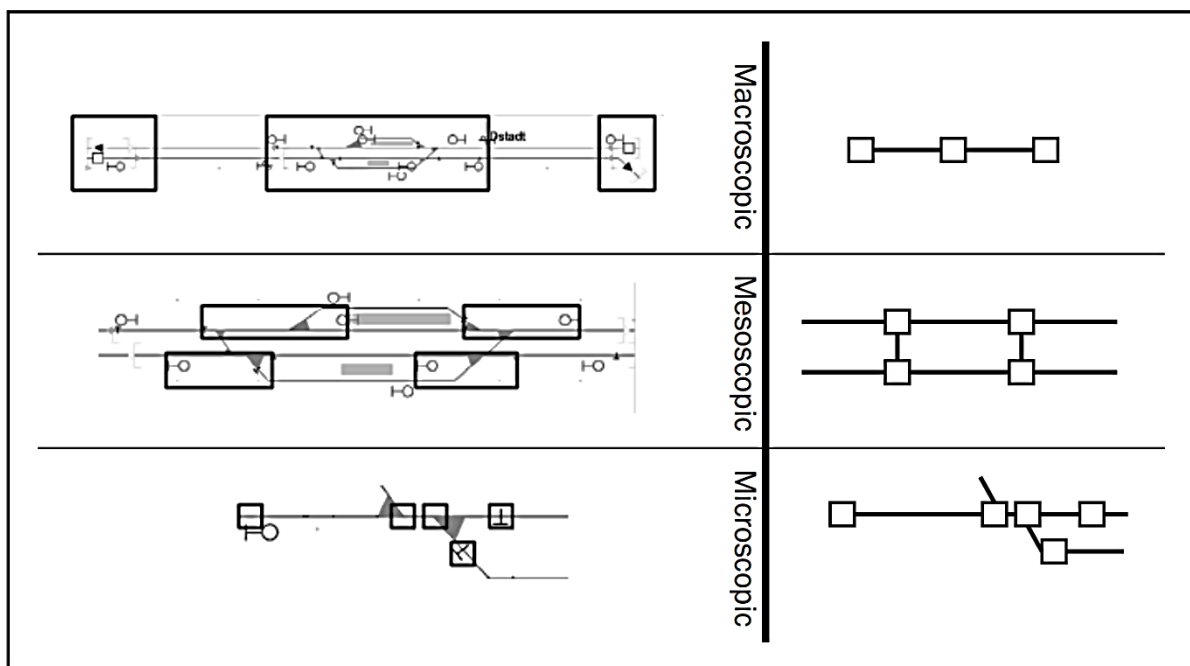


Figure 1: Visualization of infrastructure for different level of abstraction (Cui and Martin 2011)

3. APPROACH OF HYBRID MODEL

On the other hand, our presented methodology is based on a hybrid model implemented within one simulation tool (*unitary hybrid model*). That methodology supports combining submodels exploiting the microscopic and macroscopic levels of details. Microscopic simulation is typically connected to particular areas, within the frame of which detailed traffic indicators are essential for an experimenter. On the contrary, macroscopic simulation is applied within those simulator's parts where only rough operational indicators are requested. Unitary hybrid model enables to adjust the *granularity* of a part of a simulator, i.e. operational submodels for individual infrastructure submodels, thanks to the nature of the simulation. Overall computational demands of a unitary hybrid model are certainly lower than relevant demands related to a corresponding model executing pure microscopic simulation.

To create a unitary hybrid model it is necessary to manage solutions to the following problems concerning the submodels applying different level of detail: (i) construction of inhomogenous infrastructure submodels, their suitable interconnection and visualization, (ii) construction of different submodels of traffic, (iii) transformation of traffic flows on interfaces of different traffic submodels.

4. MICROSCOPIC INFRASTRUCTURE MODEL

The methodology of building unitary hybrid models focuses primarily on the construction of a track infrastructure submodel (*Figure 2*). That submodel applies the highest level of details which can be required for the given part of the railway network. Within this context the editing tool *TrackEd* (Novotny and Kavicka 2015) can be utilized. The mentioned software is specialized in (i) quick constructions of track layouts with the help of prearranged prototypes of rail objects and (ii) subsequent schematic visualizations depicting track infrastructures. The resulting submodel

is then represented by a data structure depicting a mathematical model of an undirected graph in which each node contains not only the position in a schematic plan, but also real kilometric position within the railway network.

Within the editor it is possible to define topological, metric and slope characteristics related to all tracks or their parts. Those characteristics accurately reflect either an existing situation on the spot or a planned infrastructure from the project documentation. Hence, it is possible to carry out realistic calculations (during simulation trials) concerning the dynamics of train rides. The created microscopic submodel considers: (i) tracks, (ii) switches, (iii) crossings, (iv) signal devices, (v) limit signs for train positions on tracks, (vi) platforms, (vii) isolated circuits, (viii) electrification and useful lengths of tracks, and finally (ix) speed limits valid for individual rail elements (Kubat 1999; Jirsak 1979).

In the *TrackEd* editor we can use a variety of sophisticated functionalities which help to automate some stages of a design. For example, it includes a semi-automatic calculation of a complete set of primary and alternative train routes for all possible train transfers within the investigated track layout or an automatic calculation of useful lengths of station tracks in the station which affect technologic processes in the station, etc.

5. DOUBLE-LAYER CONCEPT

The resulting *hybrid submodel/layer of infrastructure* is built over a *micro-layer*, which corresponds to the above mentioned microscopic submodel. The *hybrid layer* combines infrastructure areas applying different level of abstraction (*Figure 3*). The introduced *hybrid-layer* is composed of *micro-segments* and *macro-segments*. *Micro-segments* are represented by sub-graphs directly taken from the micro-layer. *Macro-segments* lower granularity of relevant disjoint connected sub-graphs from the micro-layer.

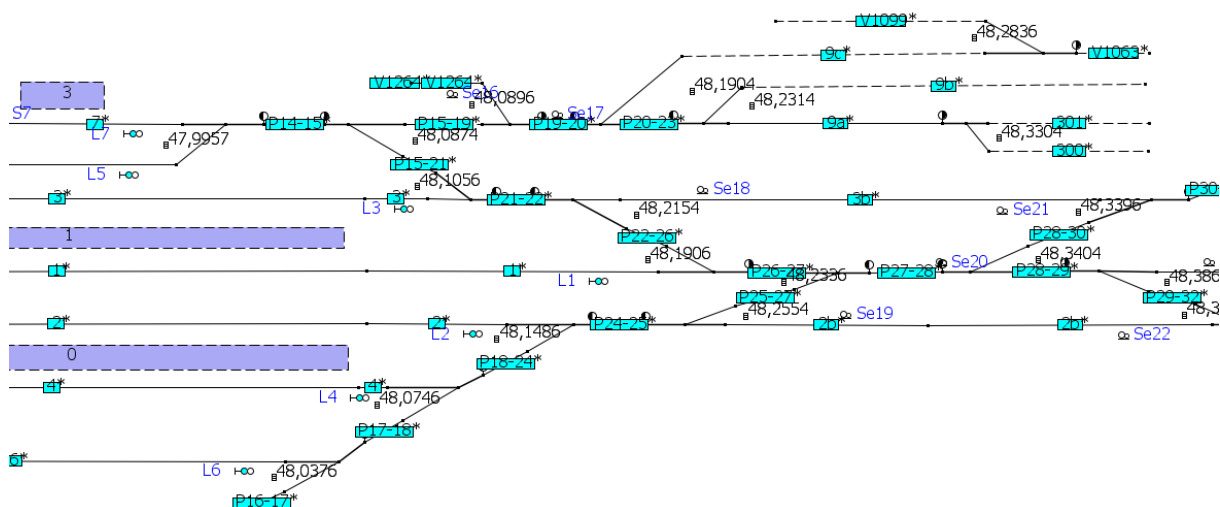


Figure 2: Schematic plan of track infrastructure in the *TrackEd* (Novotny and Kavicka 2015)

LEGEND

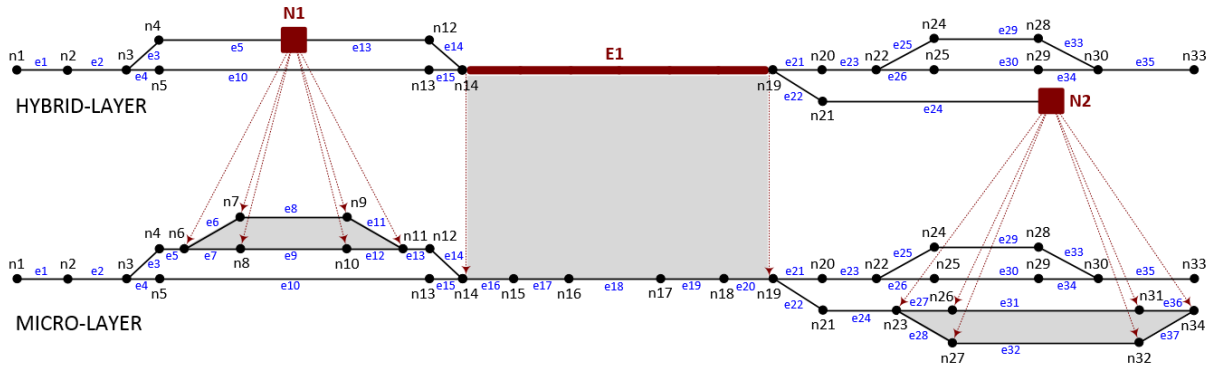
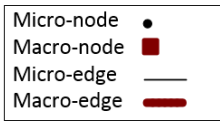


Figure 3: Double layer concept of infrastructure model

Constructions of macro-segments support creating variant configurations of hybrid submodels of railway infrastructure. It means in fact that different scenarios of simulation experiments can apply various levels of details (micro- or macroscopic) within an infrastructure submodel.

As it has already been said, a data structure in a form of a graph implemented in the operational memory and composed of two types of object (*nodes* and *edges*) is used to represent a railway network. From the viewpoint of the implementation of the unitary hybrid model, this means two different (co-existing) instances of such data structure: (i) a graph representing the primary micro-layer, and (ii) a graph representing the resulting hybrid layer of the track infrastructure (Figure 3). As a general rule, the edges or nodes in the graph of the hybrid layer can represent individual edges or nodes of the micro-layer or their aggregated area (macro-segments). When creating such objects in the hybrid layer of the infrastructure, it is necessary to first set attributes (by deriving them from appropriate encapsulated objects in the micro-layer) needed mainly for calculations within the macroscopic areas. We can thus find within the macroscopic area, for example, the average speed, the number of isolated circuits, the set of train routes defined by a technologist, etc.

When creating a hybrid model of the railway infrastructure, we must always base our model on a detailed infrastructure model, which contains detailed information (attributes) necessary for mainly the microscopic simulation and which can be then easily aggregated for the needs of the macroscopic simulation. An reverse approach, in which we only have the aggregated (macroscopic) model of the track infrastructure at our disposal, is not considered as it mainly derives detailed information about the

infrastructure for the needs of the microscopic simulation rather inaccurately. An example of such inaccuracy can be a railway stations for which we know only the coordinates of their real positions within the railway network, but we do not know the detailed information about their track layout which is important for, e.g. finding the capacity/throughput of the station, etc.

6. SEGMENTS

Two types of macro-segments (*macro-nodes* and *macro-edges*) are distinguished within the presented methodology. *Macro-edges* typically encapsulate line sequences of edges from micro-layer. *Macro-nodes* can enclose a general connected sub-graph from micro-layer (Figure 3). A typical example of an infrastructure that can be represented by rougher granularity is a railway station. Within one experiment it can be realistically specified with all basic elements (tracks, platforms, signals). Another experiment, which does not pay attention to bottlenecks of infrastructure and/or internal station processes, can interpret that station as a point element (a macro-node). A double layer concept related to infrastructure model provides making various configurations of a hybrid layer with a fixed connection to real topological, metric and slope properties defined in the invariable micro-layer.

Concerning visual representation of the hybrid layer it is important to take into account certain visual deformations of rail elements in comparison with the original visual forms in the microscopic layer. It is caused by different graphic connections of microscopic elements to macro-segments with regard to the "native" connections within the micro-layer.

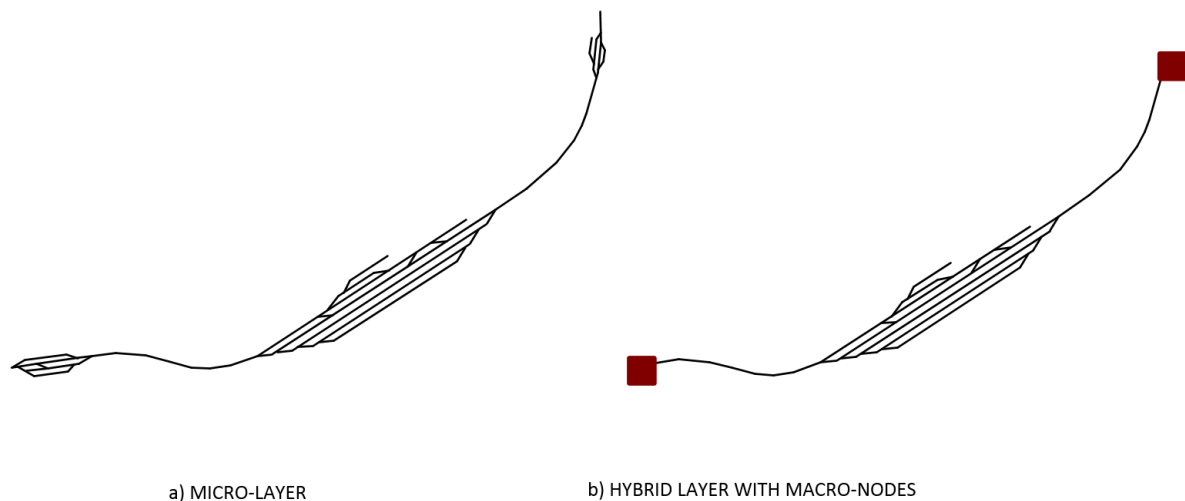


Figure 4: Macro-node illustration

7. CONSTRUCTION OF MACRO-NODE

Macro-node corresponds to the encapsulation of selected nodes and their interlocutory edges of the micro-layer into a single node of the hybrid layer. As it has been said, a typical example of a macro-node can be a railway station (*Figure 4*) which can be represented on a rougher layer of granularity and thus its detailed operational verification is not necessary.

The introduced methodology of creating a hybrid model of a track infrastructure enables to select any area of the micro-layer and its subsequent aggregation into a macro-node (and it is thus to create more of these macroscopic areas within a station *Figure 3*). An advantage of such approach then can be the detailed investigation of railway traffic only on selected station tracks, development of switches, or in a selected part of the investigated station.

Graphic position of the created macro-node in the visual schematic hybrid model of track infrastructure is always located in the middle of the selected area and thus can create visual deformation as in result. In the *TrackEd* editor it is then possible to change graphic position not only of nodes taken from the micro-nodes, but also of these macro-nodes.

8. CONSTRUCTION OF MACRO-EDGE

In the railway network there are parts of tracks where there is no track crossing, convergence into one or divarication into more tracks. Such parts of tracks are called an open track and it connects two traffic landmarks (places with special significance). In other words, it is a route between two adjacent stations which can usually be modelled using several smaller edges with component nodes. Within simulation modes focused mostly on operational verification of stations, such open tracks do not have to be principal for the experimenter, thus it is not necessary to conduct any macroscopic traffic simulation in this part of the railway network. For these reasons it is important to enable the encapsulation of such component edges along with their border nodes into a single, so called, *macro-edge*

(*Figure 5*) and to conduct above it only macroscopic traffic simulations, when applying the described methodology. A macro-edge is then used to encapsulate a liner connection of edges in the micro-layer, or rather the sub-graph, in which each inner node is interlocutory only with two edges (a liner sub-graph).

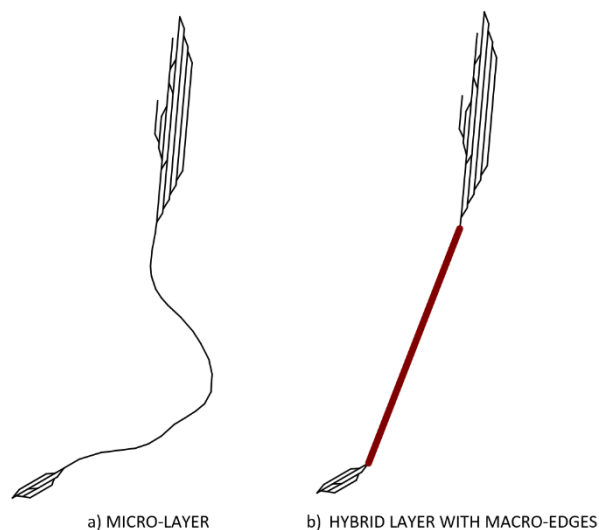


Figure 5: Macro-edge illustration

If we are considering a single track, it can be for the macro-segments modelled as one macro-edge. If we are considering a multitrack and at the same time there is no crossing, convergence or divergence of tracks, it is necessary to model each track individually (i.e. encapsulate each track into one macro-edge). If a multitrack is represented as macro-edges, it is necessary to manage their graphic representation so no visual overlay occurs in case of the same initial or end node.

Generally, we can use the macro-edge in any part of the railway network not only for open tracks, but it also has to be represented as encapsulation of a liner subgraph. In this way we can treat all station tracks, which is not evaluated in detail in the aim of the investigation within

the station, and only information about the traffic density is sufficient.

9. HYBRID INFRASTRUCTURE MODEL

When creating a hybrid track model by applying macro-edges or macro-nodes above the defined areas (while respecting the abovementioned rules), one must also consider suitable setting of borders of the selected macro-segment while considering switches, crossings, platforms, isolated circuits, and other traffic landmarks (places with special traffic significance). Generally, creating a hybrid model means that creating macro-segments above the defined infrastructure segment is limited not only by its track layout, but also by other subsystems (e.g. security devices system, etc.)

Combining any macro-segments is allowed when creating a hybrid track layer (Figure 6). It is then possible to connect macro-nodes with macro-edges. However, such connections always have only one common microscopic element (node) within their encapsulated infrastructure submodels. That means that it is not possible for two border nodes of one edge of the micro-layer encapsulated into a macro-layer to also occur in a macro-node.

It is presupposed a static configuration of all macro-segments before an execution of simulation experiments. Thus, dynamic changes of levels of details related to track layouts are not supported during simulation trials. The resulting track model can also have the characteristics of a microscopic model when it does not include any area encapsulated into a macro-segment. And thus the resulting hybrid layer is identical with the input micro-layer. The resulting model can be entirely macroscopic if each edge or node of the micro-layer also belongs into any of the defined macro-segments. However, for the resulting hybrid infrastructure model it is typical to combine the areas applying different levels of detail.

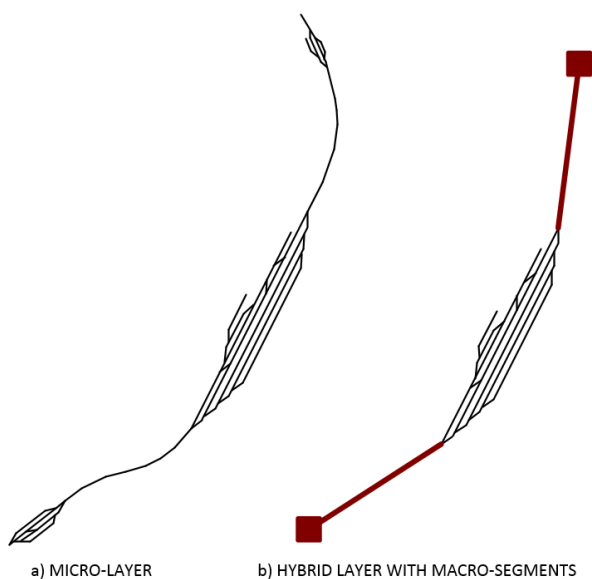


Figure 6: Combination of macro-segments

For the needs of repeated use of the defined configuration of the hybrid layer, the *TrackEd* editor enables saving such configuration into a predefined template and the use it again in the future.

10. TRAFFIC MODEL

Because of combining macro-segments with areas composed of microscopic elements within the hybrid layer, it is necessary to apply different traffic models (implementing various levels of abstraction). Diverse traffic models are connected with different traffic indicators - e.g. detailed riding features of individual rail vehicles or average rates of traffic flows within the observed areas (Cenek 2004), etc. As an example can be mentioned a macroscopic traffic model, which observes individual vehicles on a high level of abstraction (applying constant average speeds etc.). Another approach can be based on the theory of traffic flow exploiting the analogy with the flows of liquids (Krivy and Kindler 2003).

Because several different submodels coexist within a hybrid model, it is necessary to solve the transformation of traffic flows, i.e. it is necessary to unambiguously define the information about railway traffic on the interface of microscopic and macroscopic submodels in order to maintain information consistency. The main problem is the loss of a large amount of information when switching the rolling stock from the microscopic to the macroscopic submodel. The result is absence of detailed information about individual rolling stock which later switches from the macroscopic area back into the microscopic one. These attributes must be unambiguously defined for the microscopic simulation from aggregated information (e.g. average speed or the intensity of the traffic flow) and thus establish the current speed, the acceleration, the occupancy of isolated circuits, and intervals between trains. Of course, the current status of traffic and traffic situation in the relevant surroundings must be considered.

When the rolling stock switches to or from the macroscopic submodel, it is necessary to decide if the switching is even feasible based on the occupancy of the area to which the train is switching. If the switch is directed onto the microscopic rolling element, the vacancy is evaluated based on the occupancy of the relevant isolated circuit. In the switch is directed onto the macroscopic area, such situation is ambiguous and in some extent it depends on the traffic model applied for macro-segments, i.e. it can depend on the intensity of the traffic flow, vacant capacity of the segment, etc.

For rolling stock within macroscopic areas and mainly to maintain consistency of traffic simulation, it will be necessary to select a mechanism of derivating aggregating information. An example of such can be attributing to rolling stock average speed which can consider maximal allowed speed within all rolling elements encapsulated in a macro-segment or just in those in which the train will be moving.

11. PERSPECTIVE OF DEVELOPMENT

The next stage of development related to a scalable railway traffic simulator (based on the unitary hybrid model) will be focused on a selection and advancement of traffic models utilized within microscopic and macroscopic parts of a hybrid infrastructure model. Special attention must be paid to transformations of traffic flows when they transit from the microscopic to the macroscopic elements and vice versa (Burghout 2004). It is presumed that series of simulation experiments will be carried out over the created hybrid infrastructure model within an integrated development environment belonging to *TrackEd* tool.

Prospects of further development can be extending the editor by functionality, thanks to which macro-edges created above a multi open track are encapsulated into one aggregated macro-edge.

Other prospective option of development can be extending the unitary hybrid model by mesoscopic areas enabling more detailed traffic simulation than it is possible in macroscopic areas, but above more detailed infrastructure model.

12. CONCLUSION

The article deals mainly with the explanation of basic phrases of methodical approach to the construction of a scalable infrastructure model.

Unitary hybrid model (currently in development in the *TrackEd* editor) combines areas constructed on different levels of detail (microscopic and macroscopic) and thus differentiates between a micro-layer and a hybrid layer. To create a hybrid layer, the approach of decreasing the level of detail for specified areas of the micro-layer (macro-segments) is applied. Two types of macro-segments, used to encapsulate various areas of a microscopic model of track infrastructure and thanks to which it is possible to create variant configurations of a scalable simulation model, are introduced.

The scope of future development also considers the selection of traffic models for microscopic and macroscopic parts of the hybrid model and transformation of traffic flows on their interface.

REFERENCES

- Cui Y., and Martin U., 2011. Multi-scale Simulation in Railway Planning and Operation. PROMET Traffic. Available from: <http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/186> [Accessed 6 July 2016].
- Burghout W., 2004. Hybrid microscopic-mesoscopic traffic simulation. Doctoral Dissertation. Royal Institute of Technology.
- Kettner M., Sewczyk B., and Eickmann C., 2003. Integrating microscopic and macroscopic models for railway network evaluation. Association for European Transport, Germany. Available from: <http://abstracts.aetransport.org/paper/index/id/1731/confid/9> [Accessed 6 July 2016].
- Novotny R., and Kavicka A., 2015. Model of a railway infrastructure as a part of a mesoscopic traffic simulator. To appear in: The 27th European Modeling & Simulation Symposium. Bergeggi, Italy.
- Gille A., Klemenz M. and Siefer Th., 2008. Multiscaling Analysis. UIC, Germany. Available from: http://www.uic.org/cdrom/2008/11_wcrr2008/pdf/O.2.1.2.2.pdf [Accessed 6 July 2016].
- Simcon, 2014. Villon. Simcon. Available from: <http://www.simcon.sk/en/tools/villon> [Accessed 6 July 2016].
- Radtke Alfons, Bendfeldt Jan-Philipp, 2011. Handling of railway operation problems with RailSys. Institute of Transport, University of Hanover, Germany. Available from: <http://www.railway-research.org/IMG/pdf/235.pdf> [Accessed 6 July 2016].
- Huerlimann Daniel, Nash Andrew, 2010. OpenTrack - Simulation of Railway Networks: User's Manual Version 1.6. ETH Zurich: Institute for Transportation Planning and Systems.
- Krivy I., and Kindler E., 2003. Modeling and simulation I - II. Ostrava: PrF OU. [in Czech]
- Cenek P., 2004. Management traffic processes in logistics systems. AT&P journal, 13-15. [in Czech]
- Sewczyk B., Kettner M., 2001. Network Evaluation Model NEMO. Institute of Transport. Germany, University of Hanover. Available from: http://www.uic.org/cdrom/2001/wcrr2001/pdf/poster/3_3/014.pdf [Accessed 6 July 2016].
- Kubat Bohumil, Fliegel Tomas, 1999. Railway buildings 30. Prague: CVUT. [in Czech]
- Jirsak Zbynek, 1979. Railway stations and junctions. Bratislava: Alfa. [in Czech]
- Montero L., Codina E., Barcelo J., and Barcelo P., 2001. Combining macroscopic and microscopic approaches for transportation planning and design of road networks. Transportation Research. Available from: https://www.inrosoftware.com/assets/pres-pap/international/ieug98/octet_stream_7.pdf [Accessed 6 July 2016]
- Magne L., Rabut S., and Gabard J.F., 2000. Towards an hybrid macro-micro traffic flow simulation model. Informs String Conference, Salt Lake City. Available from: http://www.researchgate.net/publication/237560732_Towards_an_hybrid_macro-micro_traffic_flow_simulation_model.pdf [Accessed 6 July 2016]
- Hansen Ingo, Pachl Jorn, 2008. Railway timetable and traffic. Hamburg: Eurailpress.
- Casas J., Perarnau J., and Torday A., 2011. The need to combine different traffic modelling levels for effectively tackling large-scale projects adding a hybrid meso/micro approach. Procedia - Social and Behavioral Sciences. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S187704281101411X> [Accessed 6 July 2016]

HLA LISTENER – A GUI DRIVEN MAPPING OF HLA STANDARD SERVICES

Mostafa Ali^(a), Yasser Mohamed^(b)

^(a)PhD Candidate, University of Alberta

^(b)Associate Professor, University of Alberta

^(a)MostafaAli@ualberta.ca

^(b)yal@ualberta.ca

ABSTRACT

High Level Architecture (HLA) is a distributed simulation framework that offers integration, reusability, and flexibility. Nonetheless, it has been criticized for its complexity and overwhelming services. Consequently, it has been rarely used on a day-to-day basis in industrial world. A very steep learning curve is one of the main challenge that faces a new user of HLA distributed simulation. The user has to be familiar with different HLA services (e.g. objects and time management) beside a programming language to develop a basic simulation. In order to overcome this challenge, we developed a tool with friendly graphical user interface that encapsulates all HLA services. This tool allows a new user to explore services as described in the standards without the need for programming language. In addition, the interface can be used by an experienced simulator to debug existing federation quickly and efficiently.

Keywords: Distributed Simulation, HLA, HLA Listener

1. INTRODUCTION

Simulation provides an affordable means for teaching, analyzing, and testing complex physical system that would usually require a huge investment and manpower to perform in a real-world environment. For example, flight simulation has been widely used in commercial and military aviation for training potential pilots (Page, 2000).

Over the years, Different simulation paradigms such as discrete event, agent based, system dynamic simulations have been proposed to address specific domains and applications. Each simulation approach advocates a modelling system that simplifies real-world system. For example, discrete event simulation depends on event scheduling basis to simulate real world scenarios (Goti, 2010). While Agent-based simulation relies on a set of autonomous agents that interacts with each other and with the environment to simulate systems like socio-economic science (Helbing, 2012; Jennings, 2000; Wooldridge, 1997).

On the other hand, distributed simulation takes a unique approach by suggesting decomposing the simulation domain into separate manageable components that interact with each other to provide a complete simulation

solution. Distributed simulation manages interactions and interoperation between different simulation components; however, it does not intervene with any of the component internal mechanism. Each component can use different simulation type internally to model its scope without affecting another component. This powerful flexibility allows combining multiple simulation model into one scenario easily and effectively.

High Level Architecture (HLA) is a popular distributed simulation standard. HLA has been originally developed by the Department of Defense in the United States (Borshchev et al., 2002; Fujimoto, 2003). It is currently regulated by IEEE, and the latest version is 1516-2010 known as HLA evolved.

Distributed simulation has been used in different domains; like logistical processes, traffic management systems, and freight transportation (Schulze et al., 1999; Zacharewicz et al., 2011); optimizing supply chain from manufacturing through distributor to end customer (Turner et al., 2000); tunneling (AbouRizk, 2010); fleet optimization in earthmoving operations (Ali et al., 2014), manufacturing (Hibino et al., 2002), 3D virtual environment for marine port environments (Bruzzone and Longo, 2013), and combat simulation (Ham et al., 2014).

Although distributed simulation has many potential beneficial application in industry such as supply chain and digital factories (Taylor et al., 2002), it is not fully embraced in industry comparing to research and military domains (Boer et al., 2006a; Lendermann et al., 2007; Taylor et al., 2002). Poor understanding of industrial needs, complexity of distributed simulation, ambiguities in HLA standards, and steep learning curves are main reasons that hinder applying distributed simulation on day-to-day basis in industrial domain (Boer et al., 2008, 2006b; Taylor et al., 2002).

A non-expert user who wants to investigate HLA distributed simulation has to spend substantial time to gain knowledge in three areas: 1) Distributed simulation; 2) Understanding HLA standards (“IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-- Federate Interface Specification,” 2010); and 3) Using a programming language to write a simple start-up distributed simulation scenario. This steep learning curve turns practitioners away from

embracing distributed simulation. This shows the necessity for providing simplified approaches that allow non-expert to develop distributed simulation with minimum knowledge requirements.

This paper shows our effort to develop an intuitive graphical interface that encapsulates all services provided in HLA distributed simulation. Graphical User Interface dramatically improves reception, and learning (Gerhardt-Powals, 1996; Staggers and Kobus, 2000), as it simplifies complex concepts into an intuitive user interface that is familiar to anyone with a minimal computer knowledge.

The developed prototype, named HLA Listener, allows non-experts to try distributed simulation with almost zero-learning-curve. Through the interface, the user can create a new federation, join an existing one, and interacts with other federates without the need for a programming language.

Although HLA Listener is not sufficient to develop a complete real-world scenario simulation, it can be used for two scenarios. First, it provides a first step for learning and understanding basic HLA distributed simulation concepts such as: Federation management, Object management, and Time management. In addition, it can be used by experts to debug existing federation through scripting window as will be shown later.

The remainder of the paper is structured as follows: The next section will discuss the structure of HLA Listener and discuss briefly its different components; then a sample use of HLA Listener in training and debugging is provided. Finally, we will conclude by providing our vision for potential improvements for this tool.

2. HLA LISTENER STRUCTURE

HLA Listener aims to provide an intuitive interface for all HLA distributed services to allow a beginner to make sense of the distributed simulation without requiring knowing much about HLA standards or mastering any programming languages. HLA Listener is written in Java which is one of the two Application Programming Interfaces (API), along C++, that are considered part of the HLA standards. Despite that the HLA standard encourages third-parties to develop APIs for other programming languages, we found that most of the commercial and open-source RTI implementations support these two APIs only. Java Virtual Machine® allows running HLA Listener on different operating systems without having to rewrite the code.

HLA Listener consists of five main components, as shown in Figure 1, **HLA Interface** which is the Java API that ships with HLA standard and it allows running the simulation with any HLA-compatible RTI implementation; **FOM parser** retrieves the simulation FOM and provides object classes, interaction classes, attributes, parameters, and data types to the other components; **RTI ambassador** is responsible for sending requests from the user to the RTI, these requests represent different services provided in the HLA standards such as “create a federation” and “Publish object class”; **Federate ambassador** receives callbacks,

such as “Discover object instance” and “Time regulation enabled” from the RTI and displays them to the end user; **Scripting module** allows the automation of RTI and Federate ambassador tasks by writing java scripts as will be shown later.

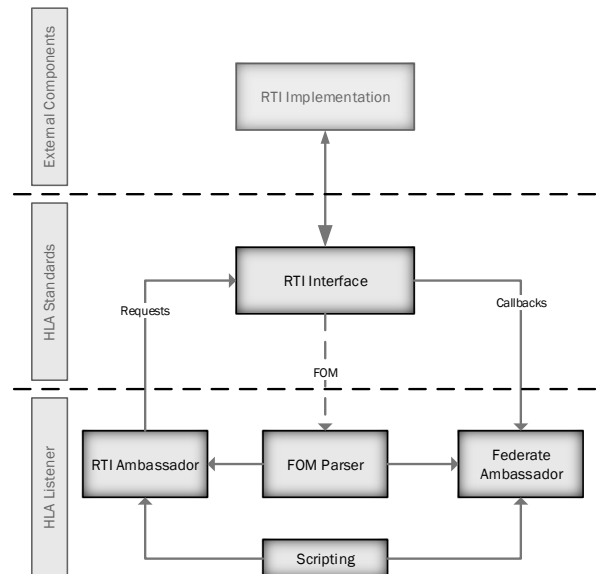


Figure 1 HLA Listener Structure

2.1. Dynamic Link Compatibility

One of the key issues with earlier versions of HLA distributed simulation was the overhead effort required to run the simulation with different RTI implementations; as each RTI implementation has different interface specifications and the simulator has to readjust the simulation to comply with the RTI. The latest version of HLA standards (“IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)- Federate Interface Specification,” 2010) resolves this problem through Dynamic Link Compatibility (Möller et al., 2008). Dynamic Link Compatibility enables the simulator to develop a simulation using a specific RTI implementation that can be replaced with another implementation without requiring any changes in the simulation’s internal structure.

In addition, Dynamic Link Compatibility allows selecting the RTI implementation dynamically at runtime. Using this feature, HLA Listener can work with any HLA-compatible RTI implementation. When launching HLA Listener, it will ask the user to select RTI implementation to run the simulation; we tested it with two commercial and one open-source RTI implementation, and it works as expected without having to change any line of code. The following sample java code demonstrates Dynamic Link Compatibility.

```
public class DynHLA {
    //List all available RTI
    implementations
    RtiFactoryFactory.getAvailableRtiFa
    ctories().stream().forEach((rti) -> {

        System.out.println(rti.rtiName());
    });
}
```

```

private final String preferredRTI =
"RTI A"; // the name of the RTI
implementation
//Work with a specific RTI
implementation
RtiFactory rtiFactory =
RtiFactoryFactory.getRtiFactory(prefer
edRTI);
}

```

2.2. FOM Parser

HLA Standard defines Federation Object Model (FOM) “A specification defining the information exchanged at runtime to achieve a given set of federation objectives. This information includes object classes, object class attributes, interaction classes, interaction parameters, and other relevant information” (“IEEE Standard for Modeling and Simulation (M\&S) High Level Architecture (HLA)-- Federate Interface Specification,” 2010). Latest HLA standards support modular FOMs (Möller et al., 2007) which allows different federates to add one or more FOM files. Those FOMs along MOM and Initialization Module (MIM) are merged into one simulation FOM according to merging rules specified in HLA standards.

This flexibility means that FOM module could change during the simulation model and a well-designed federate should track those changes during simulation execution. HLA Listener tracks any changes in simulation FOM during the simulation execution by subscribing to “HLAcurrentFDD” attribute in “HLAobjectRoot.HLAmanager.HLAfederation” class, which is part of the standard MIM. By subscribing to this attribute, RTI will send a “Reflect attribute value” callback to HLA Listener whenever the simulation FOM gets updated; the callback contains the updated simulation FOM in XML format.

Because HLA Listener is a general-purpose tool and the FOM is processed solely based on the callback provided by the RTI. FOM parser component processes the FOM to get the following information:

1. **Date types:** FOM parser extracts all data types: basic, simple, enumerated, array, fixed record, and variant record. Those data types are used in many HLA services such as “Update attribute value”
2. **Update rate:** FOM module contains different update rate (in Hz) that specifies the maximum rate for sending updates to a specific federate when used with “Best-effort attribute”. Update rate is used in many HLA services such as “Subscribe Object Class Attributes”.
3. **Transportation type:** There are two basic transportation types “Reliable” and “Best-effort”; however, a FOM might define additional transportation types.
4. **Dimensions:** It defines intervals for each attribute in a format [0, upper value], which is used for Data Distribution Management (DDM) service.
5. **Object Classes and Attributes:** a FOM contains a hierarchy of object classes and their attributes, these classes are used in exchange information between federates. FOM parser retrieves all classes that are used in many services such as “Publish Object classes attribute” and “Register Object Instances” services.
6. **Interaction Classes and Parameters:** Interaction classes and their parameters are very similar to Object Classes and they are used in “Publish Interaction Class” and “Send Interaction” services.

2.3. HLA standard mapping

Although HLA Listener simplifies HLA distributed simulations’ services, it does not obfuscate it. The user can still map all graphical components to the standards.

In order to encourage the user to utilize and lookup the standards, each graphical component in HLA Listener has been numbered according to its section number in the standards (“IEEE Standard for Modeling and Simulation (M\&S) High Level Architecture (HLA)-- Federate Interface Specification,” 2010). The section number and text in standards are displayed for Menu, Menu items, dialog boxes’ titles, requests, and callbacks as in Figure 3.

3. HLA LISTENER EXAMPLE

This section provides an example of using HLA Listener to explore HLA distributed simulation. It does not provide a complete demo of its capability but rather a short introduction of the main steps in a distributed simulation life cycle as shown in Figure 2. An interested reader can come up with more advanced scenario to use HLA Listener to explore advanced services in the HLA standards.

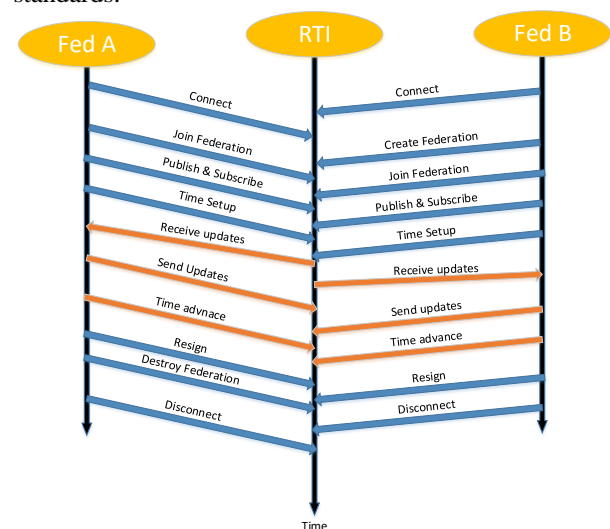


Figure 2 A typical life cycle of HLA distributed simulation.

When a user runs HLA Listener, it will ask for a jar file for the HLA-compliant RTI to connect to through

Dynamic Link Compatibility. All RTI implementations (commercial and open-source) would work exactly the same way. If the link is successful, the main interface for HLA Listener will be shown as in Figure 3. Each menu item represents a section in the HLA distributed simulation standards. The left part shows a history of requests and callbacks while the right section gives more details about the selected request/callback.

The latest version of the standards introduced the concept of the “Connection Mode” which flags each federate either connected or disconnected. Any HLA service, such as create or join federation, must run by a connected federate. Consequently, the first step is connecting to RTI through “Connect service” in Figure 4 (a).

One of the connected federates creates the federation execution by providing its name, a set of FOM modules, and federation logical time type (integer or float) as in Figure 4 (b). Once created, all federates join the federation through the “Join federation execution” as Figure 4 (c). Each federate has the chance to add additional FOM modules during joining the federation. Time management is a key service in HLA standards; however, a beginner might find it hard to understand all of its associated concepts such as “Look ahead”, “Time regulation / constrained”, and “Advance in time”. HLA Listener can be used to explore all scenarios related to the time management as in Figure 4 (d).

In order to exchange data between federates, each federate should declare object classes and interactions that it will publish or subscribe to. RTI uses these declarations to manage the transfer of updates from publishing federate to subscribing ones. A federate might publish / subscribe many object classes and interactions and an object class / interaction can be published or subscribed by many federates. The declaration menu in HLA Listener can be used to publish / subscribe as in Figure 5 (a). A federate that publishes an object class should register object instance as in Figure 5 (b) and a subscribing federate will receive “Discover Object Instance” callback.

Publishing data in the federation is done through “Update Attribute Values” and “Send Interaction” services, while receiving the data is done through “Reflect Attribute Values” and “Receive Interaction” services. “Update Attribute Values” service, shown in Figure 5 (c), requires providing values for owned published attributes for an instance along a user-supplied tag and optional time stamp.

After sending all updates and processing all received values, a federate should request advance in time through “Time Advance Request” as in Figure 5 (d), “Time advance Request Available”, “Next Message Request”, “Next Message Request Available” services. Time advance tells the RTI that the federate will not send updates with time stamp less than requested time advance plus look ahead value (Figure 4 (d)). When all regulating federates request to advance a certain point in time, RTI will send “Time Advance Granted” to all federates requesting time advance.

This cycle of sending / receiving updates, request time advance, and time advance granted may be repeated many time during federation execution. When the simulation is complete, each federate should exit gracefully by invoking “Resign Federation Execution” service as in Figure 5 (e), afterwards, one federate should destroy the federation by invoking “Destroy Federation Execution” service as in Figure 5 (f). Finally, all federates should disconnect from the RTI.

The previous demo briefly explained the workflow for a beginner user who wants to explore HLA distributed simulation capabilities without going into too much details. However, HLA Listener can be also used by professionals to test and debug existing federations. If, for example, a federate is expected to receive specific values at certain point of time but it did not, then we have to determine if there is a problem with the sending or receiving federate. Typically, a minimal federate will be used to replace either the sending or receiving federate to determine the problem source. Instead of writing this minimal federate from scratch, HLA Listener can be used for this purpose. A professional user can use the interface to imitate sending or receiving the values. Additionally, the script window, Figure 6, can be used to write a minimal code for testing and debugging.

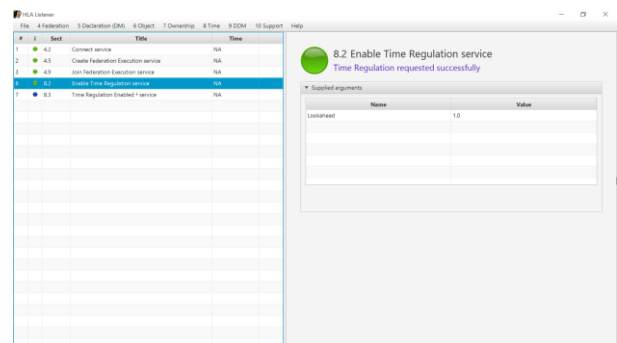


Figure 3 The main interface of HLA Listener.

4. LESSON LEARNED AND FUTURE WORK

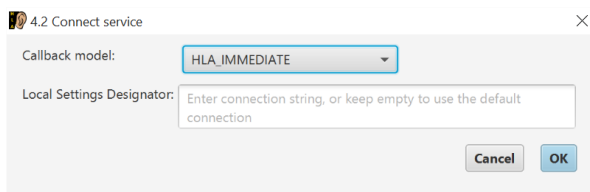
The comprehensive investigation of HLA services and its standards during the development of the HLA Listener revealed a number of issues that has to be addressed in the next version of HLA standards. These issues are related to ambiguities and discrepancies in the HLA standards as follows:

1. HLA standard (1516.1 section 4.3) states that “Disconnect service” should raise “Not connected” exception if a disconnected federate invokes “Disconnect service”. However, the accompanying Java API does not include this exception.
2. According to the standards, multi-dimensional array is supported. However, it was left to the simulator to determine how to encode and chain different array’s dimensions (i.e. by row or column basis) which will severely affect simulation integration and reusability as

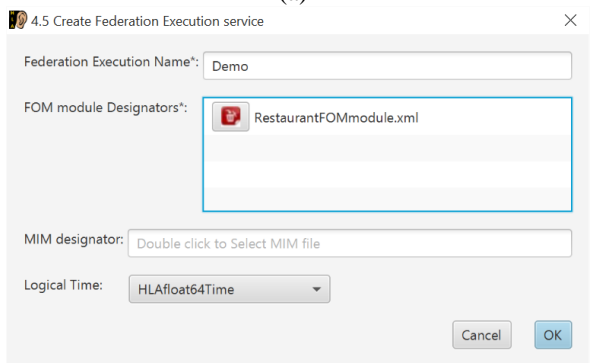
different simulators might opt to different encoding preferences.

3. There are not enough learning resources that explain in details the HLA distributed simulation.

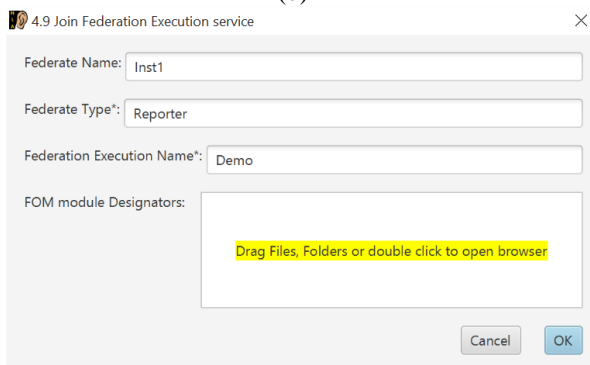
For future work, we are planning to add the capability to record user interaction with HLA Listener and auto generate the corresponding code. This will reduce developing time and effort significantly and allows the development of more advanced and complex federation with minimal coding effort. Another proposed enhancement would be the ability to add custom basic data types by manually defining its encoding and decoding factories.



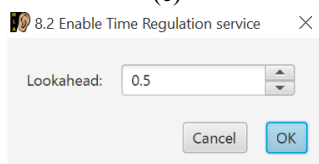
(a)



(b)

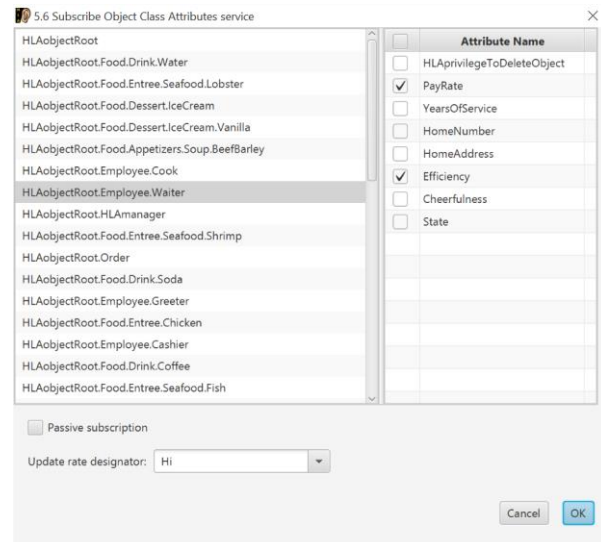


(c)

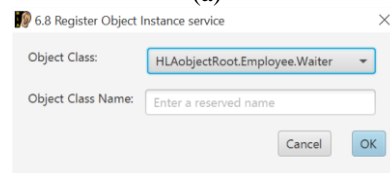


(d)

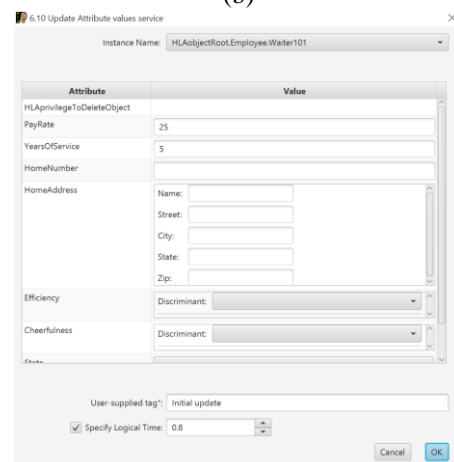
Figure 4 A typical federation execution will start with a) Connect, b) Create Federation (if not exist), c) Join Federation Execution, and d) Enable time regulation / constrained.



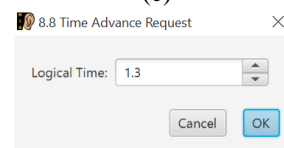
(a)



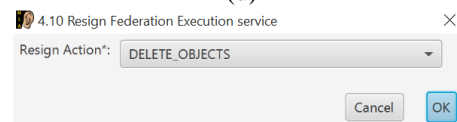
(b)



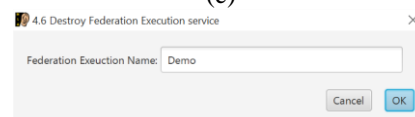
(c)



(d)



(e)



(f)

Figure 5 a) Publish / subscribe object classes and interactions, b) Register / discover object instance, c) Send / receive updates, d) Advance in time and repeat

steps c) and d) for the simulation execution lifecycle, then e) Resign federation, f) Destroy federation and / or disconnect.

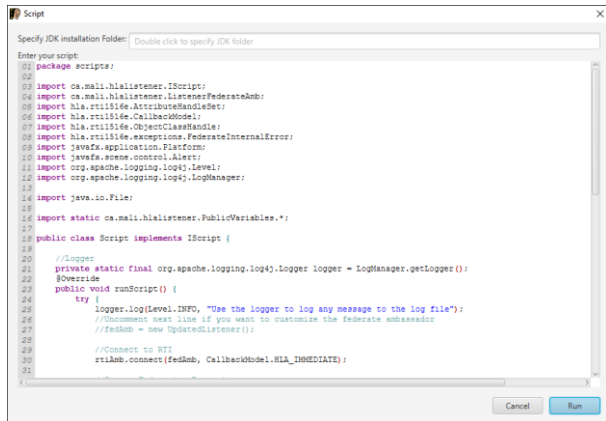


Figure 6 The script window can be used to interact with RTI through code.

5. HLA LISTENER SOURCE CODE

HLA Listener is published publicly under Apache License agreement. The source code and binaries can be obtained through its GitHub web page: <https://github.com/EMostafaAli/HLaListener>.

6. CONCLUSION

Distributed simulation is a powerful and flexible framework that enhances flexibility, reusability, and integration of large scale simulations. Nonetheless, it has a very steep learning curve and its standard is written in rigid technical style. This complexity turns away industrial world from utilizing this simulation framework. In this paper, we presented “HLA Listener”, a tool that encapsulates all services provided in the HLA distributed simulation through a graphical user interface. HLA Listener was developed with two use cases in mind. The first usage is oriented to a beginner who wants to explore HLA capabilities without going into much details about programming. Through the interface, the user is able to invoke calls and receive callbacks from the RTI, it can also be used to run a simulation with other federates. In addition, HLA Listener can be used by an experienced simulator to test and debug existing federates as it has a console to write code along the interface. We plan to extend HLA Listener by adding the capability to record user interactions with the interface and automatically generate the corresponding code.

REFERENCES

AbouRizk, S.M., 2010. Role of Simulation in Construction Engineering and Management. *Journal of Construction Engineering and Management* 136, 1140–1153. doi:10.1061/(ASCE)CO.1943-7862.0000220

Ali, M., Fagiar, M., Mohamed, Y., AbouRizk, S.M., 2014. Beyond classic models—design and development of a comprehensive earthmoving simulator, in: 14th International Conference on

Construction Applications of Virtual Reality in Construction and Conference on Islamic Architecture. Sharjah, UAE.

Boer, C.A., Bruin, A.D., Verbraeck, A., 2006a. Distributed Simulation in Industry - A Survey Part 1 - The Cots Vendors, in: *Simulation Conference, 2006. WSC 06. Proceedings of the Winter*. Presented at the Simulation Conference, 2006. WSC 06. Proceedings of the Winter, pp. 1053–1060. doi:10.1109/WSC.2006.323194

Boer, C.A., Bruin, A.D., Verbraeck, A., 2006b. Distributed Simulation in Industry - A Survey Part 2 - Experts on Distributed Simulation, in: *Simulation Conference, 2006. WSC 06. Proceedings of the Winter*. Presented at the Simulation Conference, 2006. WSC 06. Proceedings of the Winter, pp. 1061–1068. doi:10.1109/WSC.2006.323195

Boer, C.A., Bruin, A. de, Verbraeck, A., 2008. Distributed simulation in industry - a survey Part 3 - the HLA standard in industry, in: *Simulation Conference, 2008. WSC 2008. Winter*. Presented at the Simulation Conference, 2008. WSC 2008. Winter, pp. 1094–1102. doi:10.1109/WSC.2008.4736178

Borshchev, A., Karpov, Y., Kharitonov, V., 2002. Distributed simulation of hybrid systems with AnyLogic and HLA. *Future Generation Computer Systems, Selected Papers presented at the 6th Int. Conf. on Parallel Computing Technologies (PaCT-2001)* 18, 829–839. doi:10.1016/S0167-739X(02)00055-9

Bruzzone, A.G., Longo, F., 2013. 3D simulation as training tool in container terminals: The TRAINPORTS simulator. *Journal of Manufacturing Systems* 32, 85–98. doi:10.1016/j.jmsy.2012.07.016

Fujimoto, R.M., 2003. Distributed simulation systems, in: *Simulation Conference, 2003. Proceedings of the 2003 Winter*. Presented at the Simulation Conference, 2003. Proceedings of the 2003 Winter, p. 124–134 Vol.1. doi:10.1109/WSC.2003.1261415

Gerhardt-Powals, J., 1996. Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction* 8, 189–211. doi:10.1080/10447319609526147

Goti, A., 2010. Discrete event simulations. Sciyo, Rijeka.

Ham, W.K., Kwon, Y., Park, S.C., 2014. Combat Simulation Framework Including Continuous Detection System. *International Journal of Simulation Modelling* 13, 395–408. doi:10.2507/IJSIMM13(4)1.262

Helbing, D., 2012. Agent-Based Modeling, in: Helbing, D. (Ed.), *Social Self-Organization, Understanding Complex Systems*. Springer Berlin Heidelberg, pp. 25–70.

- Hibino, H., Yura, Y., Fukuda, Y., Mitsuyuki, K., Kaneda, K., 2002. Manufacturing Modeling Architectures: Manufacturing Adapter of Distributed Simulation Systems Using HLA, in: Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers, WSC '02. Winter Simulation Conference, San Diego, California, pp. 1099–1107.
- IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-- Federate Interface Specification, 2010. . IEEE Std 1516.1-2010 (Revision of IEEE Std 1516.1-2000) 1,378. doi:10.1109/IEEESTD.2010.5557728
- Jennings, N.R., 2000. On agent-based software engineering. *Artificial Intelligence* 117, 277–296. doi:10.1016/S0004-3702(99)00107-1
- Lendermann, P., Heinicke, M.U., McGinnis, L.F., McLean, C., Strassburger, S., Taylor, S.J.E., 2007. Panel: distributed simulation in industry - a real-world necessity or ivory tower fancy?, in: Simulation Conference, 2007 Winter. Presented at the Simulation Conference, 2007 Winter, pp. 1053–1062. doi:10.1109/WSC.2007.4419704
- Möller, B., Löfstrand, B., Karlsson, M., 2007. An overview of the HLA evolved modular FOMs, in: Simulation Interoperability Workshop, Spring.
- Möller, B., Morse, K.L., Lightner, M., Little, R., Lutz, R., 2008. HLA evolved—a summary of major technical improvements, in: Proceedings of 2008 Spring Simulation Interoperability Workshop, 08F-SIW-064.
- Page, R.L., 2000. Brief history of flight simulation. *SimTecT 2000 Proceedings* 11–17.
- Schulze, T., Strassburger, S., Klein, U., 1999. Migration of HLA into Civil Domains: Solutions and Prototypes for Transportation Applications. *SIMULATION* 73, 296–303. doi:10.1177/003754979907300506
- Staggers, N., Kobus, D., 2000. Comparing Response Time, Errors, and Satisfaction Between Text-based and Graphical User Interfaces During Nursing Order Tasks. *Journal of the American Medical Informatics Association* 7, 164–176. doi:10.1136/jamia.2000.0070164
- Taylor, S.J.E., Bruzzone, A., Fujimoto, R., Gan, B.P., Strassburger, S., Paul, R.J., 2002. Distributed simulation and industry: potentials and pitfalls, in: Simulation Conference, 2002. Proceedings of the Winter. Presented at the Simulation Conference, 2002. Proceedings of the Winter, pp. 688–694 vol.1. doi:10.1109/WSC.2002.1172948
- Turner, S.J., Cai, W., Gan, B.P., 2000. Adapting a supply-chain simulation for HLA, in: Fourth IEEE International Workshop on Distributed Simulation and Real-Time Applications, 2000. (DS-RT 2000). Proceedings. Presented at the Fourth IEEE International Workshop on Distributed Simulation and Real-Time Applications, 2000. (DS-RT 2000). Proceedings, pp. 71–78. doi:10.1109/DISRTA.2000.874065
- Wooldridge, M., 1997. Agent-based software engineering. *IEE Proceedings - Software Engineering* 144, 26–37. doi:10.1049/ip-sen:19971026
- Zacharewicz, G., Deschamps, J.-C., Francois, J., 2011. Distributed simulation platform to design advanced RFID based freight transportation systems. *Computers in Industry, Special Issue: Grand Challenges for Discrete Event Logistics Systems* Grand Challenges for Discrete Event Logistics Systems 62, 597–612. doi:10.1016/j.compind.2011.04.009

AUTHORS BIOGRAPHY

Mostafa Ali is a PhD candidate at the Department of Civil and Environmental Engineering at the University of Alberta. His e-mail address is MostafaAli@ualberta.ca.

Yasser Mohamed is an Associate Professor in the Department of Civil and Environmental Engineering at the University of Alberta. His research interest includes modeling and simulation of construction processes and the integrating of simulation models and knowledge engineering tools into construction management and decision making processes. His e-mail address is yaly@ualberta.ca.

FORECASTING CONSTRUCTION DURATION OF WATERMAIN UPGRADE PROJECT BY USING DISCRETE EVENT SIMULATION

Dongwoo Kim^(a), Yi Zong^(b), Yasser Mohamed^(c), Simaan AbouRizk^(d)

^{(a),(b),(c),(d)} Department of Civil & Environmental Engineering, University of Alberta, Canada
^(a)dongwoo@ualberta.ca, ^(b)yzong@ualberta.ca, ^(c)yaly@ualberta.ca, ^(d)abourizk@ualberta.ca

ABSTRACT

Practitioners in the construction industry often rely on their experience or historical data of their organizations to estimate construction duration. However, variations in site conditions are not usually taken into consideration. This paper presents a Discrete Event Simulation (DES) model that can help decision-makers accurately estimate durations of underground watermain construction projects and review key schedule milestones in detail. A case study is conducted on an actual project of a municipality in Canada. Different site conditions and the municipality's design criteria are taken into consideration to accurately reflect real world conditions. Chainage, a distance measuring system commonly used to design and locate linear infrastructure, is used as a key attribute to control the overall model and provide status of model. In addition, compaction test results that may affect the schedule at certain probability are modeled as well to measure the possible delays caused by such failures.

Keywords: discrete event simulation, construction scheduling, project management, construction management

1. INTRODUCTION

Installing, or upgrading underground watermain is one of the most common design and construction works of municipalities, civil engineers, and civil contractors. Typically, a section of watermain is a linear network of multiple pipes which requires various resources such as crews and equipment to work together. Practitioners in the construction industry often use their own experiences or historical data collected by their organizations to estimate construction schedule. However, different site conditions are often ignored in this conventional way.

With a rapid urban expansion, water distribution network becomes an important issue for urban planning and infrastructure construction. In the City of Toronto, many pipes are beyond their theoretical service life of about 80 years (City of Toronto 2009). It is also predicted by the United States Environmental Protection Agency (U.S. EPA) that in the United States alone, \$138 billion will be required to replace and maintain the existing drinking water systems during the next 20 years (Selvakumar, Clark and Sivaganesan

2002). Rehan, Knight, Unger and Haas (2013) developed causal loop diagrams and a system dynamics model for financially sustainable management of urban water distribution networks and discussed on how the developed system dynamics water model can be used by water utilities to achieve a variety of utility short and long-term objectives. As a result, Asnaashari, McBean, Gharabaghi and Tutt (2013) concluded that the water industries in cities need an intelligent system that can combine all recorded data to analyze the complex relationships in the data and to assist the processes of decision-making for asset management.

With the utilization of Discrete Event Simulation (DES), this paper presents an effective tool that can help decision-makers to more accurately estimate construction schedule of underground watermain construction and enable them to review detailed schedule of activities of a project.

2. LITERATURE REVIEW

Simulation technology has been implemented in construction industry for many years. In 1976, Halpin (1997) introduced the cyclic operation network (CYCLONE) to construction industry and since then, many scholars did further study based on Halpin's work, including but not limited to INSIGHT (Paulson 1978), RESQUE (Chang and Carr 1987), UMCYCLONE (Ioannou and Ioannou 1990), and DISCO (Huang, Grigoriadis and Halpin 1994).

Simphony is a useful simulation tool which was developed by University of Alberta (Hajjar and AbouRizk 1999), and it provides a framework for developing General Purpose Simulation (GPS) and Special Purpose Simulation (SPS) templates to help users create models based on their knowledge of simulation as well as the construction domain (Moghani, Sander, AbouRizk and AbouRizk 2011). Simphony.NET supports general purpose modeling constructs (e.g. CYCLONE) which can be used to model different construction processes, and allows users to build models utilizing abstract elements such as activities, queues, and resources. Hajjar and AbouRizk (1997) introduced special purpose simulation (SPS) and implemented it with Simphony. Special purpose simulation was designed to facilitate adoption of simulation by industry and uses a visual object-oriented modeling environment to provide a simple way for non-

expert simulators to control simulation. AbouRizk and Mohamed (2000) discussed Symphony as an integrated environment for developing Special Purpose Simulation Tools. Samples of the templates were developed using Symphony were presented to highlight some of the powerful features that can be easily included in templates by a developer. In 2002, with the chosen three construction methods in earthmoving, aggregate production and site dewatering, Hajjar and AbouRizk updated Symphony as Symphony.NET, which allows for quick, flexible analysis of various construction plans; for instance, it is a simple matter to change the soil type, the number of backhoes, or the number of crews and observe the outcome. Symphony uses a hierarchical, modular approach which considerably simplifies the development and use of complex and large simulation models (Hajjar and AbouRizk 2002). In this research, simulation model was developed by Symphony. NET 4.0 which is the latest version of Symphony.

Discrete event simulation (DES) is a particular type of dynamic simulation which could be processed by advancing time in discrete activities based on critical events. An event in the context of DES could be defined as an instant of time when an important state change occurs in the system (Pidd 1988; Labban, AbouRizk, Haddad and Elserly, 2013). The simulation model generally starts with a given event, which triggers other events, until a termination point is met. The methodology and algorithms behind DES, which concerns “the modeling of a system as it evolves over time by a representation where state variables change only at a countable number of points in time” (Law and Kelton 1982) provides an alternative approach to project control. By predicting the future condition of a real construction system following a computer model which is based on real life statistics and operations (Lu 2003), DES is reliable and could be advisable for decision-makers. Nowadays, DES has become prevalent in various fields, such as manufacturing, telecommunication, supply chain, and finance systems. A simulation model can be used as a flexible tool to estimate the significance of different site conditions (Agbulos Mohamed, Al-Hussein, AbouRizk and Roesch 2006). In general, construction process of watermain installation is linear and could be repeated when site is changed. Each task in this process has distinct duration

and the project duration is accumulated as the task moves from one to another. Therefore, the proposed methodology in this paper is generating DES analysis with the implementation tool as Symphony.Net 4.0 (General Purpose Simulation), in order to simulate project duration and track the construction process. A simulation model of Symphony can have multiple runs, which allows a user to run scenarios and compare the results generated by statistical inputs and output.

3. CASE STUDY: WATERMAIN UPGRADE PROJECT

3.1. Project Description

A municipality in British Columbia, Canada, finished a upgrade project for approximately two kilometers of existing water distribution lines, located in its one of Northern neighborhoods. The project was to install new watermain lines and abandon existing ones that exceed their service life. The actual construction started in winter of 2012 and finished in spring of 2013, and a rough construction schedule had been estimated by using their historical data prior to the construction. However, actual construction schedule was not tracked and recorded properly due to the frequent interruptions from the city’s other construction projects where the same crew was utilized, and also due to the long holiday break in the middle of project. This study focuses on simulating project schedule and various activities based on observations of the construction inspector of the project, and developing a more accurate tool that can be used to forecast the construction schedule of similar projects.

The project included eight different sites within an area of approximately 400 metre radius. Typical plan view of a site is shown in Figure 1. Each site has its own conditions which have to be considered during construction schedule planning. Those conditions include frequency of traffic, steepness of slope, and seismic area. Each site has different length of pipes, and different number fire hydrants and service connections to install. Activity durations vary based on these factors because different construction methods, that are specified in the Design Criteria further discussed in the next subsection, have to be applied.

Pipe size varies between 4 inches to 8 inches (100mm

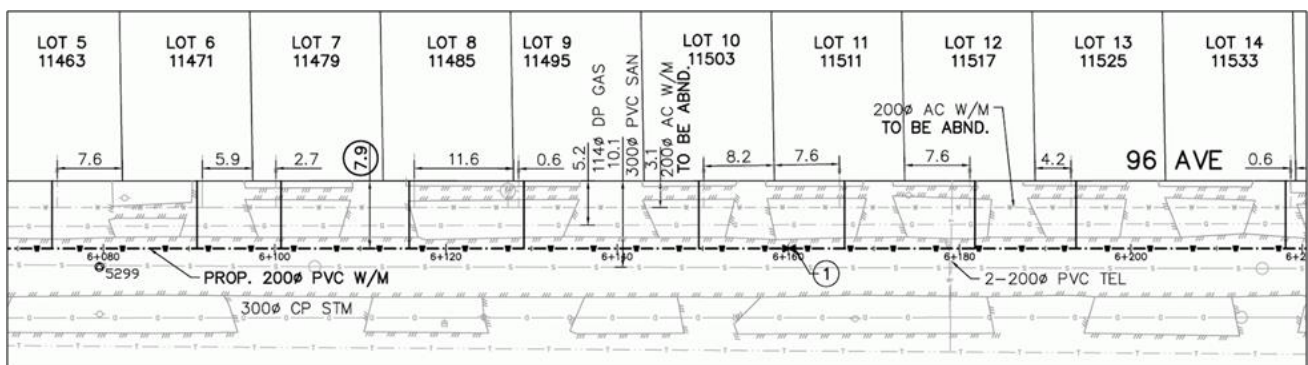


Figure 1: Typical Plan View of Watermain Construction Drawing (City of Surrey 2012)

to 200mm diameter). Ductile iron (DI), high-density polyethylene (HDPE), or polyvinyl chloride (PVC) pipes are used. Typical pipe length is 20 feet, which is 6.1 meters.

The city utilized its own workforce (hereafter referred to as the city crew) for the watermain pipe installation part, and hired third party contractors for pavement saw-cutting, traffic control and surface restoration. The city crew included three labourers, and one backhoe with its operator. Survey layout was done by the survey department of the city, and a geo-technical engineering firm was responsible for compaction tests.

3.2. Design Criteria Review

Design Criteria Manual released by the city's engineering department (City of Surrey 2004) was reviewed to reflect more detailed installation of water pipe installation. Key findings are as follows:

- **Grade:** When the slope of a water main equals or exceeds 10%, pipe anchor has to be used. Also, ductile iron pipe has to be used.
- **Fire hydrant:** Hydrants shall not be spaced more than 200 meters apart. Quantity and locations of hydrants are normally determined during the design phase.
- **Seismic Area:** If the project site is located within the seismic area, all pipeline shall be restrained so that they will not pull out when subjected to extension forces. Also, to minimize soil-pipe interaction, pipe shall be wrapped with polyethylene (baggy) such as is commonly used for corrosion protection. The intent of the wrapping is not to provide corrosion protection.

The conditions of sites are summarized in Table 1.

3.3. Construction Process and Activities

Typical construction process of the underground watermain installation can be summarized as shown in Figure 2. Prior to initializing actual construction, the city sends its survey crew to conduct the survey layout based on construction drawings prepared by a design firm. Once the survey layout of the first site is complete, a saw-cutting contractor is brought in to saw-cut paved area according to the survey layout marked on site. These two activities can proceed to the next sites without any dependency on the city crew's pipe installation.

Once the saw-cutting of the first site is done, the city crew can initiate their activities. First, the crew mobilizes in a location among the eight sites that they can access to the trailer from any of sites throughout the whole project. Therefore, mobilization only happens once at the beginning of main cycle of the first site.

Before starting a main cycle of a site, the crew has to determine whether traffic control is necessary for safety by checking the traffic frequency during the day time and the alignment of pipes on the road. If needed, traffic control crew is hired. Once the traffic control crew is in place, the backhoe can start removing pavement, or topsoil, and excavating trench. Whether the area is paved or not, the difference of activity duration is assumed to be negligible because the pavement is already saw-cut. A haul truck is also required to load the excavated earth, but it is ignored in this study. If trench is excavated enough as shown in Figure 3, then two laborers prepare pipe bedding and level it so a pipe can be installed as designed. Once the pipe bedding is ready, a pipe is laid down by backhoe using chainlink, and one laborer is required to assist with this activity. If the site is located in the seismic area, then joint restraints and pipe wrapping have to be applied to the pipe before connecting it to proceeding pipe. If the pipe is ready for connection, then two laborers connect the pipe to the proceeding pipe and seal it. If the slope of pipe equals or exceeds 10%, then pipe anchoring has to

Table 1: Conditions of eight sites

	Site (total of 8 sites in the project)							
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8
Conditions								
Total project length(m)	99	247	325	170	452	156	137	269
Start of chainage (m)	1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000
End of chainage (m)	1,099	2,247	3,325	4,170	5,452	6,156	7,137	8,269
Traffic Control req'd?	No	No	No	No	Yes	No	No	No
Length (m) of trench where slope > 10%	63.8	59.2	200.0	0.0	0.0	100.0	0.0	0.0
Start of chainage (m) w/slope > 10%	1,020.0	2,123.0	3,070.0	0.0	0.0	6,055.8	0.0	0.0
End of chainage (m) w/slope > 10%	1,083.8	2,182.2	3,270.0	0.0	0.0	6,155.8	0.0	0.0
Seismic area?	Yes	Yes	Yes	Yes	No	Yes	Yes	No
No. of service connection?	4	13	26	13	22	15	14	14
No. of fire hydrant?	1	1	1	0	2	1	1	1
Length (m) of paved area	99	172	325	15	452	156	137	79
Length (m) of landscaped area	0	75	0	155	0	0	0	190

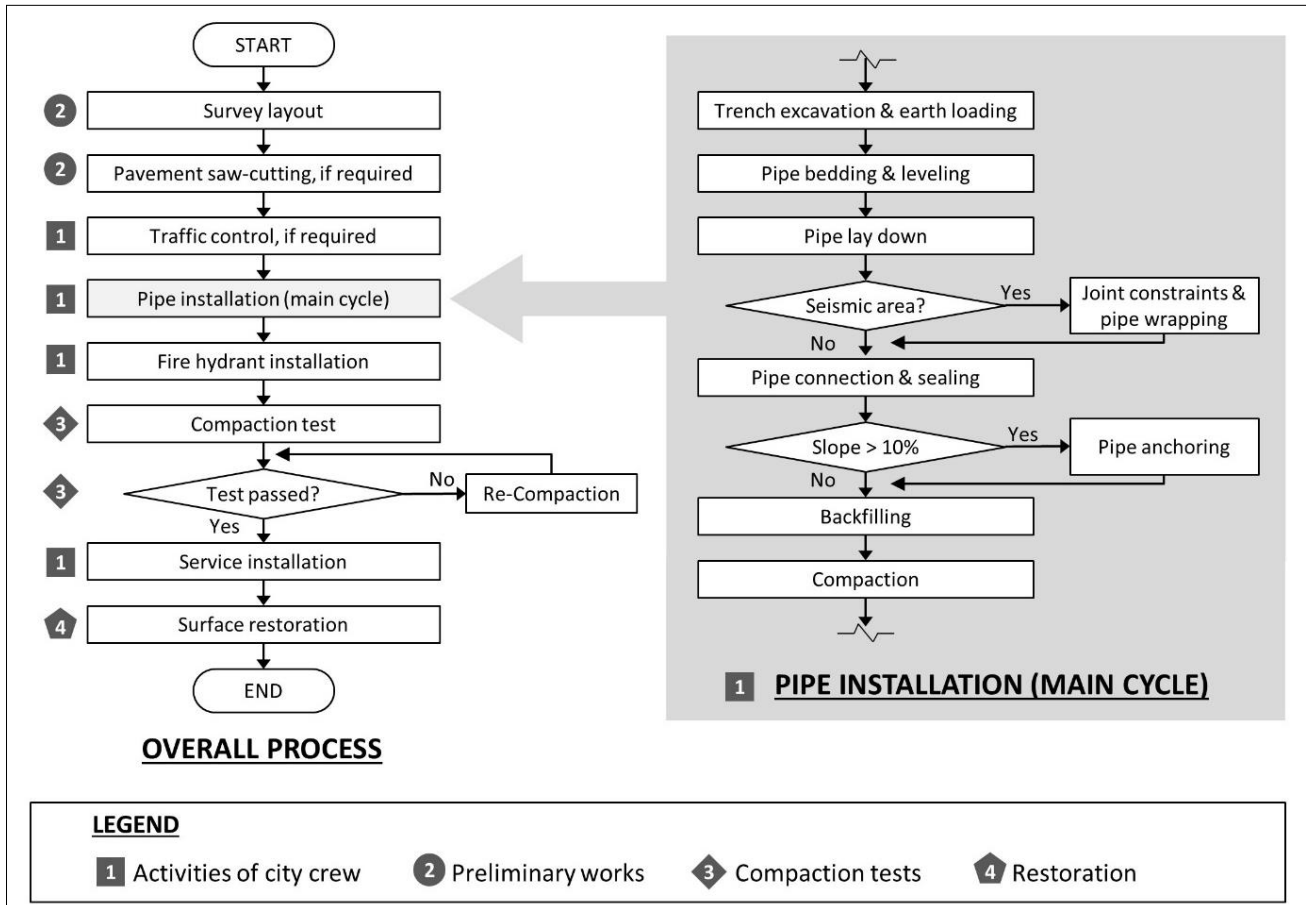


Figure 2: Construction Processes

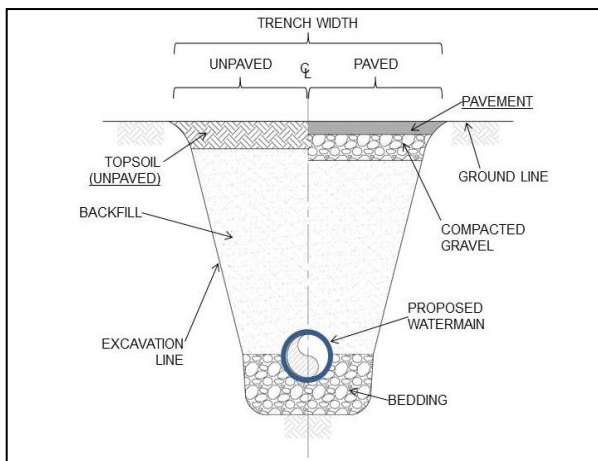


Figure 3: Typical Watermain Trench Detail

be applied by using concrete and tie-rods. When pipe installation is complete, the backhoe can backfill the trench with assistance of one laborer. The backhoe is also responsible for compacting the backfilled area with its vibratory plate installed on the opposite side of its body. This series of main cycle happens repetitively at every one pipe length (20ft, or 6.1m) until the city crew reaches the end of a site. Once the city crew completes this series of main cycles for a site, then the crew can move on to the next site. As previously mentioned, sites are saw-cut by a third party contractor independently;

therefore, no delay is expected when the crew moves to the next site.

While the city crew is working on the other sites, the project manager of the city requests a geo-technical engineering firm to conduct compaction test. The geo-tech firm sends a technician at the end of every second sites (i.e. when the main cycles of 2nd, 4th, 6th, and 8th sites are complete), and conduct compaction tests. Then the firm analyzes the test results collected by its technician, and forwards the result to the project manager of the city. If the trenches are found to be compacted to the level satisfying the city's standard, no further action is required. If the test is found to be unsatisfactory, then the project manager asks the city crew for re-compaction. Then, the backhoe is sent back to the site of which compaction was not done properly, and redo the compaction. It happens only when the crew moves between different sites. For this project, the compaction is assumed to be failed in the possibility of 10% (one fail out of ten tests), and the impact from the test failures will be measured in the simulation.

Once the city crew finishes the main cycles of all eight sites, then it moves back to each site to complete fire hydrant installation and service connections from the newly-installed watermain to each property. Installation of service connection can start only when all compaction test results are satisfactory. The city crew demobilizes when all the above activities are complete.

Surface restoration, either pavement or landscaping depending on existing site conditions, starts when all the service connections are done, and it is done by third party contractors that are hired by the city. The whole project ends when surface restoration is complete. Durations of activities are summarized in Table 2. When sample observations are not available, a simulator can use subjective judgment of an expert knowledgeable about the work being modeled. A simulator often uses a triangular (low, high and mode estimates) or uniform (low and high estimates) distribution in such case (AbouRizk and Halpin 1992). In this paper, opinion from the project inspector was used to determine activity durations, and they are

assumed to be triangular distributions for simplicity of the simulation model. However, note that mode values only are shown in the table for clarity. For instance, the mode value of Mobilization activity is assumed to be 4 hours and presented in the table, although low (minimum) estimate of 3.5 hours and high (maximum) value of 4.5 hours are also used in the actual simulation model.

4. MODEL DEVELOPMENT

The algorithm for the simulation follows the flowchart, Figure 2, previously presented in the project description. As the first step of model development, all the parties involved were identified, namely city crew,

Table 2: Description of Construction Activities

Name of activity	Duration	Done by	Remarks
Before main cycle			
Survey layout	2 hrs/100m	City's survey crew	Independent activity
Pavement saw-cutting	1 hrs/30m	Saw-cutting contractor	Independent activity
Mobilization	4 hrs	Whole city crew	One time item
Main cycle (mainly by city crew)			
Traffic control	0.5 hours	Traffic controller contractor	One time item, at the beginning of a site, only when applicable
Excavation & earth loading	6.5 min/3.05m	Backhoe	
Pipe bedding & leveling	6.5 min/6.1m	Laborer(s)	
Pipe laydown	7.0 min/6.1m	Backhoe & Laborer(s)	
Joint Restraint & Pipe wrapping	Add 50% time to 'Pipe connection & seal'	Laborer(s)	Applicable only to seismic area
Pipe connection & sealing	15.0 min/6.1m	Laborer(s)	
Pipe anchoring	25 mins / each pipe layout	Laborer(s)	For where slope is greater than 10% only, Each at end of every pipe (6.1m)
Backfill	7.5 min/6.1m	Backhoe	
Compaction	5.0 min/6.1m	Backhoe	
Moving to Next Site	30 min	Whole city crew	
<i>*Once the city crew completes the series of main cycles of a site, the crew moves to the next site.</i>			
City crew's work outside of Main Cycle			
Installing fire hydrant	4 hr/ea	Backhoe & laborer(s)	Happens at the end of a site when all main cycles are complete
Installing service Connection	1 hr/ea	Backhoe & laborer(s)	Can start after fire hydrant are all installed and passing compaction test
De-Mobilization	4 hrs	Whole city crew	One time. Happens after all the service connections are complete
Other's work after each Main Cycle			
Compaction test	4 hours/each time	Geo-tech technician	Happens at end of site 2, 4, 6, and 8. If failed, backhoe is captured to redo the compaction
Contacting geo-tech firm, & Receiving test result	16 hr/ea	City crew & Geo-tech	Can overlap with other activities
Re-Compaction	45 min/site	Backhoe	
Service connection	1 hr/ea	Backhoe & laborer(s)	Can start only after passing all the compaction tests
Surface restoration, Pavement	4 min/m	Paving contractor	Can start only when the city crew is ready for demobilization
Surface restoration, landscaping	1.5 min/m	Landscaping contractor	Can start only when the city crew is ready for demobilization

survey crew, traffic control crew, pavement saw-cutting crew, geo-technical engineering firm, pavement and landscaping crew.

Instead of developing a model for each crew or firm's activity, they are grouped into four groups to simplify the simulation model. For instance, pavement saw-cutting always follows survey layout, in other words, they have finish-to-start relationship. Their activities take place independently prior to city crew's work, and do not interrupt other activities. Therefore, they can be grouped. Through analysis of the whole construction process from the perspective of model design, four different sub-models were created in the main model, as shown in Figure 2 and Table 3, to simulate the watermain upgrade project using General Purpose Simulation template in Symphony.Net 4.0.

Table 3: Four sub-models

Model Description	Role in the main model
Activities of city crew	To simulate the activities of the city crew
Preliminary works	To simulate the preliminary works (survey layout & Pavement saw cutting) conducted independently by third party contractors
Compaction Test	To simulate compaction tests and their results. It captures backhoe resource when the test is failed.
Restoration	To simulate restoration works (either pavement or landscaping) that can start at the end of all activities of the city crew

Coding plays important role in this model. Global and local attributes, GX and LX respectively, are designed to carry information of different entities and current simulation time to avoid redundant elements and make the model more effective. The different conditions of each site, that are previously summarized in Table 1, are coded as shown below in the model so each site can be processed differently in the simulation. The activities of the city crew is mainly controlled by the LX(1), which is the start of chainage and is also used as the current chainage of an entity during the process. LX(1) increases by 6.1 (length of one pipe) at the end of each cycle for this purpose.

```
//Initiate each site{
Redim LX(20) //'Expand number of local
attributes
```

```
LX(0)=1 //'Site number
LX(1)=1000 //'Start of chainage (m)
LX(2)=1099 //'End of chainage (m)
LX(3)=0 //'Traffic control required?
(1:T, 0:F)
LX(4)=1 //'Seismic area? (1:T, 0:F)
LX(5)=99 //'Length(m) of paved area
LX(6)=0 //'Length(m) of landscaped area
LX(7)=4 //'Number of service connections
LX(8)=1 //'Number of fire hydrant
Return True};
```

Durations of optional activities, such as traffic control, Joint Restraint, Pipe wrapping and Pipe anchoring, are processed also based on these attributes of an entity. Example of optional activity and its coding is shown in Figure 4 and the following coding.

```
//Control Pipe anchoring {
//'LX(1) is current chainage
If LX(1)+6 > 1020 and LX(1) < 1083.8
Return True
Else If LX(1)+6 > 2123 and LX(1) < 2182.2
Return True
Else if LX(1)+6 > 3070 and LX(1) < 3270
Return True
Else if LX(1)+6 > 6055.8 and LX(1) <
6155.8
Return True
Else
Return False
End if };
```

Key activities, including start/end of main cycle, corresponding with geo-tech, compaction tests and their results, optional activities, etc, are monitored by using 'trace' element, which provides comments whenever an entity passes the elements. Examples of comments generated by trace elements are shown below, followed by corresponding codings. Note that the time unit in this model is minute.

Preliminary work (Survey layout & pavement sawcut) for Site No.1 is complete at time 316.8
Crew starts main cycles of site no.1 at time 316.8
Crew completed mobilization and is starting main cycle at time 552.4
Chainage 1000 to 1006.1 is complete at time 614.1
Pipe anchor has been applied at chainage 1018.3 (slope > 10%)
Crew finished FH installations of site1 at time 2089.7
Crew starts main cycles of site no.2 at time 2123.8
Geo-tech arrived for compaction tests of site 3 and site4 at time 12215.9

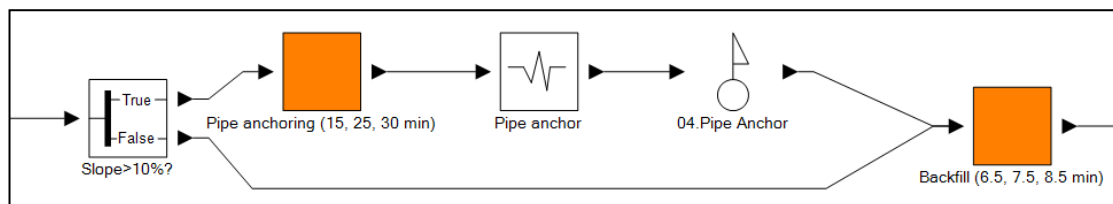


Figure 4: Conditional Decision Element for Installation of Pipe Anchoring

Table 4: Schedule of Key Activities

Key activity	Mean of Event End time*			Standard deviation
	minutes	hours	days	minutes
Mobilization	556	9.3	1.2	12
End of main cycles for all 8 sites	22,570	376.2	47.0	60
Installation of service connections for all 8 sites	30,929	515.5	64.4	485
Demobilization	31,228	520.5	65.1	537
End of restoration works	38,953	649.2	81.2	679

* Measured from the beginning of survey layout work

Compaction test FAILED for site3 and site4. Result received at time 13197.4

Backhoe is captured for recompaction of site3 and site4 at time 15910.7

Crew is starting service connections of site 1 at time 22610.6

Crew finished service connections of site 1 at time 22853.2

Crew demobilized at time 31035.7

Landscape restoration is finished at time 33778.9

Pavement restoration is finished at time 39016.2

Project is complete at time 39016.2 (Total of 81.2 working days)

```
//Coding of trace element for event of
compaction test fail{
Return "Compaction test FAILED for site"
& LX(1) & " and site" & LX(1) + 1 & ".
Result received at time " & Timenow};
//Coding of trace element for project
completion{
Return "Project is complete at time " &
CSTR(Timenow) & " (Total of " &
Timenow/60/8 & " working days)";
```

A resource is used in the model to control the activities of backhoe rather than to actually model the utilization of resources. For example, when a compaction test is failed, the backhoe is sent to the site for re-compaction. However, this activity only happens when the city crew moves between different sites, and therefore the backhoe is captured at the beginning of each site and released only at the end of each site. Higher priority is used for re-compaction activity to contact geo-tech engineering firm as early as possible and therefore to minimize delay in the project schedule. Another resource is modeled for three laborers of the city crew to check utilization of them.

5. SIMULATION RESULT ANALYSIS

Each simulation model was executed for 100 runs to provide users with statistical output such as project duration and cycle times. Also, impact of the 10% failure probability of compaction test on the total project duration was analyzed.

5.1. Project Duration

Based on 100 runs of simulation, the total project duration is found to be the mean of 38,953 minutes

(649.2 hours, or 81.2 working days assuming 8 hours per day) and standard deviation of 679 minutes (See Figure 5). Schedules for other key activities are modeled as shown in Table 4. 95% confidence interval for the mean of total project duration was calculated to be between 38,824 minutes and 39,082 minutes. Although the total project duration and event end times of key activities can be used as effective information for schedule estimate, the user cannot know when the failures of compaction test happen and where they impact on the overall project schedule.

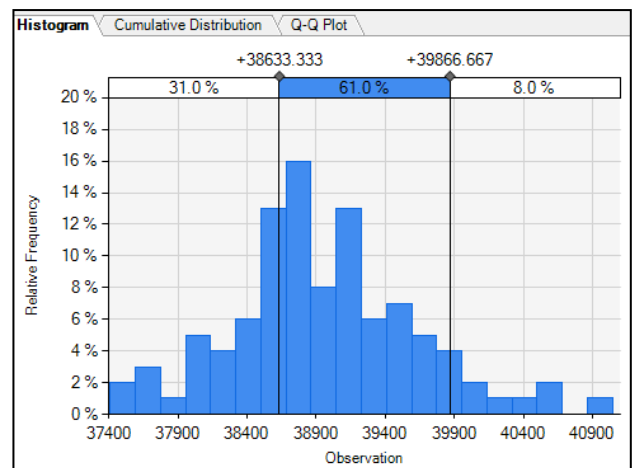


Figure 5: Total Project Duration (minute)

5.2. Cycle Time per One Pipe Length

Cycle time per each pipe length can be more desirable for more accurate schedule estimate, because it is not impacted by the 10% of probability of compaction test failure. Cycle times of this model are measured at the end of each main cycle. Although each site has different conditions and therefore the cycle time can vary, DES analysis can be used as an effective tool for schedule estimate with all the different conditions taken into the consideration. The result of cycle time is shown in the histogram, Figure 6. The mean of cycle time was measured to be 64.3 minutes per each pipe length (6.1m, 20 feet) with standard deviation of 0.14 minutes. 95% confidence interval for the mean of cycle time was calculated to be between 64.27 minutes and 64.33 minutes.

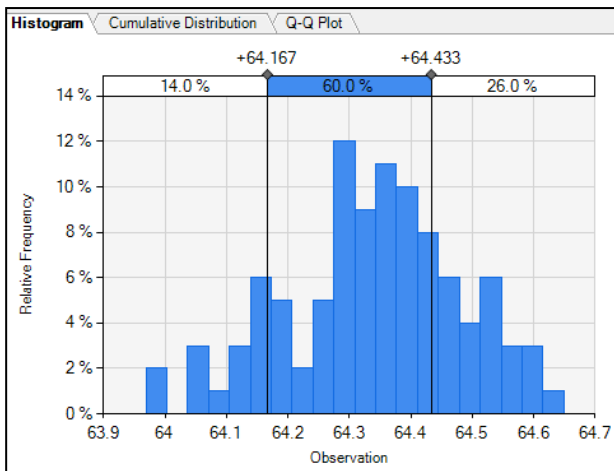


Figure 6: Cycle Time per One Pipe Length (minute)

5.3. Impact from the 10% Failure of Compaction Test on Project Schedule

As previously discussed, impact on the total project duration that can be caused by 10% failure probability of compaction test can be analyzed in this model by checking the result from 100 runs. For this purpose, two different scenarios were tested, Case 1) with 10% probability and Case 2) with 0% probability of test failure. The impact on project schedule was measured by subtracting the total project duration of Case 2 from that of Case 1. The result of this analysis is shown in Table 5.

Table 5: Calculating Impact of 10% Probability of Compaction Test Failure

	Case 1	-	Case 2	=	Impact
Minutes	38,953		38,827		126

Because the duration and number of compaction tests were modeled based on number of sites, not length of sites, the impact can be understood as a percentage factor of project duration. In this case, the overall impact on the total project duration caused by 10% failure probability of compaction test can be calculated as 0.42%. ($126 \text{ minutes} / 30,373 \text{ minutes} * 100 = 0.42\%$) Note that the only duration between the end of mobilization and the end of service connection installations for all 8 sites are used in this calculation, because the other activities outside the duration happen independently and therefore do not need to be considered.

5.4. Utilization of Resources

Utilization of resources is not discussed in this study because the City usually hires, or utilizes, enough number of laborers and equipment and therefore resources do not play important role in the model.

5.5. Summary of Results

Based on the historical data and his own experience, the foreman of the city uses '40 meters per day' as his own 'rule-of-thumb', which is a typical pipe installation

distance per day (m/day). This rule can be used as an expert's opinion for the purpose of validating the model developed in this study. Because his rule-of-thumb is used to estimate construction schedule of main cycles only, the duration only between the end of mobilization and the end of main cycles of all 8 sites was considered for the validation. Total length of the project is 1,855 meters and the duration was modeled, based on 100 runs, to be 22,005 minutes, which is 45.8 days. Thus, the average distance of pipe installation based on the model is calculated to be 40.5 meters per day, and it is found to be significantly close to the foreman's rule-of-thumb, 40 meters per day.

The result demonstrates that the model developed in this study can be considered satisfactorily accurate in forecasting duration of main cycles of watermain construction. Other parts of model can be validated in the further study.

6. LIMITATIONS

One limitation of this study is that the simulation model cannot consider every potential scenario such as bad weather condition, employees' vacation and unexpected delays. For instance, the actual crew of the construction experienced a unexpected delay that was caused by heavy rain during winter season. The crew had to keep dewatering trenches throughout the pipe installation, which significantly affected the construction schedule. Also, the crew sometimes encountered old underground utility lines which had not been recorded in the municipality's as-built drawings.

The simulation model presented in this study assumes a continuous work flow. Numbers of days are calculated by converting the simulation time, i.e. minute, to days using 8 hours per day. In reality, however, additional minor activities can happen at the beginning and end of a day or before and after a lunch break, which include equipment inspection, site cleaning, and moving between site and site office.

Finally, before actually implementing this type of simulation for schedule estimation, each simulation model needs to be validated and gain support from the municipality, which a simulation model would be implemented for, and its in-house experts including construction crew supervisor.

7. CONCLUSION

Underground watermain construction is one of the most common design and construction works of municipalities, civil engineers, and civil contractors. However, practitioners usually use their own experiences or historical data collected by their organizations to estimate construction schedule. This paper presents a Discrete Simulation Model developed in Symphony.Net 4.0 that can help decision-makers to more accurately estimate construction schedule and to review the milestones of schedule in detail. DES analysis was also presented to provide users with statistical output such as project duration and cycle times. The result of this study demonstrates that the

simulation model can be considered satisfactorily accurate in forecasting duration of main cycles of watermain construction.

REFERENCES

- AbouRizk, S., & Halpin, D., 1992. Statistical properties of construction duration data. *Journal Of Construction Engineering And Management*, 118(3), 525-544.
- AbouRizk, S., & Mohamed, Y., 2000. Symphony: an integrated environment for construction simulation. In *Proceedings of the 32nd conference on Winter simulation*, pp. 1907-1914. December 10-13, Orlando (Florida, USA).
- Agbulos, A, Mohamed, Y., Al-Hussein, M. AbouRizk, S., & Roesch, J., 2006. Application of lean concepts and simulation analysis to improve efficiency of drainage operations maintenance crews. *Journal Of Construction Engineering And Management*, 132(3), 291-299.
- Asnaashari, A., McBean, E. A., Gharabaghi, B., & Tutt, D., 2013. Forecasting watermain failure using artificial neural network modelling. *Canadian Water Resources Journal*, 38(1), 24-33.
- Chang, D. Y., & Carr, R. I., 1987. RESQUE: A resource oriented simulation system for multiple resource constrained processes. In *Proceedings of the PMI Seminar/Symposium*, pp. 4-19. October 2-7, Milwaukee (Wisconsin, USA).
- City of Toronto. 2009. Watermains and watermain breaks. Water supply: fact sheets. Available from: http://www.toronto.ca/water/supply/system/water_mains.htm. [Accessed: January 2011].
- City of Surrey, 2004. British Columbia. "Design Criteria Manual".
- City of Surrey, 2012. British Columbia. "Kennedy North Distribution Mains Upgrade" Construction Drawings.
- Hajjar, D., & AbouRizk, S. 1997. Development of an object oriented framework for the simulation of earth moving operations. In *Intelligent Information Systems, 1997. IIS'97. Proceedings* (pp. 326-330). IEEE.
- Hajjar, D., & AbouRizk, S., 1999. Symphony: An Environment for Building Special Purpose Construction Simulation Tools. *Winter Simulation Conference*, pp. 2998-1006. December 5-8, Phoenix (Arizona, USA).
- Hajjar, D., & AbouRizk, S. M., 2002. Unified modeling methodology for construction simulation. *Journal of Construction Engineering and Management*, 128(2), 174-185.
- Halpin, D. W., 1997. CYCLONE-method for modeling job site processes. *Journal of the construction division*, 103(ASCE 13234 Proceeding).
- Huang, R., Grigoriadis, A. M., & Halpin, D. W., 1994. Simulation of cable-stayed bridges using disco. *Proceedings Of The 26th Conference Winter Simulation*, pp.1130-1136. December 11-14, Orlando (Florida, USA).
- Ioannou, P. G., & Ioannou, C. P. G., 1990. UM-CYCLONE discrete event simulation system user's guide.
- Labban, R., AbouRizk, S., Haddad, Z., & Elserly, A., 2013. A discrete event simulation model of asphalt paving operations. *Proceedings Of The 2013 Winter Simulation Conference*, pp. 3215-3224. December 8-11, Washington (D.C., USA).
- Law, A. M., & Kelton, W. D., 1982. *Simulation modeling and analysis*. New York : McGraw-Hill.
- Lu, M., 2003 Simplified discrete-event simulation approach for construction simulation. *Journal Of Construction Engineering And Management*, 129(5), 537-546.
- Moghani, E., Sander, H., AbouRizk, H., & AbouRizk, S. M., 2011. Analyzing transit tunnel construction strategies using discrete event simulation. *Proceedings Of The 2011 Winter Simulation Conference (WSC)*, pp. 3505-3515. December 11-14, Phoenix (Arizona, USA).
- Paulson, B. C., 1978 Interactive graphics for simulating construction operations. *Journal of the Construction Division*, 104(1), 69-76.
- Pidd, M., 1988. *Computer simulation in management science*. Chichester ; New York : Wiley.
- Rehan, R., Knight, M. A., Unger, A. J. A., & Haas, C. T., 2013. Development of a system dynamics model for financially sustainable management of municipal watermain networks. *Water research*, 47(20), 7184-7205.
- Selvakumar, A., Clark, R. M., & Sivaganesan, M., 2002. Costs for water supply distribution system rehabilitation. *Journal of water resources planning and management*, 128(4), 303-306.

AUTHORS BIOGRAPHY

Dongwoo Kim is a Master's of Science (M.Sc) student of the Hole School of Construction Engineering at University of Alberta, Canada, and also currently working for the City of Edmonton in Canada. Prior to it, he has worked as a civil/municipal engineering designer in Canada, and as an urban planner in South Korea.

Yi Zong is a Master's of Science (M.Sc) student of the Hole School of Construction Engineering at University of Alberta, Canada.

Dr. Yasser Mohamed is an associate professor of the Hole School of Construction Engineering at University of Alberta, Canada. His research interest focuses on modeling construction processes using discrete-event simulation techniques.

Dr. Simaan Abourizk is a professor of the Hole School of Construction Engineering at University of Alberta, Canada. His research focuses on developing innovative information technologies for modeling, analyzing, and optimizing operations in the construction and natural resource extraction industries.

A FLEXIBLE AND GRAPHICAL OPERATIONAL EFFECTIVENESS EVALUATION TOOL

Wei LI ^(a), Shenglin LIN ^(b), Zhizhao LIU ^(c), Ping MA ^(d), Ming YANG ^(e)

Control and Simulation Center, Harbin Institute of Technology, Harbin 150080, China

^(a)frank@hit.edu.cn, ^(b)lin_44627079@yeah.net, ^(c)liuzhizhao2007@hotmail.com
^(d)pingma@hit.edu.cn, ^(e)myang@hit.edu.cn

ABSTRACT

To solve the reusability of evaluation information and extensibility of evaluation algorithms, and improve the efficiency of operational effectiveness evaluation of weapons, a flexible, graphical and simulation-based evaluation tool is developed. The problem analysis of effectiveness evaluation and solutions are given firstly. A process model for effectiveness evaluation is presented, which is flexible and suitable for different evaluation tasks and objects and a primitive-based graphical interface is designed to improve the usability of this tool. And the reusability mechanism of evaluation information and extensibility interface of evaluation algorithms are designed. Then, the design and implementation of the evaluation tool are described and an example of simulation-based operational effectiveness evaluation of an air-defense system is illustrated how to use the tool. Finally, the conclusion and future work are summarized.

Keywords: operational effectiveness, simulation-based evaluation, flexible evaluation, graphical interface

1. INTRODUCTION

Operational effectiveness is used to describe the degree of weapon system accomplishing the missions in the specific operational environment. It is an important comprehensive index for evaluating the quality of weapons and the basic reference for developing, equipping and using the weapons. Through evaluating operational effectiveness of weapons, the periods and cost of design and manufacture of weapons can be reduced and the comprehensive performance of weapons can be improved. The simulation-based evaluation method of weapon operational effectiveness gains the operational effectiveness value by performing simulation experiments and processing the simulation result data. Because the simulation-based method has the advantages of safety and economical and many different operational conditions can be simulated, it is more popular for evaluating the weapon operational effectiveness (Jiao 2013).

With the development of the weapon technology and more and more complex situation of warfare, the simulation-based effectiveness evaluation for weapon system has been raised new requirements, which

involve multiple scenario schemes, large amount simulation experiments and data, complex data management and analysis. So the auxiliary tool needs to be developed to support the simulation-based evaluation process for operational effectiveness of weapons. For assisting the operational effectiveness evaluation, some researchers have provided the effectiveness evaluation frames and tools. U.S. RAND gave a mixed interactive evaluation frame based on multi-attribute decision-making and exploratory modeling (Pinder 2000). Li and Matthew proposed an effectiveness evaluation framework of the automatic target recognition and classification system (Li 2014, Matthew 2015). According to the characters of Anti-torpedo Torpedo Weapon System (ATTWS), the effectiveness evaluation system of ATTWS is established, and then a method based on the cloud model is proposed to achieve the effectiveness evaluation for ATTWS (Xu 2015). A distributed denial of service (DDoS) protection effectiveness evaluation system (DPEES) in Linux system is proposed (Qu 2012). An effectiveness evaluation system (EES) is proposed to eliminate the possible risk existing in the development of automatic test systems (ATS) (Wang 2010). The specific evaluation tools for the operational effectiveness of electronic warfare system (Lancon 2011, Qi 2014) and the ship system (Akyuz 2014) are also developed. In addition an integrated software environment of simulation experiment design, analysis and evaluation is researched in our previous work (Li 2016).

The existing effectiveness evaluation frames and tools are not applicable for the general evaluation objects and their complex evaluation procedures are only suitable for the professional, not easy to learn and apply for general users. Besides, the practicality and human-computer interaction (HCI) characteristic of these tools need to be strengthened further. So a flexible and graphical operational effectiveness evaluation tool, HIT-OEET (Harbin Institute of Technology Operational Effectiveness Evaluation Tool), is designed and implemented, which can be applied to various evaluation objects and has good reusability and extensibility. The remainder of this paper is organized as follows. The existing problems of effectiveness evaluation and solutions are analyzed in Section 2. Section 3 describes design and implementation of HIT-

OEET. And an example of simulation-based operational effectiveness evaluation is used to illustrate the efficiency and usage of HIT-OEET in Section 4. Finally, the conclusion and the future work are summarized.

2. THE PROBLEM ANALYSIS AND SOLUTIONS

In the simulation-based operational effectiveness evaluation applications, many difficulties are faced, such as the different evaluation tasks and objects, complex components of weapon system, plenty of operational scenarios and simulation experiments, a large quantity of complex simulation data, and many evaluation algorithms. So the main problems of the effectiveness evaluation to be solved are summarized below:

- **Flexible evaluation (FE).** The evaluation tool could apply for different effectiveness evaluation objects, which include signal equipment, the weapon system and system-of-systems (SoS), in different operational scenarios. At this point, the evaluation process is not constant and could be constructed flexibly according to the different evaluation requirements.
- **Human-computer interaction (HCI).** Recent researches indicate that HCI techniques can contribute much to an efficient fulfillment of the evaluation task. So a suitable evaluation tool must provide a good interface of HCI for users from different fields and assist to accomplish evaluation tasks effectively.

- **Reusability and extensibility (R&E).** In different evaluation tasks and objects, much evaluation information is overlapped such as, evaluation index system, data extraction methods, evaluation index arithmetic, integration evaluation methods. So the practicality and efficiency of the evaluation tool can be further improved through the storage and reuse of the evaluation information. Besides, more and more evaluation algorithms are developed and applied newly in different evaluation tasks and the evaluation tool must integrate these new and existing evaluation algorithms.

Based on the problem analysis, relevant solutions are developed. First, a process model of different evaluation tasks and objects is presented to implement the flexible evaluation. Considering the HCI characteristic of this tool, the primitive-based graphical interface is designed. Besides, the reusability of evaluation information and the extensible interface of evaluation methods are researched. The detailed solutions are given as follows.

2.1. The Process Model of Effectiveness Evaluation

According to the problems analysis above, the general procedures of operational effectiveness evaluation need to be researched. So we extract the common points of different evaluation tasks and construct a process model of effectiveness evaluation in form of IDEF0 (see figure 1). This model contains the primary and detailed selection of evaluation indexes, simulation experiment execution, value and ranking evaluation of operational effectiveness. The detailed illustration of each procedure is given below:

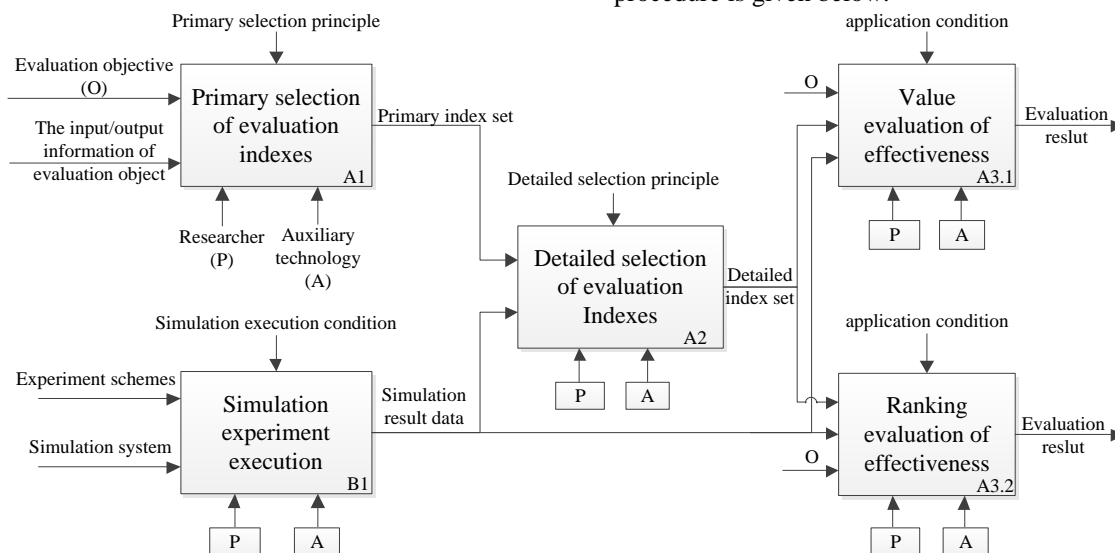


Figure 1: Process Model of Effectiveness Evaluation

I) Primary selection of evaluation indexes (A1): According to evaluation purposes and objects information, researchers could select the available evaluation indexes from the evaluation index model library. With respect to the indexes without extraction model, we need to construct them based on the input and output data information of

evaluation object, and then, the primary indexes set are obtained.

II) Detailed selection of evaluation indexes (A2): Multiple sample values of primary indexes could be achieved via the simulation result data from B1 and the primary index set from A1. After that, the sample values would be conducted through the

detailed selection method to acquire the detailed index set.

III) Value evaluation of effectiveness (A3.1): The evaluation result of each detailed index can be obtained through computing the simulation result and scenario data with evaluation index arithmetic and evaluation integration methods.






IV) Ranking evaluation of effectiveness (A3.2): The sequential effectiveness evaluation results of different operational scenarios can be obtained by conducting the simulation result and scenario data with integration evaluation methods.





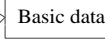
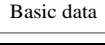
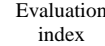
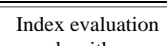
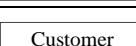
V) Simulation experiment execution (B1): The simulation result data could be acquired through executing experiments schemes on the simulation system and these data is used to stored and analyzed.

2.2. The Graphical Interface Design of HIT-OEET

HIT-OEET is developed to assist users from different fields to accomplish evaluation tasks better, then a suitable HCI interface needs to be designed and applied. So a primitive-based graphical interface is proposed in the effectiveness evaluation process. Some relevant primitives are defined based on the evaluation activities from the process model (Section 2.1) and three kinds of primitives are designed (see table 1) which include index system establishment, basic data extraction and evaluation algorithms setting. The primitives of index system establishment are used to build the index system of effectiveness evaluation, including root index, middle index and bottom index. The primitives of basic data extraction are used to construct the condition and attribute of interested data and these data would be extracted from the scenario and experiment result database. The primitives of evaluation algorithms setting involve the evaluation arithmetic of bottom indexes and integration evaluation methods and the evaluation results could be obtained via these algorithms.

Table 1: Primitives Definition in Effectiveness Evaluation Process

Primitive type	Primitive form	Primitive description
Index system establishment		Root index
		Middle index
		Bottom index
Basic data extraction		Condition expression described the experiment data attribute
		“And” expression between conditions

Evaluation algorithms setting		“Or” expression between conditions
		“Not” expression of condition
		1*1 dimension data
		n*1 dimension data
		Output basic data
		Input basic data
		Output evaluation index
		Evaluation algorithm
		Extended evaluation algorithm

2.3. The Reusability and Extensibility Design of HIT-OEET

For solving the repeated construction of much overlapped evaluation information for different evaluation objects and improving the evaluation efficiency, the information of each evaluation object should be stored and reused. Due to large amount of data to be stored and diverse data formats, we select relational database as the container of evaluation information and the storage format of data tables and the interaction interface between HIT-OEET and database needs to be designed. The database design will be given in Section 3.2. Here, we will take the information of evaluation algorithms as an example to illustrate the interface design (see figure 2). The evaluation algorithm interface contains index information and algorithm information. The index information describes the attribute of evaluation indexes such as index name, index type and primitive position etc. and the algorithm information is used to record the parameter information of algorithms and evaluation value of indexes. The evaluation results of bottom indexes could be either equal to the basic data directly or obtained via the evaluation arithmetic and the evaluation results of integrated indexes (middle and root indexes) could be achieved through integration evaluation methods.

Besides, more and more evaluation algorithms are being developed and applied for different evaluation tasks and objects and these new and existing algorithms all need to be integrated in evaluation tool. If we update the tool repeatedly with the fast development of these evaluation algorithms and this will generate high cost and lead to the worse practicability of the tool. So we design and implement the extended interface of evaluation

algorithms based on Dynamic Link Library in HIT-OEET. It consists of three parts including algorithm description, parameter configuration and algorithm execution (see figure 3) and the detailed description of each procedure is given as follows.

I) Algorithm description (A1): This procedure is used to give the information description of extended algorithms including algorithm usage, input/output variable attribute and application condition.

II) Parameter configuration (A2): Based on the algorithm description and actual input/output variables, the parameter information such as parameter name, parameter type and description could be given. Then, users could set the parameter value for the extended algorithm.

III) Algorithm execution (A3): Based on the input variable and parameter setting, the calculation results could be achieved through the algorithm execution.

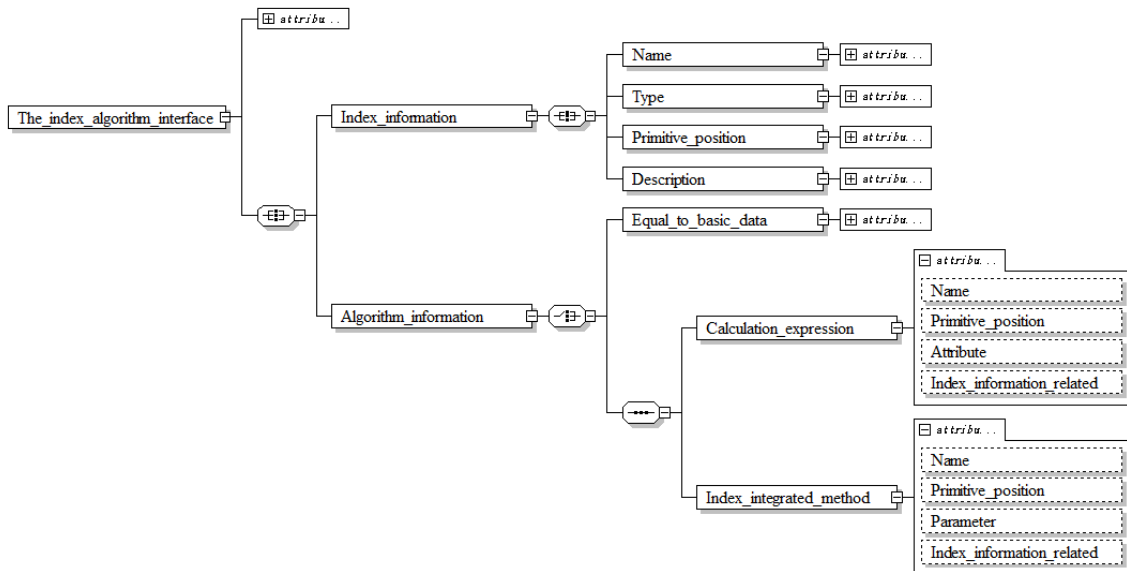


Figure 2: Interaction Interface of Evaluation algorithms

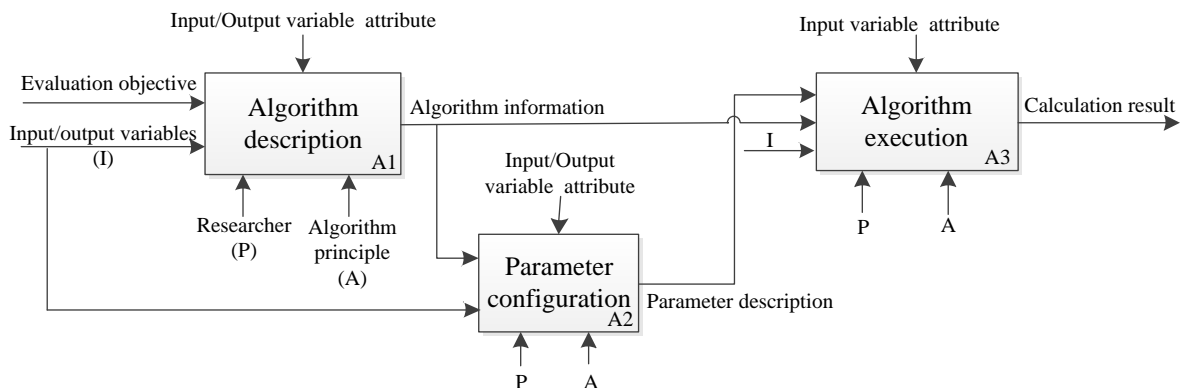


Figure 3: Execution Procedures of Extended Algorithm

3. THE DESIGN AND IMPLEMENTATION OF HIT-OEET

According to the researched content above, the functions of HIT-OEET would be conceived as below: establishing evaluation project, setting evaluation content, evaluating weapon operational effectiveness, analyzing evaluation result and generating evaluation report. This tool can also assist the users to set the information of data extraction methods, evaluation index arithmetic and indexes integration methods. Based on the described functions above, HIT-OEET is designed and implemented including component and structure design, database design and software implementation.

3.1. Component and Structure Design of HIT-OEET

On the basis of the necessary functions of HIT-OEET and modular design principle, the structure framework and components of HIT-OEET are designed firstly. This tool consists of six modules including interface of HCI, evaluation project information management, evaluation execution, database management, evaluation process information management and evaluation process monitoring (see figure 4). The functions and interaction relation of these components are designed as follows.

a) The HCI interface provides the operation panel for users to implement evaluation tasks and the graphical operation manner is integrated in it.

- b) The module of evaluation project information management is used to manage and set the related information, which includes personal information of estimators, evaluation project information and evaluation phase information.
- c) The evaluation execution module is used to execute the evaluation operations and set the information of evaluation index system, data extraction methods, evaluation index arithmetic and integration evaluation methods. The evaluation results can be obtained via this module.
- d) The module of evaluation process information management is used to provide the interface for storing and extracting the evaluation information including evaluation project and evaluation execution. This information will be packed and transmitted to database management module, and also displayed by the module of evaluation process monitoring.
- e) The evaluation process monitoring module is used to monitor the user operation and evaluation process information. All evaluation information is all displayed in this module, if some errors such as user operation and setting information appear and then the tool would give the caution signal.
- f) The database management module consists of scenario database, evaluation database and experiment result database, and is used to store the information, which contains scenario data, simulation data and evaluation process in a certain format. The storage format of data is designed in Section 3.2.

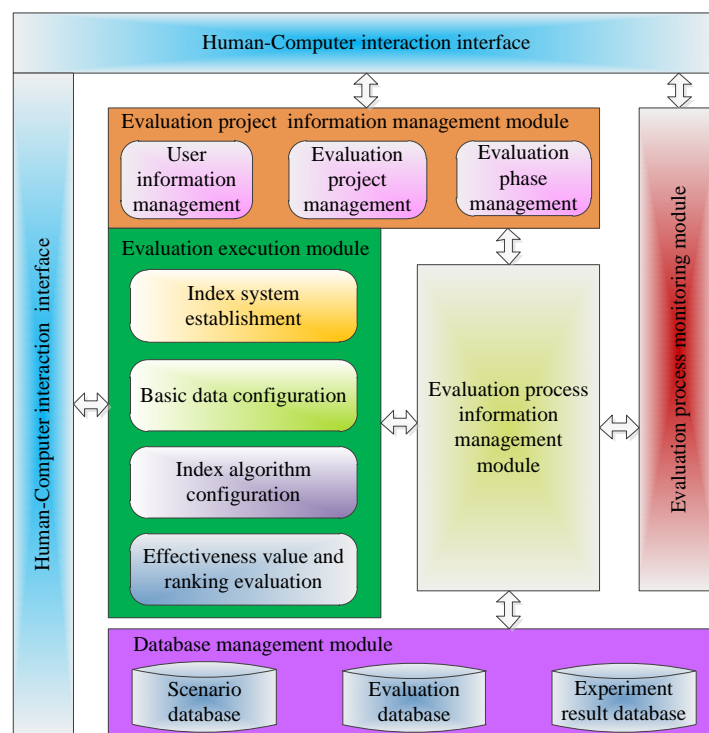


Figure 4: Component and Structure design of HIT-OEET

3.2. The Database Design of HIT-OEET

The simulation-based operational effectiveness evaluation of weapons is a data-intensive work, which involves large amount of operations of evaluation information, scenarios and simulation data. So a data container needs to be provided in evaluation tool. Relational database could be used to store the data due to the support of superior data size and different storage formats. According to the different sources of data, there are three databases are provided in HIT-OEET: scenario database, evaluation database and experiment result database. The scenario and experiment result database are used to store the scenario and simulation result data respectively, and provide the data source for effectiveness evaluation. The evaluation database is responsible for the storage and management of evaluation information and closely related to evaluation

process. So this paper mainly introduces the design of evaluation database.

According to the structure design of HIT-OEET above, there are six kinds of evaluation information need to be stored, which includes evaluation project information, user information, index system information, data extraction information, evaluation index arithmetic and evaluation indexes integration, evaluation result information (see figure 5).

3.3. Implementation of HIT-OEET

Based on the design thought and functions description above, HIT-OEET is implemented based on C++ language, Matlab and ACCESS through utilizing the technologies of ADO, ActiveX. The main interface and each function component of HIT-OEET are shown in Figure 6, which contains user and project management,

index system setting, data extraction configuration, evaluation index arithmetic, evaluation indexes

integration, evaluation results display and extended algorithm interface etc.

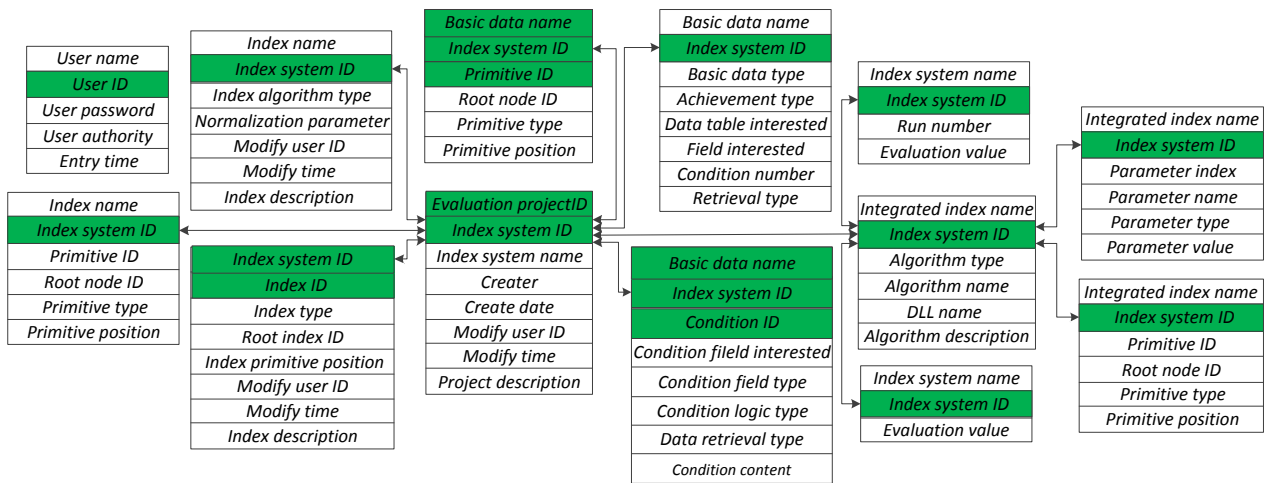


Figure 5: Structure of The Evaluation Database

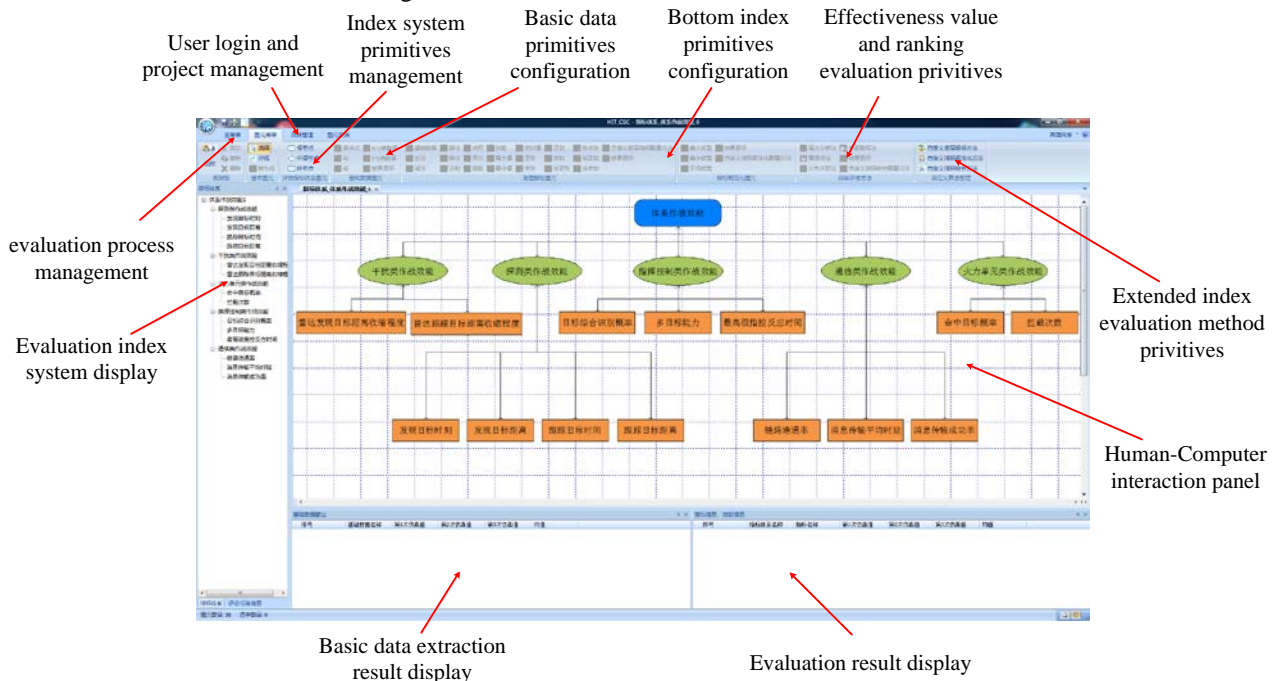


Figure 6: Main Interface of HIT-OEET

4. APPLICATION

In order to demonstrate the efficiency and usage of HIT-OEET, an example of simulation-based operational effectiveness evaluation of an air-defense system is given. The system consists of many entities, including aircrafts, missiles, radars, regional command and control system, etc. The operational scenario is that blue aircrafts violate the red area and the blue planes will be intercepted by red aircrafts and missiles when the objects are found by red radar. Our goal is to evaluate the operational effectiveness of red air-defense system. According to the process model proposed in Section 2.1, we need to establish the indexes system of effectiveness evaluation firstly. Considering application domain of the air-defense system, we construct the evaluation indexes system which consists of four middle indexes and eight bottom indexes, including investigation

ability, command and control capability, communication ability, Interception ratio, hitting ratio, etc. (see figure 7).

Second, we need to configure the bottom and middle indexes based on the basic data. Take the interception ratio as an example and it's computation formula is given as follows.

$$\eta = (NP + NM) / NE \quad (1)$$

where η is the interception ratio of red side. NP is the number of intercepted planes. NM is the number of intercepted missiles. NE is the number of entities which need to be intercepted.

All necessary basic data could be determined from Eq. (1) and then extracted from experiment results (see figure 8). After that, the interception ratio could be

configured based on these basic data and Eq. (1) in

Figure 9.

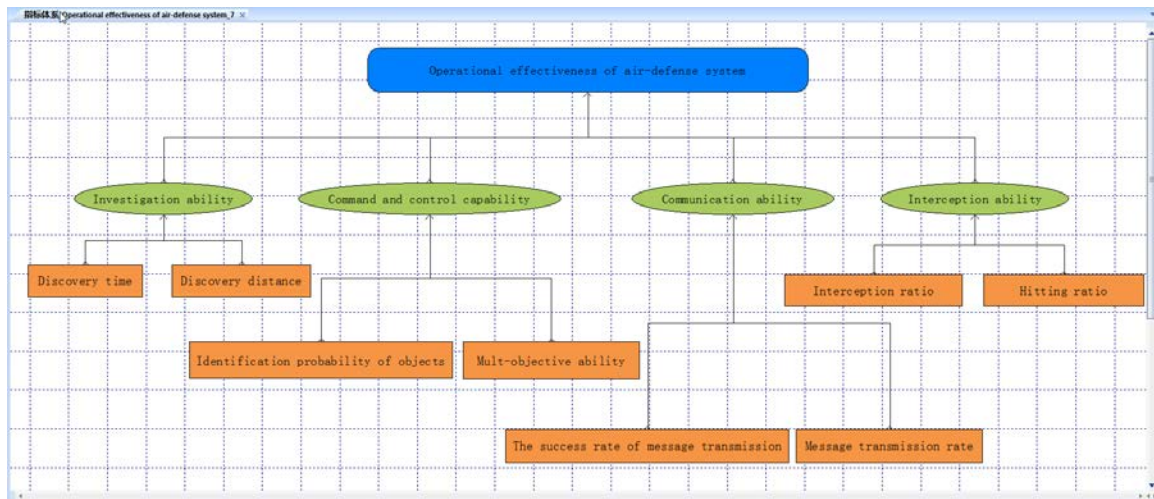


Figure 7: Evaluation Index System of Operational Effectiveness of an Air-defense System

So, the middle and root indexes could be configured layer by layer based on the indexes system. Finally, the configuration of root index, Operational effectiveness of air-defense system, is given in Figure 10. In the example, we adopt weighted average method to synthesize the middle indexes and each weight of four middle indexes is 0.25.

After accomplishing all configuration of basic data extraction information, evaluation index arithmetic and integration evaluation methods, the ultimate effectiveness results of an air-defense system could be computed via the information of indexes configuration.

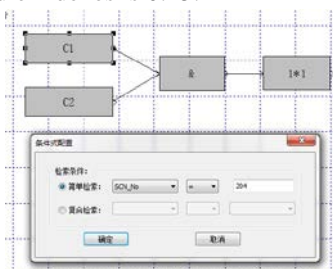


Figure 8: Condition expression setting

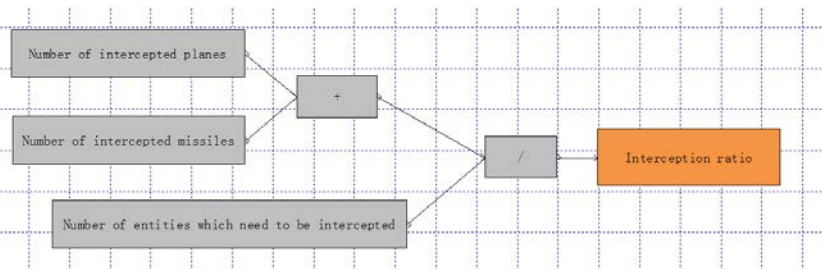


Figure 9: Setting information of interception ratio

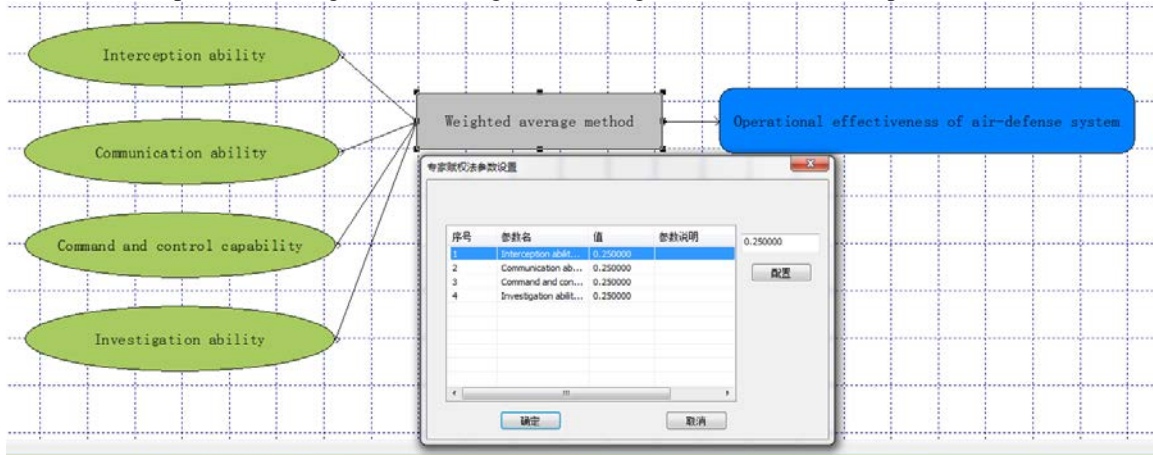


Figure 10: Setting information of indexes integration

5. CONCLUSION

This paper focuses on the development of a simulation-based operational effectiveness evaluation tool, HIT-OEET. In order to solve the existing problems of effectiveness evaluation and improve the evaluation efficiency, a process model of effectiveness evaluation is presented and reusability of evaluation information

and extensibility of evaluation algorithms are clearly enhanced. Also a primitive-based graphical interface is designed to improve the HCI characteristic of the tool. The proposed process model is regarded as a guideline for the simulation effectiveness evaluation and the graphical interface based on primitives makes HIT-OEET suitable for users from different fields better.

Depending on the reusability design of evaluation information and the extensibility design of evaluation algorithms, the efficiency of the operational effectiveness evaluation is improved greatly. Based on the researched content above, HIT-OEET is designed and implemented.

In the example, HIT-OEET is validated to accomplish the operational effectiveness evaluation of weapons well and the usage of this tool is shown. In addition, the graphical interaction interface primitive-based makes the evaluation process more straightforward than existing evaluation tools.

Due to the more and more complex effectiveness evaluation objects in the future such as, Sps, multiple components of weapon system, etc. and the evaluation tasks may be accomplished by several users together in different places. So a networked, multi-node and simulation-based effectiveness evaluation tool for complicated weapon system needs to be developed in future work.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (Grant No. 61403097) and the "Fundamental Research Funds for the Central Universities" (Grant No. HIT. NSRIF. 2015035).

REFERENCES

- Jiao S., Li W., Ma P. et al, 2013. The simulation evaluation system for weapon operational effectiveness based on knowledge management. *International Journal of Modeling, Simulation and Scientific Computing*, 4 (4), 13500201-17.
- Pinder J.D., 2000. Evaluating Future Force Options for the U.S. Army. Thesis (PhD). RAND Graduate School.
- Li Y.P., Li X., Wang H.Q., 2014. A compact methodology to understand, evaluate, and predict the performance of automatic target recognition. *Sensors*, 14(7): 11308-11350.
- Matthew S., Brian R. 2015. An evaluation of open set recognition for FLIR images. *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 9476N1-10. May 20-22, Baltimore, MD, USA.
- Xu H., Kang F.J., Li D., 2015. Method of effectiveness evaluation for ATTWS based on cloud model. *Journal of Information & Computational Science*, 12(12): 4749-4757.
- Qu H.P., Chang L.N., Ma L. et al, 2012. DPEES: DDoS protection effectiveness evaluation system. *International Conference on Data Engineering and Internet Technology*, pp. 155-161. March 15-17, Bali, Indonesia.
- Wang L., Fang Y.W., Xu X. et al, 2010. Effectiveness evaluation system of ATS. *Systems Engineering and Electronics*, 32(12): 2608-2611+2720.
- Lancon F, 2011. Naval electronic warfare simulation for effectiveness assessment and softkill

programmability facility. *Conference on Modeling and Simulation for Defense Systems and Applications VI*, pp. 8600N1-10. April 25-27, Orlando, Florida, USA.

- Qi Z.F., Wang G.S., 2014. Grey Synthetic Relational Analysis Method-Based Effectiveness Evaluation of EWCC System with Incomplete Information. *3rd International Conference on Civil, Architectural and Hydraulic Engineering*, pp. 2109-2412. July 30-31, Hangzhou, China.
- Akyuz E., Metin C., 2014. A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. *Safety Science*, 68, 169-179.
- Jiao S., Yang M., Li W., 2013. Parameterized simulation evaluation technology for operational effectiveness of weapon, *Transactions of Beijing Institute of Technology*, 33(12), 1269-1273.
- Li W, Lu L.Y., Liu Z.Z. et al, 2016. HIT-SEDAES: An integrated software environment of simulation experiment design, analysis and evaluation. *International Journal of Modeling, Simulation, and Scientific Computing*, 7(3).

AUTHORS BIOGRAPHY

WEI LI is an associate professor at Harbin Institute of Technology (HIT), and received the B.S., M.E. and Ph.D. from HIT in 2003, 2006 and 2009 respectively. His research interests include simulation evaluation, simulation data analysis, and distributed simulation. His email is frank@hit.edu.cn.

SHENGLIN LIN is a Ph.D. student at HIT, and received the M.E. from Harbin Engineering University in 2015. His research interests include operational effectiveness evaluation and VV&A. His email is lin_44627079@yeah.net.

ZHIZHAO LIU is a Ph.D. student at HIT, and received the M.E. from HIT in 2012. His research interests include simulation optimization and operational effectiveness evaluation. His email is liuzhizhao2007@hotmail.com.

PING MA is the corresponding author. She is a professor at HIT, and received Ph.D. from HIT in 2003. Her research interests include distributed simulation technique and VV&A. Her email is pingma@hit.edu.cn.

MING YANG is a professor and the director of control and simulation center at HIT. Also he is the vice editor-in-chief for *Journal of System Simulation* and editor for *International Journal of Modeling Simulation and Science Computing*. His research interests include system simulation theory and VV&A. His email is myang@hit.edu.cn.

A TREE-SEARCH BASED HEURISTIC FOR A COMPLEX STACKING PROBLEM WITH CONTINUOUS PRODUCTION AND RETRIEVAL

Sebastian Raggel^(a), Beham Andreas^(b), Fabien Tricoire^(c), Michael Affenzeller^(d)

^(a,b,d) Heuristic and Evolutionary Algorithms Laboratory, University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria

^(b,d) Institute for Formal Models and Verification, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria

^(c) Department of Business Administration, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

^(a)sebastian.raggel@fh-hagenberg.at, ^(b)andreas.beham@fh-hagenberg.at, ^(c)fabien.tricoire@univie.ac.at
^(d)michael.affenzeller@fh-hagenberg.at

ABSTRACT

We present a real world steel stacking problem featuring non-instantaneous crane movements, continuous production and retrieval and stacking constraints based on the dimensions as well as temperature of the slabs. An exact Branch & Bound solver as well as three tree-search based heuristics is developed. Random benchmark instances derived from the real world problem are used to evaluate the performance of the heuristic solvers and compare them to the exact solver.

Keywords:

Stacking Problem, Branch & bound, Heuristic

1. INTRODUCTION

Stacking problems arise in many sectors of industry in a great number of variants. A good overview of the different problem variants including solvers and complexity considerations is given in (Lehnfeld & Knust, 2014). They differentiate loading, unloading, premarshalling and combined problems. According to their categorisation the problem considered in this paper is a combined loading unloading problem.

The work most closely related to this one is probably (Rei & Pedroso, 2013). They also describe a stacking problem with continuous production and delivery but with instantaneous crane movements and without any capacity or stacking constraints. They develop a stochastic tree search algorithm in order to tackle big problem instances.

The Stacking Problem considered in this paper arises in the operation of a steel factory. Steel slabs are casted according to a fixed schedule and have to be put onto a delivery stack according to a fixed list of delivery lots. The slabs can be moved by a crane which can only move a single slab at a time and only access the slabs on

top of a stack. Every crane movement requires time to perform and the stacking is subject to some restrictions.

Figure 1 shows a possible layout of a stacking area with a single caster at the left, nine partially used buffer stacks in the middle, and a crane moving a slab to the handover stack at the right.

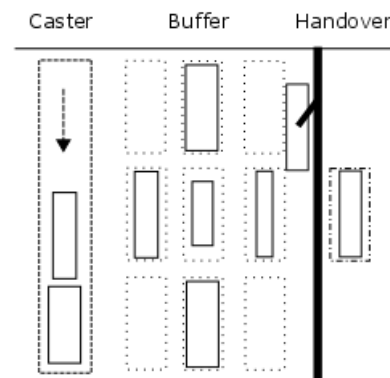


Figure 1: Example of a stacking area

If crane movements were instantaneous and the last slab was already casted this problem can be viewed as a variant of the Block Relocation Problem with additional stacking constraints. In this case the caster is simply viewed as another stack. The original authors of the BRP (Kim & Hong, 2006) use a Branch & Bound algorithm and a heuristic rule based on the expected number of additional relocations, called ENAR. Improvements on this approach come from (Ünlüyurt & Aydın, 2012). Two binary linear programming are introduced in (Caserta, et al., 2012) where one only considers strictly necessary moves and the other one also allows so called voluntary moves. Those models are improved upon by (Petering & Hussein, 2013) and (Expósito-Izquierdo, et al., 2015).

The remainder of this article is organised as follows. In Section 2 we describe the problem in detail. In Section 3 we present an exact Branch & Bound algorithm as well as a heuristic approach. Section 4 compares the two solvers in terms of solution quality and runtime using randomly generated problem instances derived from real world problem instances. Finally we discuss the results of the experiments and present an outlook on further research.

2. PROBLEM DESCRIPTION

There is a set of K casters $C = \{c_k\}_{k=1}^K$ which produce N steel slabs $S = \{s_i\}_{i=1}^N$ of different dimensions according to a fixed schedule. The slabs are to be placed onto a single handover stack h where they are picked up according to fixed delivery lots. A delivery lot consists of the due-time and the order in which the slabs of that lot should be in. At the due-time all slabs of a lot must be on the handover stack so a transporter can deliver them all to their destination. It is important to note that if the casting schedule and the order of the slabs in the transport lots match up the problem becomes trivial, but this is not always possible.

When a slab comes out of a caster and cannot be put directly onto a delivery stack can be put on one of M so called buffer stacks $B = \{b_j\}_{j=1}^M$. There is a single crane which can transport a slab from the top of a buffer stack or front of a caster to a handover or buffer stack using a certain amount of time given that none of the stacking constraints described in 2.1 are violated. Note that the crane is not allowed to take a slab from the handover stack or to put a slab back onto the caster.

The objective is to deliver all the slabs according to the given delivery lots while performing as few crane movements as possible.

2.1. Stacking constraints

The height of the buffer stacks is limited but since the slabs have different dimensions this does not translate to a fixed amount of slabs per stack.

In order to guarantee stability of the stack as well as to prevent deformation of the slabs there is a limit on the allowed length and width difference between each slab and all the slabs beneath it in the stack. Note that this constraint only has to be checked when a slab is placed on a buffer stack because the delivery lots are fixed.

Another factor that determines if a slab can be stacked onto another is the temperature difference between the two. This constraint is necessary because the cooling is vitally important for the steel quality and reheating a slab by placing a hotter slab on top can mean that the slab is no longer usable. The temperature of a slab is calculated using a simple logarithmic cooling scheme starting at the casting temperature and time.

2.2. Time constraints

Every slab has a production timestamp and a delivery timestamp. The most basic temporal constraint is that a slab cannot be moved before its production or after its

delivery. Since the caster can only hold a certain number of casted slabs and the casting schedule is fixed, there is a time window in which a slab can be taken out of the caster. This window is between the casting time of the slab and the casting time of the $n+1$ slab. Otherwise the caster would have to be stopped which should be avoided at all costs.

Similarly, there is a time window for putting slabs on the handover stack, which begins with the delivery time of the previous transport lot and ends with the delivery time of the slab.

3. SOLVING THE STACKING PROBLEM

The solution of a stacking problem consists of a list of crane movements called moves from here on. A move is described as a tuple of the source and the target. Since there are different types of locations we distinguish between five different kinds of moves:

- *Put* (C, B)
- *Remove* (B, h)
- *PutDirect* (C, h)
- *Relocate* (B, B)
- *Delivery* (h, -).

Put moves take the first slab from a caster and put it on buffer stack. *Removal* moves are taking a slab from a buffer stack and putting it on the handover stack. Of course when a slab from the caster can be directly put onto the handover stack it would be nonsense and even potentially impossible to put it on a buffer stack first. This is what a *PutDirect* move is for.

A *Relocation* move is one that moves a slab from one buffer stack to another. The BRP literature differentiates two kinds of *Relocation* moves, namely forced and voluntary moves (Caserta, et al., 2012). If the slab that has to be removed next does not lie on the top of a stack the slabs above it have to be relocated. This is called a forced move. Not considering voluntary moves can lead to improved run time at the cost of potentially worse solutions.

Delivery moves are special in a number of ways. They always move all the slabs of a transport lot at once. We considered them to be instantaneous because we are not concerned with the actual delivery process but instead only care when the handover stack is free to use again.

All moves that put slabs on a buffer stack are subject to the stacking constraints described in 2.1. A slab can only be put on the delivery stack if the stack is either empty or already contains all the slabs that are in lower positions in the same delivery lot.

The combination of the current time and the creation and delivery times of all slabs determine which moves are valid. On the other hand the exact time at which each move is performed is irrelevant as long as the constraints are respected. Given a valid sequence of

moves, it is easy to calculate a time window within which each move must be performed.

Table 1: Example1

Creation time	Due-time/ Position	Initial position
0	20/1	$b_1/1$
0	10/0	$b_1/0$
10	20/1	c_1

Table 1 shows an example with a single caster, two buffer stacks and three slabs. Figure 2 shows the same sequence of moves and therefore the same solution distributed in time in three different ways.

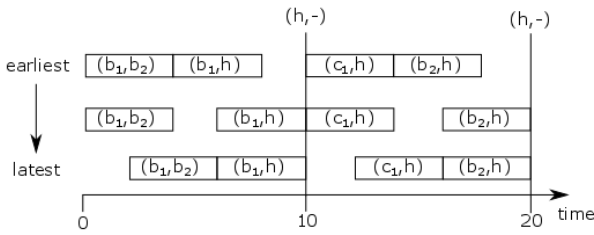


Figure 2: Solution of Example1

3.1. Branch and bound

In order to perform a branch and bound search we need a lower bound. In order to get a valid lower bound we consider that, for every remaining lot rl we need one *Delivery* move. Every slab that is not already at the handover stack needs to be moved at least once. If a slab lies on top of a slab that has to be put on the handover stack before, the upper slab must be moved at least two times. So if $lb(x)$ is the sum of minimal required moves of all slabs at location x and m is the number of moves already performed a total lower bound is:

$$LB = m + \sum_{b \in B} lb(b) + \sum_{c \in K} lb(c) + rl$$

Note that this lower bound is only valid if there is only a single handover stack because if there are multiple handover stacks a slab that is blocking another slab could potentially be placed directly on another handover stack instead of a buffer stack.

We perform a depth first search using all possible moves but we use the same order as described below in order to improve the chances of finding good solutions earlier.

3.2. Heuristic tree search

Due to the time and stacking constraints every move has the potential to generate a situation where there is no valid solution. While choosing a valid move in a BRP will never lead to an invalid solution. So instead of using a heuristic that aims to choose the best move at every step, we propose a heuristic that uses a list of all "good" moves ordered by quality. So in most cases the

first move on the list should be the one used in the final solution, but if there is a move that results in a situation where there is no solution, alternative moves can be applied.

To find such a list of moves let us make some observations. Placing a slab on the handover stack is preferable to placing it on a buffer stack because once it is on the handover stack the crane never has to touch that slab again. Moving a slab from a buffer stack can always be done immediately. Moving a slab from the casters or the handover stack might involve waiting for the casting or delivery time of the slab. Therefore *Remove* moves should be done before *PutDirect* moves and *Relocate* moves before *Put* moves. So the moves that will be tried are in the following order:

1. If a caster must be cleared (see 2.2)
 - a. Return all *PutDirect* moves.
 - b. Return safe *Put* moves.
 - c. Return remaining *Put* moves.
2. If a delivery must be performed (see 2.2) return the *Delivery* move.
3. Otherwise return a list of all
 - a. *Removal* moves
 - b. *PutDirect* moves
 - c. *Safe Relocation* moves
 - d. *Safe Put* moves
 - e. *Delivery* move
4. If there are no safe moves return a forced move.

A safe move is one that does not increase the lower bound LB described in Section 3.1. *Removal* and *PutDirect* moves are always safe because moving a slab away from the handover is not allowed. This means that if the move were unsafe it is also forbidden because it cannot lead to a feasible solution. *Put* and *Relocate* moves are considered safe, when any slab of the source stack has to be moved to the handover stack before the topmost slab and none of the slabs of the target stack have to be handed over before the topmost slab of the source stack.

If there are multiple safe *Put* or *Relocation* moves that will put the same slab on different empty buffer stack only one of them is considered in order to reduce the total count of possible moves. This is however only valid if the crane movement times between all the stacks are equal and the height limits of the buffer stacks are also equal.

Based on this list of good moves we define three different heuristic solvers. The first, we will call it H1, uses the list of good moves as defined above to do a depth first search and return the first solution found.

Note that a solution of H1 is a valid upper bound for the B&B solver described in Section 3.1.

The second solver uses a two stage approach where in the first step a breadth first search is used to generate level after level of the search tree, using all possible moves, until a level is generated that has more than a certain number of nodes. Then each of these partial solutions is completed using H1. We call this solver H(n) where n is the minimum number of nodes that we generate in the first step. It would also be possible to specify the number of levels to generate, but this gives us better control over the runtime. One nice feature of this is that the second step can be parallelized because the partial solutions are totally independent of each other.

Finally we use the B&B solver as described in Section 3.1 but instead of considering all moves we only consider the moves specified before. From now in order to distinguish between the two variants we will refer to the exact solver as E B&B and the heuristic one as H B&B.

4. EXPERIMENTS

We evaluated the solvers presented above using a set of randomly generated test instances of different sizes. The test instances have two casters that are used to produce a varying number of lots with four slabs per lot. The production dates of the slabs, as well as the due dates of the lots are randomly chosen. Then a random start date is chosen and all the slabs are put either in a caster or in one of the M buffer stacks depending on their production dates. Using this schema we generate eight different scenarios with 10 random instances each. The scenarios have 4, 8, 16 or 32 lots and 4 or 8 buffer stacks.

We perform three different experiments using this test instances. In Section 4.1 we compare the runtime and solution quality of our heuristic solvers. Afterwards we compare the best solution of the H B&B solver against the optimal solution found by the E B&B in order to judge the quality of our heuristic solutions.

Finally because the real world application of the solvers requires us to be able to find a solution quickly we compare how the heuristics perform when given a time limit of one minute. So the search is aborted after one minute and the best solution reached so far is reported. The reason we need to be able to find solutions quickly is that it can easily happen that the real-world circumstances change in such a way that our solution is no longer valid. If we require hours to come up with a new solution this is a problem.

All solvers were implemented in C# using the .NET Framework 4.6.1 using the Task Parallel Library. All tests were run on a Dell Latitude E6540 with an Intel i7 4810MQ CPU @2.80GHz and 16 GB RAM running Windows 7.

4.1. Comparing the heuristics

Table 2 shows a comparison of the average solution quality of the heuristic solvers for eight different scenarios with ten instances each. The N and M columns are the number of slabs and the number of buffer stacks respectively. LB is the average lower bound as described in Section 3.1. The remaining three columns contain the number of moves more than lower bound the algorithms described in Section 3.2 required on average.

The problems get harder the more slabs they contain. Interestingly the problems are also harder if they contain fewer buffer stacks. This may seem counter intuitive because having more stacks also means there are more possible moves. What increases the difficulty is that the likelihood of producing a new conflict when putting a slab on a buffer stack is higher the fewer buffer stacks there are.

Table 2: Best Solutions of Heuristic Solvers.

N	M	LB	H1	H(50)	H B&B
16	4	26.5	0.6	0.6	0.5
16	8	25.8	0.4	0.3	0.3
32	4	55.0	5.8	4.9	2.6
32	8	54.3	3.0	2.6	1.5
64	4	112.3	12.7	11.0	4.4
64	8	111.6	9.5	9.0	4.0
128	4	226.0	22.6	21.9	15.9
128	8	225.2	19.0	18.7	14.4

Even the simple H1 manages to find solutions within 10% of the lower bound for the majority of problem instances. As described above H(50) runs H1 on all the nodes of the first level of the search tree which has more than 50 nodes. For the problem sizes we looked at this means evaluating all possible combinations of the first 2-3 moves. This brings an improvement over H1 for 25 out of 80 problem instances.

Table 3: Runtime of Heuristics in Seconds.

N	M	H1	H(50)	H B&B
16	4	0.00	0.01	0.00
16	8	0.00	0.03	0.00
32	4	0.00	0.03	550.85
32	8	0.00	0.82	782.58
64	4	0.00	0.07	2624.04
64	8	0.00	0.58	2528.75
128	4	0.01	0.36	3600.00
128	8	0.16	9.78	3600.00

H B&B is never worse than H(50) and outperforms it on 54 out of 80 problem instances but as can be seen in Table 3 is the runtime much higher. The runtime of H1 is less than 0.1 seconds for all but one problem instance with 128 slabs and 8 buffer stacks. H(50) takes less than 10 seconds to run for all instances except the one mentioned above. The H B&B solver fails to complete in less than one hour for at least one instance in every problem class except for the two with only 16 slabs. For the two biggest problem classes no instance can be solved in less than one hour.

4.2. Optimality gap

In Figure 3 we can see that of our 80 test instances only 33 could be solved by the exact B&B within one hour. H1 solved 21 instances to optimality while H(50) managed to solve 23. The H B&B solver found the optimal solution in 30 cases. The solutions to the remaining three instances all required one more move than would have been optimal. Additionally there are two instances where the E B&B managed to find better solutions than the H B&B despite not running to completion.

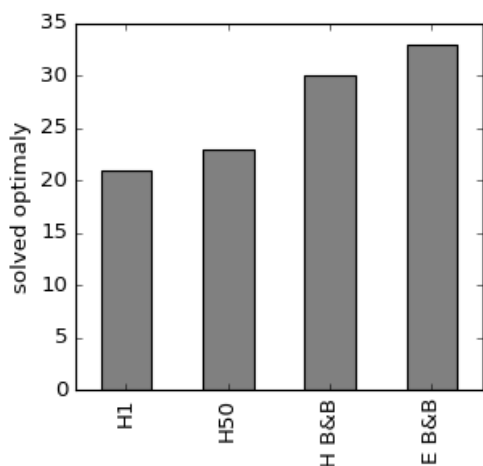


Figure 3: Number of Problem Instances Solved Optimally

It is interesting to look at situations where the heuristic is unable to find the optimal solution. One problem is that sometimes it is necessary to perform otherwise useless moves in order to enable save moves that are vital for the final quality.

Another point is that our heuristics manage contain rules about which moves are “good” and which are not based on general considerations about slab stacking as well as the temporal constraints. The stacking constraints described in Section 2.1 however are only used to forbid moves that would violate them completely but not in deciding which legal move would be better. This can lead to poor decisions where a stack is blocked by a slab that cannot be stacked upon.

4.3. Comparing the heuristics with time limit

In Figure 4 and Figure 5 we show the best solutions found divided by the lower bound for all problem categories after one hour and one minute respectively.

The two smallest problem sizes are the same because they were all solved in less than one minute. Five of the twenty problems with eight lots could not be solved. The one minute solution was worse in three instances by at most three moves.

For the problems with 16 and 32 lots the additional runtime made more of a difference. But even in those larger problems was the biggest improvement that an additional 59 minutes of runtime could bring was eight moves.

What the graphs do not show is that for 49 out of 80 problem instances the quality did not differ at all between the run with a one hour time limit and a one minute time limit. All of this suggests that the order we choose the moves in is good since the heuristic manages find the good solutions early.

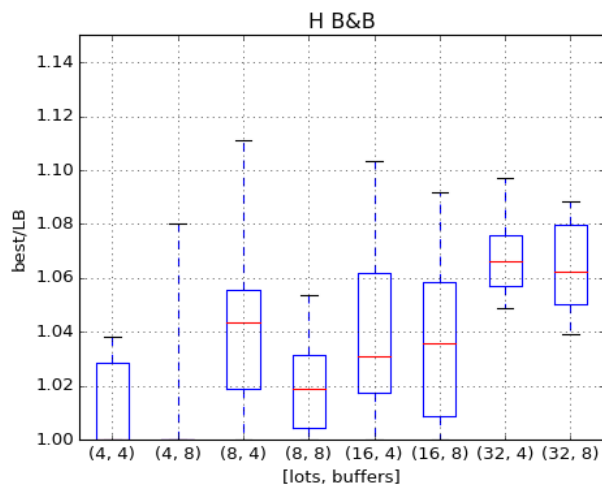


Figure 4: Best Solutions of H B&B after 1 hour

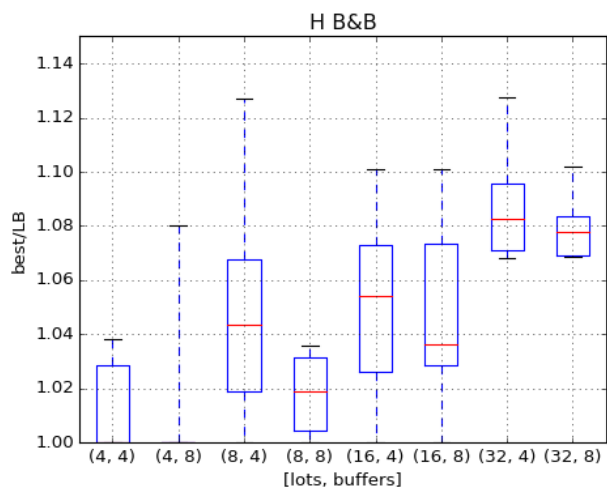


Figure 5: Best Solutions of H B&B after 1 Minute

5. CONCLUSION

The exact solver is unable to solve all but the smallest problems in reasonable time and where it did manage to

find a solution there was barely an improvement over the heuristic.

We showed that our heuristic works and that finding a better solution than the first one which is returned by H1 takes considerably more effort for all but small problems. Running H1 from multiple starting points can improve the result, but the time is better spent on running the H B&B for as long as possible.

While it is good to give the H B&B solver as much time as possible, because it could find better solutions, it is often not necessary to run it to completion because the heuristic aims at looking at the most promising moves first.

6. MANAGERIAL INSIGHTS

Efficiently stacking the slabs coming out of the caster and preparing transport lots previously required an experienced worker. We showed that automatization is indeed possible using the approach presented in this paper. Our heuristic is able to deliver high quality solutions in a reasonable runtime. In addition to this work we are also currently working on optimization algorithms for the transport lot building as well as the transportation and on integrating

In addition to freeing human resources and reducing labour costs are these optimizers also useful for evaluating the impact of planned changes on the performance of the system.

7. OUTLOOK

We are working on applying our results to the real world case. One difficulty is that sometimes it is necessary to bend the rules laid out in Section 2.1 and 2.2 slightly in order to find a solution at all. If that is the case our algorithms cannot find an answer. There are a few possible solutions to this problem. One approach is to relax the constraints and try again if no solution was found. This is problematic because if the algorithm uses all the time we allotted only to be started again afterwards we may not have a solution in time. The better solution would be to allow but penalize violating the constraints. We are currently investigating how we can accomplish that.

ACKNOWLEDGMENTS

The work described in this paper was done within the COMET Project Heuristic Optimization in Production and Logistics (HOPL), #843532 funded by the Austrian Research Promotion Agency (FFG).

REFERENCES

Caserta, M., Schwarze, S. & Voß, S., 2012. A mathematical formulation and complexity considerations for the blocks relocation problem. *European Journal of Operational Research*, 219(1), pp. 96-104.

Expósito-Izquierdo, C., Melián-Batista, B. & Moreno-Vega, J. M., 2015. An exact approach for the Blocks

Relocation Problem. *Expert Systems with Applications*, 42(17), pp. 6408-6422.

Forster, F. & Bortfeldt, A., 2012. A tree search procedure for the container relocation problem. *Computers and Operations Research*, 39(2), pp. 299-309.

Kim, K. H. & Hong, G.-P., 2006. A heuristic rule for relocating blocks. *Computers & Operations Research*, 33(4), pp. 940-954.

Lehnfeld, J. & Knust, S., 2014. Loading, unloading and premarshalling of stacks in storage areas: Survey and classification. *European Journal of Operational Research*, 239(2), pp. 297-312.

Petering, M. E. & Hussein, M. I., 2013. A new mixed integer program and extended look-ahead heuristic algorithm for the block relocation problem. *European Journal of Operational Research*, 231(1), pp. 120-130.

Rei, R. J., Kubo, M. & Pedroso, J. P., 2008. Simulation-based optimization for steel stacking.. In: H. A. Le Thi, P. Bouvry & T. Pham Dinh, eds. *Modelling, Computation and Optimization in Information Systems and Management Sciences*. Berlin: Springer, pp. 254-263.

Rei, R. J. & Pedroso, J. P., 2012. Heuristic search for the stacking problem.. *International Transactions in Operational Research*, 19(3), pp. 379-395.

Rei, R. J. & Pedroso, J. P., 2013. Tree search for the stacking problem. *International Transactions in Operational Research*, 203(1), pp. 371-388.

Ünlüyurt, T. & Aydın, C., 2012. Improved rehandling strategies for the container retrieval process. *Journal of Advanced Transportation*, 46(4), p. 378)393.

OPTIMIZATION OF PAINT SHOP DEPARTMENT USING DISCRETE EVENT SIMULATION

Pavel Raska^(a), Zdenek Ulrych^(b)

^(a) Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia, Univerzitni 22, 306 14 Pilsen

^(b) Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia, Univerzitni 22, 306 14 Pilsen

^(a)praska@kp.v.zcu.cz, ^(b)ulrychz@kp.v.zcu.cz,

ABSTRACT

The paper deals with a realistic simulation case study focused on the optimization of the production processes of products staining. The simulation study is focused on the validation of various options for increasing the total production quantity of the production system. The goal of the simulation study is to validate the planned investment and also verify other possible measures leading to an overall increase of the production of the paint shop department.

Keywords: optimization, paint shop, discrete event simulation, simulation model

1. INTRODUCTION

Using simulation in the optimization of the production processes is well known (Boysen, Scholl, and Wopperer, 2012; Frank, Laroque, and Uhlig, 2013; Gopalakrishnan, Skoogh, and Laroque, 2014; Jahangirian, Eldabi, Naseer, Stergioulas, and Young, 2010; Longo, Massei, and Nicoletti, 2012; Rabbani, Ahmad, Baladi, and Khan, 2013; Scholl, Laroque, and Weigert, 2014; Su, Fu, Tan, and Hu, 2010). It is possible to achieve various levels of performance of the whole system by choosing various configurations of the production systems (Ulrych, Votava, Raska, and Horejsi, 2013). Simulation can verify which measures should be implemented in a given company with a given structure and production volume. The paper deals with the modelling and simulation of a paint shop. Designing a paint shop is usually difficult, if we need to ensure its maximum throughput. Using simulation models to describe paint shop optimization problems is covered in many papers, e.g. (Arinez, Biller, and Meerkov, 2010; Cheng and Park, 2010; Lemessi, Schulze, and Rehbein, 2011; Lemessi, Rehbein, and Rehn, WSC 2012, 2012; Li, Blumenfeld, and Marin, 2007).

The Department of Industrial Engineering and Management at the University of West Bohemia has been dealing with discrete event simulation and optimization of production processes for several years. Various results of simulation studies have been published e.g. (Horejsi, Horejsi, Latif, and Ulrych, 2011; Raska and Ulrych,

EMSS 2014, 2014; Votava, Ulrych, Edl, Korecky, and Trkovsky, 2008).

The second area of interest we are intensively engaged in, is the use of various algorithms applicable to optimization of the discrete simulation, e.g. (Raska and Ulrych, 2015).

2. SIMULATION MODEL DESCRIPTION

This paper deals with the optimization of a paint shop department where various types of products of different materials are dyed. The products range in size from tens of centimetres to approximately four meters long.

The modelled production system consists of a series of workplaces. Transportation between workplaces is performed by chains strung on defined paths. The chains designed for towing a girder are approximately two kilometres long. Chains are used for conveying the hanging girder, which carries the products. Each type of product passes through different workplaces of the paint shop department according to its painting program. There are dozens of painting programs.

2.1. Workplaces in the Paint shop Department

Basic modelled painting workplaces are:

- Hanging on girder - the products are hung on the girder. Number of hung products depends on the size of the product and product type. Selecting a painting program determines the product path, sequence of workplaces and processing times. Furthermore, the chemical preparation program has to be selected
- Sandblasting – this workplace is only used for some of the painting programs. Sandblasting time depends mainly on the size and number of parts
- Chemical preparation - products pass through several successive chemical preparation chambers. The processing times in different chambers depend on the selected chemical preparations program (some of products just pass through the chamber)

- Drying furnace - drying time depends on the selected painting program. All the products hung on more girders are dried according to the maximum capacity of the drying furnace
- Paint booths – there are several paint booths at the workplace. Each paint booth applies a specific powder coating (base coat, second and another layer, or colour variants). It is possible to ensure partial interchangeability of the individual paint booths.
- Kiln - firing time depends on the selected painting program. All the products hung on more girders are fired according to the maximum capacity of the kiln. Girders in the kiln cannot overtake each other.
- Cooling – there must be some time for cooling the products after the firing phase. The container for this purpose is located behind the kiln.
- Removing - The last workplace in the circuit is where the product is removed. Products are removed from the girder and the girder is sent back to the container, which is located in front of the workplace where products are hung on the girder.

The basic scheme of material flow of the paint shop workplaces is shown in Figure 1. Green arrows represent input/output of material to/from the paint shop department workplaces. Before each workplace is a buffer defined by its maximum capacity.

Depending on the painting program, the girder can pass through some workplaces several times.

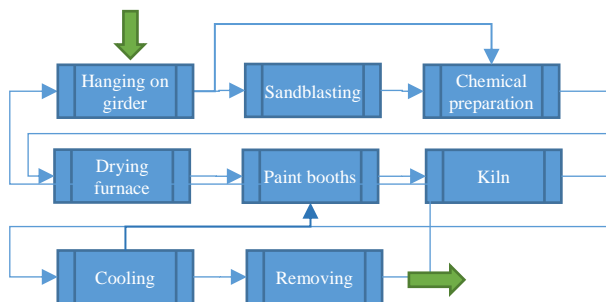


Figure 1: Material Flow of the Paint Shop Workplaces

3. BUILT SIMULATION MODEL

The simulation model was built by Plant Simulation software version 12.1. It was created hierarchically using the principles of object-oriented programming. The following figure shows an example of a class of sandblasting workplace where there are two sandblasting chambers - Figure 2. Thus it is possible to sandblast the two girders at once. This sample class is used in the model several times depending on the required number of this workplace in the system that is being modelled.

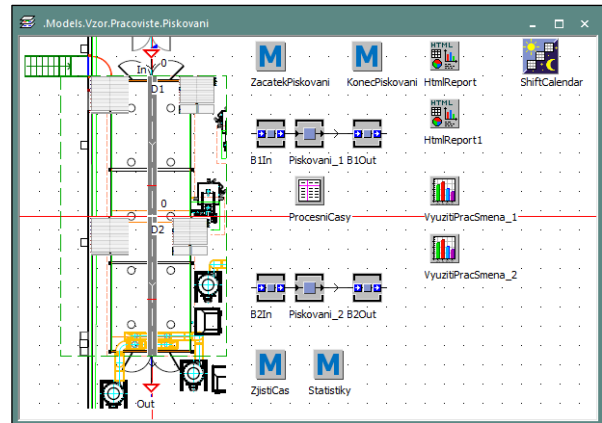


Figure 2: Sandblasting Workplace Class

The most difficult class model used in this model is a class representing a section of the path of girder - Figure 3. This sample class path is inherited from the "Track" class that is used for modelling a path for e.g. forklifts or Automated Guided Vehicles. We can see only a small part of the defined user attributes and the programming method, which assures the behaviour of the girder in the circuit according to the set attributes of the selected path and according to the selected painting program of the girder. Built-in user-defined methods can provide:

- Stopper control - whether the girder stops e.g. when reaching the maximum capacity of the next section; whether a girder can start moving - free capacity of the next section; defined rules of the crossroad (FIFO, according to priority, etc.); the time needed to hook a girder on a hook located on a chain
- Girder rotation – girder can keep going longitudinally or transversely (usually in the magazine)
- Collecting statistics - especially statistics on blocking path
- Controlling of intersections - which way a girder will go
- etc.

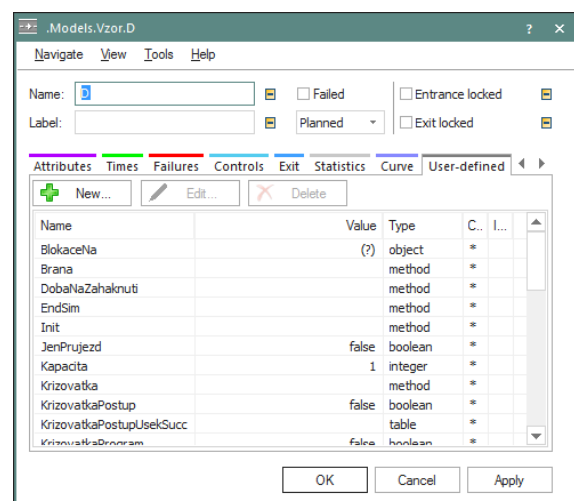


Figure 3: Class for Modelling Path

The following figure shows only a small part of the simulated system, including the defined animations - Figure 4. The animation shows the following attributes:

- The girder - which type of product is hung on the girder (colour); the size of hung products (size of coloured rectangle)
- Stopper on the path - the chart containing the percentage of blocking on the girder (girder cannot continue moving on defined path);
- Buffers - histogram of buffer occupation
- Checkboxes - used for setting a combination of simulated variants

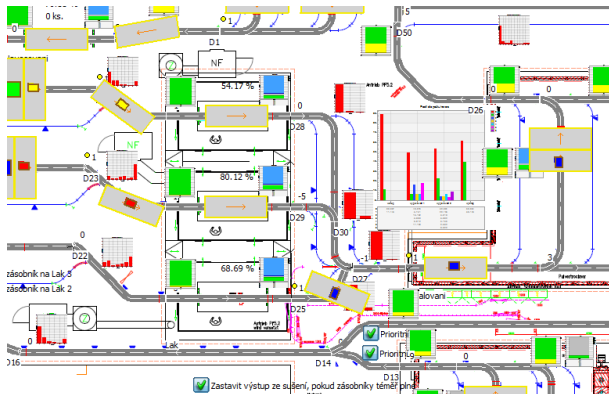


Figure 4: Animation of Simulated System

A detailed description of the simulation model and the values of the input parameters cannot be given due to the legal restrictions of the contract.

4. SIMULATION OF PRODUCTION OPTIMIZATION

The goal of this study was to find the appropriate measures to increase the production flow in the paint shop. Measures to be simulated can be defined as follows:

- Resize selected buffers before workplace (estimated investment)
- Possible changes to the path (planned investment)
- Logic of intersection control (path priority)
- Logic of possible path change according to current occupancy lanes on the path
- Change the logic of defining the path according to the painting program
- Change the logic of selection of paint booths according to current buffers occupancy located before paint booth than the selection of paint booths specified in the painting program.

4.1. Modelled measures

The goal of this simulation study was to find ways to increase the flow of products through the paint shop department. We tested a number of possible measures, which could lead to an increased flow rate:

- Resize selected buffers before workplaces - identification of possible buffers where capacity

can be increased, considering their area. These measures focus on changing the positions of the stoppers and the control software of the product path. The product path remains the same. Using these measures, it is possible to place more girders between the stoppers instead of one girder

- Possible changes to the path - the influence of planned investment in the creation of new transport paths connecting some workplaces. It is possible to shorten the length of the transport paths for some painting programmes
- Logic of intersection control (path priority) – testing the control system of selected intersections where multiple paths from the various workplaces are merged into one path. We tested the effects of the advantages of individual input paths. Tested variants were:
 - FIFO Method – right of way of first girder to arrive at an intersection
 - Priority path - each path received a priority for entry of the girder into an intersection
- Logic of possible path change according to current occupancy of lanes on the path - there were a number of duplicate workplaces with various transport paths. The painting programme clearly identifies the target workplace. We tested the control of sending the girder to a duplicate workplace if the buffer before the workplace specified by the painting programme is full (the buffer of the duplicate workplace cannot also be full)
- Change the logic of defining the path according to the painting programme – we tested the impact of a predefined path change on the painting programme. We mainly focused on an alternative workplace specified in the painting programme and the change of the selected path to this workplace
- Possible change of the logic of paint booth selection according to the fullness of the buffer located in front of a paint booth instead of definition in the painting programme – some paint booths have a partial interchangeability. We tested the impact on overall production of the selection of an alternative paint booth if the buffer in front of a paint booth is full (the buffer of an alternative paint booth is partially filled)

We tested other measures such as the impact of the shift model on individual workplaces, impact of the production structure of products on the overall production capacity, etc.

4.2. Conclusions from the Simulation Model

A simulation model was validated on the historical data from a company. We validated whether the model showed the same outputs as the already implemented production. We set the following parameters for validation of the model:

- The structure and the volume of products entering the workplace at a defined time according to the data from the information system
- Setting workplaces
 - The shift model
 - Times of operation
 - Number of duplicate workplaces
- The logic of control of transport paths

A pre-defined standardized HTML report of each simulated variant is automatically exported at the end of the simulation. This report contains settings and all relevant results from the simulation.

The following table shows the percentage change of performance of the modelled system according to the selected production program (different structure, number and size of products). Option 0 is the default state. The table shows that the proposed measures may lead to a 21% increase in production for a certain structure of products.

Table 1: Percentage Change of Performance of the Modelled System - First Part

Variant	0	1	2	3	4	5	6	7
Product structure 1	100	104.9	104.9	104.9	103.7	104.9	104.9	103.7
Product structure 2	100	100.5	100.7	100.9	101.3	99.8	99.9	100.0
Product structure 3	100	100.5	100.7	100.9	101.3	99.8	99.9	100.0
Product structure 4	100	107.1	106.8	106.9	107.4	108.0	108.7	108.0
Product structure 5	100	100.0	100.5	100.3	100.0	104.8	105.0	105.0
Average	100	102.6	102.7	102.8	102.7	103.4	103.7	103.4

Table 2: Percentage Change of Performance of the Modelled System - Second Part

Variant	8	9	10	11	12	13	14	15
Product structure 1	117.5	117.5	117.5	117.5	117.5	117.5	117.5	117.5
Product structure 2	119.9	119.7	120.4	120.9	121.0	121.3	121.0	121.3
Product structure 3	119.9	119.7	120.4	120.9	121.0	121.3	121.0	121.3
Product structure 4	111.9	112.0	112.5	115.1	114.7	114.4	114.7	114.4
Product structure 5	100.6	99.8	100.0	106.8	106.7	106.7	106.7	106.7
Average	114.0	113.8	114.2	116.2	116.2	116.3	116.2	116.3

5. CONCLUSION

The simulation study described here focuses on testing and design alternatives that lead to an increase in the

overall production of a paint shop department. The simulation study supports the proposed investment in the production system and the possible expected effect of these investments. The second group of simulation options focuses on the possible effects associated with organizational measures (shifts setting - not described in this paper) and also on the measures related to the change of the control logic of the automated programs that control the movement of the girders through the entire system that is modelled.

Simulation experiments demonstrate that the correct setting of the control logic of the production system usually has the most fundamental effect on the overall production of the entire system.

The results of the simulations proved the possibility to increase production by 21%. The overall increase of production is very dependent on the structure of the product type, size and quantity of products hung on girders. Recommendations from the results of the simulation study were implemented in the paint shop department.

ACKNOWLEDGMENTS

This paper was created with the subsidy of the project SGS-2015-065 “The development of parameters of the sustainable manufacturing system” carried out with the support of the Internal Grant Agency of University of West Bohemia.

REFERENCES

- Arinez, J., Biller, S., and Meerkov, S., 2010. Quality/quantity improvement in an automotive paint shop: A case study. *IEEE Transactions on Automation Science and Engineering*, 7 (4), pp. 755-761.
- Boysen, N., Scholl, A., & Wopperer, N., 2012. Resequencing of mixed-model assembly. *European Journal of Operational Research*, 216, pp. 594-604.
- Frank, M., Laroque, C., & Uhlig, T., 2013. Reducing computation time in simulation-based optimization of manufacturing systems. 2013 Winter Simulations Conference (WSC), pp. 2710 - 2721. Washington DC (USA).
- Gopalakrishnan, M., Skoogh, A., & Laroque, C., 2014. Simulation-based planning of maintenance activities by a shifting priority method. *Proceedings of the Winter Simulation Conference 2014*, pp. 2168 – 2179, Savannah (GA, USA).
- Horejsi, P., Horejsi, J., Latif, M., and Ulrych, Z., 2011. Automatic generator of lift dispatcher system model. *Annals of DAAAM for 2011 and 22nd International DAAAM Symposium "Intelligent Manufacturing and Automation: Power of Knowledge and Creativity"*, Danube Adria Association for Automation and Manufacturing, DAAAM, Vienna (Vienna, Austria).
- Cheng, S., and Park, S., 2010. The optimization research of train carriage painting. 2010 International

- Conference on Machine Learning and Cybernetics, Volume 3, pp. 1258 – 1263, Qingdao (China).
- Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., and Young, T., 2010. Simulation in manufacturing and business: A review. *European Journal of Operational Research*, 203(1), pp. 1-13.
- Lemessi, M., Rehbein, S., and Rehn, G., 2012. Semi-automatic simulation-based bottleneck detection approach. 2012 Winter Simulation Conference, pp. 1 - 12. Institute of Electrical and Electronics Engineers Inc.
- Lemessi, M., Schulze, T., and Rehbein, S., 2011. Simulation-based optimization of paint shops. 2011 Winter Simulation Conference, WSC 2011, pp. 2346-2357. Phoenix (Phoenix, AZ, United States)
- Li, J., Blumenfeld, D. E., and Marin, S. P., 2007. Manufacturing System Design to Improve Quality Buy Rate: An Automotive Paint Shop Application Study. *Automation Science and Engineering*, 4(1), pp. 75 - 79. January.
- Longo, F., Massei, M., and Nicoletti, L., March 2012. An application of modeling and simulation to support industrial plants design. *International Journal of Modeling, Simulation, and Scientific Computing*, 3(1).
- Rabbani, M. J., Ahmad, F. M., Baladi, J., & Khan, Y. A., 2013. Modeling and simulation approach for an industrial manufacturing execution system. Third International Conference on System Engineering And Technology (ICSET), pp. 26-31, Shah Alam (Malaysia).
- Raska, P., and Ulrych, Z., 2014. Hierarchical approach to developing a logistic discrete event simulation model using Automated Guided Vehicles. 26th European Modeling and Simulation Symposium, pp. 205-211. Bordeaux (France).
- Raska, P., and Ulrych, Z., 2015. Comparison of optimisation methods tested on testing functions and discrete event simulation models. *International Journal of Simulation and Process Modelling*, 10(3), pp. 279-293.
- Scholl, W., Laroque, C., & Weigert, G., 2014. Evaluations on scheduling in semiconductor manufacturing by backward simulation. *Proceedings of the Winter Simulation Conference 2014*, pp. 2552 - 2560. Savannah (GA, USA).
- Su, Z., Fu, Y., Tan, G., & Hu, Y., 2010. Application of Discrete-Event Simulation in Distribution Center Design Procedure: Framework & Cases Study. 2010 International Conference on E-Product E-Service and E-Entertainment (ICEEE), pp. 1-4, Henan (China).
- Ulrych, Z., Votava, V., Raska, P., Horejsi, P., 2013. Simulation of production systems and processes. Pilsen, Czech Republic: SmartMotion.
- Votava, V., Ulrych, Z., Edl, M., Korecky, M., and Trkovsky, V., 2008. Analysis and optimization of complex small-lot production in new manufacturing facilities based on discrete simulation. 20th European Modeling and

Simulation Symposium, pp. 198-203. Amantea (CS, Italy).

AUTHORS' BIOGRAPHIES

Pavel Raska is Doctor at the Department of Industrial Engineering and Management at the University of West Bohemia in Pilsen (Czech Republic). He holds M.Sc., Ph.D. in Mechanical Engineering at the same university. His research interests are oriented towards discrete event simulation, simulation optimization, modelling and simulation tools (ARENA, Plant Simulation) and working on practical simulation projects for companies.

Zdenek Ulrych is Associate Professor at the Department of Industrial Engineering and Management at the University of West Bohemia in Plzen and he is also a research worker in the Regional Technological Institute at the University of West Bohemia in Pilsen (Czech Republic). He holds M.Sc., Ph.D. and doc. in Mechanical Engineering at the same university. His research interests are oriented towards discrete event simulation, optimization in the simulation, modelling and simulation tools (ARENA, Plant Simulation), design and development of software and working on practical simulation projects for companies.

NO MORE CONFLICTS: THE DEVELOPMENT OF A GENERIC AIRPORT MODEL IN A SEQUENCE-OPTIMIZATION FRAMEWORK

P. Scala^(a), M. Mujica Mota^(b), D. Delahaye^(c)

^{(a),(b)}Amsterdam University of Applied Sciences- Aviation Academy
1097 DZ Weesperzijde 190, Amsterdam, The Netherlands

^(c)Ecole Nationale de l'Aviation Civile
7 avenue Edouard Belin 31055, Toulouse, France

^(a)p.m.scala@hva.nl, ^(b)m.mujica.mota@hva.nl, ^(c)daniel@recherche.enac.fr

ABSTRACT

Components of the airport airside such as runways, taxiways and aprons, have a significant impact in the total capacity of the airport system, where capacity is usually considered as maximum number of air traffic movements or number of passengers accommodated in a given period of time. Operations on the airside impact in the propagation of delay and consequently in the perceived level of service by passengers the terminal buildings. This paper put the focus on the airside operations at airports. A methodology for modelling operations on the ground and the successive optimization is proposed. The methodology presented in this paper is generic enough in the sense that it can be applied to any airport. The objective of this work is to come up with a generic tool that can be used by air traffic controllers in order to minimize conflicts on the ground and consequently increase the airport capacity.

Keywords: Modelling, Optimization, Airport Ground Side, Sliding Window

1. INTRODUCTION

Nowadays, with the constant increase of the air traffic demand, many airports are almost on the edge of their declared capacity. Figures from the previous year show that European flights have increased for the 1.5% in 2015 compared to 2014 (EUROCONTROL, 2016). These figures indicate the growth of air traffic, therefore, airports need to be able to accommodate this traffic without incurring in congestion situations, meaning that the capacity of airports needs to be increased. Airport capacity, considered as number of air traffic movements (landings and take offs), is mainly constrained by the runway system (Idris et al., 1998). In this context, the work of the air traffic controllers is to find a good balance between the rate of arrivals and departures in order to ensure safety and the smooth flow of aircraft on the airport airside surface. When priority is given to landings, the apron system is called to accommodate a big amount of aircraft and as a consequence it will release a smaller number of aircraft. Following this strategy, capacity of taxiway and apron system will be particularly exploited, with the risk that, at a certain point, they will

not be able to accommodate other aircraft, causing congestion situations. On the other hand, a strategy that prioritizes departures, causes the opposite effect, that is the release of gate capacity and consequently, the increase of queues at the runway holding points.

The research community has widely focused on problems about airside congestion, they came up with new methodologies providing good results. In literature we can find works about the optimization of surface operation where the main objective is to avoid congestion on the ground (Montoya et al. 2010; Simaiakis et al., 2014; Khadilkar and Balakrishnan, 2014). Many of them focused on the scheduling of the take off times in order to optimize the departure flow (Gupta et al. 2009; Rathinam et al., 2009; Pujet et al. 1999; Simaiakis and Balakrishnan 2009; Sandberg et al., 2014; Simaiakis and Balakrishnan 2015). A further objective that is often investigated by the different researchers is related to the reduction of fuel consumption (Simaiakis and Balakrishnan 2009; Simaiakis et al., 2014; Sandberg et al., 2014). Another branch of the research on ground operations focuses on gate assignment problem (Bolot 2000; Dorndorf et al., 2007; Kim and Feron 2012; Narciso and Piera 2015), here the main objectives can be to find the optimum number of gates required to absorb the traffic or to make a robust gate assignment, meaning that the gap between two aircraft using the same gate is maximized.

This work is based on the algorithm proposed in (Liang et al., 2015; Ma 2015) work, where a sliding window approach (Hu and Chen 2005; Zhan et al., 2010; Furini et al., 2015; Toratani et al., 2015) is applied to the sequencing and scheduling of aircraft flows in the airspace. The main contribution of this work is that, it addresses the problem at the airside implementing the aforementioned algorithm. The different problem tackled leads to come up with a different mathematical model that takes into account different decision variables. Moreover, in this work the decisions to be taken are operational so the time frame needs to be short enough to be used as a real-time application. Unlike the previous works done so far, in this paper all the components of the airside such as runway, taxiway and gates are considered together. All the operations that involve these

components are modeled, furthermore, an optimization algorithm is proposed in order to detect and solve all the conflicts. The optimization process was carried out implementing a meta-heuristic, Simulated Annealing (Kirkpatrick et al. 1983), with the objective of detecting and resolving ground conflicts.

This approach can be used for every airport since its general characteristics make it flexible and suitable to different airside configurations. The algorithm is tested using real data from Paris Charles De-Gaulle Airport. The main results are related to number of conflicts detected.

2. METHODOLOGY

In this work a methodology on how to model and optimize ground operation for a generic airport is presented. It is based on the work of (Liang et al., 2015; Ma 2015), and it consists in the implementation of a Sliding Window algorithm together with the use of a meta-heuristic for optimizing ground operations. In figure 1 the main steps of the methodology are shown.

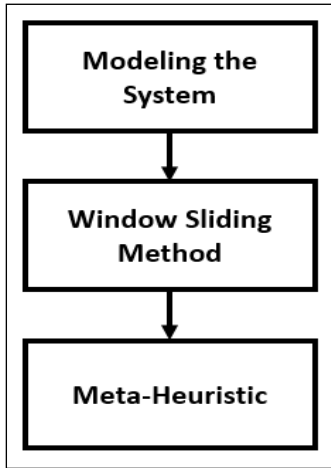


Figure 1: Methodology for modelling and optimizing airport airside operations

The first step consists in the modeling of the system, in other words, which level of abstraction to use to model the different operations. In the second step the optimization model is constructed using a sliding window approach, this approach leads to tackle the problem dynamically and to obtain a more robust solution. In the last step we make use of a metaheuristic, the simulated annealing, in order to solve this N-P Hard optimization problem. The potentiality of this meta-heuristic are described in detail in next section, but in general it allows to find a global optimum searching in the whole state space.

2.1. Modeling the Airport Airside

In the present work the emphasis is put on ground operations. As a first step the airport airside is modelled, with its components. The components that were modelled are listed below:

- Runways
- Taxiways

- Gates

Each one of these components has some generic parameters, in this way, every airport with a different airside configuration can be modelled. For instance, the component “runway” is characterized by the runway name, “taxiway” component is characterized by an entry, an exit and a length and the “gate” component is characterized by a gate number and type. To each one of the components some operations are associated, for example the “runway” component has associated landings and take offs, the “taxiway” component has taxi-in operations and taxi-out operations and the “gate” component has turnaround operations. In figure 3 the objects and their main operations are represented.

In next sections these components and their operations are explained in detail.

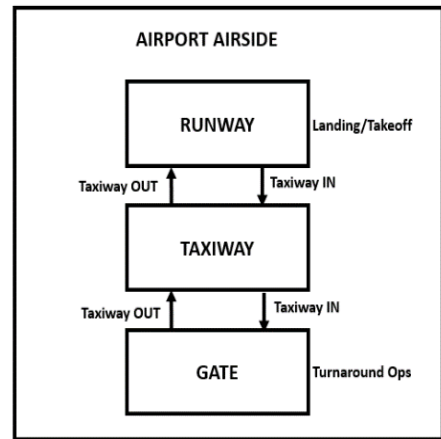


Figure 2: Components of the airside with their main operations

2.1.1. Runway

The runway component is the bottleneck of the airport ground system. Due to safety reasons it can accommodate only one aircraft at a time and even more because separation rules has to put in place based on aircraft ICAO wake turbulence category (WTC) (ref.). Therefore, runway component is mainly characterized by its name. Operations on the runway, like landings and take offs, need to respect separation rules and capacity constraints. Conflicts for this component were evaluated calculating the entry time of the trailing aircraft $T_{In}^{t,r}$ and exit time of the leading aircraft $T_{Out}^{l,r}$ of each aircraft $l, t \in \mathcal{F}$ on the runway $r \in \mathcal{R}$ for a generic operation (landing or take off). If the entry time of the trailing aircraft is less than the exit time of the leading plus the separation minima required $s_{l,t}$ then a conflict is detected (1).

$$R_{l,t}^r(\vec{x}) = \begin{cases} 0, T_{Out}^{l,r} + s_{l,t} < T_{In}^{t,r} \\ 1, otherwise \end{cases} \quad (1)$$

where, $R_{l,t}^r(x)$ represents the number of conflicts between a trailing and a leading aircraft $l, t \in \mathcal{F}$ on the runway $r \in \mathcal{R}$.

2.1.2. Taxiway

The taxiway network is represented by a set of different available taxiway routes. Each pair runway-gate as a specific taxiway route set available $tw \in TW(r, g)$ where tw is a taxiway route that belongs to the taxiway route set $TW(r, g)$ for every pair runway $r \in \mathcal{R}$ and gate $g \in G$. Each taxiway route is constituted by links and nodes. The choice of taxiway route is a decision variable of the optimization problem. On the taxiway network aircraft have a certain speed and a minimum separation is required between two aircraft. We can identify two different types of taxiway operation, “taxiway in” which is the pattern that leads the aircraft from the landing runway to the gate, and “taxi out”, which is the pattern that leads the aircraft from the gate to the departing runway. Conflict on the taxiway routes are detected on links and nodes. Regarding node conflicts they are detected using the same concept for runway conflicts (2).

$$N_{l,t}^n(\vec{x}) = \begin{cases} 0, T_{out}^{l,n} + g_{s_{l,t}} < T_{in}^{t,n} \\ 1, otherwise \end{cases} \quad (2)$$

where, $N_{l,t}^n(x)$ is the number of conflicts detected by two aircraft $l, t \in \mathcal{F}$ on the node $n \in \mathcal{N}$. A conflict is detected when the entry time for the trailing aircraft $T_{in}^{t,n}$ is less than the exit time of the leading aircraft $T_{out}^{l,n}$ plus the minimum separation between two aircraft on the ground $g_{s_{l,t}}$. Concerning the conflicts on links, the model makes sure that the aircraft sequence is respected along the links. Conflicts are calculated in the following way (3):

$$L_{l,t}^{me}(\vec{x}) = \begin{cases} 0, D_{l,t}^{me} \geq t_{g_{l,t}} \\ 1, otherwise \end{cases} \quad (3)$$

where, $L_{l,t}^{me}(x)$ is the number of conflicts detected by two aircraft $l, t \in \mathcal{F}$ on the entry link $me \in \mathcal{M}$. In this case a conflicts are detected both at the link entry and link exit $mx \in \mathcal{M}$ every time the distance between two aircraft $D_{l,t}^{me}$ is less than the minimum separation between two aircraft on the taxiway $t_{g_{l,t}}$ with $l, t \in \mathcal{F}$.

2.1.3. Gate

The gate is a parking position for aircraft. In an airport there can be many terminals, and each terminal has some gates where he aircraft can park. At each gate an aircraft undergoes the turnaround operations such as boarding, deboarding, fueling and so on, and then leaves to a new destination. The main parameter associated to the gate component are:

- gate number, which is an identifier of the gate and it is unique,
- gate type, if it is able to host all type of aircraft or only small and medium aircraft.

When an aircraft parks at the gate a “time in” is tracked and then as it leave the gate a “time out” is assigned. In this context, a conflict is detected when for each gate $g \in$

G the “time in” $T_{in}^{f,g}$ for each aircraft $f \in \mathcal{F}$ at a specific gate g , is less than “time out” $T_{out}^{i,g}$ for another aircraft $i \in \mathcal{F}$ and $f \neq i$ at the same gate g (4).

$$G_{f,i}^g(\vec{x}) = \begin{cases} 0, T_{in}^{f,g} > T_{out}^{i,g} \\ 1, otherwise \end{cases} \quad (4)$$

The turnaround time for making all the turnaround operation has been estimated as 45 minutes.

2.2. Sliding Window Approach

The Sliding Window method, or Rolling Horizon method, is based on the Receding Horizon Control (RHC) approach. It consist in the evaluation of the system in an extended time horizon, by dividing it in intervals of the same size (windows), and shifting the window using a specific time shift. In each window decisions are made, these decision will affect the decision that will made in the next window. This fact makes this approach to be dynamic. For each aircraft, four different status are identified: Complete, On-Going, Active and Planned. These status depend on the “Time In” and “Time Out” of the aircraft in the system and on the “Starting” and “Ending Time” of the current window. In figure 3 this approach is better explained.

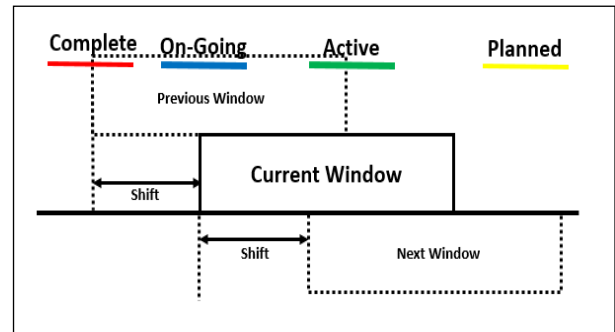


Figure 3: Sliding window approach

For instance, the status “Complete” happens when the aircraft “Time Out” is smaller than the “Starting time” of the current window, the status “Planned” happens when aircraft “Time In” is greater than the “Ending time” of the current window. These two status will not be taken into account when decisions are made. On the other hand, aircraft with the status “On-Going” and “Active” will be taken into account when decisions have to be made, since they are partially or entirely inside the current window.

2.3. Implementation of a Meta-Heuristic

Finally a meta-heuristic can be applied to conduct the optimization process and find an optimal or close to optimal solution. In this work the meta-heuristic Simulated Annealing was applied. Simulated annealing is a local search algorithm, and thanks to his hill-climbing movements it allows to search into the all state space in order to look for a global optimum avoiding to

be trapped into a local optimum. Its main parameters are temperature T and a cooling schedule. Starting from the temperature T , it generates neighbor solutions that are compared with the current one. Every neighbor solution that improves the objective function is accepted, but also solutions that degrade the objective function are accepted, and the acceptance rate for them follows a certain probability. In this way the algorithm tries to explore all the state space. At each temperature there is a number of transitions to be made, and with the cooling schedule the temperature is decreased. Therefore, if the temperature decreases slowly, the algorithm will perform better because it will search in a bigger portion of the state space. At the same time a high temperature requires a greater computational time which might be too long. In the same way higher initial temperatures are recommended as long as they do not affect too much the computational time. So the main parameters to set are the initial temperature, the number of transitions for each temperature and the cooling schedule.

The objective of the optimization process was the resolution of conflicts for every of the objects modelled. Conflicts have been explained in detail in the previous section.

The decision variables taken into account for the optimization process are:

- Taxiway-In route, given a set of available routes from the runway to the assigned gate
- Taxiway-Out route, given a set of available routes from the gate to the assigned runway
- Pushback Time, a delay assigned to the aircraft, when it is parked at the gate, before it can start taxi out operations.

The objective of the optimization model is to minimize the objective function that in this case is computed by the number of conflicts detected by the main components of the airside airport that were modelled, runway, taxiway and gate (5).

$$Z(\vec{x}) = \sum_{r \in \mathcal{R}} \sum_{l, t \in \mathcal{F}, l \neq t} R_{l,t}^r(\vec{x}) + \sum_{n \in \mathcal{N}} \sum_{l, t \in \mathcal{F}, l \neq t} N_{l,t}^n(\vec{x}) + \sum_{n \in \mathcal{N}} \sum_{l, t \in \mathcal{F}, l \neq t} (L_{l,t}^{me}(\vec{x}) + L_{l,t}^{mx}(\vec{x})) + \sum_{g \in \mathcal{G}} \sum_{f, i \in \mathcal{F}, f \neq i} G_{f,i}^g(\vec{x}) \quad (5)$$

3. SCENARIO AND RESULTS

In order to test the potentiality of the methodology it was taken as a case study Paris Charles de Gaulle Airport. As a first stage of the research, it was not modelled the entire ground surface, but only two of the four runways and one of the three terminals that constitutes the entire airport airside. In Figure 4 there is a representation of the portion of airside modelled in this work.

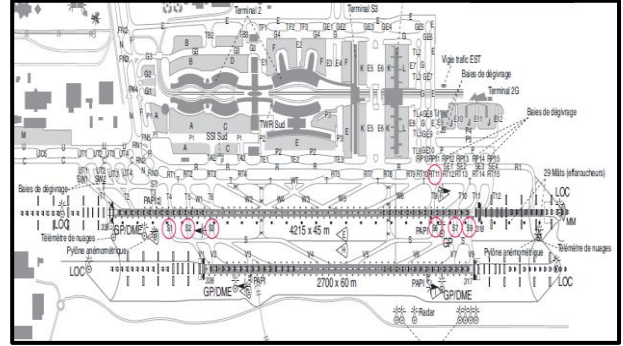


Figure 4: Part of the airside modelled of Paris Charles de Gaulle Airport

In table I all the main characteristics of the airside components are listed, whereas in table II shows the main parameters of the sliding window and the simulated annealing.

Table I: Airside components characteristics

Number of Runway	2 (one for landing and one for departures)
Taxiway network	336 taxiway routes set
Number of Gates	168

Table II: main parameters of the sliding window and simulated annealing

Sliding window method	
Window length	30 min
Shift	5 min
Simulated annealing	
Initial Temperature	Increasing the temperature until the acceptance rate of neighbor solutions reaches the 80%
Transitions	200
Cooling schedule	$\alpha=0.95, T_{i+1} = T_i \alpha$

3.1. Scenario

A timeframe of two hours of a specific day of operations in Paris Charles de Gaulle Airport was tested. It was chosen the peak time of the time from 6:00 A.M. to 8:00 A.M. In total there are 58 air traffic movements split in 22 arrivals and 36 departures.

3.2. Results

After running the optimization mode we obtained results about conflicts related to the airside components. We run first the optimization model without the use of the simulated annealing algorithm and then implementing it. In tables III and IV the results are shown.

Table III: results scenario without the implementation of the algorithm

Scenario without the implementation of the algorithm		
Computational time		44.326 sec.
Runway conflicts		45
Taxiway conflicts	links	0
	nodes	0
Gate conflicts		0

Table IV: results scenario with the implementation of the algorithm

Scenario with the implementation the algorithm		
Computational time		50,314 sec,
Runway conflicts		31
Taxiway conflicts	links	0
	nodes	0
Gate conflicts		0

In the scenario without the implementation of the algorithm there are 45 conflicts detected on the runway and 0 for the other components. In the scenario where the algorithm is implemented the number of conflicts is 30 and 0 for the other components. Looking at these results it can be seen that the algorithm improves the initial solution by 34%. Nevertheless, it does not give a conflict free solution.

4. CONCLUSIONS

In the methodology presented in this paper, the airport airside is modelled. The approach employed allows to obtain a high level of flexibility and abstraction that permits to model any airport airside. Furthermore, with the implementation of a sliding window algorithm and the use of simulated annealing meta-heuristic, we aimed at resolving conflicts for the airside components. Preliminary results show that the initial solution is improved with the use of this methodology, although not all conflicts have been solved. The solution that was found was tested in a time frame of two hours and it was achieved in 50,314 seconds, which makes this methodology able to be used in a dynamical environment where decision has to be taken in real time. The proposed methodology can be a very useful tool, when operational decisions have to be made for ground operations. The algorithm implementation needs further refinements in order to give a free conflict solution, this could be achieved tuning the sliding window parameters and the simulated annealing parameters. In the future would be good to compare the same methodology by using a different meta-heuristic and moreover introduce simulation approaches to test the feasibility of the optimized solution.

REFERENCES

- Bolat A., 2000. Procedures for providing robust gate assignments for arriving aircrafts. *European Journal of Operational Research*, 120, 63-80.
- Dorndorf U., Drexl A., Nikulin Y., Pesch E., 2007. Flight gate scheduling: State-of-the-art and recent developments. *Omega*, 35, 326-334.
- EUROCONTROL, 2016. Industry Monitor. Issue N180. 01/02/2016
- Furini F., Kidd M.F., Persiani C.A., Toth P., 2015. Improved rolling horizon approaches to the aircraft sequencing problem. *Journal of Scheduling*, 18 (5), 435-447.
- Gupta G., Malik W., Jung Y.C., 2009. A Mixed Integer Linear Program for Airport Departure Scheduling. *Proceedings of AIAA Aviation Technology, Integration, and Operations Conference*, pp. 1-13. September 21-23, Hilton Head, (South Carolina, USA).
- Hu X., Chwn W., 2005. Receding Horizon Control for Aircraft Arrival Sequencing and Scheduling. *IEEE Transactions on Intelligent Transportation System*, 6 (2), 189-197.
- Idris H.R., Delcaire B., Anagnostakis I., Hall W.D., Pujet N., Feron E., Hansman R.J., Clarke J.P., and Odoni A.R., 1998. Identification of flow constraint and control points in departure operations at airport systems. *Proceedings of AIAA Guidance, Navigation and Control Conference*, pp. 947-956. August 10-12, Boston, (Massachusetts, USA).
- Khadilkar H. and Balakrishnan H., 2014. Network Congestion Control of Airport Surface Operations. *Journal of Guidance, Control, and Dynamics*, 37 (3), 933-940.
- Kim S.H., Feron E., 2012. Impact of gate assignment on gate-holding departure control strategies. *Proceedings of Digital Avionics Systems Conference (DASC)*, pp.4E3-1 – 4E3-8, October 14-18, Williamsburg, (Virginia, USA).
- Kirkpatrick S., Gelatt C.D., Vecchi M.P., 1983. Optimization by Simulated Annealing. *Science* 220 (4598), 671-680.
- Liang M., Delahaye D., Xu X., 2015. A Novel Approach to Automated Merge 4D Arrival Trajectories for Multi-parallel Runways. *Proceedings of EIWAC 2015, 4th ENRI International Workshop on ATM/CNS*, 1-6. November 17-19, Tokyo, (Japan).
- Ma J., 2015. Aircraft merging and sequencing problems in TMA. Master thesis. ENAC.
- Montoya J., Woord Z., Rathinam S., Malik W., 2010. A Mixed Integer Linear Program for Solving a Multiple Route Taxi Scheduling Problem. *Proceedings of AIAA Guidance, Navigation, and Control Conference*, pp. 1-15. August 2-5, Toronto, (Ontario, Canada).
- Narciso M.E., Piera M.A., 2015. Robust gate assignment procedures from an airport management perspective. *Omega*, 50, 82-95.
- Pujet N., Declaire B., Feron E., 1999. Input-output modeling and control of the departure process of

- congested airports. Proceedings of Guidance, Navigation, and Control Conference and Exhibit, pp. 1835-1852, August 9-11, Portland, (Oregon, USA).
- Rathinam S., Wood Z., Sridhar B., Jung Y., 2009. A Generalized Dynamic Programming Approach for a Departure Scheduling Problem. AIAA Guidance, Navigation, and Control Conference, pp. 1-12, August 10-13, Chicago, (Illinois, USA).
- Sandberg M., Simaiakis I., Balakrishnan H., Reynolds T.G., Hansman R.J., 2014. A Decision Support Tool for the Pushback Rate Control of Airport Departures. IEEE Transactions on Human-Machine Systems, 44 (3), 416-421.
- Simaiakis I., Balakrishnan H., 2009. Queuing Models of Airport Departure Processes for Emissions Reduction. Proceedings of Guidance, Navigation, and Control Conference, pp. 1-26, August 10-13, Chicago, (Illinois, USA).
- Simaiakis I., Khadilkar H., Balakrishnan H., Reynolds T.G., Hansman R.J., 2014. Demonstration of reduced airport congestion through pushback rate control. Transportation Research Part A: Policy and Practice, 66, 251-267.
- Simaiakis I., Balakrishnan H., 2015. A Queuing Model of the Airport Departure Process. Transportation Science, 50 (1), 94-109.
- Toratani D., Ueno S., Higuchi T., 2015. Simultaneous Optimization Method for Trajectory and Sequence for Receding Horizon Guidance in Terminal Area. SICE Journal of Control, Measurement, and System Integration, 8 (2), 144-153.
- Zhan Z., Zhang J., Li Y., Liu O., Kwok S.K., Ip W.H., Kaynak O., 2010. An Efficient Ant Colony System Based on Receding Horizon Control for the Aircraft Arrival Sequencing and Scheduling Problem. IEEE Transactions on Intelligent Transportation System, 11 (2), 399-412.

HIGH-PERFORMANCE WEB SIMULATION

Jan Voracek^(a)

^(a) Faculty of Electrical Engineering and Informatics, University of Pardubice

^(a) jan.voracek@student.upce.cz

ABSTRACT

This paper demonstrates a possibility of creating high-performance web simulations. It utilizes Emscripten and asm.js to create a performance-optimized simulation that runs in a web browser. The paper provides a short introduction into used technologies and compares the performance of optimized simulation of fluid dynamics with usual approach of creating web simulations.

Keywords: computer simulation, web-based simulations, javascript, compilation.

1. INTRODUCTION

Computer simulation is an effective approach to study the behaviour of various systems over time. Simulated systems are often quite complex, and therefore have usually high demands on performance of the environment where they are executed. When creating a simulation model, it is necessary to take this into account and use corresponding level of abstraction. For example, we can replace complex input systems (systems from which data is entered into our model) with random number generators with corresponding probability distribution.

Sometimes, however, even abstraction from everything unessential may not be enough. Some simulation systems are computationally intensive themselves. The selection of appropriate technologies for the implementation of the simulator can help here.

Generally, we can follow rules saying that compiled languages provide higher performance than interpreted and that languages translated into native code provide higher performance than languages compiled into bytecode and run in a virtual machine (Java VM, CLR).

This paper, however, focuses on the web simulation and introduces possibilities of creating performance-optimized web simulations to the reader.

2. WEB SIMULATION

Web technologies are already used for several years in the field of simulation and are still gaining popularity. They are often utilized thanks to their portability. All you need to run a simulation is a modern web browser. You can run the simulation on a desktop computer, laptop, tablet, smartphone or smart TV (Voráček, 2015).

Lower performance is often referred as the biggest disadvantage of web simulation (Karták, 2015). We will see how it is possible to mitigate this disadvantage in the following chapters.

3. JAVASCRIPT

JavaScript is the only language that can run in the web browser¹ (on the client side). It is an interpreted language with a dynamic type system. This means that the source code is not compiled but runs directly.

All variables in JavaScript are dynamic. It means that their type can be changed by

¹ In Dartium web browser you can run scripts written in the Dart language. However,

Dartium is not intended for common use. See www.dartlang.org/tools/dartium/.

assigning another value and their true type is determined on run time. When you assign a one to a variable, it will have some numeric type. Whether it will be integer or floating point number depends on JIT (Hanenberg, 2010).

JIT (just-in-time optimization) is a process when the interpreter executing the source code tries to optimize the code during the execution (Sanghoon, 2012).

4. ASM.JS

Asm.js is a strict subset of JavaScript that can be used as a low-level, efficient target language for compilers. This sublanguage effectively describes a sandboxed virtual machine for memory-unsafe languages like C or C++. A combination of static and dynamic validation allows JavaScript engines to employ an ahead-of-time² (AOT) optimizing compilation strategy for valid asm.js code (Herman, 2014).

As JavaScript is a language with a dynamic type system, asm.js uses type annotations to indicate the type of variable. For example:

- `var x = f()|0;` tells that x is an integer,
- `var y = +f();` tells that y is a double,
- `var z = f()>>>0;` uses output as unsigned int.

Another very important part of asm.js specification is typed arrays. Software written in JavaScript usually uses simple, generic arrays (in fact they are maps). However, JavaScript contains also typed, fixed arrays which have better both memory and time complexity, for example:

- `Uint8Array` – array of 8-bit unsigned integers,
- `Int32Array` – array of 32-bit unsigned integers,
- `Float64Array` – array of 64-bit floating point numbers.

See the JavaScript reference for more.

5. EMSCRIPTEN

Emscripten is a source-to-source compiler that runs as a back end to the LLVM (Low Level

Virtual Machine) compiler and produces a subset of JavaScript known as asm.js described in the previous chapter (Zakai, 2011).

Emscripten also do some optimization of the code, for example:

- **Variable nativization:** Converts variables that are on the stack – which is implemented using addresses in the HEAP array (see Zakai, 2011 for more) – into native JavaScript variables.
- **Relooping:** Recreate high-level loop and if structures from the low-level code block data that appears in LLVM assembly.

Emscripten's compilation approach is to generate "natural" JavaScript, as close as possible to normal JavaScript on the web, so that modern JavaScript engines perform well on it (Zakai, 2011).

However, there are some limitations of Emscripten:

- **64-bit Integers:** JavaScript numbers are all 64-bit doubles, with engines typically implementing them as 32-bit integers where possible for speed. A consequence of this is that it is impossible to directly implement 64-bit integers in JavaScript, as integer values larger than 32 bits will become doubles, with only 53 significant bits (Zakai, 2011).
- **Multithreading:** JavaScript has Web Workers, which are additional threads (or processes) that communicate via message passing. There is no shared state in this model, which means that it is not directly possible to compile multithreaded code in C++ into JavaScript (Zakai, 2011).

6. LLVM

An LLVM is a compiler or a compiler infrastructure which compilers can be implemented in.

A simple diagram illustrating the LLVM workflow is shown in Figure 4. An LLVM

² Ahead-of-time optimization is an approach used by compiled languages where the optimization runs within the compilation step.

consists of a front-end and a back-end. The front-end translates high-level programming languages, e.g. C, C++, Java, Python, into a LLVM-IR (intermediate representation), which is a low-level programming language similar to assembly languages and is a language independent code. The back-end then translates the LLVM-IR into the architecture- and hardware-specific code (Lattner, 2004).

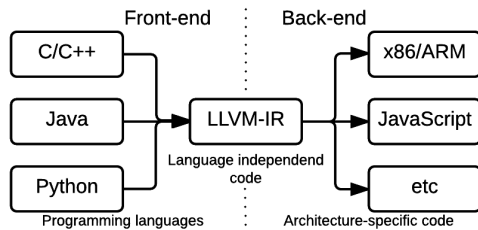


Figure 1 – LLVM architecture

7. PERFORMANCE

A computation of the Fibonacci sequence was chosen as a simple test for initial performance comparison. You can see the source codes below. The experiment was made with plain JavaScript (executed in NodeJS), for C++ (compiled by clang) and with asm.js (also executed in NodeJS).

This microbenchmark is indeed very simple; however, even this simple code can demonstrate that the performance asm.js is quite higher than hand-written JavaScript.

```
#include <iostream>

long fib(int x) {
    if (x < 2) {
        return 1;
    } else {
        return fib(x - 1) + fib(x - 2);
    }
}

int main() {
    long result = fib(50);
    std::cout << result << std::endl;
    return 0;
}
```

Source code 1 – Computation of 50th number of Fibonacci sequence in C++

Code written in JavaScript is not very different. Besides the data types they are substantially identical.

```
function fib(x) {
    if (x < 2) {
        return 1;
    } else {
        return fib(x - 1) + fib(x - 2);
    }
}

var result = fib(50);
console.log(result);
```

Source code 2 – Computation of 50th number of Fibonacci sequence in plain JavaScript

In the following example you can see the code generated by Emscripten. It is a JavaScript code compiled from C++ code (Source code 1). You can see, for example, the type annotations mentioned in chapter 4 or that the AOT optimization removed one of the recursive calls.

```
function __Z3fibL($x) {
    $x = $x|0;
    var $0 = 0, $1 = 0, $2 = 0, $3 = 0, $4 = 0,
        $5 = 0, $accumulator$tr$lcssa = 0,
        $accumulator$tr1 = 0, $x$tr2 = 0, label = 0, sp
    = 0;
    sp = STACKTOP;
    $0 = ($x|0)<(2);
    if ($0) {
        $accumulator$tr$lcssa = 1;
        return ($accumulator$tr$lcssa|0);
    } else {
        $accumulator$tr1 = 1;$x$tr2 = $x;
    }
    while(1) {
        $1 = (($x$tr2) + -1)|0;
        $2 = (__Z3fibL($1)|0);
        $3 = (($x$tr2) + -2)|0;
        $4 = (($2) + ($accumulator$tr1))|0;
        $5 = ($3|0)<(2);
        if ($5) {
            $accumulator$tr$lcssa = $4;
            break;
        } else {
            $accumulator$tr1 = $4;$x$tr2 = $3;
        }
    }
    return ($accumulator$tr$lcssa|0);
}
```

Source code 3 – Computation of 50th number of Fibonacci sequence – source code generated by Emscripten

On the following graph you can see the comparison of execution times. You may notice that the code compiled from C++ to JavaScript is quite faster than hand-written JavaScript.

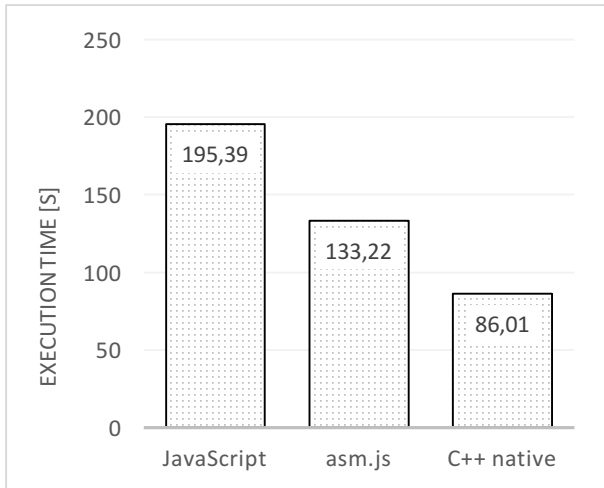


Figure 2 – Comparison of execution time of calculation of 50th number of Fibonacci sequence

As you can see on Figure 3, also other benchmarks confirm that JavaScript code compiled from C++ is almost always faster than hand-written JavaScript.

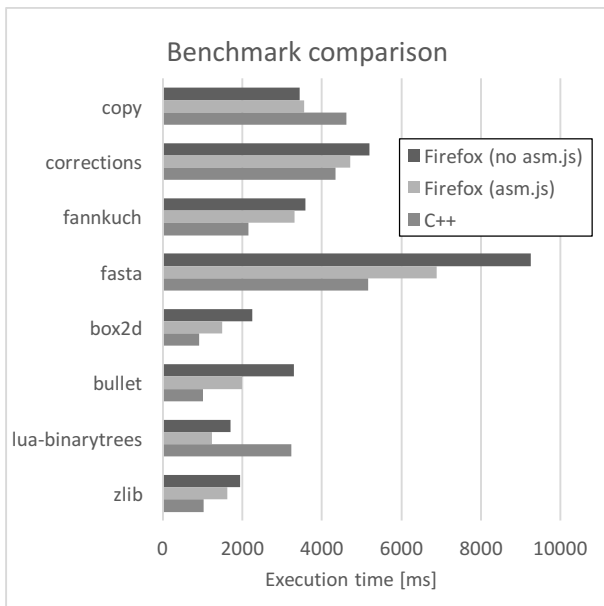


Figure 3 – Comparison of benchmark results [Data source: arewefastyet.com on March 24, 2016]

8. CASE STUDY

In 2013 Daniel Schroeder from Department of Physics at Weber State University, Utah has implemented a simple fluid dynamics simulator in JavaScript based on Lattice-Boltzman methods. You can see this simulator on Figure 4.

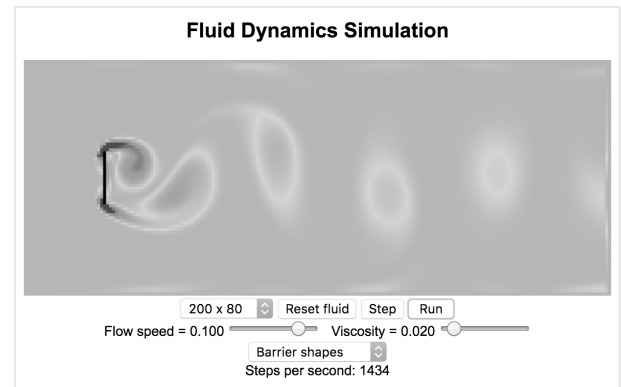


Figure 4 – Fluid dynamics simulator

This simulator was chosen as a reference for the comparison. It was reimplemented in C++ and transpiled back to JavaScript using Emscripten. It was chosen because fluid dynamics is not entirely trivial in terms of mathematical calculations and it is simple to animate.

As you can see on Figure 5, there is a difference in how many simulation steps per second are executed. The original JavaScript implementation performs 1050 steps per second while the C++ implementation compiled to JavaScript performs 1434 steps per second (average number of steps after 5 minutes of simulation). That is 36.57% higher performance.

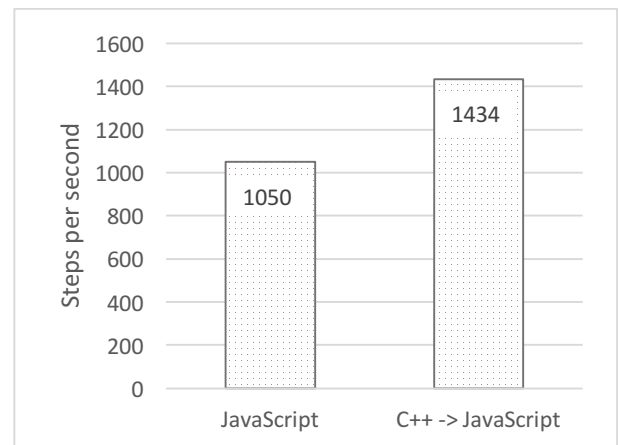


Figure 5 – Performance test of fluid dynamics simulators

9. CONCLUSION

Performance optimization is an important and frequently discussed topic, especially in the context of web-based simulation. Results presented in this paper indicate that implementing simulators in languages with strong type system like C++ and their

compiling to JavaScript is worth considering. JavaScript optimized by Emscripten gave often better performance than hand-written JavaScript.

There is also a lot of mathematical and physical libraries for C++ and this approach allows you to use them easily in your web-based simulators.

REFERENCES

- Hanenbergs S., 2010. An experiment about static and dynamic type systems: doubts about the positive impact of static type systems on development time. In Proceedings of the ACM international conference on Object oriented programming systems languages and applications, pp. 22–35, New York, NY, USA.
- Herman D., Wagner L., Zakai A., 2014. Specification of asm.js. Available from: <http://asmjs.org/spec/latest/> [November 2015].
- Karták Š., 2015. Web simulation as a platform for training software application. In Proceedings of the 27th European Modeling and Simulation Symposium (EMSS 2015), pp. 70-78, Bergeggi, Italia.
- Lattner C., Adve V., 2004. LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation. In Proceedings of the international symposium on Code generation and optimization: feedback-directed and runtime optimization (CGO '04), pp. 75-, Washington, DC, USA.
- Sanghoon J., Jaeyoung C., 2012. Reuse of JIT compiled code in JavaScript engine. In Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC '12), pp 1840-1842, New York, NY, USA.
- Voráček J., 2015. Web Simulation with Support of Mobile Agents. In Proceedings of the 27th European Modeling and Simulation Symposium (EMSS 2015), pp. 32-35, Bergeggi, Italia.
- Zakai A., 2011. Emscripten: an LLVM-to-JavaScript compiler. In Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion (OOPSLA '11), pp. 301-312, New York, NY, USA.

INTERACTIVE DISTRIBUTED SIMULATION REALISED AS A WEB APPLICATION

Stepan Kartak

University of Pardubice, Faculty of Electrical Engineering and Informatics

stepan.kartak@student.upce.cz

ABSTRACT

This paper is focus on a practical use of a web browser and its modern technologies for realisation of distributed web-based simulations. A web application is realised in the environment of a web browser, providing basic software support and additional services for a user-friendly, simple to realise (above all simply programmable) interactive, distributed, discrete simulation. The user is freed from having to deal with the process of creating a simulation core, synchronisation methods etc., and can fully concentrate on the logic of the solved problem. The paper also describes the realisation of the simulation in a web browser, the appropriate classification and extent of applications which are suitable for this service, as well as the positives and negatives of a web-based simulation realisation.

Keywords: Distributed Simulation, Web-based simulation, HTML5, WebRTC

1. INTRODUCTION

This work focuses on use of a web browser to realise a user-friendly distributed simulation. The aim is to create a web-based application which would provide the user with basic functionality (simulation core, synchronisation, interactive approach ...) for realisation of logical processes, composition and configuration of a distributed simulation model, central management and record of data. The user is then freed from tackling the aforementioned fundamental problems, allowing concentration on programming which is the actual problem to be solved.

Web browsers are more than suitable for such an application. Browsers have contained functions allowing solving problems of this classification without third-party plugins since the year 2012 (Kartak, 2015). However, in spite of the progress web browsers have made in the last few years, creating distributed simulation was only possible after making a number of compromises. For this reason, we will operate with a distributed discrete simulation made up of a maximum of 20 logical processes (theoretically, the number is unlimited).

A web browser is suitable for the realisation due to many reasons – one of them being that it is available on

practically every device which can connect to a computer network. Algorithms of the solution were programmed in today's well-known programming language JavaScript. This fact is an advantage as well – high availability.

The foundation of the solution is based on previous work, covering distributed web-based simulation (Kartak 2014; 2015). The previous work was aimed at simulators (simulator applications for testing or education, of workers / dispatchers). This solution contained many compromises (e.g. unshared state space), and is surpassed by the introduced work.

2. WEB APPLICATION CHARACTERISTICS

The web application allows the user to define a distributed, synchronised space, made up of logical processes available in the form of web pages.

There are three main parts of solution:

- Distributed and synchronised (memory) space: visible to all logical processes it contains. Configurable.
- Logical process type: prearranged web page, which represents a logical process after it has been configured
- Logical process: an instance of the logical process type; a concrete webpage engaged in the simulation.

The realisation itself presents a user with an editor (Fig. 1), i.e. a 2D area which represents a configurable, shared, distributed space, into which logical processes are incorporated. Any number of logical processes may be placed here (realised as a placeholder icon in the editor, with a practical maximum of 20 logical processes). Every logical process represents a workstation or user space. Every logical process has access to the global state of the whole simulation. The application itself provides services of a simulator and automatic synchronisation of logical processes. Furthermore also provides access to configuration of the 2D space, in which logical processes are placed, and in which the simulation unfolds. All functionality has been designed with interactive behaviour in mind.

The logic itself is programmed by the user alone (i.e. the user programs the behaviour of logical processes). Since the simulator itself and synchronisation methods are prepared in the form of a library, the user must create only discrete activities, and in some cases an animated output.

Due to the limited computational capacity of web browsers (see chapter 3) and high requirements on network infrastructure, use on local networks for discrete distributed simulation is expected.

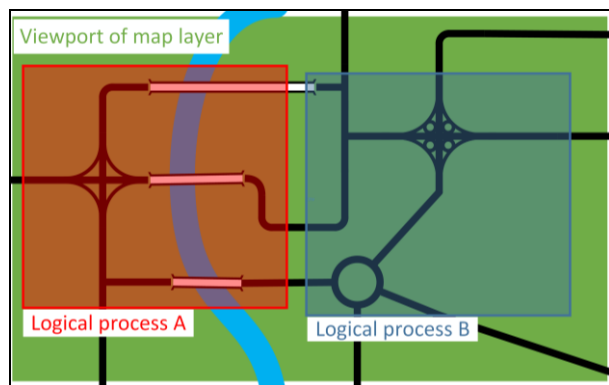


Figure 1: Detail of idea of simulation editor, example of simulation model (the road infrastructure) defined from two logical processes with defined playground 2D area (part of a map).

Considered application/simulation characteristics that can be realised this way:

1. A foundation for „simulator applications“, i.e. applications (interactive simulations) for testing (training, education) of workers / dispatchers. Dispatcher (i.e. one logical process, serviced by a computer/workstation operator). To illustrate, imagine a railway station dispatcher in a region (where there are more station dispatchers). In the same way, a simulation of a technological process / production can be realised. In this case, logical processes represent work areas of individual operators in the technological process. In this scenario, the shared space would represent the scheme of a production process, in which the stations of dispatchers would be located (i.e. logical processes).
2. A simple creation of a multiplayer game, where logical processes represent the area for individual players, and the allocated (configurable) 2D space of simulation represents the game area.
3. Distributed space for data exchange in the frame of a workgroup. This case does not represent an actual simulation. Only the synchronisation methods would be utilised, to keep the memory space for all logical processes up to date.

The primary classification of a possible use can be generally specified as a distributed system which requires an interactive approach, which does not have high computational requirements, and which is based on discrete events.

Interactive approach is represented by user input (keyboard, mouse), which alters the logical process (and the distributed simulation as a whole). The execution of the user input is realised in relation with the online animation output, i.e. the user influences the immediate state of the simulation that is displayed (see Image 9 - screenshot of use case example).

The aim of this work is not to create competition for extensive standards such as DIS, HLA, TENA etc. (IEEE 1278.1-2012; Kuhl et al. 2007), and such a thing is not even possible due to the limitations of a web browser (see chapter 3). The aim is to utilise the availability of a web browser to build a simple, yet functional and rationally available (if we disregard the programming of the necessary logical behaviour) distributed interactive simulation.

3. WEB BROWSER ENVIRONMENT SPECIFICATION

A contemporary web browser provides sufficient functional support for the realisation of any application, which would have been, until recently, considered a solely desktop application. To illustrate, take working with files, the 2D and 3D graphic drawing possibilities, and direct server to client communication (WebSocket technology), as well as direct peer-to-peer connection between browsers (WebRTC, W3C 2016). This solution is based on these features. All these new functions are commonly denoted as HTML5.

In spite of the new functions mentioned above, web browsers are still an imperfect platform for demanding applications. The limiting factors essential for this work are as follows:

JavaScript is a programming language which web browsers use to add dynamic behaviour to web pages. It is a weakly (loosely) typed, prototype language, which is compiled at runtime of the script. JavaScript also contains several functions, which complicate optimisation of the compiled code, although this condition is constantly improving with new versions of web browsers. Libraries which compile C/C++ code into a well-optimised JavaScript code also exist. However, a second, greater problem persists – the “single-threaded” (this term is only partially accurate, but covers the essence of the problem, which is why the term is used throughout the text) approach.

A single threaded event-based system of processing user code is a critical problem for a distributed simulation application. These characteristics of JavaScript practically mean that all operations are

executed synchronously in a single thread. There is no concurrent multithreaded solution available to users. (Processes the user cannot influence, such as output data rendering, network communication etc. are done asynchronously by the browser.) This does not present a problem to many algorithms that the browser usually realises. But in general, this single-threaded situation makes creating algorithms more difficult, because some problems require parallel operation (in terms of both multi-threaded and multi-processor processing), due to effectivity or time requirements of individual tasks. Specific reasons why this fact presents a serious problem for the implementation distributed simulation are defined in chapter 4.

4. IMPLEMENTATION OF A DISTRIBUTED SIMULATION IN A WEB BROWSER

The previous chapter covers JavaScript and its functionality. This chapter focuses on a specific description of the simulator and simulation topology, in relation to the previously mentioned problems, and explains how the synchronisation was done.

4.1. Single-threaded web application

The simulator must process four independent critical tasks that are constantly running: (i) simulation core executing discrete events, (ii) network communication (accepting and sending messages) with other members of simulation, (iii) user input processing and (iv) animation output.

These four tasks are usually realised as parallel tasks in classic desktop applications. This is not possible in JavaScript. Because of this fact, the simulator is realised as a series of cyclically repeated instructions (only essential steps are listed):

1. Evaluation of incoming messages.
2. User input processing.
3. Execution of available (especially in relation to the logical process synchronisation) discrete events.
4. Animation output calculation:
 - entities' position,
 - collisions and other interactions between entities,
5. The information about state changes are sent to all other logical processes.
6. Situation rendering onto the animation output.
7. Cycle repeats from step 1.

The algorithm described hints at a problem, which originates in the event-based nature of JavaScript. The event-based approach means that any sequence of commands may be interrupted by another task (user simulation input, accepting messages, etc.). An event-based approach is not a problem in itself, as after an event is processed, the interrupted "main" algorithm continues. The problem is, that these "unexpected" events naturally take some time to be processed, and are thus slowing down the computation, which negatively

influences the synchronisation of logical processes, and the animation fluidity as well.

4.2. Logical process interconnection topology

Considering the above stated problems with computational capacity of web browsers, realisation takes place as a purely peer-to-peer simulation. Communication between processes takes place directly, without a server. This implementation was chosen in order to: (i) minimise infrastructure expenses and (ii) lower communication load (i.e. latency in communication between individual browsers). The server is used only for setting up the connection and "bookkeeping" of the simulation state.

A one-on-one connection is realised between every logical processes (with the WebRTC technology). During the initialisation process of the simulation, a connection is made between every logical process. The connections are realised, above all, to establish and ensure a global memory state. All changes of the state of a logical process are sent to every other logical process (in fact, changes are broadcasted), and it is up to each individual logical process whether or not the accepted data will be processed, and how. This technique is inspired by the DIS standard.

These general broadcasts are also utilised for synchronisation purposes.

4.3. Logical process synchronisation algorithm

The optimistic methods of synchronisation are usually more suitable for interactive simulation, as they do not require strict logical process runtime synchronisation. This synchronisation in turn, makes the calculation (and, by the same rule, animation as well) smoother (i.e. there is no waiting for "slower" logical processes). To ensure a fluent performance (with animation fluidity in mind), the conservative approach is not effective, as it requires a short lookout (look-ahead) to ensure a fluent animation, which increases communication load.

The synchronisation solution was designed with interactivity (the user influences the immediate state displayed in the form of an animation – see chapter 2) of the application in mind.

In the end, a "two-level" synchronisation method was chosen:

1. For basic synchronisation, the Conservative synchronization technique of sending null messages with a look-ahead (Chandy-Misra-Bryant Distributed Discrete-Event Simulation Algorithm, Fujimoto 2000) is used – a specific implementation is defined in the previous work (Kartak 2015). This method is applied, above all, to start the simulation, and to intercept greater-than-average fluctuations (delays) in network communication. Look-ahead of 120 ms is used.

2. For precise synchronisation, time readings of state messages sent periodically (every 40 ms) by every logical process to all other processes in simulation (broadcast message sending). These messages are labelled *broadcast*.
 - The animation output displays a continuous animation of the logical process state. The animation is scaled in terms of time, with 1 second of animation corresponding to X seconds of real time. Considering the user experience and possibilities, a 1:1 ratio is used, where one second of animation corresponds to one second of real time.
 - Considering point one, synchronisation of animation outputs between individual processes is not ensured (look-ahead 120 ms).
 - Information from broadcast messages are used for precise synchronisation (chapter 4.2), which are realised as discrete activities (they are part of the basic synchronisation).

The application allows two ways of synchronisation (depends on the nature of the simulation), which differ primarily in the animation – simulation core relationship (executes discrete activities and synchronisation):

1. **Fully conservative synchronisation:** Time order of activities is strictly upheld, i.e. the simulator awaits null messages if necessary. Animation is executed only in the safe time between discrete activities. This way of synchronisation is completely safe from the time perspective, and allows for acceptable framerates of animation, considering the network latency when transferring the null message request and response averages between 10 – 15 ms on a local area network, and 25 FPS gives 40 ms between each rendering of the animation scene. A disadvantage is FPS limitation and slight “slacking” of the simulation, caused by necessary waiting for synchronisation (during this wait, the animation may not proceed – the simulation time is not advancing).
2. **Minimal synchronisation:** The animation output is not fixed on animation activities. The simulation core works with regard to the time of the animation output, i.e. activities corresponding to the time of the animation output. In this case, the conservative synchronisation method is used primarily for preliminary synchronisation. This method of synchronisation is still in development, but provides more computing power to simulations with focus on interactive behaviour instead of on perfect synchronisation (there is no waiting

for synchronisation). As an example (see chapter 6 - use case) the achieved framerate is between 70 and 80 FPS (the minimalistic scene - around 50 animation activities), three times more than the first method. The user can then observe a perfectly continuous animation, eventually there is more space for demanding calculations.

Due to the above stated facts, use only on local networks is assumed, where the latency of WebRTC messages varies between 10 and 20 ms, which is a state allowing fluent operation, assuming the extent of the application.

5. BASIC DESCRIPTION OF THE WEB APPLICATION

The web application is divided into five parts. The administrative interface is made up of four parts: (i) model configuration, (ii) simulation management, (iii) visualisation centre, and (iv) an initialisation (signaling WebRTC server, W3C 2016) server. The fifth part is represented by a software library for logical processes. The individual parts are described in the following sub-chapters.

5.1. Software library for logical process implementation

The simulation model is made up of logical processes, which are hereby represented by web pages. The logic of these processes is implemented by users. A JavaScript library is available, which provides:

- Connection of the logical process to the administration interface.
- Synchronisation of a running simulation.
- Basic functional support for animation output.
- Auxiliary classes and functions extending the standard JavaScript functions and commonly available JS libraries with practical classes (working with time) and data structures (priority queues etc.).

5.1.1. Basic structure of logical processes

A logical process is made up of 6 parts (for an UML diagram see Image 2, for graphic illustration of relations see Image 3):

1. Simulation core: operates simulation activities, ensures synchronisation. Includes:
 - Calendar: priority queue for simulation activity planning.
 - Environment: contains environment and state information related to the simulation (primarily activity handler).
 - Modules: any named data structure, usually auxiliary, available to all dependent parts (usually activity handler).

Used, among others, for the text report of simulation states.

2. Simulation activity: specified the type of activity, time of execution and any other additional information
3. Activity Handler: execution of given activity type
4. ConnectionRegister: logical process communication realisation layer
5. Animation Activity: Described a graphic element for animation rendering. One of the modules of the simulation core.
6. AnimationManager: renders a scene based on the animation activities

Other program parts that are not critical for the execution of a logical process:

7. SettingsManager: contains a description of the simulation configuration.
8. EntityManager: contains information about entity types and individual entities.
9. ActionManager: describes interactions and eventual reactions of individual entity types.

The solution as a whole works under several basic premises:

- All simulation and animation activities can be serialised.
- All simulation and animation activities can be interrupted at any time (removed from the queue or scene).

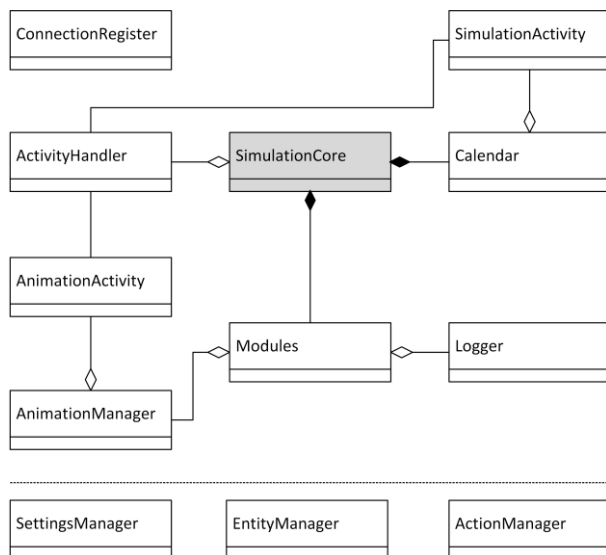


Figure 2: Basic UML schema of logical process (the simulator itself)

5.2. Model configuration

This administrative interface serves to register individual types of logical processes, and allows their subsequent use when building the distributed simulation model.

The basic approach to create a distributed simulation is as follows:

1. A user registers a custom logical process type.
2. The visual editor provides a 2D space (the distributed simulation space). Figure 3.
3. In this area, the chosen logical process types are placed (the logical process types are reusable in the frame of a simulation model, i.e. one type of a logical process can be used several times in a single distributed simulation). The user configures logical processes, and eventually the simulation itself (i.e. configuration of the 2D shared space) – the specification is done by XML for testing purposes at the moment.
4. Simulation is ready to run. Every logical process has a unique link, which can be sent to a user, who will then operate it (dispatcher, production operator, player, etc.).

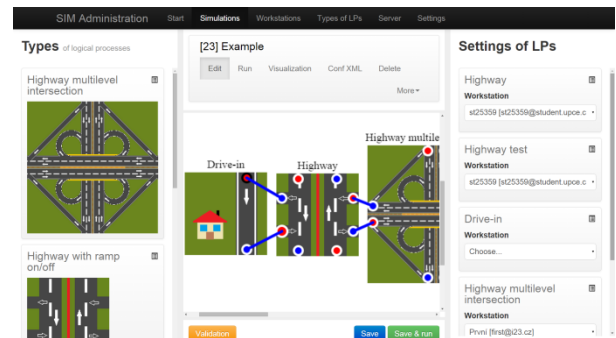


Figure 3: Administration web interface, visual editor; blue lines are network connections between logical processes

5.3. Simulation management

To simplify the organisation and launch of a distributed simulation, a module for central management of all logical processes is available (realised as a standalone application).

5.3.1. Application characteristic

JavaScript Remote control (JSRC) is an application generally designed to control web pages by text commands in bulk. In this case, it serves to control logical processes. Primarily, it is used to load a logical process, create individual WebRTC connections between all logical processes and launch of all logical processes. It is also used for simulation state accounting. A detailed description of this application is published in previous paper (Kartak, 2015). A brief description follows.

Web pages are identified as *workstations* within JSRC. Workstations are defined by:

- name,
- initialization page (page that will be controlled).

These workstations are grouped into categories of workstations (*workstation groups*). And just within these groups it is possible to enter mass commands (nevertheless it is possible to specify a workstation subset that will receive the command).

Really the workstation corresponds to the logical process and one group of workstations corresponds to the whole model.

The commands are entered by text form (figure 4, Kartak 2015), moreover there is an option to create advance prepared command sets at “one click” (figure 5, Kartak 2015).

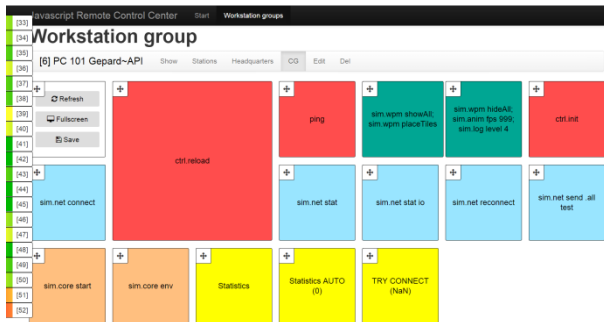


Figure 4: JSRC: Prepared command set, one square is user-defined command (or commands), prepared for touch-devices

The commands are user-defined JS functions of any kind. The use is virtually unlimited. Commands are easily and quickly extensible.

5.3.2. Incorporation in the application, API description

In reality, a workstation corresponds to a logical process, and a whole model corresponds to a workstation group.

As stated before, JSRC is an independent application on an independent server, with other applications (here: administrative interface, logical processes, central visualisation) communicating through API, which allows:

- add groups of stations (WG_ADD),
- add stations to groups (WE_ADD),
- add commands (COMMAND_ADD),
- remove commands for processing (COMMAND_FETCH),
- send responses (results) of commands (COMMAND_RESULT),
- get information about commands (primary results, COMMAND_GET),
- get information about station groups (WG_INFO),
- get information about very stations (WE_INFO).

An usage scheme in the simulation is displayed at figure 6, Kartak 2015.

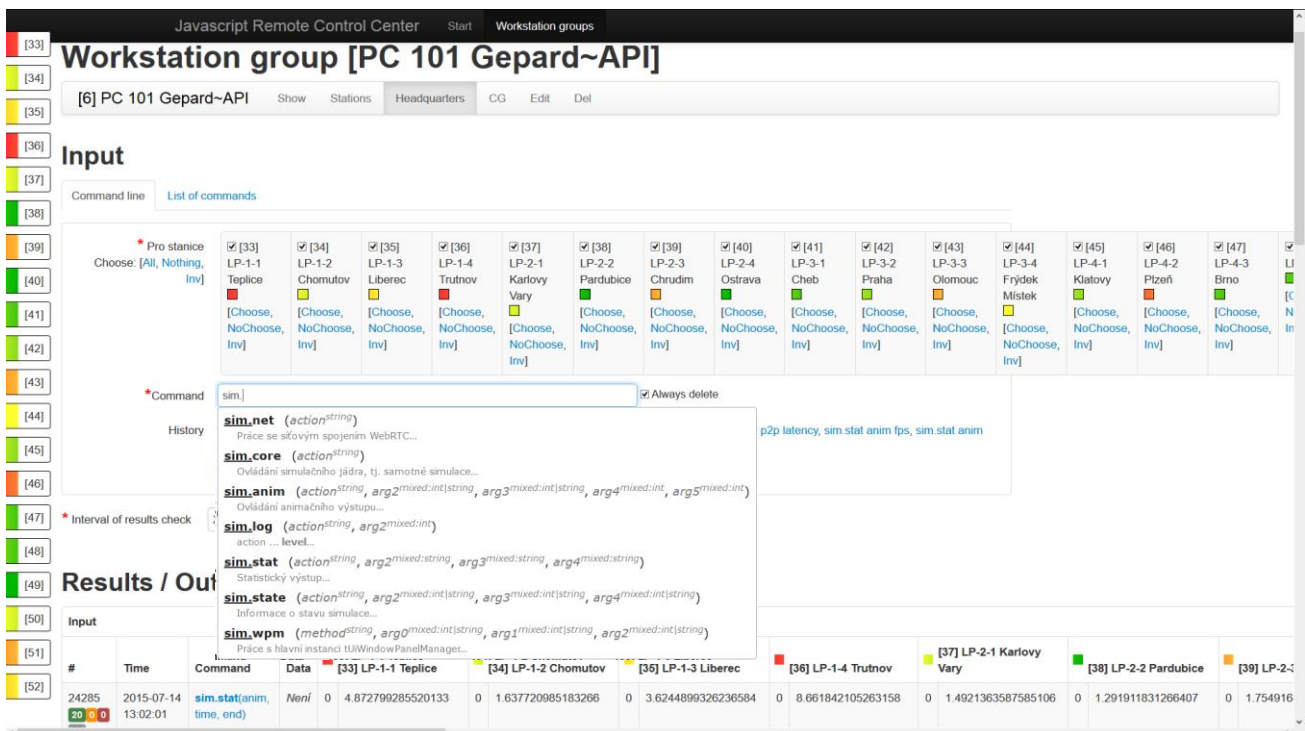


Figure 5: Remote Control web interface

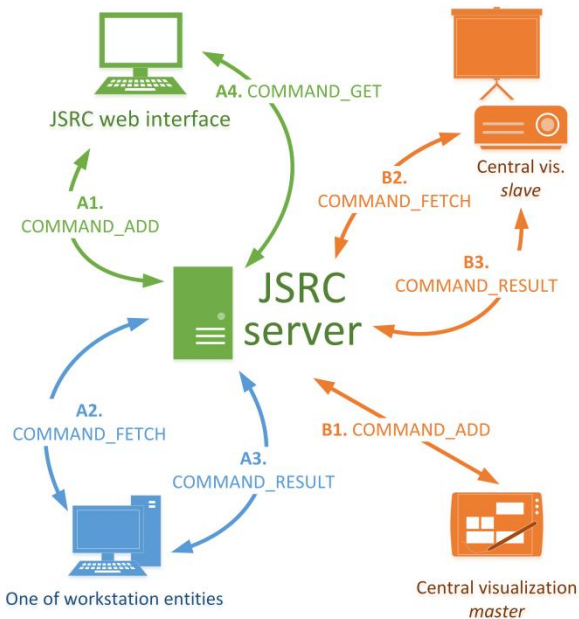


Figure 6: The scheme of usage JSRC API in the simulation

A JS library is available, which implements API calls, which then in turn executes specified actions on the client side (e.g. „Launch logical process“) or on the server side (e.g. „Add Workstation“).

5.4. Central visualisation

The administration interface is extended by the central visualisation (CV). This module allows (i) recording of animation output (Image 7) and (ii) capture screenshots for static preview of logical process state (Image 8).



Figure 7: Logical process animation replay in central visualisation

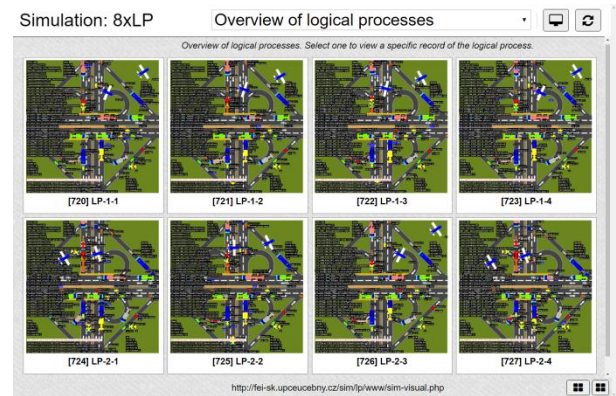


Figure 8: Overview of logical process in simulation in central visualisation

Animation output recording is realised as a recording of the animation activities. These activities are distributed in batch into a player (see Image 7). Due to time delays lower server load with increased number of logical processes, it is possible to:

- Observe the animation output of a chosen logical process “on-line” with a 0 to 2 second delay (batch update of the player). Due to conservative method of synchronisation and used look-ahead, it is possible to observe the same scene as the workplace operator (under optimal conditions). Appropriate for e.g. supervising worker during dispatcher training etc.
- Run an animation recording at any time.

Detailed description is stated in previous paper (Kartak 2015).

5.5. Initialisation server

The initialisation (signaling) server is designed to establish a peer-to-peer connection between all logical processes. It is a simple web application, which allows monitoring of connection state (unconnected, negotiation, connected). Meaningful only before a simulation starts.

6. TESTING AND USE CASE

For testing, a scenario from previous developing activities was chosen (Kartak 2015). For a logical process scene (screenshot) see Image 9. It is a typical two-level cloverleaf highway interchange.

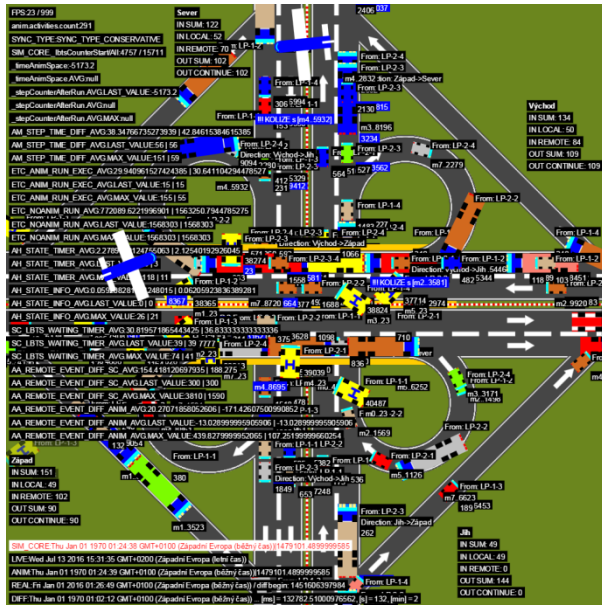


Figure 9: Screenshot of simulation, the scene of two-level cloverleaf highway interchange is consist of work of 8 logical processes, debug level

The scene is shared between all logical processes, with every logical process generating „its own“ vehicles at

driveways. All vehicles are displayed in all other logical processes. There are several vehicle types, differentiated by colour, speed and appearance. All of the aforementioned elements are used to create an environment that generates activities and tasks for processing, but are still considered to be „static“, non-interactive parts (the user cannot influence them).

The interactive element is represented by a single vehicle, which reacts to user input (keyboard – cursor arrows – changing directions, spacebar – stopping the car, mouse click – sets a specific destination). All changes are transferred to all logical processes, and all logical processes display the shared scene. To further test the dynamic behaviour and possible interactions between individual entities, mouse-based controls check for eventual collisions with other entities at the given destination. The solution was tested on a configuration of 8, 12 and 24 logical processes (each on an individual PC) at a company local area network. The critical computing power pointers are stated in table 1. Measurements took place on identically configured PCs (Intel® Core™ i3-3240 CPU @ 3.40 GHz, 4 GB RAM, Windows 7 64bit, only one application running – Google Chrome browser, version 51).

Table 1: Results of tested use case, lookahead 120 ms, critical data for simulation run with focusing on animation output and interactive approach

Test	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13
LP count	8	8	8	8	12	12	12	12	20	20	20	20	20
Sync method	CONS	CONS	TIME	TIME	CONS	CONS	TIME	TIME	CONS	CONS	TIME	TIME	TIME
User interaction	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	NO	YES
Animation FPS	23	23	29	27	24	21	27	25	6	8	8	13	21
Sync request count	1256	675	186	142	161	321	137	44	6123	4820	108	119	22
Sync waiting time [ms]	9	8	8	8	8	9	8	8	9	9	8	8	10
Animation 1 frame draw time (AVG) [ms]	31	42	34	36	38	40	40	45	182	195	119	54	49
Animation 1 frame draw time (MAX) [ms]	60	62	37	51	55	62	70	68	383	360	150	71	56
Animation activity count in animation scene	266	292	215	245	270	245	278	266	1050	1120	758	480	300
FPS per animation activity	0,08	0,08	0,13	0,11	0,08	0,08	0,08	0,07	0,005	0,007	0,1	0,03	0,07

Result notes (table 1):

- Two types of synchronization were tested (see chapter 4.3) - marked as CONS and TIME.
- User interaction (mouse click) was programmatically generated. Calculated by uniform distribution (min 400 ms, max 3000ms) between simulated interactions.
- One animation activity consists from 10 graphics base elements / primitives in average.
- Results were collected after 5 minutes (real time) run.

Summary of results:

- *Sync waiting time [ms]*: The time required to send a request for synchronization and receive all required data is almost constant.
- User interaction generated broadcast messages and this is reason of decreasing number of synchronization requests. Negative is lower FPS, because every interaction must be handled.
- Critical for fluent performance (with animation fluidity in mind) is count of animation activities. FPS is directly dependent on their number (row *FPS per animation activity*).

7. CONCLUSION

The presented solution is still being developed, but the class of applications, for which it will be beneficial, can already be defined. Specifically, interactive distributed simulations, which do not require calculation-heavy operations (e.g. real-life object representing entity movement dynamic calculations, expansive animation scenes). Due to the single-thread nature of JavaScript, simulations with large amounts of user input in short time intervals (generally tens in a second) are not suitable, because JavaScript will be busy with dispatching these events, and will not have time left for calculation of other critical parts (activity calculation, animation).

Considering the aforementioned facts, the primary motivation for use of web-based simulation is the availability of the runtime environment – web browser – on any computer or modern device connected to a computer network. JavaScript is very well supported by modern-day browsers, and is extensible and well known. This comfort of availability and simplicity is not without a cost – when compared to native applications, the scripts are slow.

The solution as a whole is still not finished, Lots of optimisations could be done on the administrative interface, and integration of the individual parts of the solution, which are at this time resolved by external applications (JSRC, RTCC, VC), primarily because of ongoing development. Also, if the JavaScript runtime seems too slow, the logical process software library can always be optimised at source code level.

REFERENCES

- Fujimoto, Richard M. Parallel and distribution simulation systems. New York: Wiley, 2000. Print.
- Kartak, Stepan, and Antonin Kavicka. "WebRTC Technology as a Solution for a Web-Based Distributed Simulation". Proceedings of the European Modeling and Simulation Symposium 2014. Genova: Università di Genova, 2014, s. 343-349.
- Kartak, Stepan. "Web Simulation as a Platform for Training Software Application". Proceedings of the European Modeling and Simulation Symposium 2015. Genova: Università di Genova, 2014, s. 70-78.
- Kuhl, Frederick, Judith Dahmann, and Richard Weatherly. Creating computer simulation system: an introduction to the high level architecture. Upper Saddle River, NJ: Prentice Hall PTR, 2000. Print.
- Tropper, Carl. Parallel and distributed discrete event simulation. New York: Nova Science, 2002. Print.
- Hridel, Jan, and Stepan Kartak. "Web-based simulation in teaching". The European Simulation and Modelling Conference 2013. EUROSIS-ETI, 2013. Print.
- The Institute Of Electrical And Electronics Engineers, Inc, 2012, 1278.1-2012: IEEE Standard for Distributed Interactive Simulation - Application Protocols. New York; IEEE. 2012.
- W3C, 2016. WebRTC 1.0: Real-time Communication Between Browsers. Available from: <https://www.w3.org/TR/webrtc/> [Apr 2016].

PROCESS SIMULATION AS SUPPORT IN LEARNING THEORETICAL-PRACTICAL WITHIN THE LEAN MANUFACTURING LABORATORY OF UTFV

(a) Carlos Quintero Aviles, (b)Wendy Guerrero Coronel, (c)Omar A. González
Manzanares

(a) Engineering Faculty, UNAM
(b) Technological University Fidel Velázquez
(c)Technological University Fidel Velázquez

(a)carlosqa1@comunidad.unam.mx (b)aurea.wen@yahoo.com.mx

ABSTRACT

This paper presentation is a study of the application of lean manufacturing tools in a production-training system in an educational practice, which involves the development of single-acting pneumatic cylinders. Then the process is simulated to analyze the behavior activities and thus help improve production system modeled through alternatives and the application of lean tools. The paper begins with an overview of the areas and activities of the system, the methodology is followed by the statement of the problem and then simulation models created in the APE software and subsequent results are presented. Finally, some conclusions and suggestions for improvement are presented about system activities that can serve real systems with similar characteristics.

Keywords: lean manufacturing, productive improvement, didactic system, modeling and simulation.

1. INTRODUCTION

The productive-didactic process in which the project takes place consists of the elaboration of pneumatic cylinders of simple effect, these are made up of, a piston rod, a packing, a piston, a cover of two different diameters, a cylinder, and packages, these are elaborated in four models different under specifications from a client.

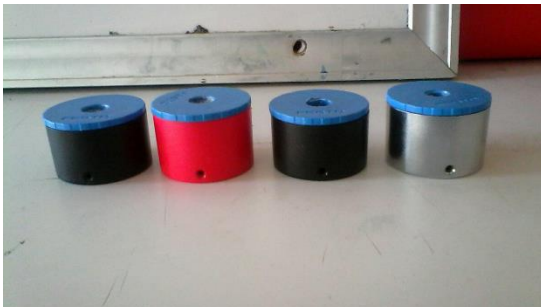


Figure 1: pneumatic cylinders models

1.1 DEVELOPMENT OF PRODUCTIVE- TRAINING SYSTEM

The practice was carried out in the University Technological Fidel Velázquez (UTFV, Mexico) in Lean Manufacturing laboratory. The objective of the practice consists of simulating the production of the pneumatic cylinders through a productive-didactic system. The Process of the manufacture of the cylinders begins after receiving order from a client, thus beginning all the internal organization of the production system to satisfy the demand of the client in specifications and certain times.

The areas involved in this production system are:

Client: should buy cylinders from the production company under certain specifications and timelines, if the received product does not meet the specifications the product should be return to the company for exchange.

Production control: get the customer purchase order and generates a work order (W.O.), send this order to warehouse.

Warehouse: responsible for supplying raw materials (R.M.) to production, assembly 1 and 2, requested R.M to suppliers and sends orders to press for punch out lids

Suppliers: provides R.M. to Warehouse for the cylinders production.

Production: receive the R.M. and only performs the thread to place the cylinder valve.

Assembly 1: assembles the stem, spring, punched out lid and cylinder

Functional Test: Test the functionality of the cylinder, if it works then it will be assembled if it doesn't work it will be sent back to the company for corrections.

Assembly 2: after being approved by functional test, the valve of the cylinder will be set.

Re-Work: to be inspected by returns or defects of manufacture, after correcting the errors gives back the product to them to the area from where it comes.

Press: it makes the perforation of covers in two diameters by warehouse orders

Packing and Quality: it makes an inspection of the finished product, to be approved places it within its packing, if it is sent to re-work is not approved.

Accounting: it makes the delivery and collection of product to the client by the product that this approves and orders of purchase of MP with warehouse.

- Display existing waste in production processes and their impact on cost, productivity, delivery times and reliability.

- Recognize the importance that plays in a company the staff participation in processes to identify opportunities for improvement.

- To generate in students the ability to identify opportunities for improvement in a company or a real production system, through educational tools

2.2 STAGES OF DEVELOPMENT AND PROCESS IMPROVEMENT

Those responsible for carrying out this study are student from The University Technological Fidel Velázquez (UTFV, Mexico) in productive systems engineering it is necessary to work with a group of 15 people on average for its development. The practical is to produce 60 cylinders in 45 minutes. Four phases are made to study and analysis the product it also help detected errors for improvement.

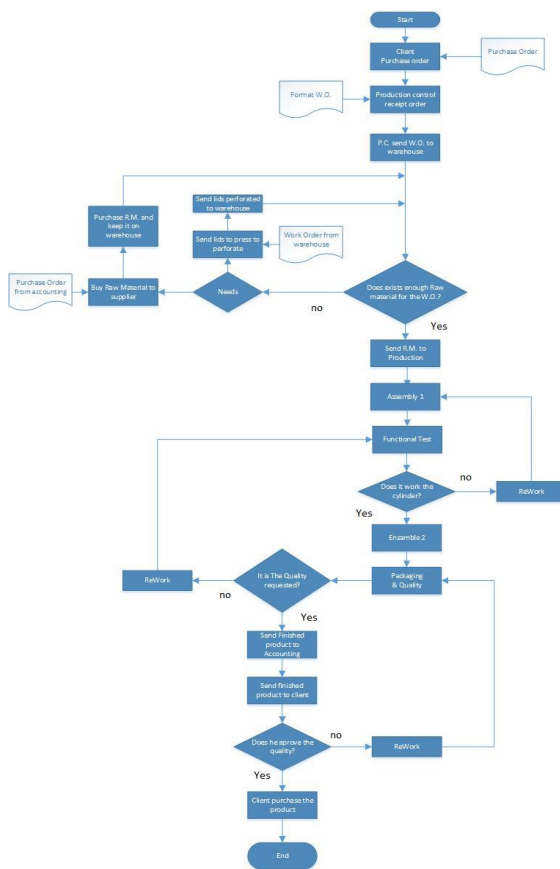


Figure. 2 Flow diagram of the production process in the development of a cylinder.

2. METHODOLOGY OF THE DEVELOPMENT OF PRACTICE

2.1 OBJECTIVES THEORETICAL PRACTICAL OF THE SYSTEM

- Learn the techniques of lean manufacturing and applying them.

Phase	Description of improvements to be implemented
1	The purpose of the organization is identified, roles are assigned and identify operational steps, processes, work areas, anomalies and opportunities for improvement
2	Identify the 7 wastes, identify the value chain, implement a redistribution of plant, implementation of 5S
3	Introduce Lean accounting TPM and SMED Value Stream Mapping (VSM) Just In Time (JIT) Reorganize by product families elimination of inventories
4	Performing a balancing lines Manufacturing cell formation. Implementation of Kanban 7 waste reduction production system PUSH to PULL

Table 1. Phases of development of practice and implement actions for continuous improvement.

2.3 EXPECTED RESULTS AFTER EACH PHASE

Two main results of the implementation of improvements in practice development is expected after each phase:

- A decline in the use of materials, constantly the levels of raw material inventory in warehouse and product in process are levels above the capacity of customer satisfaction, that is to say, you have more raw material to make products that really sell this, what generates that at the end of the practice and perform the analysis of economic, this point is one of the most part that affects the finances of the organization that produced the cylinders.
- Improvement of the indicators of productivity, economic and quality: each practice will pass through four stages of improvement, at the end of each of these phrase, it will be evaluated by the indicators mentioned above, and compared with each previous stage. The last stage of each practice is the one that shows the best performance for the application of different Lean tool methodologies.

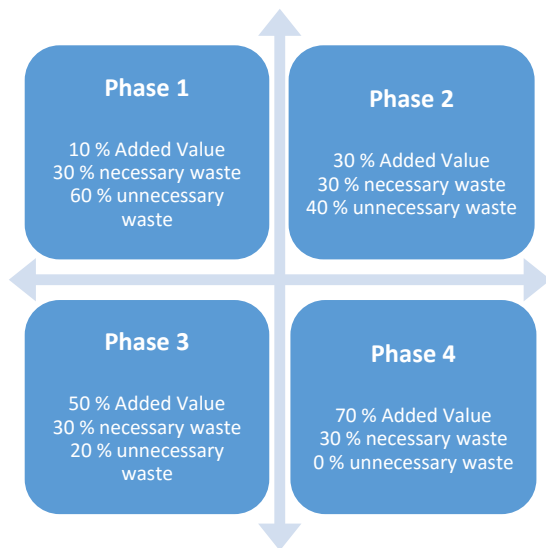


Figure 3. Expected results after applying lean tools in each of the phases. “Waste is anything that adds cost, but not value, to a product” (Wang, 2008)

3. IMPLEMENTATION OF SIMULATION IN THE STUDY AND DEVELOPMENT OF PRACTICE

3.1 OBJECTIVE

Introduce simulation in the process of teaching and learning within academic activities in the workshop lean manufacturing by designing the model with simulation software SIMIO to assist students in achieving the objectives of the practice. “Simulation is an effective method to recognize these problems with consuming least cost and time” Hafezalkotob (2014)

3.2 SPECIFIC OBJECTIVES

Simulation is a technique to model a real-life or hypothetical situation on a computer so that it can be used for analyzing the behavior of system (Senem, B., 2011.)

- Modeling and Simulation of the production process of pneumatic cylinders based on each phase of the system.
- Analysis of behavior of the production system for enhancement the learn process, decision-making and improvement proposals skills on students
- Compare and evaluate the results in groups which will use the simulation whit respect to those who have not made use of it.

3.3 JUSTIFICATION

“The first step of applying simulation in the real industrial problem is recognition of the simulation parameters” Hafezalkotob (2014).

This study on the implementation of simulation in conjunction with lean manufacturing in the learning process of students in Engineering Production Systems, will reveal the impact of both tools as complementary elements in their education, else helping at the same time to strengthen skills of study and understanding of processes, analysis of causes and effects, identifying variables and above all generate in them a technique oriented systems approach.

In terms of development of the practice as a production process, we also seeks the performance improvement of the metrics of historical data of practices realized previously in which both instruments simultaneously not intervene. Proving as this way, that simulation is a powerful tool with different guidelines for finding solutions and improvement.

4. MODELS AND SIMULATION OF DEVELOPMENT PHASES

Then we introduce four scene corresponding to each of the four phases involved in the development of practical and production metric analysis is performed.

4.1 PHASE 1 MODEL

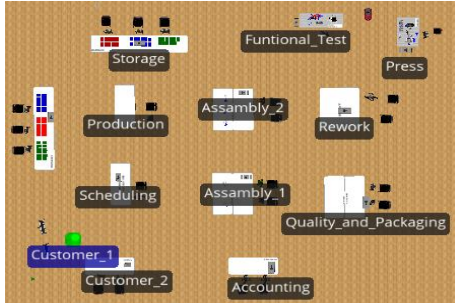


Figure 4. Layout phase 1



Figure 4.1 Model phase 1

Results	Phase 1	
	Results obtained	Desired results
Produced cylinders	13	60
Cylinders delivered time	0	60
Average delivery time	900 s	45 s
Earnings	64	960
Rework coefficient	12%	0%

Table 2. Results of Phase 1

During the first phase the students form a group and they choose a leader to start a teamwork and produce 60 cylinders in 45 minutes to deliver them to the consumer on a specific time and in due time which is the main objective. At this stage roles are assigned to participants to manage the workspace in which he or she will be responsible and should work in the respective processes to meet the target. After the implementation, the objectives will be measured and placed in a format called "summary

of results" where the number of produced and delivered on time cylinders will be shown, also will be shown the profits, if the objectives have been produced in time and the coefficient labor. Thus the quality and productivity is measured.

4.2 PHASE 2 MODEL



Figure 5. Layout phase 2



Figure 5.1 Model phase 2

Results	Phase 2	
	Results obtained	Desired results
Produced cylinders	27	60
Cylinders delivered time	1	60
Average delivery time	400 s	45 s
Earnings	380	960
Rework coefficient	10%	0%

Table 3. Results of Phase 2

In the second phase an analysis will be prepared to identify the seven wastes and the value chain, the students propose strategies to redistribute plant to help waste disposal, they begin by ordering the work area and standardize the process through the application of 5 S's methodology.

4.3 PHASE 3 MODEL

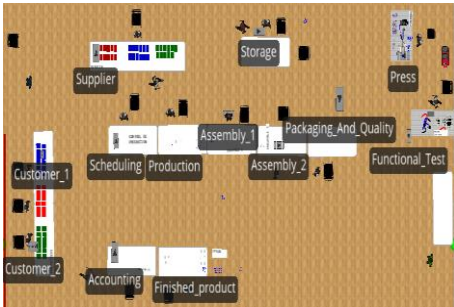


Figure 6. Layout phase 3.



Figure 6.1. Model phase 3.

Results	Phase 3	
	Results obtained	Desired results
Produced cylinders	42	60
Cylinders delivered in time	11	60
Average delivery time	82 s	45 s
Earnings	629	960
Rework coefficient	5%	0%

Table 4. Results of the phase 3

In this phase, participants begin to introduce the Lean accounting, reducing inventory levels; They begin to apply kanban cards to identify the merchandise and the processes; apply the TPM, then reduce tooling change through the SMED and begin to meet the objectives of JIT, this phase may take between two and three production runs.

4.4 PHASE 4 MODEL

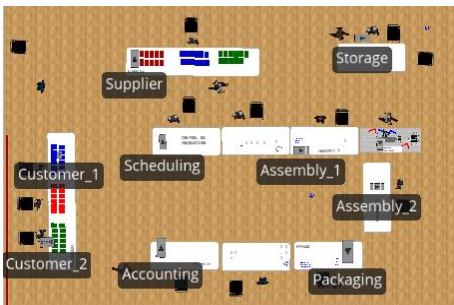


Figure 7 Layout phase 4.

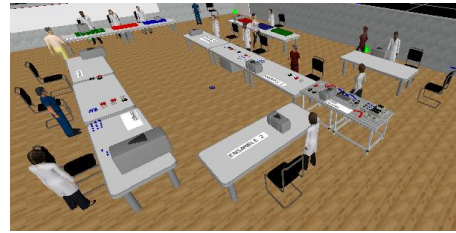


Figure 7.1 Model phase 4.

Results	Phase 4	
	Results obtained	Desired results
Produced cylinders	67	60
Cylinders delivered in time	51	60
Average delivery time	47 s	45 s
Earnings	960	960
Rework coefficient	0%	0%

Table 5. Results of the phase 4

Kaizen is a cyclical methodology is continuous improvement, once the methodology of "small big improvements results" is adopted, at this stage make a value stream mapping to observe the status quo, and identify areas of opportunity to observe accurately and numerical results , time in the value chain and continues with the JIT and design manufacturing cells to meet the objective: to produce 60 cylinders 45 minutes, delivering on time, obtaining the necessary profits and 0% in the rate of re work, 100% quality. Being a 100% dependable enterprise.

4.5 CONTINUOUS IMPROVEMENT IN THE FOUR STAGES

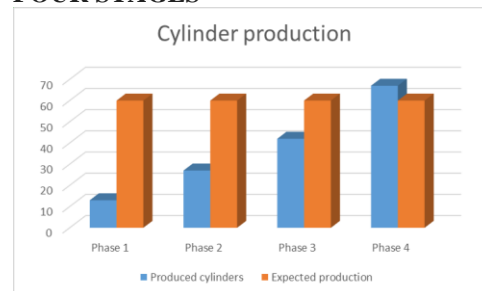


Figure 9. Cylinder production

Cylinder production in phase 4, the goal is met.

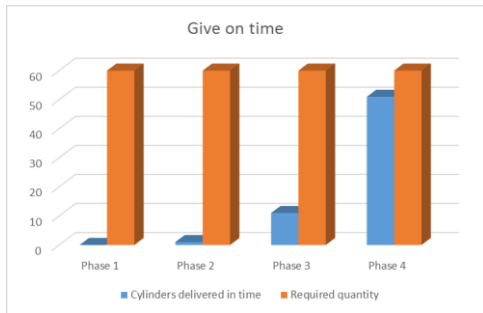


Figure 10. Cylinder production and delivered in time required.

Give on time of cylinder in phase 4, it has not met the target, but the result was close



Figure 11. Increased earnings.

The earnings were sufficient and business investment recovered and got the 100% gain for each pneumatic cylinder sold.



Figure 12. Quality of the finished products

Quality is measured by the coefficient of Rework obtained by counting defective cylinders. The graph shows that in the last phase no defective cylinders, thereby complying with the principle of "zero defects" lean work accomplished team.

5. CONCLUSIONS

Finally, we present some conclusions and proposals for improvement about the activities of the system which is just one of the objectives of the simulation processes. The improvements that we

propose can serve for real systems with similar characteristics.

- Simulation is a very valuable tool that allows us to see the changes that we can make without working with the real system, avoiding stoppages in the processes and delays in the production, making more optimal the optimization process proposed by the continuous improvement or the philosophy of lean manufacturing.

- Include a simulation software in the workshop of Lean Manufacturing of the UTFV helped the students learn a little about the benefits and applications of this tool within a training practice and also within a production system, these benefits and activities, students can propose them once they begin their professional career, and initiating continuous improvement in the companies where they work.

5.1 RECOMMENDATIONS

- The students that participated in this project have not yet work with a simulator, so it is recommended to have an introduction to the simulator in the same area where they practice.

- Standardize and document through a manual of procedures all the characteristics of the process such as: the times, the movements of each one of the operations in the stages of the system, so that no matter who performs the operation, the times are similar and thus to achieve a line-balancing, which is a step subsequent to the application of lean manufacturing with success.

- Analyze those activities that could be integrated in a same stage of the process, and simplify operations. (p. e. The packing of the parts in practice may be made by the area assembly II).

REFERENCES

Hafezalkotob, A., Ketabian, H., Rahimi, H., 2014. *Balancing the Production Line by the Simulation and Statistics Techniques: A Case Study*. Research Journal of Applied Sciences, Engineering and Technology, 7(4): 754-763.

Bozickvic, R., Radosevic, M., Cosic, I., 2012. *Integration of Simulation and Lean Tools in*

Effective Production Systems – Case Study. Journal of Mechanical Engineering, 58(2012)11, 642-652.

Wang, S., Koch, D., Ren, Y., 2008. *A Hands-on Kanban Simulation Kit for Lean Manufacturing*. Proceedings of the 2008 IAJC-IJME International Conference, Paper 199, ENG 105.

Senem, B., 2011. Assembly Line Balancing in Garment Production By Simulation. <http://www.intechopen.com/books/assembly-line-theory-and-practice/model-sequencing-and-assembly-line-balancing-in-garment-production-by-simulation-worker-transfer-syst>

Kelton, W. David, Smith, Jeffrey S., Sturrock, David T. *Simio and Simulation: Modeling, Analysis, Applications* 2nd. Ed. 2012

F. De la Mota, E. Cortes. *Simulation notes* Universidad Nacional Autónoma de México, Facultad de Ingeniería, México, 2006.

Kenett, S. Ron, Shelenyahu Z., 2000. *Estadística Industrial Moderna*. 2da edición. Thomson editores.

Dub, M., Pipan, G., Hanzálek Z. 2002. Sotock Optimization of a Kanban-based Assembly Line. *International Conference on Flexible Automation and Intelligent Manufacturing 2002, Dresden, Germany..*

M.C. Ortiz F et al. Comparison of Standard Costing System and the Theory of Constraints Control Flow of Materials Using a simulation model, Volume, No. 1, México 2008.

Báez O. et al. *Methodology decision support for production plan in flexible manufacturing systems*, Volume 2, No. 1, 2008.

Dr. Eduardo Marroquín Prado, *A scheme to monitor and control the statistical and technical capacity of a process* Volumen 5, México 2011.

research is simulation and optimization of productive processes.

Wendy Guerrero Coronel, She studied A Top Technical Degree in Graphic Arts Area Industrial Processes and Production Systems Engineering at the University Technological Fidel Velazquez, she was accepted to study Master degree in Manufacturing Engineering, she is currently teaching in the University Technological Fidel Velazquez. She is currently researching on Lean Manufacturing.

Omar Alejandro González Manzanares: He studied a Top Technical Degree on Industrial Process at the University Technological Fidel Velazquez. He is currently studying Production Systems Engineering, and has participated in local conferences. His line of research is simulation and productive processes.

AUTHOR'S BIOGRAPHY

Carlos Quintero Aviles, studied Industrial Engineering in the ITTLA, Technological Institute of Tlalnepantla. He teaches at UTFV University and has participated in national and international conferences. He Studied a Master's Degree in Operational Research in the UNAM. His line of

RADIO FREQUENCY SAFETY DEVICE APPLIED TO CHAIR CHILDREN IN CARS

Robson da Cunha Santos^(a), Karina Barros de Oliveira^(b), Marcos Antônio Cruz Moreira^(c), Severino Joaquim Correia Neto^(d), Gerson Gomes Cunha^(e)

^(a,b,c,d)Fluminense Federal Institute, Campus Macaé, Engineering and Automation Control
164 km Amaral Peixoto Road, Brazil

^(a)Estácio de Sá University, General Alfredo Bruno Gomes Martins Highway, s/n -Braga - Cabo Frio / RJ, Brazil

^(e)Federal University of Rio de Janeiro, Alberto Luiz Coimbra Institute Graduate Engineering and Research, Civil Engineering Program/COPPE/UFRJ

^(a)rcunha@iff.edu.br ^(b)adudakaka@hotmail.com ^(c)macruz@iff.edu.br ^(d)scorreia@iff.edu.br ^(e)gerson@lamce.ufrj.br
[mail](mailto:gerson@lamce.ufrj.br)

ABSTRACT

The article aims to present a prototype device able to identify and warn those responsible when there is a possibility of forgetting of a child inside the car, thus avoiding an injury to the health of the child or even death. To this effect electronic sensors able to detecting the presence of the child on the chair have been used, microcontrollers (Arduino Uno, Arduino Nano) programmed to perform the data processing, wireless communication module (XBee) to perform the transfer of data between the chair, and a specific key chain, and ultimately actuators responsible for a possible warning against forgetting a child inside the car. Generally speaking, this article comprises concepts involving instrumentation and programming. This paper has the effect of academic results, nowadays there are not many studies concerning the proposed theme and its application is highly viable because of the relatively low cost, making it accessible to all social classes.

Keywords: Instrumentation, programming, Microcontrollers, XBee, child care

1. INTRODUCTION

Rush of day-to-day marks society and the constant change of scenery happens especially in large cities. These changes occur constantly and it is necessary to one adapt itself in real time, since the change of routine can give rise irreparable accidents such as the forgetting of a child in a closed car, damaging the child's health or taking this child death.

In recent years, news emerged of several cases of forgotten children inside vehicles, facts that shocked society. For this reason, the idea of creating a device that detects this possible occurrence arises. To design the device, the use of two microcontrollers was necessary, as well as two wireless communication modules, sensors and child safety car seats.

Is it worth it? In Brazil, there are no official statistics on such occurrences. However, in the United States, KidsandCars NGO which deals in all types of accidents involving children and cars, has counted 550 cases have been counted since 1998 (Buzzinaro 2016).

One must consider that the issue concerns children who are not able to care for themselves, and would have had a whole life ahead. There is only one life, there are many lives that have to deal with this loss, a father, a mother, a brother who will always miss the child who was in the car. So, this mistake is unacceptable.

2. OVERVIEW

Thinking about the safety of children, a control prototype with a baby chair was developed by adapting Arduino boards, XBee modules, two pressure sensors, one for the seat and one for the locking of the chair seat belt actuators as LEDs, buzzers and vibration motor.

The following is the operation of the two parts of the system, the first chair and the second key chain.

2.1. Chair Operation

The chair system is composed of: 1 Wireless communication module (XBee Series 2 Pro); 1 USB adapter for XBee Explorer; 1 Arduino Uno Rev 3; 1 Buzzer; Leds; 2 Pressure sensors (buttons Normally Open); 1 Key ON / OFF; 9 V battery.

The process starts with the positioning of the chair in the car without the need for car modification for installation of the device. It will just a need to connect the battery charger in the car's cigarette lighter.

After correctly position and connect the chair is then triggered the ON / OFF switch which in turn activates the chair system. Therefore, the control variables responsible for managing communication between the chair and the keychain are initialized.

The system will check if there is "Child in the car," if there are not triggered the alarm and return to the state sends signal. If there is a new check will be performed to determine if the child has been forgotten or not.

However, if the child has been not forgotten, the alarms will not be triggered and the system returns to step "send signal", otherwise the alarm will be triggered.

2.1.1. Operation Diagram - Chair

Following it is shown the operating diagram of the chair and its respective steps.

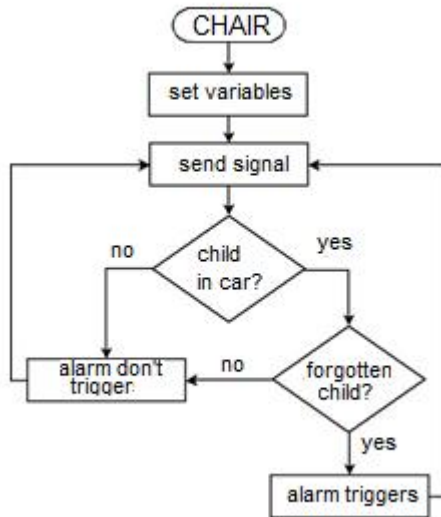


Figure 1: Chair - Operating Diagram

- Set variables: Step responsible for setting the initial conditions of the system;
- Send Signal: Step responsible for reading the signals obtained from the chair sensors and sending signals to the keychain module;
- Child in car: Step comparative data obtained from pressure sensors;
- No Triggers Alarm: Failing a child forgetting, not the actuators will be triggered;
- Forgotten Child: Step comparative where sensors identify if the child has or has not forgotten;
- Alarm triggers: Identifying a possible forgetting, all actuators will be triggered;

2.2. Keychain Operation

The keychain system is composed of: 1 Wireless communication module (XBee Series 2 Pro); 1 USB Explorer Adapts pain for XBee; 1 Arduino Nano, 1 Buzzer; Motor Vibratory; Leds; 1 Key ON / OFF; 1 9v battery; 4 buttons.

The process starts with the key drive ON / OFF which in turn will activate the keychain system. Then beginning the control variables responsible for managing communication between the chair and the keychain.

The system checks whether the keychain is enabled or not. If ceste is not, the system will be redirected to step sends signal and the keychain will only be reactivated if the password is entered correctly, or if it is more than 1 minute near the chair. If enabled it will check whether there is "Child in the car," if there are not triggered the alarm and return to the state sends signal. If there is a new check will be performed to determine if the child has been forgotten or not.

If the child is not forgotten, the alarms will not be triggered and the system returns to step sends signal. If it is forgotten, the alarm will be triggered.

2.2.1. Operating Diagram - keychain

The keychain operating diagram and their steps are shown in the following figure.

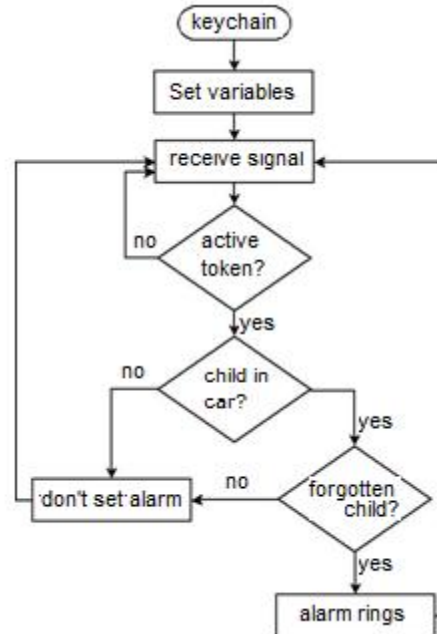


Figure 2: keychain Operation Diagram

- Starts variables: Step responsible for setting the initial conditions of the system;
- Receives signal: Step receiving the data sent by the chair and it is also responsible for the activation Keychain with the buttons;
- keychain Enabled: Checks the current state of the Keychain control variable.
- Child in the car: Step comparative data obtained from pressure sensors, via wireless communication;
- No Triggers Alarm: Actuators are not triggered if the child is not inside the car.
- Forgotten Child: Step comparative where sensors identify if the child was or was not forgotten;
- Alarm triggers: All actuators will be triggered, because there is possibility of a child have been forgotten inside the car.

2.3. Module Wireless XBee Communication

The name XBee module refers to a family of radio frequency modules developed by Digi International. These are responsible for managing a wireless network using the IEEE 802.15.4 protocol (The Institute Electrical and Electronics Engineers).

The XBee is widely used in control systems, as aside from present reliability data transmission. It also presents low cost and low power consumption, which is important for applications in difficult remote access systems where the use is necessary of a battery.



Figure 3: XBee Series 2 module Pro

2.4. XBee USB Adapter

The adapter is used to regulate the input and output voltages in order to protect the XBee modules, once these have a supply voltage of 3.0 - 3,4V and Arduino 5V voltage.

This also offers a means of serial communication with the PC via a mini USB cable, and you can then set the XBee using X-CTU, practical and safe way.

The component has a dedicated pin called RSSI (Received Signal Strength Indicator) having an output modulated signal PWM (Pulse-Width Modulation), this in turn is read by Arduino and converted to decibels being used as a parameter to estimate distances.



Figure 4: XBee USB Adapter

2.5. X-CTU: XBee Programming Software

The X-CTU is a software developed by Digi International used to configure and test all its radio frequency products. This has compatibility with Windows operating system and has a very friendly graphical interface for handling.

This software can run in parallel, or may be open over an X-CTU a time. This feature allows the user to test and configure the communication between the radio frequency modules fast and reliably.

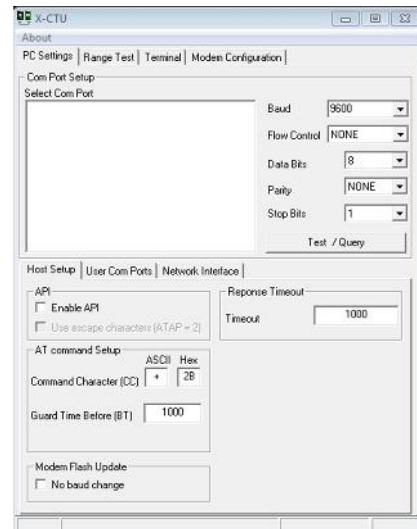


Figure 5: Programming Software and Test XBee (X-CTU)

2.6. Arduino Platform

The Arduino platform was developed with the intention of reduce costs for school projects because the prototipagens systems by this time had a high value.

Currently this technology has become quite popular and has many applications. It can be applied in simple control systems, such as opening and closing a gate, or even be implemented in an industry.

The Arduino has a microcontroller of ATMEL family, this is responsible for managing the data as a previously defined and determined programming. To insert a program in the controller chip, a compatible programming software is required, and this can be purchased on their official website (AVR 2012, Arduino 2016)

2.6.1. Uno Arduino Specifications

In this prototype, the Arduino Uno was used in the chair because this meets the requirements for the prototype. The following Figure 6 representing the Arduino board and a table of specifications.



Figure 6: Uno Arduino

Table 1: Uno Arduino Specifications

Specifications	Uno Arduino
Microcontroller	ATmega328
Voltage Operation	5V
Input Voltage	7-12V
Input Voltage Limit	6-20V
Inputs / Outputs Digital	14 (6 PWM)
Analog Inputs	6
DC For I / O	40 mA
Flash Memory	32 KB
Sram	2 KB
Eeprom	1KB
Clock Speed	16 MHz

2.6.2. Nano Arduino Specifications

To construct the keychain was necessary to save the maximum space possible, for this reason that was used in this prototype Arduino Nano. In figure and the following table is observed settings.

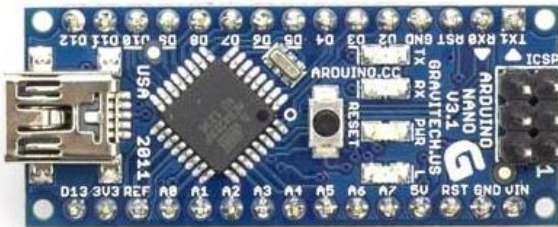


Figure 6: Nano Arduino

Table 2: Nano Arduino Specifications

Specifications	Nano Arduino
Microcontroller	ATmega328
Voltage Operation	5V
Input Voltage	7-12V
Input Voltage Limit	6-20V
Inputs / Outputs Digital	14 (6 PWM)
Analog inputs	8
DC for I / O	40 mA
Flash memory	32 KB
SRAM	2 KB
EEPROM	1KB
Clock Speed	16 MHz

2.6.3. Arduino IDE

The code needs to be entered in the Arduino memory, so it was necessary to use an IDE (Integrated Development Environment). The IDE provides the necessary tools known as toolboxes for the development of the code in WIRE programming language also able to compile and submit the code in hexadecimal to the microcontroller (Banzi 2011).

Its basic structure is composed of a function known as voidsetup () and the other as voidloop ().

The first is characterized by the initialization of variables or functions that will be used in the system.

The second function represents a loop where all the code that is inside the loop will be repeated constantly.

The IDE seen in the figure below shows a simple and friendly interface functionality. Some code examples are supplied to further facilitate the user's life and programmer, this way they will assist in the development of the prototype (McRobert, 2011).

2.7. Sensors and Actuators

To identify the presence of a child in the car, it is necessary to monitor the physical variables, in this case two components were considered: the locking of the seat belt and the pressure on the baby seat. This monitoring is done via pressure sensors, which are represented by buttons. By finding an abnormal situation the microcontroller sends a signal to the actuators in order to alert the responsible of possible forgetfulness of child or baby. There are three types of actuators used in the prototype: the sound (buzzer), light (LEDs) and mechanical (vibration motor).

3. PROPOSED SYSTEM DEVELOPMENT

To develop the prototype was necessary to use some electronic components, which are easily accessible and low cost. The following issues will be demonstrated the development stages of the safety device.

3.1. Construction of the Chair and Keychain

This topic is intended to present the development of the electrical circuits through Fritzing software, sensors and actuators of the chair, as well as their descriptions.

3.1.1. Electric Buzzer System

The Buzzer is an actuator in order to warn, sonically, because there has been a possible anomaly. It has been implemented in both the keychain as the chair and your electrical circuit can be seen in the following figure.

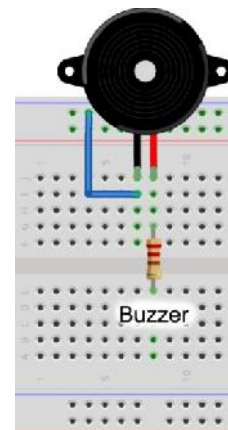


Figure 7: Electric Buzzer System

This subsystem has a track called Buzzer. It should be connected to an Arduino digital port for proper operation. moreover, it was necessary to add a resistor 220ohms to limit the system current.

3.1.2. Electrical System of Switches

The switches are touch sensitive sensors and normally open type. They were separated into two subsystems, one for the chair and the other to the keychain. In figure 8 can be seen the two pressure sensors simulating the belt sensor and the seat.

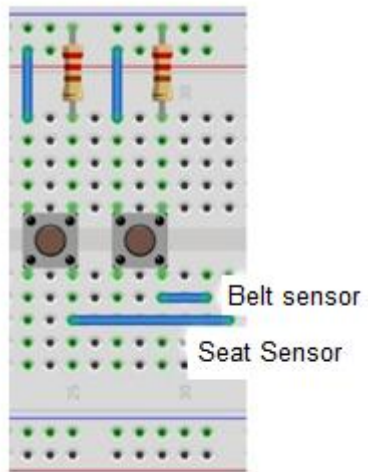


Figure 8: Belt and Seat Sensors

The second is represented by four pressure sensors. They are responsible for entering the password for the activation and deactivation of the keychain, if required. The figure below demonstrates the electric scheme.

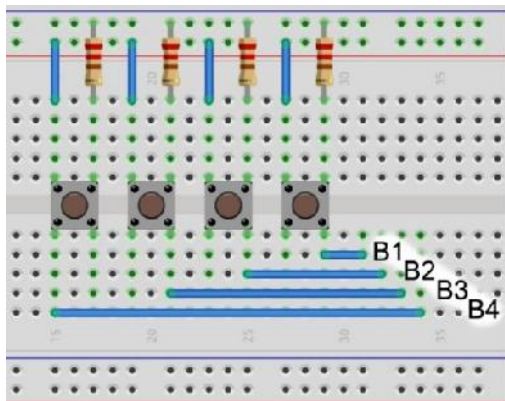


Figure 9: keychain Sensors

These systems have 10k ohm resistors that are required to limit the current Arduino and also to ensure the logic level 0 when the actuator is not actuated.

3.1.3. Electrical system of the LEDs

LEDs have intended to visually alert the current system state. It is divided into two modules, one for the seat and one for the keychain.

The first features 8 LEDs of which 6 are to draw the attention of people passing around the car and 2 to indicate the System Figure 10 connection.

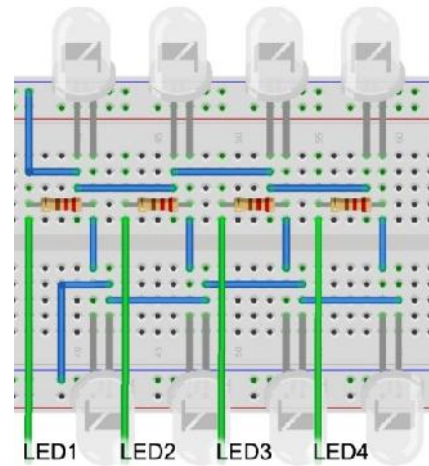


Figure 10: Set of LEDs Chair

The second contains 4 LEDs, where 2 to indicate the system connection (green LED and red LED), one for the active function on / off keychain (blue LED) and the last to indicate the alert status (Yellow LED).

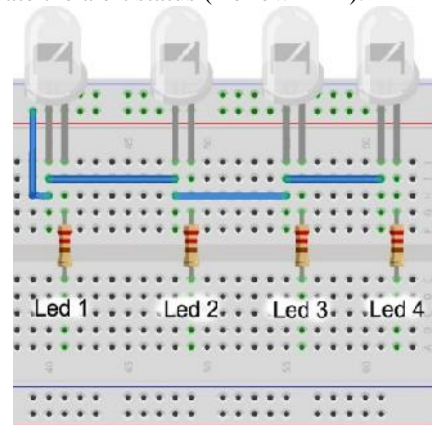


Figure 11: Set Keychain LEDs

For the system to control the current were used 220 ohm resistors in keychain LEDs and 160 ohms in LEDs chair.

3.1.4. Electrical System - Vibratory Motor

This subsystem has been implemented in keychain as mechanical warning means in case of a possible oblivion. In the following figure is shown the electrical system of the vibratory motor.

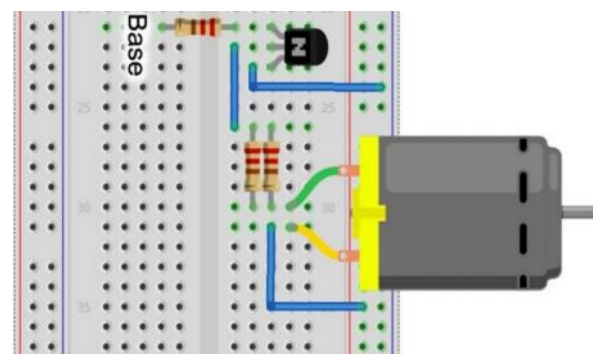


Figure 12: Set Keychain LEDs

The vibrator module requires a greater current, in this case adjusted to 60 mA, but to avoid damaging the digital output Arduino, limited to 40 mA, it was necessary to use an NPN transistor (BC-368) to perform the switching of the same.

The BC-368 transistor has three pins, which are the base, collector and emitter. The first, to be excited, is responsible for controlling the current between the collector and emitter. The second is introduced into the device to be switched with the 9V battery voltage. Finally, the third is connected to GND where the current collector and the base will be added. The resistors have base and collector, respectively, 1kohm and 220 ohms (two in parallel generating a total resistance of 110 ohms).

4. RESULTS

The tests were performed with the device inside a car simulating a baby seat in the back of it. Through physical testing were obtained some data concerning the distance in meters.

Table 3: Tests Distance in Meters

Local	Description	65dBm	10dBm	Loss of Signal
CAR (Open Place)	When the place where the car was parked, do not have significant obstacles exist for reducing the frequency between XBee modules.	24 meters	12 meters	128 meters
CAR (Closed Place)	When the distance is near a closed space, causing the keychain and the chair to communicate with interference from the walls.	16,4 meters	8,3 meters	87,6 meters

When the frequency between the XBee's arrived at approximately 65dBm, when the alarm was triggered and subsequently approximately 10 dBm, when the alarm was tripped. This occurs by the approach of key chains to the car in which the seat is installed. The table above illustrates very well the data acquired.

5. FINAL REMARKS

The proposed system has proved to be an excellent technology solution for frequent events in society. The device was built with low-cost equipment, easy purchase and installation. Initially your wireless communication module XBee showed the more complexity in the development of the project, because it requires an adapter for regular supply voltages and control, and the reason for its configuration parameters require more compared to previous models.

Due to the utilization of signal power to estimate the distance, the system did not show a good precision, but is sufficient to develop the prototype. This read signal varies with the distance and the amount of obstacles between the transmitter and the receiver.

The keychain interface is only representative because it contains components that can be used in a more sophisticated and smaller model. Regarding the use of the seat, installed sensors served to simulate the resistive sensor and the seat belt in it, because the application in a real car seat would take much time and equipment purchase and provision would be necessary. After some contacts with companies supplying of the device, they have not demonstrated interest in developing the product. However, the result of this project was satisfactory and proved to be very promising for the development of vehicle safety in the industry, despite not having any support from manufacturers.

As a proposal for future work recommended further study of signal in dBm over the distance in meters, taking into account the influences that the physical environment has on the system making it thus more accurate.

REFERENCES

- ARDUINO. Home. [USA], 2014. Available from: <www.arduino.cc>. [accessed 15 May 2016]
- AVR e Arduino: Técnicas de Projeto. 2. ed. Florianópolis: Edição dos Autores, 2012.
- BANZI, Massimo. Primeiros Passos com Arduino. São Paulo: Novatec, 2011.
- BUZZINARO, Ju. Crianças esquecidas no carro: precisamos falar sobre isso. Available from: <http://minhamaequedisse.com/2013/01/criancas-esquecidas-no-carro-precisamos-falar-sobre-isso/>. [accessed 05 July 2016].
- DIGI. Home: Products: Embedded Wireless and Wired Solutions. Minnetonka, 2014. Available from: <http://www.digi.com/xbee/>. [accessed 25 November 2015].
- EVANS, W.A., 1994. Approaches to intelligent information retrieval. Information Processing and Management, 7 (2), 147–168.
- MCROBERT, Michael. Arduino Básico. São Paulo: Novatec, 2011

THE METHOD OF NESTED SIMULATIONS SUPPORTING DECISION-MAKING PROCESS WITHIN A MESOSCOPIC RAILWAY SIMULATOR

Roman Diviš^(a), Antonín Kavička^(b)

^{(a),(b)} University of Pardubice, Faculty of Electrical Engineering and Informatics

^(a)roman.divis@upce.cz, ^(b)antonin.kavicka@upce.cz

ABSTRACT

While executing stochastic simulations, some unexpected (conflict) situations may occur thanks to which it is not possible to continue with the simulation in appliance to prescribed rules and it is necessary to decide on an alternative solution. There is a variety of tools for supporting the decision-making activities – methods of operational research, heuristic methods, etc. Another method is based on nested simulations with which the solution of a problem is searched using recursive simulations making a limited outlook into the future. At the moment when the problem occurs, alternative scenarios are created and simulated. After a certain period of time the nested simulations are stopped, their results are evaluated, the most suitable solution is selected and then the simulation continues in compliance with that solution. This contribution describes the issue of decision-making support in simulators, and it focuses mostly on the description of the method of nested simulations.

Keywords: nested simulations, decision-making support, railway transport

1. INTRODUCTION

Our research is focused on the class of transport, service, and logistic problems – main attention is paid to the area of railway transport, which has been our primary interest for a long time. Within railway transport we focus on investigating railway yards and junctions with the help of an experimental method of simulation. Exploiting simulations enable (i) to examine the current status of the infrastructure and its usage (using the same traffic flows), and (ii) to find out what results the alternative scenarios bring. Such scenarios can be represented not only by different track layouts, but also by a different structure of traffic flows (following relevant timetables). Thanks to simulations we can examine in great detail the behaviour of different traffic flows without having to apply such changes in reality. Thus, computer simulation represents an appropriate approach of examining railway systems.

2. SIMULATING RAILWAY TRAFFIC

It is possible to use various approaches and methodologies to examine, analyse, and optimize railway traffic. One issue of our interest is the examination of railway station capacity (throughput), i.e. to quantify the usage of tracks or switch zones with regard to a given traffic scenario (timetable) and to find potential bottle necks.

The capacity of a railway infrastructure can be determined in three ways: (i) by analytical methods, (ii) by graphic-analytical methods, (iii) by experimental methods, or (iv) by combination of any of the aforementioned methods. Analytical, or graphic-analytical, methods are specified for example within directives D24 (SŽDC, 2009) or UIC406 (ETF, 2013). Those methods mathematically analyse the usage of individual parts of rail infrastructure (using a special indicator called *degree of occupancy*). Alas, such methods do not enable to flexibly react to stochastic phenomena and therefore it is more appropriate to apply the experimental method of computer simulation. Simulation can use a deterministic approach (without the application of random effects), or a stochastic approach enabling the occurrence of random influences (e.g. train delays). Randomness in railway traffic can influence the overall situation greatly – rail infrastructure is limited by the number of tracks and signal and interlocking systems, so it is not possible to immediately send delayed trains to currently occupied tracks. Thus, delays can negatively influence a quality of rail traffic.

3. DETERMINISTIC VERSUS STOCHASTIC SIMULATIONS

Deterministic simulations are such simulations that do not include random inputs. A simulation trial can be repeated and its results can be exactly calculated. Deterministic traffic simulations reflecting systems of railway stations expect that all trains are on time, there are no unexpected closures, delays, or other non-standard events.

Stochastic simulations bring an element of indefiniteness to the progress of simulation trials. Thus, at least one of the inputs is supposed to imitate random influences. For that purpose pseudorandom number generators are utilized. Considering stochastic inputs,

the following phenomena in rail traffic simulations can occur: (i) random delays related to train arrivals, (ii) technical failures connected with trains or technical devices, (iii) arrivals of non-scheduled trains, etc. Let us mention an example of typical operational conflict situation involved within stochastic simulations: it is scheduled an arrival of train T_1 at the platform k , a considerably delayed train T_2 is expected to approach the station at the moment t and it is also supposed to stop at the platform k , which is currently occupied. Such collision situations can be solved in two ways: (i) within an *interactive mode* (a solution is formulated by the user), or (ii) using an appropriate technique of *automated decision support*.

4. DECISION-MAKING SUPPORTS WITHIN SIMULATION MODELS

Decision-making support in a simulator (applied during a simulation trial) can utilize a collection of methods or routines called at the occurrence of a conflict situation. The results of those methods should offer a relevant solution concerning how to proceed with the simulation experiment. There is a variety of techniques that can be used for decision-making support - such methods include:

- interactive mode of simulation,
- the method of priority planning,
- expert systems,
- methods of operational research,
- heuristic methods,
- methods of soft computing,
- methods of nested simulations etc.

4.1. Interactive mode of simulation

Interactive mode of simulation does not represent a method of automatic decision support. When a conflict occurs, the simulation is interrupted and the user is asked to define a solution. That approach enables to train users in various areas. Based on the user's inputs, the state space changes and the results are usually displayed online.

On the other hand, long-term simulations with a large number of replications and the high number of conflict situations are usually not good candidates for applying an interactive problem solving.

4.2. Priority planning

Priority planning represents a simple way of producing automatic decisions with regard to conflict situations. Before executing a simulation trial, a priority queue of possible alternative solutions is created for conflict situations. When a conflict situation occurs, the solution with the highest priority is chosen. If such a solution is not applicable, other solution is chosen based on its rank of priority until the entire priority queue is traversed. That method can be adapted for various situations (e.g. a list of alternative platform tracks for delayed passenger trains etc.).

4.3. Expert systems

Expert systems involve a group of computer programmes with the aim to provide expert advice based on using specialized routines directly developed for a certain kind of decision-making support. There are two types of expert systems, *diagnostic* ones and *planning* ones. In contrast with conventional programmes, the knowledge of expert systems is stored separately from their data, and their inference mechanism, that manages the expert system, can be based on various principles - assessing logic rules, fuzzy logic, artificial neural networks, etc. Expert systems also contain an explanatory sub-system capable of substantiating why the selected solution is the right one.

The aim of expert systems is then to substitute an expert in solving various problems using a computer programme. After the data specifying the current situation are put in, the inference mechanism is run and it calculates the solution to the problem in cooperation with its knowledge base.

Nowadays many authors deal with using expert systems being interconnected with simulations. The connection of simulation models and expert systems can be found in Masmoudi, Chtourou, and Maalej (2007). The issues of creating an expert system using simulations are discussed in Li, et al (Li, Li, Li, and Hu, 2000).

4.4. Methods of operational research and soft computing

Operational research represents a vast field dealing with various tasks and optimization issues. Tools for solving mono-criterial (so called linear programming) or multi-criterial issues belong to the field. For multi-criterial evaluations of variants, on the input there is a set of criteria influencing the quality of the solution based on various characteristics. Setting up particular values of the criteria then significantly affects the given results. There is a variety of methods of fixing those values, for example *pairwise comparison method*, the *Saaty's method* etc.

Other methods focused on finding solutions to optimization problems and realizing decision support are based on heuristic approaches. *Heuristic methods* reach relevant solutions faster than exact methods. However, their solutions are not guaranteed to be optimal (suboptimal solutions are acceptable - especially in cases if the solved task is connected with non-polynomial complexity).

The term *soft computing* includes a set of computational methods - *fuzzy logic*, *artificial neural networks*, *genetic algorithms*, and *probability calculations*. All these methods provide the potential to solve immensely complex tasks by means of fairly easy mathematical apparatus. Individual methods might not provide the optimal solution and their quality depends on their particular implementations and appropriateness for a given kind of a solved problem.

4.5. Nested simulations

Nested, or *recursive simulation* represents another methodology applicable for the needs of decision-making processes. The principle of that method is based on interrupting the *main simulation* when a conflict situation occurs and then the main simulation is cloned. Individual clones (*minor simulations*) are parameterized in such a way that various options of solutions are tested. Such nested/recursive simulations (different outlooks into the future applying a limited time horizon) are run and after a certain time period it is assessed which minor simulation shows the best results. Then the main simulation continues only with “the best” selected option. Despite the fact that the principle of nested simulations is simple and it uses a versatile simulation engine for finding solutions, several crucial technical issues have to be figured out.

5. THE METHOD OF NESTED SIMULATIONS

The method of *nested/recursive simulations* is based on the principle of using simulation trials inside a main simulation run in order to examine the results of several alternative simulation scenarios. The main instance (trial) of simulation is cloned and individual clones (minor/alternative trials) are parameterized differently. The method of nested simulations provides executing several alternative scenarios in parallel. The results of nested simulations present a broader set of solutions to the given problem.

One possible application is connected with decision-making support within simulating systems. Nested simulations are being simulated for a limited time period, their results are assessed, and the minor simulations are joined again into one instance. Then the main simulation can continue with the selected solution. That approach we apply within the frame of our research.

Another application of nested simulation is related to *multi-trajectory simulation* - a simulation experiment is divided into nested trials in critical points, and subsequently those trials can be increasingly nested. According to Gilmer and Sullivan's (1999), such a procedure is more efficient than using a higher number of replications of one simulation experiment.

5.1. The technique of nested simulations

Nested simulations allow using (i) an existing simulation engine and (ii) several simulation trials for searching solutions of occurred problems. Let us introduce a relevant procedure focused on solving critical/conflict situations:

1. A conflict situation (requiring an appropriate decision) is identified during simulation.
2. Current instance of the main simulation (S^{main}) is interrupted at the time t .
3. For the needs of nested simulations, it is necessary to set their parameters:
 - a. The criterion of optimality ($CrOpt$).

- b. The duration of an outlook into the future for the nested trials (or rather the stopping condition - $StopCond$).
 - c. The number of replications for all individual scenarios of nested simulations - $ReplCount$.
 - d. The number of alternative scenarios (minor simulations) - $ScnCount$.
4. N alternative scenarios for minor simulations are established.
5. The main simulation S^{main} is cloned and $ReplCount$ of replications is created for each i -th scenario ($i = 1 \dots N$).
6. Individual replications $S_i(j)$ are started (for $i = 1 \dots N, j = 1 \dots ReplCount$).
7. Waiting for finishing all replications $S_i(j)$ (for $i = 1 \dots N, j = 1 \dots ReplCount$).
8. Assessing the results of individual scenarios from the replications $S_i(j)$ (for $i = 1 \dots N, j = 1 \dots ReplCount$) and then selecting the scenario with the best results according to $CrOpt$.
9. The main simulation S^{main} then continues with the selected scenario from the instant t of simulation time.

The above mentioned procedure is illustrated in Figure 1. Certainly the problem how to define alternative scenarios for solving conflict situations represents a non-trivial problem which will be discussed in the full contribution.

5.2. Brief overview of the state-of-the-art

It has to be declared that not many authors pay attention to the research of nested/recursive simulations.

The authors Gilmer and Sullivan were focused in several of their articles on the efficiency of higher number of replications in contrast with multi-trajectory simulation (Gilmer and Sullivan, 1999). Their main interest is related to the military simulator *Eaglet*, which simulates the movement of military units of two armies and their mutual interactions.

Eugen Kindler (as a pioneer of nested simulation in Europe) published many articles with the focus on both, the theoretical description of nested simulations (classification, terminology, etc.) and their applications in practice (Kindler, 2010).

The issue of a planning support system is discussed by Hill, Surdu, Ragsdale, and Schafer (2000). Those authors were engaged in military planning.

Another area of applied nested simulations is connected with scalable simulation models, which allow applying both a macroscopic and a microscopic level of investigation within the frame of one simulator (Bonté, Duboz, Quesnel and Muller, 2009). Another area of exploiting nested simulations is financial a risk management – e.g. Gordy and June (2010).

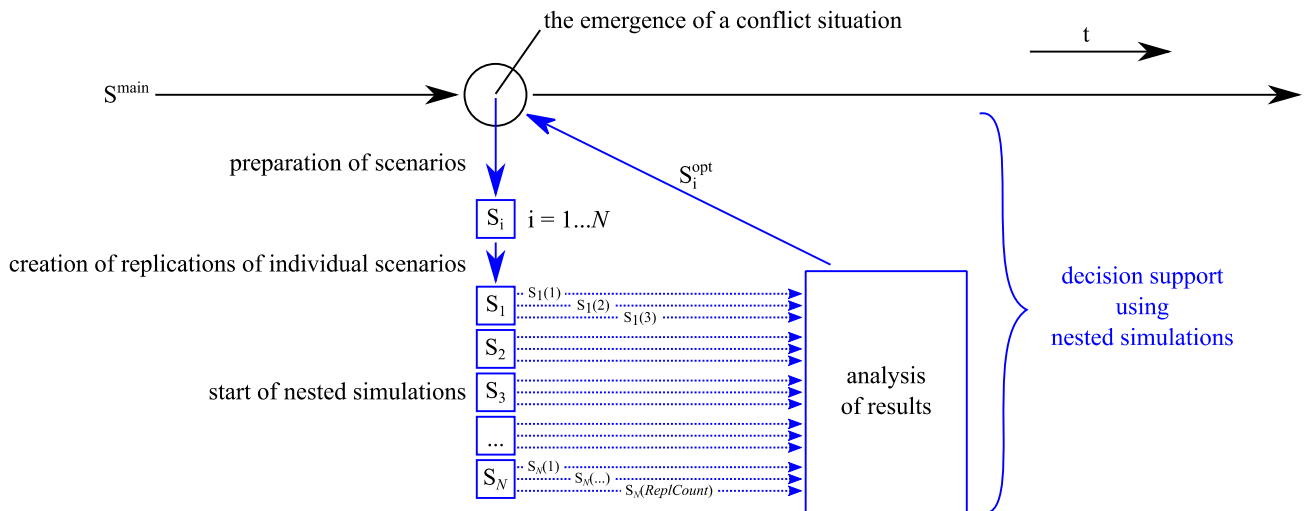


Figure 1: Illustration of one replication belonging to the main simulation - occurrence of one conflict situation is depicted

5.3. Executing nested simulations

Before the execution of nested simulations, it is necessary to fix a set of parameters. That set can differ depending on the type of the conflict situation that has occurred.

The first parameter specifies the number of alternative scenarios (or rather their minimal and maximal permissible count, *ScnCount*), which determines how many various alternatives will be examined in the nested simulations.

For every alternative scenario it is necessary to get statistically processed results (based on the outcomes from *ReplCount* replications). Individual nested simulation trials must be terminated after a certain period of time in order to assess their results. Stop condition *StopCond* dictates after what elapsed simulation time (or under which conditions) the nested simulations will be terminated. After terminating all nested simulations, their assessment is carried out.

The last parameter is the criterion of optimality (*CrOpt*) – that function evaluates the results of individual scenarios, which were executed as nested simulations and it selects the scenario providing the best solution. The main simulation then continues using the selected scenario.

5.3.1. Computational complexity

A separate and complex issue of nested simulations is related to applicable implementation techniques. Implementations are connected with potential time-consuming computational tasks that are influenced by the numbers and lengths of executed simulation trials. Thus, overall computational complexity is influenced by several factors:

- the number of alternative scenarios,
- the number of replications,
- the lengths (time durations) of replications,
- the number of conflict situations that occur in the main simulation and which require making

decisions based on the outlooks of nested simulations,

- the number of replications of the main simulation.

Apart from these factors, it is also necessary to consider the possibility of occurred conflict situations within the nested trials. Such a phenomenon can cause recursive run of other nested trials and cause in fact an exponential growth of the problem complexity. One way how to avoid such a problem is to terminate the nested simulation exactly at the moment when a conflict situation occurs inside it. That approach will be considered in our case study.

Executing a large number of nested simulations can take a lot of time even if the modern computers are utilized. The individual nested simulation trials do not affect each other and therefore they can be executed in parallel. A simple way is to execute the calculations in separate processes or threads and allow the nested simulations to use more processor cores within one computer. The most demanding tasks can allocate relevant computation on the GPU or use parallel processing on more computers (either within a distributed grid structure or a cloud).

Allocating computations into more processes, threads, or computers means that the simulator must be able to save the status of the simulation, copy it, and then prepared it for further parallel processing. For distributed methods (grid/cloud) it is then necessary to select a suitable way of data transfer and synchronization - shared file storage, communication over network sockets, etc.

5.4. Case study

Our research focus is related to traffic simulations mainly reflecting railway systems. Hence, an application of nested simulations as a decision-making support within the above mentioned kinds of simulators was chosen to be tested.

5.4.1. Simulation tool MesoRail

The simulation tool *MesoRail* (Diviš and Kavička, 2015) is a mesoscopic simulator of railway traffic, which focuses on examining traffic characteristics of railway stations on a mesoscopic level. The mission of that simulator is to allow processing simulation studies in shorter time than it is usual with the help of simulation tools applying a microscopic level of details. Our current research is concentrated on decisions about assigning substitute platform tracks to delayed trains. The appropriate decisions are supposed to be taken with the help of nested simulations.

The method of nested simulations was selected because of several reasons. The first reason is the ambition to use the tested simulation engine as a tool participating in decision-making processes. Another reason is that the principle of nested simulations allows simulating potential traffic progress directly, thus the results of the selected scenario are immediately transparent, and the best solution is then selected.

5.4.2. Simulation scenario

The problem of assigning substitute platform tracks was selected as a case study of deploying nested simulations as an automated decision support in a simulator. In stochastic simulations, there are delays occurring for individual trains (upon entry to the simulator and during the simulation itself) and thus conflict situations may arise. Standard platform track for an arriving train can already be occupied and so the role of the decision support is to select a substitute platform track. The selection of a substitute track also affects the situation in the station and it can cause more conflict situations. The decision support algorithm should ideally minimize the subsequent conflicts and it should also keep the station throughput at maximum.

Our goal was to apply and test the method of nested simulations. The method of priority planning was selected as a competitive method. A list of alternative targets (tracks in the station) was defined for each train route. During the run of the simulation, an available route with the highest priority is automatically selected. The priorities are static and fixed for all simulation experiments.

The parameters of the nested simulation method were selected as following:

- optimality criterion – the total of delays for all trains in the simulation,
- the number of replications of nested simulation – 1
- the number of alternative scenarios – maximal possible number given the situation,
- the time duration of the nested simulation – 5 minutes,
- the possibility to conduct recursive nested simulations – no.

For the first experiments with the nested simulations method, we created a small infrastructure of a railway station – 5 platform tracks, 2 passable station tracks, and a double track leading from east to west. The default scenario includes freight and passenger trains arriving to the station in a 10 minute interval from both directions. Occupancy of platform tracks for a given timetable is depicted in Figure 2.

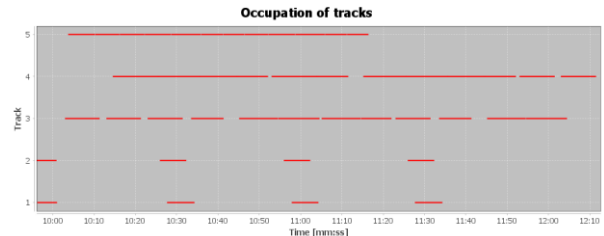


Figure 2: Occupation of tracks in the case study simulation scenario

To assess the quality of the decision support, stochastic simulations were conducted - the trains were assigned a random delay on the point of entry into the simulation. Exponential division of probability of the average of 5 minutes for passenger trains and the average of 15 minutes for freight trains was selected.

Then, a series of experiments was conducted: (i) without delay, (ii) with delay only for passenger trains, (iii) with delay for all trains. To assess the quality, the total of delay of all trains in the simulation was used as a criterion. Example of running nested simulations within *MesoRail* simulator is shown on Figure 3, four nested simulations are trying to find replacement station track for delayed incoming train. The results of individual experiments are illustrated in Table 1.

To conclude from the shown results, the method of priority planning showed the best results for the conducted case study, second best was the nested simulations method, and the longest delays occurred without decision support (in this scenario, trains wait for the original station track to be vacant). Alternative II shows a great difference between both methods; the method of priority planning effectively used vacant station tracks. Nevertheless, the method of nested simulations does not show such great results for the selected parameters. Alternative III depicts a more complex situation in which all trains have been assigned a delay. The method of priority planning again shows the best results, method of nested simulations is in second place, differences between both methods remains nearly similar.



Figure 3: Running main simulation with four nested simulations (on left side of image)

Table 1: Results of simulation experiments

Scenario	Method	Average value of sum of delay additions	Minimum	Maximum
I	no decision support	100,0 %		
	priority planning	97,2 %		
	nested simulations	98,2 %		
II	no decision support	124,8 % ± 14,7 %	103,7 %	151,4 %
	priority planning	94,4 % ± 12,0 %	71,8 %	112,0 %
	nested simulations	111,2 % ± 16,4 %	85,4 %	134,5 %
III	no decision support	131,3 % ± 20,6 %	94,9 %	159,4 %
	priority planning	97,5 % ± 14,2 %	79,1 %	117,8 %
	nested simulations	121,1 % ± 11,7 %	100,2 %	138,8 %

6. CONCLUSION

The issue of solving conflict situation in stochastic simulations was introduced in this article. Such situations can be solved manually or automatically by using a decision support based on various methods. Next, the method of nested simulations, which uses nested simulations to find the best solution, was introduced. The method of nested simulations was implemented into the Mesoscopic simulation tool *MesoRail* and tested on a case study. The results of the method were compared to the method of priority planning. From the collected data it is obvious that the method of priority planning now shows better results than the method of nested simulations. Nevertheless, the method of nested simulations offers a wide variety of parameterization and application. In the next development phase of the *MesoRail* tool, the effect of various parameter combinations on gained results will be tested further. An interesting yet very complex and computationally demanding task is the option to perform recursive nested simulations and thus search for an optimal solution in great detail. Using such approach with an appropriate selection of parameters should allow to gain even better results than with the static method of priority planning.

REFERENCES

- ETF, 2013. Capacity leaflet (UIC 406). Available at: http://www.uic.org/etf/codex/codexdetail.php?lang_ue_fiche=E&codeFiche=406 [accessed 6 April 2015].
- SŽDC, 2009. D24 – The regulation for determining capacities of railway lines. Available from: <http://www.szdc.cz/dalsi-informace/dokumenty-apredpisy/provoznotechnicke.html?page=detail&docid=1%3B%237eacf179-1551-468f-949a-641e9600c22d> [accessed 6 April 2015].
- Diviš R., Kavička A, 2015. Design and development of a mesoscopic simulator specialized in investigating capacities of railway nodes. Proceedings of the European Modeling and Simulation Symposium, 52-57. September 21-23, Bergamo, Italy.
- Gilmer J. B., Sullivan F. J., 1999. Multitrajectory simulation performance for varying scenario sizes [combat simulation]. In: WSC'99. 1999 Winter Simulation Conference Proceedings. 'Simulation - A Bridge to the Future' (Cat. No.99CH37038), 1137-1146. December 5-8, Phoenix, AZ, USA.
- Kindler E., 2010. Nested Models Implemented in Nested Theories. In: Proceedings of the 12th WSEAS International Conference on Automatic Control, Modelling & Simulation, 150-159. May 29-31, Catania, Sicily, Italy.
- Hill J. M. D., Surdu J. R., Ragsdale D. J., Schafer J. H., 2000. Anticipatory planning in information

- operations. In: SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. 'Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions' (Cat. No.00CH37166), 2350-2355. October 8-11, Nashville, Tennessee, USA.
- Bonté B., Duboz R., Quesnel G., Muller J. P., 2009. Recursive simulation and experimental frame for multiscale simulation. In: Proceedings of the 2009 Summer Computer Simulation Conference, 164-172. July 13-16, Istanbul, Turkey.
- Gordy M. B., Juneja S., 2010. Nested Simulation in Portfolio Risk Measurement. *Management Science* 56:1833-1848.
- Li H., Li Z., Li L. X., Hu B, 200. A production rescheduling expert simulation system. *European Journal of Operational Research* 124:283-293.
- Masmoudi W., Chtourou H., Maalej A. Y., 2007. Labor and machine sizing through a Simulation-Expert-System-based Approach. *Simulation Modelling Practice and Theory* 15:98-110.

ANALYSIS OF UNCERTAINTY IN ENGINEERING DESIGN OPTIMIZATION PROBLEMS

Philipp Fleck^(a), Michael Kommenda^(b), Michael Affenzeller^(c), Thorsten Prante^(d)

^{(a), (b), (c)} Heuristic and Evolutionary Algorithms Laboratory,
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg i. M., Austria

^{(b), (c)} Institute for Formal Verification,
Johannes Kepler University, Altenberger Straße 69, 4040 Linz, Austria

^(d) V-Research GmbH,
CAMPUS V, Stadtstraße 33, 6850 Dornbirn

^(a) philipp.fleck@fh-hagenberg.at, ^(b) michael.kommenda@fh-hagenberg.at,
^(c) michael.affenzeller@fh-hagenberg.at, ^(d) thorsten.prante@v-research.at

ABSTRACT

In this paper, we analyze popular benchmark instances in the field of engineering design optimization regarding the robustness of published solutions. First, we implement selected benchmark problems with HeuristicLab and show the advantages of having a framework that enables rapid prototyping for optimization and analysis. Then, we show that many solutions quickly become infeasible when considering uncertainty like production inaccuracies. Based on these findings, we motivate why robust solutions for engineering design are important and present methods for measuring, identifying and visualizing robustness. Finally, we present how solutions can be compared and selected using a novel robustness measure.

Keywords: engineering design optimization, constraint handling, uncertainty, robustness

1. INTRODUCTION

The increasing complexity of today's engineering tasks have made computer systems an omnipresent part of the engineering design process. These tools help human experts to make meaningful design choices in order to develop products or systems that satisfy a complex set of requirements. Particularly, finding the optimal set of decisions for a given engineering design problem is an ongoing challenge.

Engineering design problems are often formulated as constrained optimization problems with a design vector $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, an objective function $f(\mathbf{x})$, inequality constraints $g_j(\mathbf{x}) \leq 0$ and equality constraints $h_k(\mathbf{x}) = 0$ with $f, g_j, h_k: \mathbb{R}^n \rightarrow \mathbb{R}$. The range of each design value x_i is either given explicitly, as $\min_i \leq x_i \leq \max_i$, or as inequality constraints; for example $1 \leq x_1 \leq 5$ is equivalent to $g_1 = 1 - x_1 \leq 0$ and $g_2 = x_1 - 5 \leq 0$.

A variety of benchmark engineering design optimization problems exists for evaluating and comparing different optimization techniques. In this paper, we cover four well studied benchmark problems, with formal descriptions of each problem given in the Appendix:

- Pressure Vessel (PV) (Sandgren 1988)
- Speed Reducer (SR) (Golinski 1973)
- Tension/Compression Spring (TCS) (Belegundu 1982)
- Welded Beam (WB) (Ragsdell and Phillips 1976)

We show how to implement the selected benchmark problems with the open source optimization framework HeuristicLab (Wagner et al. 2014) and demonstrate the advantages and simplicity of using a framework for rapid prototyping and compare our solutions to solutions from relevant literature. Then, we discuss uncertainty and present methods to quantify uncertainty and finally, we introduce analysis methods to compare the robustness of different solutions.

2. OPTIMIZATION OF ENGINEERING DESIGN PROBLEMS

2.1. Solving Engineering Design Optimization Problems with HeuristicLab

HeuristicLab (Wagner et al. 2014) is an open source framework for heuristic optimization that is freely available at <http://dev.heuristiclab.com> and offers a wide range of popular optimization algorithms and benchmark problems. Users can easily implement their own optimization problems, using the built-in scripting environment, which allows specifying the objective function of optimization problems in C# code directly in HeuristicLab (Scheibenpflug et al. 2015).

To implement a new problem, the user has to create a *Programmable Problem (single-objective)*, where some exemplary code is already provided. After having implemented the problem, the code is compiled at runtime and the problem can be solved using one of the many algorithms. A fully specified optimization problem requires three parts that the user has to implement (for illustration see Figure 1):

1. Specification whether the objective should be maximized or minimized, which is done by implementing the *Maximization* property.
2. An appropriate *Encoding* for the optimization problem. Code for popular encodings is already provided in the *Initialize* method, for instance the *RealVectorEncoding*.
3. The objective function, which is implemented in the *Evaluate* method and has to return the quality of a given solution.

Different forms of evaluating the objective are possible, ranging from implementing code for a simple analytical model to executing a full-scaled simulation. The benchmark optimization problems used in this paper specify analytical objective functions and are therefore quite easy to implement. For instance, the objective function for the tension/compression spring problem,

$$f(x) = (x_3 + 2) \cdot x_2 \cdot x_1^2, \quad (1)$$

translates to the following source code (without constraint handling):

```
public double Evaluate(
    Individual ind, IRandom r) {
    var x = ind.RealVector();
    double x1=x[0], x2=x[1], x3=x[2];
    return (x3 + 2) * x2 * x1*x1;
}
```

Optionally, the *Analyze* method in the *Programmable Problem* can be implemented to gather additional information during the optimization, for instance recording the best feasible solution found. Figure 1 shows a screenshot of a fully specified optimization problem, including constraint handling.

Furthermore, HeuristicLab can be extended by developing new plugins, which allows the user to implement new encodings or new algorithms (see Scheibenpflug et al. (2015) for more details).

2.2. Constraint Handling

Many optimization problems specify different kinds of constraints. However, optimization algorithms usually do not handle constraints explicitly and require them being incorporated into the objective function. A common way of handling constraints are *penalty functions*, where the original constrained objective function is transformed into an unconstrained objective function by adding a penalty for infeasible solutions (Coello Coello 2002).

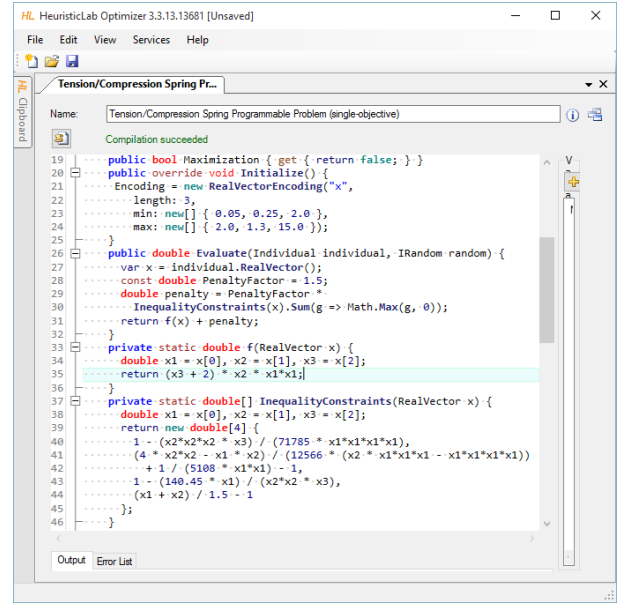


Figure 1: Screenshot showing the implementation of the tension/compression spring problem with HeuristicLab's integrated scripting environment.

A new unconstrained objective function $\phi(x)$ is often formulated as

$$\phi(x) = f(x) \pm \left(\sum_{j=1} r_j G_j + \sum_{k=1} c_k H_k \right) \quad (2)$$

where r_j and c_k are constants called *penalty factors*. G_j and H_k are functions of the inequality and equality constraints $g_j(x)$ and $h_k(x)$ respectively and are commonly defined as

$$G_j = \max(0, g_j(x))^\beta \quad \text{and} \quad H_k = |h_k(x)|^\gamma \quad (3)$$

where β and γ are usually 1 or 2. The penalties are added if the problem is minimized, or subtracted if maximized. All problems solved in this paper are minimization problems with only inequality constraints; further, we simplify problem specification by only a single penalty value r . The resulting objective function used further in this paper is

$$\phi(x) = f(x) + r \sum_{j=1} \max(0, g_j(x)) \quad (4)$$

For each optimization problem, the penalty factor r is determined by obtaining a typical objective value for this problem and multiplying it by a factor to make it intentionally worse. We obtained the typical objective value by calculating the mean of 10,000 random solutions and then multiplying it by 5. For each problem, the resulting penalty value is stated in Table 1.

2.3. Single-Objective Optimization

We use the Covariance Matrix Adaption Evolution Strategy (CMA-ES) to solve the engineering design optimization problems. The CMA-ES is a population-

based evolutionary algorithm, specifically designed for real-valued, single objective optimization problems (Hansen et al. 2003).

The basic idea of the algorithm is to sample promising areas in the search space with a multivariate normal distribution $\mathcal{N}(\mathbf{m}, \sigma^2 C)$, defined by the mean vector $\mathbf{m} \in \mathbb{R}^n$ and the covariance matrix $\sigma^2 C \in \mathbb{R}^{n \times n}$ where $\sigma \in \mathbb{R}$ defines an additional scaling factor (called the step-size). Figure 2 illustrates different types of distributions by different covariance matrices.

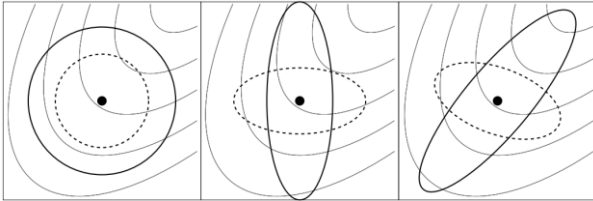


Figure 2: Different distributions given by the same mean vector but different covariance matrices, symbolized by density-ellipsoids: The identity covariance matrix resulting in circular distributions (left). A diagonal covariance matrix resulting in axis-aligned distributions (middle). A symmetric, positive-definite matrix capable of describing axis-independent distributions (right). The background shows the gradients of the search space towards the optimum in the upper right. Figure adapted from Hansen (2005).

The algorithm maintains and updates the parameters of the multivariate normal distribution so that it increases the probability of moving towards promising areas in the search space. Creating new sample points and updating the parameters is done iteratively, until a termination criteria is met. The following briefly describes the main parts of the CMA-ES. An extensive description is given in Hansen et al. (2003).

- (1) In the beginning, the first mean vector \mathbf{m} is determined by randomly creating λ vectors and calculating their mean. The initial step-size σ is given by the user, usually as a vector to setup axis-scaling for each dimension. The matrix C is initialized with the identity matrix.
- (2) At each iteration, λ samples are created with the multivariate normal distribution $\mathcal{N}(\mathbf{m}, \sigma^2 C)$, with the mean vector \mathbf{m} describing the current center of search and the covariance matrix $\sigma^2 C$ describing the form of the distribution and thus the search direction. The step-size σ controls the scale, therefore the “speed” the algorithm moves through the search space.
- (3) The solutions are evaluated according to the given objective function and sorted based on their objective value.
- (4) The weighted sum of the μ best solutions is calculated to determine the new mean vector \mathbf{m} . The weights, which must sum to one, are either equal or decrease to favor better solutions.
- (5) The step-size σ is updated based on the $1/5^{\text{th}}$ success rule already used in a standard evolution strategy

(Schumer and Steiglitz 1968). The matrix C is updated based on the gradient of the quality, approximated by the sampled solutions.

2.4. Results

We implemented and optimized the four selected engineering design optimization problems with the techniques described in the previous sections. For all problems we used the CMA-ES with a population size (λ) of 50 and 25 individuals (μ) for calculating the new mean (log-scaled weighting). After a maximum of a 100 iterations, the algorithm terminates. Table 1 shows the initial step-size σ and penalty factor r for each problem.

Table 1: Parameters used for the different problems.

Probl.	Initial σ	r
PV	(2, 2, 63, 63)	65000
SR	(0.33, 0.033, 3.67, 0.33, 0.33, 0.33, 0.167)	20000
TCS	(0.65, 0.35, 4.33)	50
WB	(0.63, 3.3, 3.3, 0.63)	70

Table 2 contains a comparison between our results and the results of various literature, showing that the results of the CMA-ES are equal or close to the best-found solutions from the literature. This demonstrates that common benchmark optimization problems in the fields of engineering design can be solved with little implementation effort, when using an appropriate framework that allows easy problem specification and offers proper optimization algorithms. Otherwise, implementing the problem specification and optimization algorithms can be time consuming.

While recalculating the quality of the solutions from the literature for validation, we observed that several solutions slightly violate some of the constraints. We suspect that the published solutions are rounded to some degree, causing the small infeasibilities in our calculations. Hence, we introduced a tolerance value for evaluating inequality constraints, so that $g_j(\mathbf{x}) \leq 10^\epsilon$, to determine how closely the constraints are violated. Table 2 includes the number of violated constraints per solution and states the tolerance value so that the solution becomes feasible. Not all authors reported solutions for all four problems we analyze, thus some entries in the table are empty.

For readability, the numbers in Table 2 are rounded, yielding solutions with the same quality but different constraint violations. A detailed table with the precise qualities and the all solution vectors is available online at <http://dev.heuristiclab.com/AdditionalMaterial#EMSS2016>.

Additionally, all implemented problems and fully configured optimization algorithms along with the algorithm results presented in this paper are also available under the same URL.

For some problems multiple versions exist, e.g. Welded Beam, thus we only compare results that were solved using the exact same problem specification. The exact specifications used in this paper can be found in the Appendix.

Table 2: Results of selected engineering design problems solved by the CMA-ES (present study) compared with other selected literature solutions. “Quality” shows the quality (rounded) of the best solution found, “V” shows the number of violated constraints for that solution and “ ϵ ” the tolerance factor (in 10^ϵ) to make the solution feasible. The best and best feasible solution are marked bold.

Reference	Pressure Vessel			Speed Reducer			Tension/C. Spring			Welded Beam		
	Quality	V	ϵ	Quality	V	ϵ	Quality	V	ϵ	Quality	V	ϵ
Present study	6059.71	0		2994.47	0		0.0126652	0		2.38096	0	
(Akay and Karaboga 2010)	6059.71	1	-8	2997.06	1	-6	0.0126650	1	-4			
(Akhtar et al. 2002)	6171.00	0		3008.08	0					2.44260	0	
(Azad and Fernandes 2011)	6059.53	2	0	2994.32	2	-3	0.0126640	2	-4	2.38081	1	-4
(Bernardino et al. 2007)	6060.37	0		2994.47	1	-6	0.0126680	0		2.38122	0	
(Bernardino et al. 2007)	6060.14	0		2994.47	0		0.0126660	0		2.38125	0	
(Bernardino et al. 2008)	6059.85	0		2996.35	1	-6	0.0126660	1	-5	2.38144	0	
(Bernardino et al. 2008)	6065.82	0		2996.35	1	-6	0.0126838	0		2.38335	0	
(Bernardino et al. 2008)	6832.58	1	1	2996.35	1	-6	0.0126790	0		2.59610	0	
(Cagnina et al. 2008)	6059.71	1	-1	2996.35	2	-6	0.0126650	1	-4			
(Coello Coello 2000)	6288.74	0					0.0127048	0				
(He and Wang 2007)	6061.08	0					0.0126747	0				
(He et al. 2004)	6059.71	1	-10				0.0126653	1	-7	2.38096	0	
(Hu et al. 2003)	6059.13	1	-7				0.0126661	1	-8			
(Kaveh and Talatahari 2009)	6059.09	1	-4				0.0126391	1	-2			
(Kaveh and Talatahari 2010)	6059.73	0					0.0126432	1	-2			
(Lemonge et al. 2010)	6059.72	1	1	2996.35	1	-6	0.0126790	1	-5	2.38124	0	
(Mezura Montes et al. 2003)	6059.71	1	1	3025.01	0		0.0126650	0				
(Mezura Montes et al. 2007)	6059.70	2	1	2996.36	0		0.0126980	0				
(Ray and Liew 2003)				2994.74	0		0.0126692	0		2.38543	0	
(Rocha and Fernandes 2009)	6071.17	0		2994.37	0		0.0126680	0		2.38627	0	
(Rocha and Fernandes 2009)	6072.23	0		2995.80	0		0.0126670	0		2.43162	0	
(Tomassetti 2010)	6059.71	1	-8	2996.35	1	-6	0.0126650	0				
(Zhang et al. 2008)	7197.70	1	1	2994.47	0		0.0126652	1	-8	2.38096	0	

3. ROBUSTNESS IN ENGINEERING DESIGN

Our results in Section 2.4 show that small deviations can make solutions infeasible. In this section, we generalize these small deviations by applying *uncertainty* in various stages of the engineering design process to observe how the final quality and feasibility of solutions changes. Our goal is to find *robust* solutions that stay feasible for increasing uncertainty while maintaining high quality. To analyze robustness, we first define different types of uncertainty and describe how to quantify them. Then, we introduce methods for measuring and visualizing robustness that allows comparing different solutions. Finally, we introduce a novel measure that allows ranking of solutions based on their robustness.

3.1. Uncertainty characterization

Uncertainty summarizes indirect influences that can occur during the engineering design process and are not controllable by the design vector \mathbf{x} . Beyer and Sendhoff (2007) describe four types of uncertainty, based on which aspect of evaluation is influenced.

A) *Changing environmental and operating conditions* summarize controllable influences of the evaluation process that are not reflected in the design variables. Typical influences are environmental settings like operating temperature.

The objective function f becomes a function of the design vector \mathbf{x} and an additional parameter vector $\boldsymbol{\alpha}$ that quantifies all additional settings of the system,

$$f = f(\mathbf{x}, \boldsymbol{\alpha}). \quad (5)$$

B) *Production tolerances and actuator imprecision* summarize uncontrollable influences on the design vector. For instance, machine imprecision when actually manufacturing a product described by the design vector. Formally, a perturbation vector $\boldsymbol{\delta}$ is added on the design vector, resulting in an objective function

$$f = f(\mathbf{x} + \boldsymbol{\delta}, \boldsymbol{\alpha}), \quad (6)$$

where the perturbation is absolute or relative ($\boldsymbol{\delta} = \epsilon * \mathbf{x}$).

C) *Uncertainty in the system output* occurs due to imprecise measuring of the output. The observed output differs from the actual evaluation output, which is formally described by a random function

$$\tilde{f} = \tilde{f}[f(\mathbf{x} + \boldsymbol{\delta}, \boldsymbol{\alpha})]. \quad (7)$$

D) *Feasibility uncertainties* summarize effects on constraint satisfaction. Technically, this is not a separate type of uncertainty and is often modeled as type A or type B.

In the remainder of this paper, we only use uncertainty type B, *production tolerances*, because they can be specified in the same way for all engineering design problems and they are best suited to demonstrate uncertainty in this paper.

3.2. Quantifying Uncertainty

Uncertainties can be quantified *deterministically*, *probabilistically* or *possibilistically* (Beyer and Sendhoff 2007). We quantify uncertainties probabilistically, using normal distributions $\mathcal{N}(\mu, \sigma^2)$ and uniform distributions $\mathcal{U}(a, b)$ to describe the perturbation vector δ , which we define relatively to the design vector x . For instance, $\delta_i = 0.01 x_i$ increases an entry of the design vector by one percent. Quantifying the uncertainty then becomes a matter of quantifying the underlying random distributions.

Usually, uncertainties affect production processes on different levels and with different magnitudes. Therefore, each entry of the perturbation vector can be separately defined by its own probability distribution, or the whole perturbation vector is defined by a multivariate random distribution. However, for the remainder of this paper, we simply define all entries with the same distribution, except otherwise specified.

Because the perturbation is defined relative, and we do not want to introduce an intentional bias, the random distributions are parameterized to have their center at zero. This reduces the parameters of the distribution to a single value: the variance σ^2 of the normal distribution $\mathcal{N}(0, \sigma^2)$ or the range r of the uniform distribution $\mathcal{U}(-r/2, +r/2)$. We call this single parameter *uncertainty level* u . For instance, using a uniformly distributed uncertainty with an uncertainty level $u = 10^{-2}$ describes the uncertainty using $\mathcal{U}(-0.5, +0.5)$.

3.3. Measuring Robustness

Evaluating a solution at a fixed uncertainty level does not give a general measure of how a solution behaves under a broader range of uncertainty. Instead, an interval of uncertainty levels must be evaluated, with the lower and upper bounds depending on the robustness requirements of the optimization problem. If the production process is not known, we use larger uncertainty ranges, e.g. between 10^{-8} and 10^{-1} . If the expected uncertainty is known, the range should be set more precisely.

For each uncertainty level within the range, the objective value (original objective without penalties) and the constraint violations are calculated, yielding progressions of the objective values and constraint violations over increasing uncertainties. In the first step, we analyze robustness only with respect to the progression of constraint violations.

Because δ is a random variable, constraint violations are not binary anymore, but measured as a probability instead; therefore, for each constraint, a progression of the violation probability is obtained with values between zero and one. We aggregate these violation progressions by calculating the sum of all violation probabilities at each uncertainty level, resulting in a single violations

progression with each value between zero (no violations) and the number of constraints (all constraints violated). Although aggregating prevents analyzing which specific constraints were violated, this is neglected for a general measure of robustness.

Figure 3 shows an example of four different solutions for the welded beam problem, with uncertainties ranging from 10^{-8} to 1. The uncertainties on the x-axis are scaled logarithmically to be able to observe effects on lower uncertainty levels as well as on higher levels. Because we name solutions by their literature origin, we use italic font when referring to solutions to avoid confusion with regular literature cites.

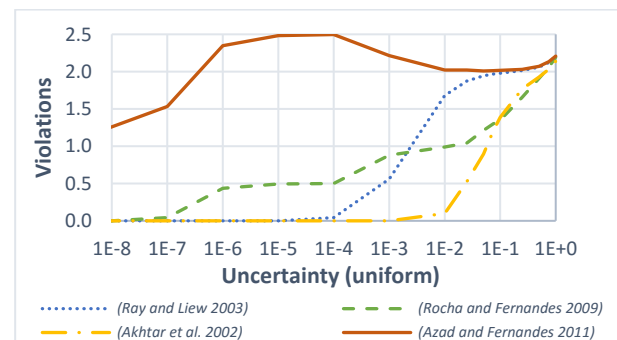


Figure 3: Progressions of violations over increasing uncertainties for selected solutions of the welded beam problem.

Typically, the violations begin to increase at different uncertainty levels for different solutions, e.g. (Rocha and Fernandes 2009) around 10^{-7} and (Ray and Liew 2003) around 10^{-4} . Robust solutions show influences only at higher uncertainty, whereas fragile solutions are affected earlier. The violations need not be strictly increasing, c.f. (Azad and Fernandes 2011). This can be the case if the solution violates constraints at lower or no uncertainty, and sufficiently high uncertainty allows it to become feasible again.

The series of violations also allows visual comparisons of different solutions. First and most important, solutions with a progression strictly lower than the progression of another solution are more robust since they violate less constraints at all uncertainty levels, cf. (Akhtar et al. 2002) versus (Ray and Liew 2003). This allows ranking solutions based on their robustness. In addition, the violations-gap of two solutions at a specific uncertainty level tells how much more robust a solution is.

Solutions with intersecting series cannot be compared unambiguously, like (Rocha and Fernandes 2009) and (Ray and Liew 2003). In such a case, unambiguity is limited to non-intersecting intervals. The smaller the interval, the more likely that multiple solutions can be compared unambiguously.

3.4. Estimate Robustness Ranking

We still want to estimate which solution is more robust over a large uncertainty interval, even though comparing solutions unambiguously is not possible when their violations progressions intersect. We calculate the area V

under the curve of the violations series over the uncertainty range. Because we do not know whether higher or lower uncertainty is more important, we weight them equally by using a log-scale on the uncertainty-axis, as in Figure 3. We use the middle Riemann sum with a log-scaled partition size to calculate the violation area

$$V = \sum_{i=1}^n \frac{v(x, u_i) + v(x, u_{i-1})}{2} (\log u_i - \log u_{i-1}), \quad (8)$$

where $v(x, u)$ yields the number of violated constraints for a solution x at the uncertainty level u and n is the number of samples drawn from the uncertainty range.

We argue that solutions with a lower area of violations are more robust over a larger uncertainty interval. This gives a single, real-valued coefficient for robustness that allows ordering and distance calculation.

However, this coefficient cannot be interpreted easily because its magnitude depends on the uncertainty range. By dividing the area by the range of the (logarithmized) uncertainty interval, it can be directly related to values calculated by a fixed uncertainty level. This relative area of violations,

$$V_{rel} = \frac{V}{\log u_{max} - \log u_{min}}, \quad (9)$$

can also be interpreted as weighted mean of violations at multiple fixed uncertainty levels, where the weights are the distances of two adjacent uncertainty levels.

With a single coefficient that describes robustness, we are able to compare quality and robustness in a scatterplot. This allows visualizing the tradeoff between robustness and quality, as shown in Figure 4 with selected solutions for the speed reducer problem. The objective is calculated similarly by the area under the objective progression. As the objective and the violations are minimized, solutions at the lower left are preferable. However, most solutions are either robust and low quality, like (Mezura Montes et al. 2003), or high quality and not robust, like (Bernardino et al. 2007). (Rocha and Fernandes 2009) would be a reasonable tradeoff.

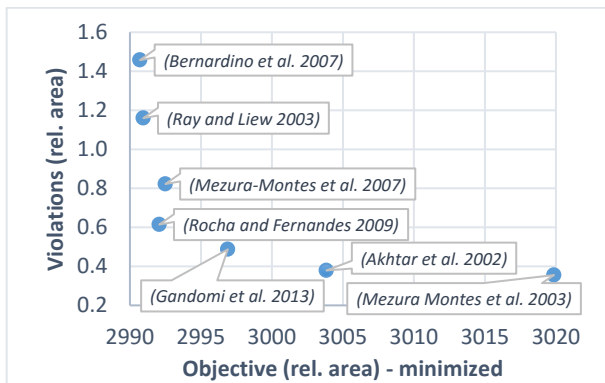


Figure 4: Comparison of objective and robustness of different solutions of the speed reducer problem by a scatterplot.

3.5. Results

In this section we present selected results from different problems for an overview. The full analysis of all problems with all solutions can be downloaded at <http://dev.heuristiclab.com/AdditionalMaterial#EMSS2016>.

For all the presented results we used uniformly distributed uncertainties, because it turned out the results are similar to using normal distributions and uniform distributions are easier to interpret due to the fixed ranges.

Concerning quality, we observed two general behaviors, demonstrated by Figure 5. First, uncertainty significantly influences the quality only at higher uncertainty levels with $u > 10^{-1}$; thus, at lower uncertainty, the order of solutions by their quality does not change at all. Secondly, the tendency whether quality increases or decreases with increasing uncertainty is the same for all solutions of a problem. Even the extend of the quality changes is very similar. This concludes that uncertainty does not significantly influence the order of solutions by their quality.

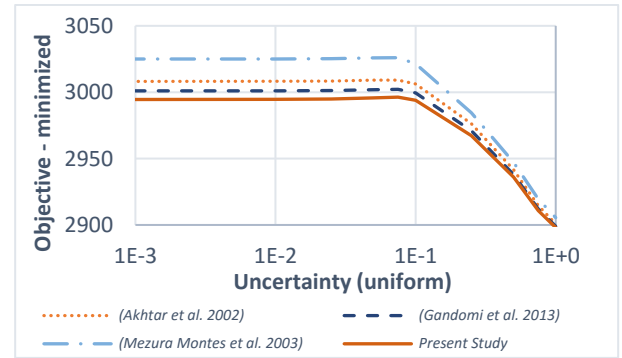


Figure 5: Quality progression over uncertainty for solutions of the speed reducer problem.

More interesting are the constraints violations, where we start with the results of the tension compression spring problem, since the results are easier to interpret. Figure 6 shows the violations progressions of selected solutions between uncertainty level 10^{-10} and 10^{-1} . Only solutions that did not violate any constraints without uncertainty were selected (see Table 2). Because there are effectively no intersections between the progressions, an unambiguous ranking for robustness can be determined. Our own solution, found by a CMA-ES, starts violating constraints early ($u < 10^{-10}$) because the solution lies very close to constraint boundaries. This also demonstrates that solutions specifically optimized for maximum quality are typically not robust. The other solutions from the literature use less significant decimal places thus are generally not as close to constraint boundaries. The most robust solution from the selected solutions in Figure 6 is (Coello Coello 2000); however, when compared by quality, this solution performs worse than all others, shown in the scatterplot in Figure 8. For quality only, (Ray and Liew 2003) would be the better choice. Choosing an appropriate compromise depends mainly on the expected uncertainty and specifics of the optimization problem.

The progressions of violations for the pressure vessel problem, shown in Figure 7, is more difficult to interpret. First of all, intersecting progressions means that not all solutions can be compared unambiguously, for instance (Akhtar et al. 2002) and (Coello Coello 2000). Some solutions, however, can be ranked unambiguously, for instance (He and Wang 2007) is more robust on all levels than (Bernardino et al. 2007), (Kaveh and Talatahari 2010) and the solution from this paper by the CMA-ES. Similar to the results of the tension compression spring, our solution is less robust than all other solutions. Choosing the most robust solution in this case depends on the expected uncertainty. If one expects small

uncertainties $u < 10^{-2}$, the most robust solution would be (Akhtar et al. 2002). If the production system usually inflicts high uncertainty ($u > 2 \cdot 10^{-2}$), (Coello Coello 2000) would be better. The estimated ranking over the whole uncertainty range, as described earlier, can be deduced from the scatterplot in Figure 9. The scatterplot also shows that (Akhtar et al. 2002) is slightly more robust than (Coello Coello 2000), but both have a significantly worse quality than all other displayed solutions. Results for the speed reducer problem and welded beam problem are omitted in this paper because the results are similar to the pressure vessel problem.

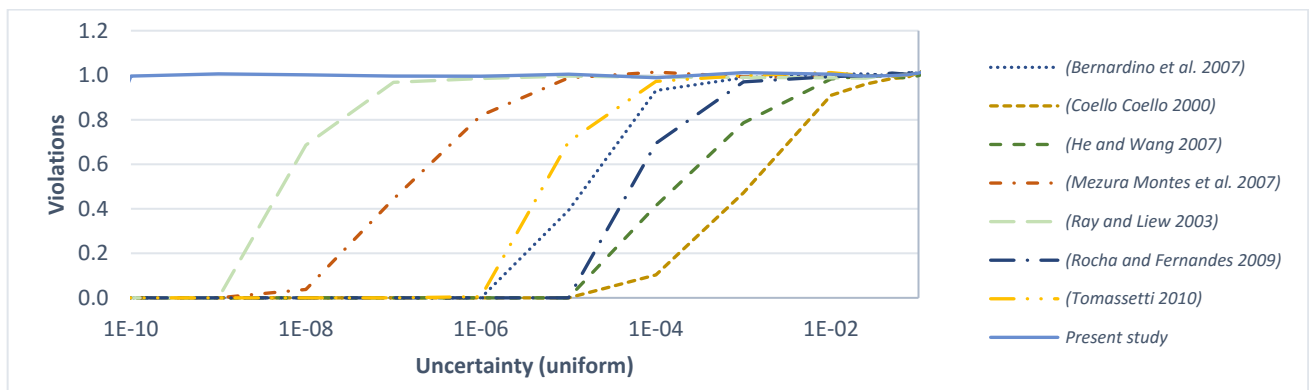


Figure 6: Violations progression over uncertainty for selected solutions of the tension compression spring problem.

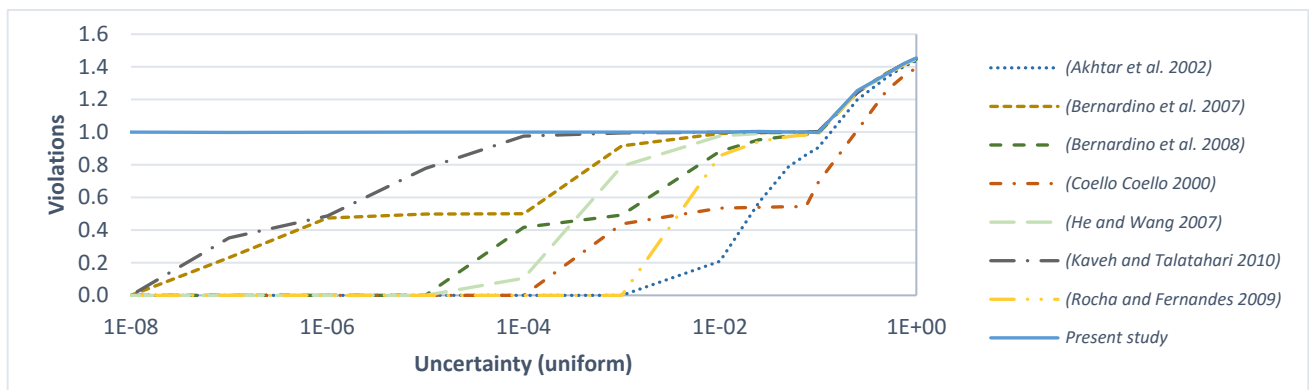


Figure 7: Violations progression over uncertainty for selected solutions of the pressure vessel problem.

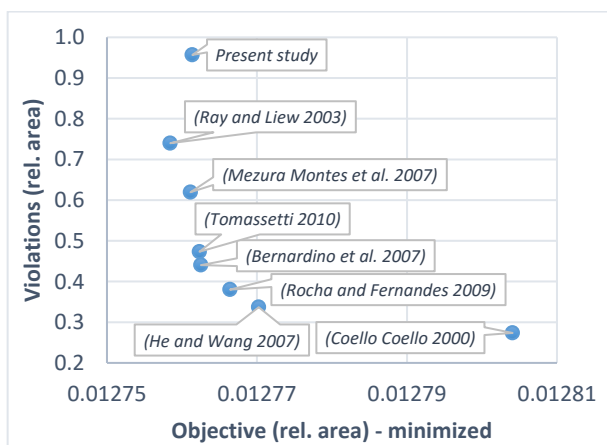


Figure 8: Scatterplot showing quality vs. violations of selected solutions for the tension compression spring problem.

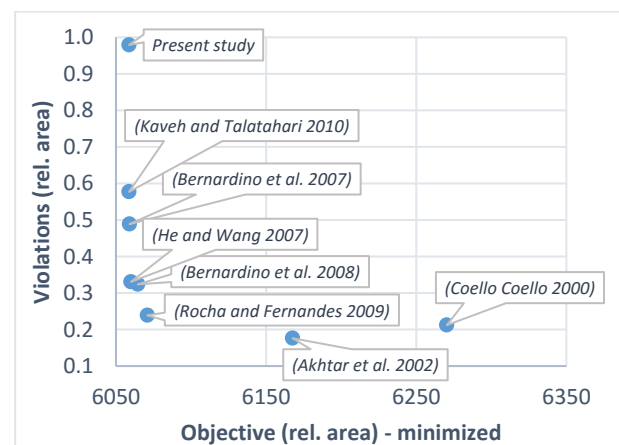


Figure 9: Scatterplot showing quality vs. violations of selected solutions for the pressure vessel problem.

4. DISCUSSION

In this paper, we motivated robustness of solutions in engineering design problems and demonstrated the necessary steps to analyze uncertainty and robustness.

In the first part of this paper, we implemented popular constrained optimization problems using HeuristicLab and demonstrated different techniques for constraint handling by transforming the problem into an unconstrained optimization problem using penalties. Then, we used the CMA-ES to find solutions for the given problems and compared our solutions to solutions from various literature sources.

In the second part of this paper, we introduced uncertainty as an influence that causes small perturbations to solution candidates. We quantified these uncertainties with random distributions and introduced uncertainty levels as a measure of the variance uncertainty inflicts on the solutions. Then, we showed that the robustness of solutions can be visualized and compared using charts that plot the progression of violated constraints over increasing uncertainty and further demonstrated how the area under the progression can be used to estimate a ranking of robustness when an unambiguous ranking is not possible. Finally, we illustrated that a solution's quality and robustness often oppose and a compromise must be picked.

A reasonable next step would be to consider uncertainty directly during the solving and optimization process, to search for solutions that are high quality and are robust. This suggests using multi-objective optimization techniques, which simultaneously optimize the objective value of a solution as well as its robustness. However, new questions arise, for instance, for which uncertainty level a solution should be optimized for? Is it possible to optimize it for a wider range of uncertainty? Which (optimization) algorithms are to use for optimizing the resulting multi-objective optimization problem?

Another interesting aspect for future research would be the analytical prediction of robustness based on the formal description of a constrained optimization problem. Our research suggested that for the same optimization problem, increasing uncertainty has the same effect on the quality for all solutions. Those correlations could also be predicted analytically by analyzing the objective function in detail. Similarly, the robustness of a solution could be predicted analytically by analyzing the constraints, how close a solution is to those constraints and how perturbation would increase the likelihood that those constraints will be violated.

For a broader empirical confirmation of our findings, additional engineering design optimization problems could be studied. A real world test scenario would also be very interesting.

We believe that uncertainty, and the goal to find robust solutions, will become a major factor in engineering problems. Thus, the study of techniques and algorithms that are able to find robust solutions while maintaining high quality will continue and increase.

ACKNOWLEDGMENTS

The work described in this paper was done within the COMET Project #843551 Advanced Engineering Design Automation (AEDA) and COMET Project #843532 Heuristic Optimization in Production and Logistics (HOPL), both funded by the Austrian Research Promotion Agency (FFG).

APPENDIX A: BENCHMARK INSTANCES

A.1 Pressure Vessel Problem

Minimize the fabrication costs of a cylindrical pressure vessel given by the thickness of the pressure vessel (x_1), the thickness of the head (x_2), the inner radius of the vessel (x_3) and the length of the vessel without heads (x_4), shown in Figure 10.

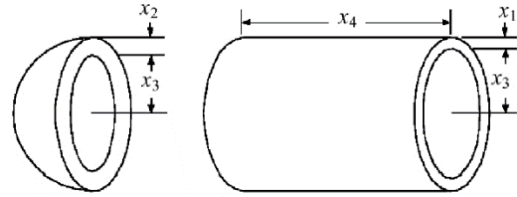


Figure 10: Schematic illustration of a pressure vessel and its parameters. Image adapted from Cagnina et al. (2008).

The pressure vessel problem is formally described as

$$f(\mathbf{x}) = 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + 3.1661x_1^2x_4 + 19.84x_1^2x_3$$

where

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \\ x_1 &= \{0.0625t \mid t \in \mathbb{Z} \cap [1; 99]\} \\ x_2 &= \{0.0625t \mid t \in \mathbb{Z} \cap [1; 99]\} \\ x_3 &\in [10; 200] \quad x_4 \in [10; 200] \end{aligned}$$

subject to $g_j(\mathbf{x}) \leq 0$ with

$$\begin{aligned} g_1(\mathbf{x}) &= -x_1 + 0.0193x_3 & g_2(\mathbf{x}) &= -x_2 + 0.00954x_3 \\ g_3(\mathbf{x}) &= -\pi x_3^2 x_4 - \frac{4}{3} \pi x_3^3 + 1,296,000 & g_4(\mathbf{x}) &= x_4 - 240 \end{aligned}$$

A.2 Speed Reducer Problem

Minimize the weight of a speed reducer given by the gear face width (x_1), the teeth module (x_2), the number of pinion teeth (x_3), the lengths of the shafts between bearings (x_4, x_5) and the diameters of the shafts (x_6, x_7), shown in Figure 11. The speed reducer is constrained to bending stress of the gear teeth, surface stress, transverse deflection of the shafts and stress in the shafts.

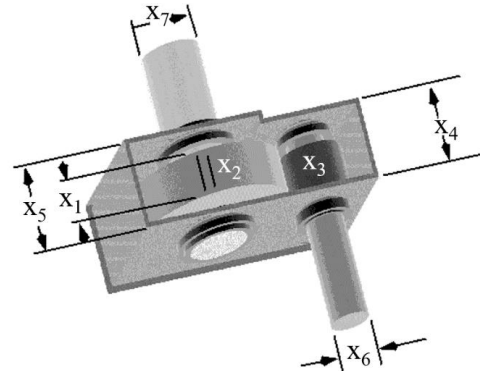


Figure 11: Schematic illustration of a speed reducer and its parameters. Image adapted from Townsend (2016).

The speed reducer problem is formally described as
 $f(\mathbf{x}) = 0.7854x_1x_2^2(3.3333x_3^3 + 14.9334x_3 - 43.0934)$
 $- 1.508x_1(x_6^2 + x_7^2) + 7.4777(x_6^3 + x_7^3)$
 $+ 0.7854(x_4x_6^2 + x_5x_7^2)$

where

$$\mathbf{x} = (x_1, x_2, \dots, x_7) \in \mathbb{R}^7$$

$$x_1 \in [2.6; 3.6] \quad x_2 \in [0.7; 0.8] \quad x_3 \in \mathbb{Z} \cap [17; 28]$$

$$x_4 \in [7.3; 8.3] \quad x_5 \in [7.3; 8.3] \quad x_6 \in [2.9; 3.9]$$

$$x_7 \in [5; 5.5]$$

subject to $g_j(\mathbf{x}) \leq 0$ with

$$g_1(\mathbf{x}) = \frac{27}{x_1x_2^2x_3} - 1 \quad g_2(\mathbf{x}) = \frac{397.5}{x_1x_2^2x_3^2} - 1$$

$$g_3(\mathbf{x}) = \frac{1.93x_3^3}{x_2x_3x_6^4} - 1 \quad g_4(\mathbf{x}) = \frac{1.93x_5^3}{x_2x_3x_7^4} - 1$$

$$g_5(\mathbf{x}) = \frac{1}{110x_6^3} \sqrt{\left(\frac{745x_4}{x_2x_3}\right)^2 + 16.9 \cdot 10^6} - 1$$

$$g_6(\mathbf{x}) = \frac{1}{85x_7^3} \sqrt{\left(\frac{745x_5}{x_2x_3}\right)^2 + 157.5 \cdot 10^6} - 1$$

$$g_7(\mathbf{x}) = \frac{x_2x_3}{40} - 1 \quad g_8(\mathbf{x}) = \frac{5x_2}{x_1} - 1$$

$$g_9(\mathbf{x}) = \frac{x_1}{12x_2} - 1 \quad g_{10}(\mathbf{x}) = \frac{1.5x_6 + 1.9}{x_4} - 1$$

$$g_{11}(\mathbf{x}) = \frac{1.1x_7 + 1.9}{x_5} - 1$$

A.3 Tension/Compression Spring Problem

Minimize the weight of a coil spring under a constant tension/compression load given by the wire diameter (x_1), the winding diameter (x_2) and the number of active coils (x_3), shown in Figure 12.

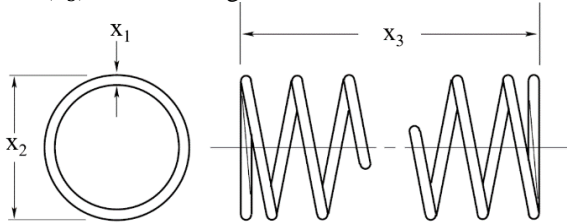


Figure 12: Schematic illustration of a tension/compression spring and its parameters. Image adapted from Bernardino et al. (2007).

The tension/compression spring problem is formally described as

$$f(\mathbf{x}) = (x_3 + 2)x_2x_1^2$$

where

$$\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$$

$$x_1 \in [0.05; 2] \quad x_2 \in [0.25; 1.3] \quad x_3 \in [2; 15]$$

subject to $g_j(\mathbf{x}) \leq 0$ with

$$g_1(\mathbf{x}) = 1 - \frac{x_2^3x_3}{71,785x_1^4}$$

$$g_2(\mathbf{x}) = \frac{4x_2^2 - x_1x_2}{12,566(x_2x_1^3 - x_1^4)} + \frac{1}{5,108x_1^2} - 1$$

$$g_3(\mathbf{x}) = 1 - \frac{140.45x_1}{x_2^2x_3} \quad g_4(\mathbf{x}) = \frac{x_1 + x_2}{1.5} - 1$$

A.4 Welded Beam Problem

Minimize the fabrication costs of a welded beam given by the thickness of the weld (x_1), the length of the welded

joint (x_2), the width of the beam (x_3) and the thickness of the beam (x_4), shown in Figure 13. The welded beam is subject to constraints on shear stress (τ), bending stress in the beam (σ), buckling load on the bar (P_c) and end deflection of the beam (δ).

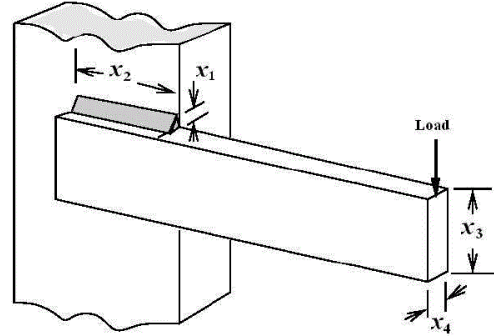


Figure 13: Schematic illustration of a welded beam and its parameters. Image adapted from Cagnina et al. (2008).

The welded beam problem is formally described as

$$f(\mathbf{x}) = 1.10471x_1^2x_2 + 0.04811x_3x_4(14 + x_2)$$

where

$$\mathbf{x} = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$$

$$x_1 \in [0.1; 2] \quad x_2 \in [0.1; 10] \quad x_3 \in [0.1; 10] \quad x_4 \in [0.1; 2]$$

subject to $g_j(\mathbf{x}) \leq 0$ with

$$g_1(\mathbf{x}) = \tau(\mathbf{x}) - \tau_{\max} \quad g_2(\mathbf{x}) = \sigma(\mathbf{x}) - \sigma_{\max}$$

$$g_3(\mathbf{x}) = x_1 - x_4 \quad g_4(\mathbf{x}) = 0.125 - x_1$$

$$g_5(\mathbf{x}) = \delta(\mathbf{x}) - \delta_{\max} \quad g_6(\mathbf{x}) = P - P_c(\mathbf{x})$$

and

$$\tau(\mathbf{x}) = \sqrt{\tau_1^2 + 2\tau_1\tau_2 \frac{x_2}{2R} + \tau_2^2} \quad \tau_1 = \frac{P}{\sqrt{2}x_1x_2} \quad \tau_2 = \frac{MR}{J}$$

$$M = P \left(L + \frac{x_2}{2} \right) \quad R = \sqrt{\frac{x_2^2}{4} + \left(\frac{x_1 + x_3}{2} \right)^2}$$

$$J = 2 \frac{x_1x_2}{\sqrt{2}} \left(\frac{x_2^2}{12} + \left(\frac{x_1 + x_3}{2} \right)^2 \right)$$

$$\sigma(\mathbf{x}) = \frac{6PL}{x_4x_3^2} \quad \delta(\mathbf{x}) = \frac{4PL^3}{Ex_3^3x_4}$$

$$P_c(\mathbf{x}) = \frac{4.013 \sqrt{\frac{EGx_3^2x_4^6}{36}}}{L^2} \left(1 - \frac{x_3}{2L} \sqrt{\frac{E}{4G}} \right)$$

$$\tau_{\max} = 13,600 \text{ psi} \quad \sigma_{\max} = 30,000 \text{ psi} \quad \delta_{\max} = 0.25 \text{ in}$$

$$P = 6,000 \text{ lb} \quad G = 12 \cdot 10^6 \text{ psi} \quad E = 30 \cdot 10^6 \text{ psi}$$

$$L = 14 \text{ in}$$

REFERENCES

- Akay B, Karaboga D (2010) Artificial bee colony algorithm for large-scale problems and engineering design optimization. *Journal of Intelligent Manufacturing* 23(4):1001–1014
- Akhtar S, Tai K, Ray T (2002) A socio-behavioural simulation model for engineering design optimization. *Engineering Optimization* 34(4):341–354
- Azad MAK, Fernandes EM (2011) Modified Differential Evolution Based on Global Competitive Ranking for Engineering Design Optimization Problems. In: *Computational Science and its Applications*, vol 6784, pp 245–260

- Belegundu AD (1982) Study of mathematical programming methods for structural optimization. PhD thesis
- Bernardino HS, Barbosa HJ, Lemonge AC (2007) A hybrid genetic algorithm for constrained optimization problems in mechanical engineering. In: IEEE Congress on Evolutionary Computation, pp 646–653
- Bernardino HS, Barbosa HJ, Lemonge AC, Fonseca LG (2008) A new hybrid AIS-GA for constrained optimization problems in mechanical engineering. In: IEEE Congress on Evolutionary Computation, pp 1455–1462
- Beyer H, Sendhoff B (2007) Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering* 196(33-34):3190–3218
- Cagnina LC, Esquivel SC, Coello Coello CA (2008) Solving engineering optimization problems with the simple constrained particle swarm optimizer. *Informatica* 32(3):319–326
- Coello Coello CA (2000) Use of a self-adaptive penalty approach for engineering optimization problems. *Computers in Industry* 41(2):113–127
- Coello Coello CA (2002) Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: A survey of the state of the art. *Computer Methods in Applied Mechanics and Engineering* 191(11-12):1245–1287
- Golinski J (1973) An adaptive optimization system applied to machine synthesis. *Mechanism and Machine Theory* 8(4):419–436
- Hansen N (2005) The CMA evolution strategy: A tutorial. *Vu le* 29
- Hansen N, Müller SD, Koumoutsakos P (2003) Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation* 11(1):1–18
- He Q, Wang L (2007) An effective co-evolutionary particle swarm optimization for constrained engineering design problems. *Engineering Applications of Artificial Intelligence* 20(1):89–99
- He S, Prempan E, Wu QH (2004) An improved particle swarm optimizer for mechanical design optimization problems. *Engineering Optimization* 36(5):585–605
- Hu X, Eberhart RC, Shi Y (2003) Engineering optimization with particle swarm. In: 2003 IEEE Swarm Intelligence Symposium, pp 53–57
- Kaveh A, Talatahari S (2009) Engineering optimization with hybrid particle swarm and ant colony optimization. *Asian journal of civil engineering* 10(6):611–628
- Kaveh A, Talatahari S (2010) An improved ant colony optimization for constrained engineering design problems. *Engineering Computations* 27(1):155–182
- Lemonge AC, Barbosa HJ, Borgesc CC, Silva FB (2010) Constrained optimization problems in mechanical engineering design using a real-coded steady-state genetic algorithm. *Mecánica Computacional* 3329(95):9287–9303
- Mezura Montes E, Coello Coello CA, Landa-Becerra R (2003) Engineering optimization using simple evolutionary algorithm. In: 15th IEEE International Conference on Tools with Artificial Intelligence, pp 149–156
- Mezura Montes E, Coello Coello CA, Velázquez-Reyes J, Muñoz-Dávila L (2007) Multiple trial vectors in differential evolution for engineering design. *Engineering Optimization* 39(5):567–589
- Ragsdell KM, Phillips DT (1976) Optimal Design of a Class of Welded Structures Using Geometric Programming. *Journal of Engineering for Industry* 98(3):1021
- Ray T, Liew KM (2003) Society and civilization: An optimization algorithm based on the simulation of social behavior. *IEEE Transactions on Evolutionary Computation* 7(4):386–396
- Rocha AMA, Fernandes EM (2009) Hybridizing the electromagnetism-like algorithm with descent search for solving engineering design problems. *International Journal of Computer Mathematics* 86(10-11):1932–1946
- Sandgren E (1988) Nonlinear integer and discrete programming in mechanical design. *Proceedings of the ASME design technology conference*
- Scheibenpflug A, Beham A, Kommenda M, Karder J, Wagner S, Affenzeller M (2015) Simplifying Problem Definitions in the HeuristicLab Optimization Environment. In: Companion Publication of the 2015 Genetic and Evolutionary Computation Conference, pp 1101–1108
- Schumer M, Steiglitz K (1968) Adaptive step size random search. *IEEE Trans. Automat. Contr.* 13(3):270–276
- Tomassetti G (2010) A cost-effective algorithm for the solution of engineering problems with particle swarm optimization. *Engineering Optimization* 42(5):471–495
- Townsend JC (2016) Golinski's Speed Reducer. <http://www.eng.buffalo.edu/Research/MODEL/mdo.test.orig/class2prob4/descr.html>
- Wagner S, Kronberger G, Beham A, Kommenda M, Scheibenpflug A, Pitzer E, Vonolfen S, Kofler M, Winkler S, Dorfer V, Affenzeller M (2014) Architecture and Design of the HeuristicLab Optimization Environment. In: *Advanced methods and applications in computational intelligence*, 1st edition, vol 6. Springer, New York, pp 197–261
- Zhang M, Luo W, Wang X (2008) Differential evolution with dynamic stochastic selection for constrained optimization. *Information Sciences* 178(15):3043–3074

GENERATION OF DISPATCHING RULES FOR JOB SEQUENCING IN SINGLE-MACHINE ENVIRONMENTS

Johannes Karder^(a), Andreas Scheibenpflug^(b), Stefan Wagner^(c), Michael Affenzeller^(d)

^{(a),(b),(c),(d)}Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria
^{(b),(d)}Institute for Formal Models and Verification
Johannes Kepler University, Altenbergerstraße 69, 4040 Linz, Austria

^(a)jkarder@heuristiclab.com, ^(b)ascheibe@heuristiclab.com, ^(c)swagner@heuristiclab.com, ^(d)maffenze@heuristiclab.com

ABSTRACT

A typical way to schedule a set of jobs is to evaluate and optimize different job sequences and then process them in the best found order. This global optimization approach can be applied for any set of jobs. Unfortunately, optimizing a subset of these jobs requires a new optimization run for this particular subsequence. In this paper we show the generation of dispatching rules that aid job sequencing in single machine environments using genetic programming and delta features. The rules are applied to sets of jobs and yield priorities depending on certain characteristics. These priorities are then used to create job orders dynamically depending on the last executed job. Once generated for specific scenarios, the rules provide on-the-fly sequence generation capability for queued subsets of jobs. Finally, we compare the performance and robustness of the generated rules against the scheduling approach.

Keywords: dispatching rules, sequence optimization, scheduling, single-machine, genetic algorithms, genetic programming

1. INTRODUCTION

We can categorize manufacturing shops into two groups: Those that have a small product catalog, but produce items in very high lot sizes and those that produce in smaller lot sizes and have a large product catalog. A long list of distinct products combined with small lot sizes can lead to volatile production scenarios, because manufacturers offering a wide variety of products might not produce all of them at the same time, but rather work on-demand and only have to produce a certain subset of their product catalog.

This paper focuses on low volume production scenarios where a single machine needs different tools to be set up to process a number of different jobs placed in a process queue. These jobs are a selection from a superset of jobs, namely the product catalog, based on existing orders. New orders and therefore additional jobs may be added to the process queue at any given time. The goal is to minimize production costs in terms of set-up time and

engage in sustainable production steering to increase long term competitiveness on the market.

To enable live scheduling for known and newly added jobs, an approach that employs dispatching rules (i.e. priority rules) is investigated. We use the concept of delta features to create rules which take job similarities into account. Whenever new jobs are added to the process queue, a predefined dispatching rule can be used to schedule the jobs. Each rule is created by means of genetic programming (Koza 1992) and specifically designed for a certain environment, i.e. a machine that is used to create products from a certain product catalog.

The rest of this paper is structured as follows: In Section 2 we provide insights into others' related work. Section 3 gives a brief definition of the addressed scheduling problem. Our methodology is explained in Section 4. Experiments, including algorithm and problem parameters, are shown in Section 5. The final Sections 6 and 7 show achieved results, as well as drawn conclusions and possible future research topics, respectively.

2. RELATED WORK

During the last centuries, scheduling problems have been widely discussed in literature. Ullman (1975) writes about NP-completeness of different scheduling problems, including the general scheduling problem and single execution time scheduling. Lenstra et al. (1977) classify scheduling problems on single, different and identical machines and experiment with various parameter to influence complexity. Both Davis (1985) and Van Laarhoven et al. (1992) use evolutionary concepts such as genetic algorithms or simulated annealing for problem solving. Cheng et al. (1999) created a tutorial survey of works on various hybrid approaches for job-shop scheduling practices using genetic methods. Dispatching rules itself have also been addressed numerous times, e.g. by Blackstone et al. (1982), Holthaus and Rajendran (1997) or Tay and Ho (2008). Beham et al. (2008) combines automated generation of dispatching rules and parallel simulation to solve scheduling problems. Comparisons between standard scheduling approaches using genetic algorithms

and the application of dispatching rules have been done by Kim et al. (2007), as well as Longo (2012). In both cases, it was concluded that scheduling performs better than dispatching. Previous work in the domain of dispatching rules for job scheduling was also done by Kofler et al. (2009), where sequence-dependent set-up costs were minimized using both scheduling and dispatching concepts. The authors suggested multiple simple priority rules and then used genetic programming to synthesize more complex rules which use different features from the specified jobs. The authors of this paper (Karder et al. 2015) also created a simulator for a specific kind of scheduling problem from scratch by using Sim# (Beham et al. 2014) with custom extensions. Experience gained from that work is used within this paper.

3. PROBLEM DESCRIPTION

The scheduling problem consists of a number of jobs that have to be scheduled on a single machine. Each job has a certain demand for tools that need to be set up inside the machine before the respective job can be processed by an operator. After the job has been processed, all tools are removed from the machine. A tool storage allows the machine to automatically load and unload most of the required tools without operator interaction. Since different jobs require different sets of tools and the storage capacity is limited, the storage layout can only be configured to support a certain number of jobs. Sets of jobs that can be executed with a particular storage configuration are referred to as job batches. When switching from one job batch to another, the operator has to manually alter the storage configuration to contain all required tools for the next batch. The time consumption of all manual tool changes is also included in the set-up time evaluation. The problem can be evaluated by using three different quality criteria:

1. Set-up Time (ST)
2. Batch Count (BC)
3. Manual Tool Change Count (MTCC)

Changing the quality criterion yields different quality values. ST and MTCC are highly correlating, meaning that lower set-up times lead to less manual tool changes and vice versa, whereas a lower set-up time usually also leads to a greater batch count. When a job sequence needs to be evaluated, a new storage layout is generated in the following way: All referenced tools (starting with job 1) are added to the storage as long as possible. If tools from a job cannot be added anymore, this job is used to start a new job batch and the storage will be altered to contain all tools for the new job batch. If replacing tools is not possible, e.g. because of storage constraints, all tools will be removed before the new ones are added.

4. METHODOLOGY

We use HeuristicLab (HL) (Wagner et al. 2014), a paradigm-independent and extensible environment for heuristic optimization, as the underlying software foundation. It features ready-to-use implementations of

many different evolutionary algorithms and problem types, including special variants of genetic algorithms and basic problem implementations that can easily be adapted to specific needs. HeuristicLab's plug-in concept allows to easily extend its functionality by implementing new plug-ins based on existing framework elements. Taking advantage of the plug-in system, we introduce a new problem type which employs a simulation model for evaluating the performance of the dispatching rules and uses its own solution encoding and respective algorithm operators to make it compatible with population-based evolutionary algorithms. To evaluate scheduling solutions, we employ a discrete-event simulation that is built upon the Sim# library. It is able to simulate every movement of the machine's main mechanical parts, which allows us to accurately compute the set-up times of the real machine.

Rule generation is done using genetic programming (GP), a concept introduced by John Koza in 1992. GP, which is included in the list of default HL plug-ins, is a method where objects in form of trees are built using evolutionary concepts such as selection, crossover and mutation. In our case, a tree represents a dispatching rule, i.e. a mathematical formula, in form of a syntax tree. This formula consists of different mathematical operations and operands, including variables and constants. An example of a simple tree is depicted in Figure 1.

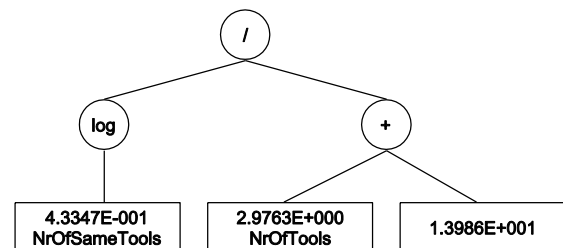


Figure 1: A random GP tree.

The size of a tree is traditionally limited by a maximum length and depth. The tree length is the total number of nodes and the depth is the number of edges from a node to the root node. The tree shown in Figure 1 has a length of 6 and a maximum depth of 2.

To evaluate formulas containing variables, it is necessary to supply a dataset which defines values for these variables. These values are also called features and a set of features supplied to a formula is called row or sample. Trees are therefore evaluated for rows in datasets. Our dataset contains one row for each job from the pool. Some features can be calculated immediately for each job. All other features – we call them delta features – must be calculated on-the-fly, because their values are state-dependent. All delta features are in a way similarity measures, which allow the rule to prioritize according to the current state. Each rule is applied m times to a random subset of n jobs from the pool. Out of this subset, one job is selected to be the first job processed by the machine. For all remaining (i.e. unprocessed) jobs of the subset, the delta features to the currently processed job

are calculated. Examples of such features would be the number of same tools, the number of same tool types or the number of same tool kinds the two jobs share. With these additional features, the rule can be used to calculate priorities for the remaining jobs. The job with the highest priority is selected to be processed next, the current job is updated and another iteration is executed until all jobs have been processed. The qualities of the resulting job sequence are calculated according to the selected quality criterion and summed up. The average quality – which should be minimized – makes up the quality of the rule. Pseudocode for this procedure can be seen in Listing 1.

```

r := rule
r_q := 0.0 // rule's summed quality
P := product catalog // all available jobs
repeat m times
  shuffle P
  J := first n elements from P // orders
  S := empty list of jobs
  c := first element from J // current job
  S.add(c)
  repeat n - 1 times
    R := J \ S
    calc. delta features from jobs in R to c
    b := best job from R according to r
    S.add(b)
    c := b
  endrepeat
  q := quality criterion for job sequence def. by S
  r_q := r_q + q
endrepeat
return r_q / m // rule's average quality

```

Listing 1: Pseudocode describing the evaluation of a single dispatching rule.

Consequently, the dispatching rule is trained to work for a certain product catalog. Subsets are used to simulate a daily or weekly amount of orders. It is important not to train with the whole catalog of products, since there will most likely never be an order containing every product and overfitting to the complete catalog itself should be avoided. Instead, more diverse batches of jobs should be used to mix different smaller orders and create a more general rule for the specific scenario.

5. EXPERIMENTS

To construct an exemplary production environment, we create a product catalog that contains 34 different jobs. These jobs are used to generate a dispatching rule. The evaluation parameters are set as follows:

$$m = 5 \text{ (sample size)}$$

$$n = 18 \text{ (group size)}$$

Additionally, 4 random subsets of jobs containing 10, 15, 20 and 25 jobs are created to reflect daily or weekly orders. The jobs are then scheduled by applying the best found rule. Afterwards, the same subsets are optimized with a standard scheduling approach using a combinatorial optimization problem, which is optimized by a genetic algorithm with offspring selection (OSGA) (Affenzeller and Wagner 2005). We then compare the best found job sequence with the sequence suggested by

the rule. All problem instances for the standard scheduling approach and the dispatching approach have been configured as shown in Table 1.

Table 1: Scheduling and dispatching problem parameters.

Parameter	Value
Scheduling	
Optimization Target	Job Order
Quality Criterion	Set-up Time
Dispatching	
Allowed GP Symbols	Add, Sub, Mul, And, Or, Not, Xor, GreaterThan, LessThan, IfThenElse, Variable, Constant
Max. Tree Depth	17
Max. Tree Length	100
Sample Size	5
Group Size	18
Quality Criterion	Set-up Time

The OSGA configurations were similar for the both approaches. Multi-operators were used to induce different operator behaviors into the corresponding optimization procedure. Constant comparison factors were set to work with the concept of weak offspring selection, which means that children are included in the next population if they outperform their weaker parent. For the dispatching approach, the mutation rate was increased by 5 % and gender-specific selection was replaced with proportional selection. A list of algorithm parameters is shown in Table 2.

Table 2: OSGA parameters for the scheduling (1) and dispatching (2) approach.

Parameter	Value
Lower Comp. Factor	0
Comp. Factor Modifier	–
Crossover	Multiple ⁽¹⁾ SubtreeSwapping ⁽²⁾
Elites	1
Max. Generations	1000
Max. Sel. Pressure	100
Mutation Probability	10 % ⁽¹⁾ 15 % ⁽²⁾
Mutator	Multiple
OS Before Mutation	False
Population Size	100
Selector	Gender-specific ⁽¹⁾ Proportional ⁽²⁾
Success Ratio	1

All experiments have been conducted on a distributed optimization cluster (HeuristicLab Hive) with 10 repetitions to produce valid empirical test results.

6. RESULTS

Table 3 shows a direct comparison between the scheduling and dispatching approaches. The scheduling

results are the set-up times of the best found job sequences. The dispatching results are the set-up times of the sequences that are generated by the best dispatching rules found for each job set.

Table 3: Test result comparison between scheduling and dispatching approach.

Job Set	Set-up Time [min]	
	Scheduling	Dispatching
js1 (10 jobs)	38.15	39.35 (+3.15 %)
js2 (15 jobs)	55.45	59.14 (+6.91 %)
js3 (20 jobs)	56.03	62.59 (+12.07 %)
js4 (25 jobs)	92.73	103.42 (+11.74 %)

As already described in previous applications (Kim et al. 2007, Longo 2012), the experiments show that the standard scheduling approach achieves better results than the dispatching rules.

By analyzing the results of the executed runs, it appears that the dispatching approach needs more adjustments to produce more robust rules. In particular, we have a look at the 10 best rules that were generated within the 10 repetitions, i.e. each best rule of each run. Figure 2 shows a box-plot of these rules' achieved set-up times, applied to the 4 different job sets. Table 4 lists more detailed measures.

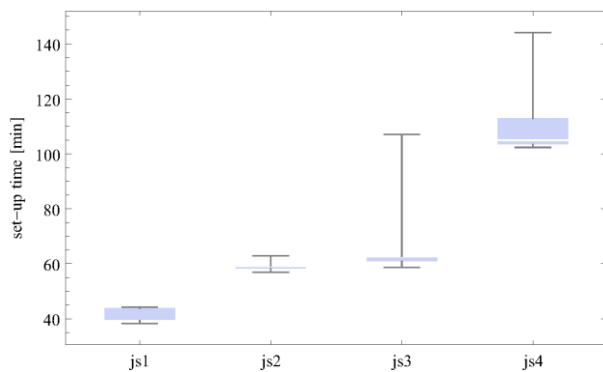


Figure 2: Box plots of set-up times.

Table 4: Statistical overview of the achieved set-up times for each job set.

Measure	Set-up Time [min]			
	js1	js2	js3	js4
Min	38.35	56.99	58.69	102.34
Max	44.26	62.91	107.06	144.12
Med	40.11	59.12	62.69	105.96
Avg	41.22	59.14	66.29	113.31
SD	2.50	1.59	14.41	16.51
Var	6.23	2.52	207.51	272.54
Q1	39.35	58.37	60.86	103.55
Q3	43.89	59.24	62.81	111.71

The rules are not totally stable, as the resulting set-up times have a standard deviation between approximately 1.5 and 16.5 minutes, depending on the job set.

The average runtime of the standard scheduling runs was around 3.5 hours with a minimum of 1.5 hours and a

maximum of 4.5 hours. Approximately the same time was consumed on average by the runs of dispatching approach. The execution time of an OSGA run can vary because normally the number of generations is not directly limited by the algorithm parameters itself, but indirectly through the performance of the algorithm during offspring creation.

7. CONCLUSION

We were able to generate rules that enable live scheduling and therefore speed up the scheduling process itself. By applying a predefined rule, respective schedules can be calculated immediately, whereas the standard scheduling approach must be executed separately for each set of jobs. It is also possible to generate multiple rules for a specific environment, which can then be applied to a subset of jobs to generate alternative sequences.

Looking at the results, it can be seen that the dispatching approach cannot reach the quality levels of the standard scheduling approach. This is expected and can be explained by two facts. First of all, the scheduling approach is purely designed to evaluate many different sequences of the respective jobs, which in the end leads to good scheduling qualities. The dispatching approach is designed to evaluate many different dispatching rules. Two distinct rules can however yield the same job sequence. To get n different job sequences, at least n different rules have to be generated. Secondly, the scheduling approach's goal is to yield the best sequence for a specific subset of jobs, whereas the dispatching approach's goal is to yield a general rule that is able to create a good schedule for any arbitrary subset of jobs from the product catalog. Additionally, the dispatching approach should be tuned to produce more robust rules. This could be done by e.g. changing the allowed GP symbols or identifying more significant features. Both approaches consume a similar amount of runtime when executed. The runtime-limiting factor is the OSGA's ability to produce better offspring.

Further research includes the generation of multiple dispatching rules for a single product catalog and the creation of a rule portfolio. This portfolio would add an additional layer of optimization in which either multiple rules are applied to create different job sequences to choose from, or additional techniques for on-the-fly rule selection could be implemented.

ACKNOWLEDGMENTS

The work described in this paper was conducted within the NPS (Sustainable Production Steering) project and funded by the Austrian Research Promotion Agency (FFG).

REFERENCES

Affenzeller M., Wagner S., 2005. Offspring Selection: A New Self-Adaptive Selection Scheme for Genetic Algorithms. Proceedings of the 7th International Conference on Adaptive and Natural Computing

- Algorithms, pp.218-221. 21-23 March 2005, Coimbra, Portugal
- Beham A., Winkler S., Wagner S., Affenzeller M., 2008. A Genetic Programming Approach to Solve Scheduling Problems with Parallel Simulation. Proceedings of the 22nd IEEE International Symposium on Parallel and Distributed Processing, pp.1-5. 14-18 April 2008, Miami, USA
- Blackstone J.H., Phillips D.T., Hogg G.L., 1982. A state-of-the-art survey of dispatching rules for manufacturing job shop operations. *International Journal of Production Research*, 20(1):pp.27-45
- Cheng R., Gen M., Tsujimura Y., 1999. A tutorial survey of job-shop scheduling problems using genetic algorithms, part II: hybrid genetic search strategies. *Computers & Industrial Engineering*, 36(2):pp.343-364
- Davis L., 1985, July. Job Shop Scheduling with Genetic Algorithms. Proceedings of the First International Conference on Genetic Algorithms and Their Applications, pp.136-140. 24-26 July 1985, Pittsburgh, USA
- Holthaus O., Rajendran C., 1997. Efficient dispatching rules for scheduling in a job shop. *International Journal of Production Economics*, 48(1):pp.87-105
- Kim I., Watada J., Shigaki I., 2008. A comparison of dispatching rules and genetic algorithms for job shop schedules of standard hydraulic cylinders. *Soft Computing*, 12(2):pp.121-128
- Kofler M., Wagner S., Beham A., Kronberger G., Affenzeller M., 2009. Priority Rule Generation with a Genetic Algorithm to Minimize Sequence Dependent Setup Costs. In: Moreno-Díaz R., Pichler F., Quesada-Arencibia A., eds. *Computer Aided Systems Theory - EUROCAST 2009*. Heidelberg, Germany:Springer, pp.817-824
- Koza J.R., 1992. Genetic programming: On the Programming of Computers by Means of Natural Selection. Cambridge, USA: The MIT Press
- Lenstra J.K., Kan A.R., Brucker P., 1977. Complexity of Machine Scheduling Problems. *Annals of Discrete Mathematics*, 1:pp.343-362
- Longo F., 2012. On the short period production planning in industrial plants: a real case study. *International Journal of Simulation and Process Modelling*, 8(1):pp.17-28
- Tay J.C., Ho N.B., 2008. Evolving dispatching rules using genetic programming for solving multi-objective flexible job-shop problems. *Computers & Industrial Engineering*, 54(3):pp.453-473
- Ullman J.D., 1975. NP-Complete Scheduling Problems. *Journal of Computer and System Sciences*, 10(3):pp.384-393
- Van Laarhoven P.J., Aarts E.H., Lenstra, J.K., 1992. Job Shop Scheduling by Simulated Annealing. *Operations Research*, 40(1):pp.113-125
- Wagner S., Kronberger G., Beham A., Kommenda M., Scheibenpflug A., Pitzer E., Vonolfen S., Kofler M., Winkler S., Dorfer V., Affenzeller M., 2014. Architecture and Design of the HeuristicLab Optimization Environment. In: Klempous R., Nikodem, J., Jacak W., Chaczko Z., eds. *Advanced Methods and Applications in Computational Intelligence*. Cham, Switzerland:Springer, pp.197-261

AUTHORS BIOGRAPHIES

JOHANNES KARDER received his Master in software engineering in 2014 from the University of Applied Sciences Upper Austria and is a research associate at the Research Center Hagenberg. His research interests include algorithm theory and development as well as production planning and logistics optimization. He is a member of the HeuristicLab architects team.

ANDREAS SCHEIBENPFLUG received his Master in software engineering in 2011 from the University of Applied Sciences Upper Austria and is a research associate at the Research Center Hagenberg. His research interests include parallel and distributed computing. He is a member of the HeuristicLab architects team.

STEFAN WAGNER received his PhD in technical sciences in 2009 from the Johannes Kepler University Linz, Austria. He is a professor at the University of Applied Sciences Upper Austria, Hagenberg Campus. He is the project manager and head developer of the HeuristicLab optimization environment.

MICHAEL AFFENZELLER has published several papers, journal articles and books dealing with theoretical and practical aspects of evolutionary computation, genetic algorithms, and meta-heuristics in general. In 2001 he received his PhD in engineering sciences and in 2004 he received his habilitation in applied systems engineering, both from the Johannes Kepler University of Linz, Austria. Michael Affenzeller is professor at University of Applied Sciences Upper Austria, Hagenberg Campus.

SIMULATING THE ASSORTING PLANNING PROCESS TO SUPPORT THE DECISION-MAKING OF GROCERY WHOLESALERS: THE MEXICAN MARKET CASE.

Salutino Flores Monroy^(a)

^(a) UNAM, Universidad Nacional Autónoma de México, Facultad de Ingeniería, División de Ingeniería Mecánica e Industrial, Departamento de Ingeniería de Sistemas

^(a)salutino.flores@gmail.com

ABSTRACT

Having the best assortment in a retail store is a complex task for the category manager in which their main job is to maximize the overall performance of each of the products cataloged, maintaining a balance between suppliers, the company and customers. The accommodations made on the shelf are made by assortment planning models in which only a single category is used to carry out this choice. However, they are overwhelmed when they have to evaluate between categories, such as complementary products. It has developed assortment planning model with dynamic programming in which based on Monte Carlo simulation using @RISK software to support the decision-making of a Mexican grocery wholesaler. Our simulation results demonstrate that increasing the scope of faces to analyze, we can give the decision maker a better picture of the assortment the store should display.

Keywords: Dynamic programming, Assortment planning, Montecarlo simulation, Retailing

1. INTRODUCTION

Nowadays, grocery wholesale in Mexico trade finds its development conditioned by a range of pressures; customers increasingly orient their preferences looking forward greater value propositions. Among the proposals, the price is a fundamental component in the evaluation of a proposed retail, the comfort in terms of proximity, quality and variety of the products, customer service, and categories selected, as well as offering services that might result in saving time.

On average we buy the same 150 products, which represent 85% of the needs at home. (Schneider & Julie, 2011) The rest are new products, where only 3% out of the other 15% missing are products that exceed sales in more than 50 million dollars. One of the great challenges trade faces is on choosing which products to exhibit, since you often ask yourself, which product do I take away in order to exhibit the new one?

How we assign products in a finite and limited space on shelves is a critical operational decision that all retailers face. This decision is directly related to the profitability of the organization as it affects operating costs and

revenues. It affects costs in the list of products we have, including transportation, purchase orders, inventory maintenance, reshuffling the product, and the possibility of falling into unavailable status because of not having quantity demanded by customers.

Businesses need to meet this consumer demand by balancing the installed capacity (furniture, shelf) depending on the variety of choices (product number) and service levels on the shelf (number of units of a product). As a central result of strategic decisions the trader should take into consideration four important aspects:

- *Listing*: Size range. What products do I select in order to satisfy customer demands?
- *Facing*: Management of space on the shelf. Is the space finite on a shelf, could it be increased? Should we remove products? How many fronts do I give to each product?
- *Replenishing*: The refill. How logistics is generated to restock the product?
- *Pricing*: The price. what price should the product selected have?

Offering a wider range can generate that appropriate service levels might not occur, and vice versa, as well as the shelf space is limited. Traders and manufacturers try to meet consumers' satisfaction with the right merchandise in the right store at the right time. (Hübner, 2011).

This task is carried out by the category manager. *Category management* is a merchandising technique that some firms – including several supermarkets, drugstores, hardware stores and general merchandise retailers – use to improve productivity. It is a way to manage a retail business that focuses on the performance of product category results rather than individual brands. It arranges product groupings into strategic business units to better meet consumer needs and to achieve sales and profit goals. Retail managers make merchandising decisions that maximize the total return of the assets assigned to them. (Bhatia, 2008). In other words, category management is a way of organizing retail product management so that suppliers, central operations and

outlet level activities are integrated into the process, in which is geared to responding effectively to consumer demand.

The category management analysis focused on the use of assets (inventories, surface and eventually employees) and results (margin) allow you to focus categories as strategic business units. Knowing the behavior of each product per store provides a more accurate building plans for each category, on which the chain suppliers and agree to coordinate their efforts basis. The margin analysis versus sales volume will allow distinguishing products according to their contribution within the category. We are not interested only on those who leave higher margin, but also those that sell well.

Retailers typically solve decision problems sequentially: determining the assortment, allocating it to the shelf and finally determining order sizes. However, there are products and categories that only have to sell if there is another who complements, for example, milk, in a study conducted in the UK it was found that only 12% of consumption is done alone, consumption is accompanied generally. (Centre for European Agricultural Studies & Institute for the Management of Dairy Companies, 1999) Therefore, when performing assortments that are carried out not fail to see the chain of influence they have and how they can benefit the store sale.

The objective of the category manager is to organize on a wiser way product assortment as well as the marketing plans to generate better profit contribution of limited display space, especially with the increase of new products.

The category manager must handle the assortment of the categories of the stores according to the following diagram:

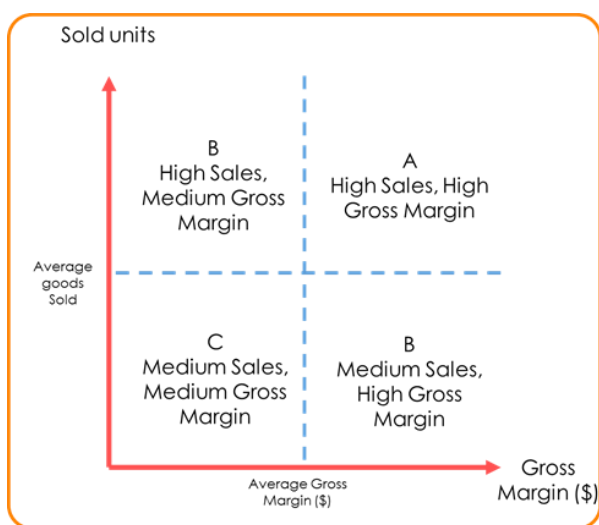


Figure 1 Classifying SKUs (Stock Keeping Units) of scorecard categories.

In Figure 1 according to (Ring & Tigert, 2002) we find that most of the catalog of a category is in the C sector, recent business deteriorates very significantly, though

subtle. A low contribution they add a cost of maintaining inventories, but above all, tie up capital or bargaining power with suppliers. This bargaining power could very well apply to products that contribute faster, allowing greater capital turnover, rather than in a low margin business is extremely significant. Most of the complementary products within this category are the first products to withdrawal the assortment the retail has to offer.

A key to achieving the desired performance in the arrangement of the shelf appearance depends on the category manager having access to an efficient decision-making system that allows managing the physical space. Traditional tool for managing space in stores is the planogram. However, software applications available on the market made the arrangement of the product by simple rules of thumb for allocation based on sales of their products.

Among the existing models in literature regarding managing shelf space, one of the first studies was done by (Hansen & Heinsbroek, 1979). In his model, the demand for each product is a function of the elasticity of space. Its model seeks forward maximizing profits for retailers liable to a limit of available space on the upper and lower quantity of each product limits, as well as in the whole of the useful fronts values. In addition, (Zufryden, 1986) presented a dynamic programming formulation for a problem with the elasticity of space and marketing variables related to demand, including the price. The article written by (Yang & Chen, A study on shelf space allocation and management, 1999) assumed a linear utility derived from the fronts of the products. They formulate a problem of accommodation space with a horizontal and vertical effect of accommodation. (Yang, 2001) proposed a heuristic model of the backpack. He found an optimal solution only in very simple versions. (Lim, Rodriguez, & Zhang, 2004) based on Yang's proposal, they work for a goal - heuristics. A hierarchical Bayes model is proposed by (van Nierop, Fork, & Franses, 2006) to estimate the interaction between the arrangement of the shelf, marketing activities and stochastic demand. They did it by working with simulated annealing. (Murray, CC; Talukdar, D; Gosavi, A, 2010) worked together to create a decision making model based on which products must be accommodated as well as the price for each item by using MINLP - solver. (Hansen, Raut, & Swami 2010) investigated a goal decision model- heuristic with which the fronts of the products are proportional to the demand generated by them.

A dynamic programming approach is proposed to select optimally between a given set of products and allocate shelving units in the whole space for the selected products on the shelves of the Mexican cash and carry wholesaler. The approach is designed to take into consideration the general shelf specifications such as the size, as well as revenues from the sale of such products.

Before going into details contained in the article, a model for the optimization and simulating parameters scenarios is described in section 2. The results of simulation are shown in section 3, and finally, in section 4 the contribution and conclusions of it are presented.

2. THE MODEL FORMULATION

We now turn to a discussion of the optimization model components and its formulation as a mathematical program. Model is based on the knapsack model, since it looks forward exhibiting products in quantity and in a way in which we can maximize revenue. The knapsack problem is a classic problem of linear integer programming.

Maximize

$$z = \sum_{i=1}^n p_i x_{ij} \quad \dots\dots\dots (1)$$

Subject to.

$$\sum_{i=1}^m w_i x_{ij} \leq W_j \quad j = 1, \dots, m \quad \dots (2)$$

$$x_{ij} \geq 0 \quad i = 1, 2, \dots, n \quad \dots (3)$$

And integer

$$j = 1, \dots, m$$

Where p_i the profit earns by each SKU i ($i = 1, 2, \dots, n$), w_i the space required in centimeters by each SKU i and W_j the total length available at the shelf are integers. Where $j = 1, \dots, m$ is the store selected. In other words, suppose you have to fill a shelf with different SKU with a revenue p_i and distance in w_i without exceeding total space given W . The problem is to find the maximum feasible allocation of items for the total value of the products on the shelf. Each item that can fit on the shelf has a size and an associated benefit. The shelf has a limited capacity.

2.1. Dynamic Programming Formulation

Dynamic programming appears to provide an effective technique to provide solution that solves the problems of accommodation in the shelf space. (Flores de la Mota, 2015) Among its advantages, we have, that it can easily handle variables for exchange, for example changing branch database to verify the maximum space you have to display products, and the best of all, is that you can handle integer solutions, you cannot divide a product in tenths of a party, in enters completely or does not enter in the shelf.

First, we consider the general problem of N items. If k_j is the number of units of an item i , the problem becomes:

Maximize

$$v_1k_1 + v_2k_2 + \dots + v_Nk_N \quad (4)$$

Subject to

$$w_1k_1 + w_2k_2 + \dots + w_Nk_N \text{ less than or equal to } W$$

$$k_1 \text{ integer not negative} \quad (5)$$

The dynamic programming model is constructed considering the following three basic elements:

1. Stage j is represented by i tem $j, j = 1, 2, \dots, N$
2. State y_j , in the stage j , is the total space assigned to stage $j, j + 1, \dots, N; y_1 = W$ y $y_j = 0, 1, \dots, W$ para $j = 2, 3, \dots, N$.
3. Alternative k_j , in stage j , is the number of unit of each item j . The value of k_j , could be as little as zero, or as big as $[W/w_i]$, where this ratio is the smallest integer maximum obtained from $[W/w_i]$.

The stages represent the items, then, we have three steps $j = 1, 2, \dots, N$. The state of step j , represents the total space of the j items, plus all the items that will be accommodated later on the shelf. The decision in step j is: ¿How many products j must be accommodated in that shelf? The answer is k_j .

Then you have the following recursive formulas:

Is $f_i(y_i)$ = optimal value of stages $j, j + 1, \dots, N$ given the state y_j .

Recursive equations call are:

$$f_N(y_N) = \max_{k_N=0,1,\dots,\lfloor y_N/w_N \rfloor} \{v_N k_N\} \quad (7)$$

$$f_j(y_j) = \max_{k_j=0,1,\dots,\lfloor y_j/w_j \rfloor} \{v_j k_j + f_{j+1}(y_j - w_j k_j)\} \quad (8)$$

2.2 Montecarlo Simulation

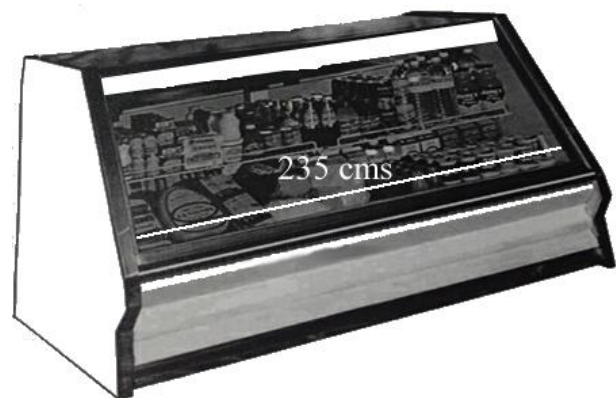


Figure 2 Selected refrigerator in which the maximum distance is 235 centimeters.

In this context, Monte Carlo simulation could be considered as a methodical way to perform the what-if

analysis, where we seek how it might behave if $w_i=1, 2, 5$ maximum fronts, how many different items we would have, and how items of classification A, B and C will behave.

Previous data would be our output variables. And as random input variables, have been based on demand information provided by the wholesaler for choosing the best choice for the main refrigerator with $W = 235$ cm in which selection within the 179 cataloged products in which w_i of each item in centimeters is known, given the average standard deviation of demand and sales of the last 12 months in which they behave on a normal way. To run the simulation @RISK software was used for performing the study.

3. RESULTS

The simulation model range has been used to understand the behavior of few different items could have our exhibit, and understand how our profitability would be affected with the choice given of 1000 range iterations.

Table 1 Risk Output Report for $W = 235$ & maximum faces by SKU = 1

Name	Graphic	Min	Mean	Máx	5%	95%	Err
Profit-earning capacity		\$135,267	\$165,275	\$197,410	\$149,085	\$181,549	\$0
Total of SKU faced		13	14.364	20	13	16	0
Total of SKU "A" faced		7	9.504	14	8	11	0
Total of SKU "B" faced		1	4.073	7	3	6	0
Total of SKU "C" faced		0	0.787	2	0	1	0

Table 2 Risk Output Report for $W = 235$ & maximum faces by SKU = 1

Name	Graphic	Min	Mean	Máx	5%	95%	Err
Profit-earning capacity		\$186,931	\$236,343	\$286,353	\$206,585	\$265,885	\$0
Total of SKU faced		7	8.6	15	7	12	0
Total of SKU "A" faced		5	6.907	10	6	9	0
Total of SKU "B" faced		0	1.557	7	0	3	0
Total of SKU "C" faced		0	0.136	2	0	1	0

Table 3 Risk Output Report for $W = 235$ & maximum faces by SKU = 2

Name	Graphic	Min	Mean	Máx	5%	95%	Err
Profit-earning capacity		\$232,177	\$316,700	\$410,037	\$266,018	\$371,217	\$0
Total of SKU faced		3	3.124	8	3	4	0
Total of SKU "A" faced		2	3.103	7	3	4	0
Total of SKU "B" faced		0	0.021	2	0	0	0
Total of SKU "C" faced		0	0	0	0	0	0

The results obtained using the simulation model range in @RISK allows us to see that, the greater the range of products, the utility may be less when having at least one front displayed, as in Table 1 in which an average of 14 SKU exhibited, we are located in the middle of the presumable utility buildable in which at least we would have nine SKUs "A" , 3 SKU "B" and at least 1 SKU "

C" displayed, and a greater utility when we have 5 possible fronts of articles, as in Table 3 in which in average 3 SKU displayed are located in the middle of probable utility buildable in which at least we would have three SKU "a" , at least 1 SKU "B" and already It not is necessary to display a SKU "C" .

3.1. Validation

It has been carried out the validation of the simulation performed taking into account the latest sales period registered to learn how has deviated from the average in which it was found that:

Table 4 Risk Output Report for $W = 235$ & maximum faces by SKU = 1 with a fixed Demand

Name	Graphic	Min	Mean	Máx	5%	95%	Err
Profit-earning capacity		\$137,405	\$165,201	\$194,827	\$149,535	\$180,321	\$0
Total of SKU faced		13	14.319	21	13	16	0
Total of SKU "A" faced		6	9.486	14	8	11	0
Total of SKU "B" faced		0	4.053	8	2	6	0
Total of SKU "C" faced		0	0.78	2	0	1	0

Table 5 Risk Output Report for $W = 235$ & maximum faces by SKU = 1 with a fixed Demand

Name	Graphic	Min	Mean	Máx	5%	95%	Err
Profit-earning capacity		\$187,704	\$236,075	\$302,852	\$205,560	\$263,878	\$0
Total of SKU faced		7	8.587	24	7	12	0
Total of SKU "A" faced		5	6.903	16	6	9	0
Total of SKU "B" faced		0	1.566	7	0	3	0
Total of SKU "C" faced		0	0.118	2	0	1	0

Table 6 Risk Output Report for $W = 235$ & maximum faces by SKU = 2 with a fixed Demand

Name	Graphic	Min	Mean	Máx	5%	95%	Err
Profit-earning capacity		\$228,338	\$316,460	\$446,847	\$265,619	\$371,299	\$0
Total of SKU faced		3	3.111	7	3	4	0
Total of SKU "A" faced		2	3.099	6	3	4	0
Total of SKU "B" faced		0	0.012	1	0	0	0
Total of SKU "C" faced		0	0	0	0	0	0

3.1.1

Table 7 Err between Table 1 and Table 4

Name	% Diff Min	% Diff Mean	% Diff Máx	% Diff 0.05	% Diff 0.95	% Diff Err
Profit-earning capacity	1.56%	-0.04%	-1.33%	0.30%	-0.68%	-
Total of SKU faced	0.00%	-0.31%	4.76%	0.00%	0.00%	-
Total of SKU "A" faced	-16.67%	-0.19%	0.00%	0.00%	0.00%	-
Total of SKU "B" faced	-	-0.49%	12.50%	-50.00%	0.00%	-
Total of SKU "C" faced	-	-0.90%	0.00%	-	0.00%	-

Table 8 Err between Table 2 and Table 5

Name	% Diff Min	% Diff Mean	% Diff Máx	% Diff 0.05	% Diff 0.95	% Diff Err
Profit-earning capacity	0.41%	-0.11%	5.45%	-0.50%	-0.76%	-
Total of SKU faced	0.00%	-0.15%	37.50%	0.00%	0.00%	-
Total of SKU "A" faced	0.00%	-0.06%	37.50%	0.00%	0.00%	-
Total of SKU "B" faced	-	0.57%	0.00%	-	0.00%	-
Total of SKU "C" faced	-	-15.25%	0.00%	-	0.00%	-

Table 9 Err between Table 3 and Table 6

Name	% Diff Min	% Diff Mean	% Diff Máx	% Diff 0.05	% Diff 0.95	% Diff Err
Profit-earning capacity	-1.68%	-0.08%	8.24%	-0.15%	0.02%	-
Total of SKU faced	0.00%	-0.42%	-14.29%	0.00%	0.00%	-
Total of SKU "A" faced	0.00%	-0.13%	-16.67%	0.00%	0.00%	-
Total of SKU "B" faced	-	-75.00%	-100.00%	-	-	-
Total of SKU "C" faced	-	-	-	-	-	-

This has already generated an important aspect in which if we reduce the catalog it could be beneficial to the wholesaler because having lower catalog, management is easier to handle, however, for customers is not necessarily the best. Among the leading causes in which a consumer goes to the point of sale, would be the catalog that the wholesaler can offer.

w = 1 w = 2 w = 5

% of SKU "A" faced	66%	80%	99%
% of SKU "B" faced	28%	18%	1%
% of SKU "C" faced	5%	2%	0%

Table 10 Percentage of SKU faced by the 3 different scenarios

This work allows the wholesaler to recognize how to balance its SKU catalog, in which if it exhibits just one front for each item, distribution tends to seem like a Pareto distribution where with the 66% of the items "A", we will be able to accommodate them maximizing out profitability with the assortment given.

4. CONCLUSIONS AND FUTURE WORK

Category management is among the biggest challenges that a commercial chain might have, in which we depend on the strategic decision of the company to choose the best assortment for each store or branch. Often, as in the case studied, we have about 180 SKU available for displaying. We have that amount of products because every area that participates in the chain has their own tastes, but thanks to the study presented, we can realize we should maintain assorting open since, as it was mentioned in the introduction, we should take into consideration the assortment to offer, the service level we

have, as well as the resupply (replenishment) and the prices we have to maximize profitability.

Dynamic programming models allow them to be flexible in placing restrictions on future work that would be necessary to try different industries to the one studied in this work as the fashion industry.

ACKNOWLEDGMENTS

I thank Osvaldo Gonzalez Molina for assistance with Dynamic Programming Model, and Hikuri Abril Isarraraz Yllescas for comments that greatly improved the manuscript.

We would also like to show our gratitude to the (Name Surname, title, institution) for sharing their pearls of wisdom with us during the course of this research, and we thank 3 "anonymous" reviewers for their so-called insights. I am also immensely grateful to Dra. Idalia Flores de la Mota & Dra. Aida Huerta Barritenos for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

REFERENCES

- Bhatia, S. (2008). Retail Management. New Delhi, India: Atlantic Publishers & Dist.
- Centre for European Agricultural Studies, & Institute for the Management of Dairy Companies. (1999). EVALUATION OF THE SCHOOL MILK. Munich: DGVI European Commission.
- Flores de la Mota, I. (2015). Programación dinámica. México, D. F.: Universidad Nacional Autónoma de México, Facultad de Ingeniería, División de Ingeniería Mecánica e Industrial, Departamento de Ingeniería de Sistemas.
- Hansen, J., Raut, S., & Swami, S. (2010). Retail shelf allocation: A comparative analysis of heuristic and meta-heuristic approaches. Journal of Retailing, 86(1), 94-105.
- Hansen, P., & Heinsbroek, H. (1979). Product selection and space allocation in supermarkets. European journal of operational research, 3(6), 474-484.
- Hübner, A. (2011). Retail category management: decision support systems for assortment shelf space, inventory and price planning. Berlin: Springer Verlag.
- Lim, A., Rodriguez, B., & Zhang, X. (2004). Metaheuristics with local search techniques for retail shelfspace. Management Science, 50(1), 117-131.
- Murray, CC; Talukdar, D; Gosavi, A; (2010). Joint optimization of product price, display orientation and shelf-space allocation in retail category management: Special Issue: Modeling Retail Phenomena. Journal of Retailing, 86(2), 135-136.
- Ring, L. J., & Tigert, D. J. (2002). An Integrated Framework for Strategic Resource. Babson Retailing Newsletter, 42 - 60.

- Schneider, J., & Julie, H. (2011, April 1). Why Most Product Launches Fail. *Harvard Business Review*, pp. 21 - 23.
- van Nierop, E., Fork, D., & Franses, P. (2006). Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements: Working Paper. ERIM Report Series. *Research in Management*, 1-35.
- Yang, M. (2001). An efficient algorithm to allocate shelf space. *European Journal of Operations Research*, 131(1), 107–118.
- Yang, M., & Chen, W. (1999). A study on shelf space allocation and management. *International Journal of Production Economy*, 60–61(4), 309–317.
- Zufryden, F. (1986). A dynamic programming approach for product selection and supermarket shelf-space allocation. *Journal of the operational research society*, 37(4), 413-422.

AUTHORS BIOGRAPHY

Salutino Flores Monroy He received a bachelor degree in Industrial Engineer with a minor in Systems by the Monterrey Institute of Technology and Higher Education (2011). Currently, he is studying the master in Research Operations in UNAM, Engineer Faculty, at Mexico City. His current research interests include inventory routing problems, simulation and Time Series.

USE OF WEB SERVICES FOR THE LOCALIZATION OF ROLLING STOCK WITH UTILIZATION TECHNOLOGY ORACLE SPATIAL

Jan Fikejz^(a), Jan Merta^(b), Jakub Oborník^(c)

^(a,b,c) Department of Software Technologies, FEI, University of Pardubice, Pardubice, Czech Republic

^(a) Jan.Fikejz@upce.cz, ^(b) Jan.Merta@student.upce.cz, ^(c) Jakub.Obornik@student.upce.cz

utilization of web services for the localization of rolling stock with utilization technology oracle spatial

ABSTRACT

This article deals with utilization of web services for the localization of rolling stock on the railway network using Oracle Spatial technology. Attention is focused on the description of the rolling stock location in a designed model of the Czech Republic railway network. Further attention is aimed on the design of web services for localization of rolling stock position on the server side with additional information for passengers.

Keywords: Railway infrastructure models, train positioning, web services

1. INTRODUCTION

Rolling stock localization is a constantly discussed topic involving a lot of companies. The problem of rolling stock localization may be divided in two main areas of interest: localization for the needs of (i) security technology (ii) information and telematics systems.

In the first case the reliability and safety are emphasized. Nevertheless, these components are often connected with higher realization costs since they often require adding further communication or identification elements/devices to railway infrastructure.

In the latter, a certain scope of inaccuracy or reduced reliability can be accepted, which often results in a significantly (lower implementation of such solutions). Rolling stock localization has recently been connected with the use of satellite navigation system (GNSS – Global Navigation Satellite System).

2. POSSIBLE TYPES OF LOCALIZATION

Generally, localisation is prone to a wide range of approaches on how to identify the position of rolling stock on a track. Put simply, localisation may be divided into the following three groups:

- localization without the use of GNSS,
- GNSS using localization,
- GNSS-based, involving further support systems.

2.1. Rolling stock localization without the use of GNSS

This type of rolling stock localization often requires complementing the rail network infrastructure with additional construction elements, which entails higher costs of the actual implementation. On the other hand, this type of localization shows a high accuracy and reliability and is often used in the railway signalling technology. Essentially, it relates to the system of:

- ETCS (Ghazel 2104; Lieskovský and Myslivec 2010),
- Automatic train control (Chudacek and Lochman 1998; Lieskovský 2004),
- Track circuits (Dorazil 2008),
- RFID.

2.2. Rolling stock localization using GNSS

When using GNSS for various application levels, it is necessary to take an indicated position error into consideration. Indicated position error is generally based on the nature of the satellite navigation. If we use systems that operate with position information on an informative level only, we can tolerate a certain error. However, such inaccuracy is unacceptable in the railway signalling technology. However, various additional systems can be implemented to eliminate error (completely or at least partially), thus making the position of the tracked object more accurate. The following systems can be listed in this group:

- EGNOS (Senesi 2012),
- Differential GPS (O'Connor 1997).

2.3. GNSS based localization involving additional support systems

As mentioned above, precise localization of rolling stock using GNSS, especially for the needs of signalling technology, is a practically impossible. Nevertheless, the position of a rail vehicle can be determined significantly more precisely with the use of additional systems. This especially concerns solutions using inertial systems (Standlmann 2006); but also less known

systems such as those based on GNSS and contactless eddy current measurement (Becker and Poliak 2008).

3. RAILWAY NETWORK MODEL

An undirected graph, as defined graph theory, is a natural candidate for a railway network model. Based on an analysis of data provided by the company SŽDC-TUDC (consisting of service regulations, passports and codebooks), sets of algorithms were subsequently created, with which it was possible to generate a three-layer model of the rail network (Fikejz and Kavička, 2011). Roughly speaking, the track can be divided into individual so called supertracks, which consist of define supra-sections (TDNU), where each supra-section contains track definition sections (TUDU) with mileposts (in hectometres). Basic aspects of the description of the rail network are collectively shown in Figure 1.

Mileposts (in hectometres) are shown in Figure 1 with the distance in kilometres and are graphically represented using gray points. TUDU is recorded using a six-digit code (163105, 163106, 16307, 173202) and are graphically represented using solid lines (red, black, orange, brown). Individual supra-sections (CLS 007, CLS008, REG023) are shown in light blue and supertracks (B421021 1 and B421021 1A) are shown in dashed lines. A place significant in terms of transportation (branch line) is symbolized by a green square.

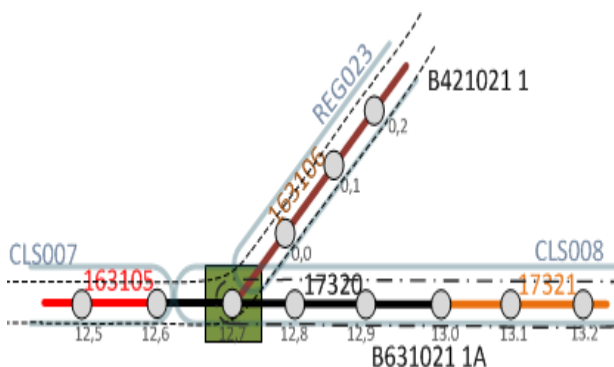


Figure 1: Basic aspects of the description of the rail network

The algorithm of a railway network model (Fikejz and Kavička, 2011; Fikejz and Řezanina 2014.) was implemented directly on the database level using PL/SQL language. However, the algorithm had to be adjusted and generalized several times since there are various nonstandard conditions in the data. These include, jumps in the mileposts (nonlinear growth of the kilometre succession between the mileposts) or change of an increasing kilometre sequence into a decreasing one. The final model includes three data layers:

- **Data-Micro**, consisting of vertices and edges,
- **Data-Mezo**, include mezo-vertices and mezo-edges

- **Data-Macro**, containing super-vertices and super-edges.

Figure 2 presents the overall concept of a complete three-layer railway network model.

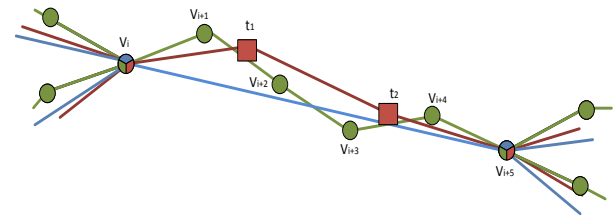


Figure 2: Illustration overall concept of a three-layer module

The data structure, non-oriented graph was finally implemented directly in the ORACLE database using the ORACLE Spatial Network Data Model (Kothuri et al. 2007) technology. This technology enables the user to build a various network representation, also involving the object scheme and the communication interface API.

The objects scheme includes metadata and network tables. The interface contains PL/SQL API (an SDO_NET packet) on the server side for the creation, control and analysis of the database network. It also includes a middle layer Java API (on client's side) for the network analysis. The actual network is then defined by means of two compulsory tables:

- Node table,
- Link table.

For the work with spatial data, ORACLE with Spatial technology defines a special object data type SDO_GEOMETRY. This enables its user to store a number of spatial information and geometric types, such as various points, arcs, linear chains or polygons.

4. LOCALIZATION

The idea of rolling stock localisation access to tracks is based on the correct pairing up of GPS position. This position is provided by communication terminals, with the nearest vertex or edge of the graph. The discovered vertex/hectometre post not only consist of a multi-dimensional key in the form of a GPS coordinate, it is also linked, through definition sections, to further information concerning the railway network infrastructure.

Considering of the situation, that the model of railway infrastructure is stored in the database Oracle, we can use the native database functions and operators. The SDO_NN (*nearest neighbor*) operator was selected in view of realising this unique rolling stock localisation approach. The aforementioned operator searches for a

geometric object that is closest to the object entered (like a point, for example). In other words, it is possible to find the nearest vertex, or more precisely the edge in a model, from the current rolling stock position (Figure 3).

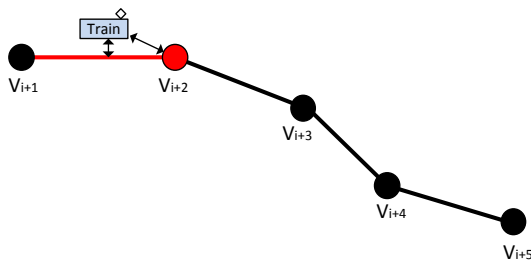


Figure 3: Main concept of localization

The actual detection of the current position of the rolling stock can be divided into the following steps:

1. **Finding the nearest vertex and edge of the graph** – from the current position of the rolling stock given the three-layer railway network model
2. **Assessment of the relevancy of incoming GPS information from the communication terminal** – verification whether the current position is not burdened by a disproportionate error (like, for example, that the distance of the rolling stock from the nearest vertex/edge is a mere few meters or tens of metres, or that the rolling stock is still assigned to the same super-edge, provided that it should still be located on it)
3. **Calculation of the exact position of the rolling stock on the edge of the model** – using perpendicular projection of the point (current rolling stock position) onto the line

The rolling stock position data was collected from the communication terminals. These communication terminals sent position information to the central database every 30 seconds (Figure 4).

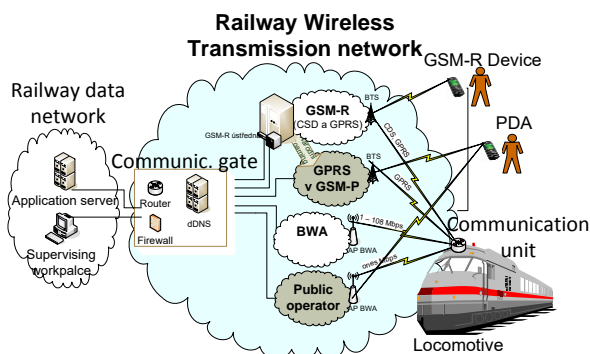


Figure 4 : Communication between the rail vehicle and dispatching centre

4.1. Searching for additional information

In addition to basic position information of the selected rail vehicle (kilometric position of selected train and definition sections), there also further additional information can be collected which is related to the railway network infrastructure. This additional information can then be used, for example, for railway information systems that are working with the position of rolling stock.

If we connect this information with the civil timetable, then we are able to find out additional information as follows:

- name and number of train under civil timetable,
- occurrence of the train within the railway station,
- previous and next railway station,
- distance in kilometres from the previous, or to the next railway station,
- current delay of the train,
- arrival time at the station.

4.1.1. Search algorithm of previous or next railway station

Searching for the previous and next railway station utilises an iterations algorithm on Data-Micro layer. This algorithm allows finding the nearest railway station, despite the fact, that within the one TUDU, there exists more stations, Figure 5.

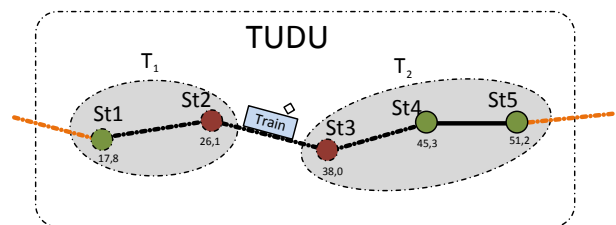


Figure 5: Finding nearest station

The concept of search algorithm consists of the following steps:

1. Sort the railway stations within the same TUDU considering the kilometric values
2. Divide the sorted railway stations into two separated subsets T1 and T2 by current position of rail vehicle
3. Select the first railway station from the subset T1 with the highest value of kilometre and the second railway station from the subset T2 with the lowest value of kilometre

As a next step, the shortest path algorithm is used for finding the real distance from the current train position to both already found railway stations and according to

the civil timetable, we are able to calculate the actual time of arrival to the next railway station – and actual train delay.

5. WEB SERVICES

For exploiting the localization position of rolling stock by other applications, web services were designed. These web services provide a basic set of information about the position of rolling stock. The main advantage of this approach is the hiding of application logic of localization mechanism from the final application. In the JAVA environment, we can use two different approaches. Their main difference is in the internal request processing, and in their architecture. The Web services were carried out using:

- REST (REpresentational State Transfer),
- SOAP (Simple Object Access Protocol).

5.1. REST

REST is very often used for designing of web services. The architecture is not related to some specific port, and the HTTP protocol is very often used. In the JAVA language, we can use, for example, the well-known JAX-RS specification, and its reference implementation in the Jersey or CXF from Apache company. The main specific is HTTP API and POJO (Plain Old Java Object) complaint.

5.2. SOAP

Unlike REST, the SOAP architecture is a protocol designed for web services with a different approach because the SOAP uses strict XML messages for the communication and is oriented to RPC (Remote procedure call). In the JAVA language we can use the well-known JAX-RS specification and its possible implementations, as Metro from GlassFish company or CXF from Apache company.

5.3. Design and implementation of web services

For both approaches the set of the localization methods were prepared, which performed selected localization tasks. It possible to use these methods, which according to the train number, is able to find out the position of selected railway vehicle on the railway network. Using the JSON protocol, this position data is then returned to the client in the final application. The concept of use of a web service is shown in Figure 6.

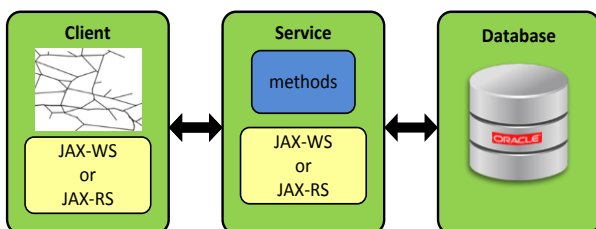


Figure 6: Concept of communication

The layer of web API is created from two different technologies (REST and SOAP). Due to this fact, we have the possibility to make a request in two ways. That means that it is only up to the client, which way is preferred.

The main web service is designed as a five separate parts, Figure 7.

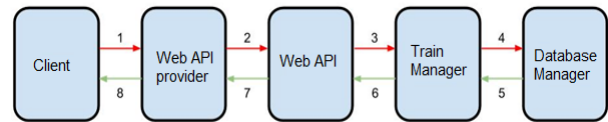


Figure 7: Concept of web API structure

Below are described the steps of communication:

1. Client is connected to the provider of the Web API,
2. After successful connection the request is sent to API for executing,
3. The web service validates the request and passes the information to *TrainManager* class (this class is responsible for all information about train),
4. *TrainManager* class at this moment does not have any information about this train, and so sends information to *DatabaseManager* class. This class is responsible for the communication with the database and performs the desired functions,
5. Information from the database is sent back to *TrainManager* class,
6. *TrainManager* class processes the information in into the desired objects and further sends them back,
7. On the base of request of input data format, the web service converts the objects and sends back the information about train position,
8. Provider sends to client the requested information (information about train position).

6. TESTING

In order to test, a supportive software application was designed with an integrated core of discrete simulation. This supporting application allows the simulation of railway traffic of rolling stock on the railway network. Based on historical records and data generated, it is possible to save the current positions of rolling stock to the database in the defined time intervals. This train position data is subsequently used by testing application for localization of rolling stock within the monitored segment of the rail network.

The testing application was designed a mobile application based on the Android platform. This testing application uses a designed web service, which is based

on the input parameter/s (for example the train number). This provides a complex set of information on the selected train position. Information is shown about:

- current GPS position,
- definition sections (TUDU),
- current kilometre positions,
- route,
- current speed,
- name of the next railway station,
- distance to the next railway station.
- estimated time of arrival,
- current delays.

The testing application also allows us to view the civil timetable for the selected train vehicle and the service which is available on board the train. Also, the actual train position is shown on the interactive base map. The testing application for the Android platform is shown in Figure 8.

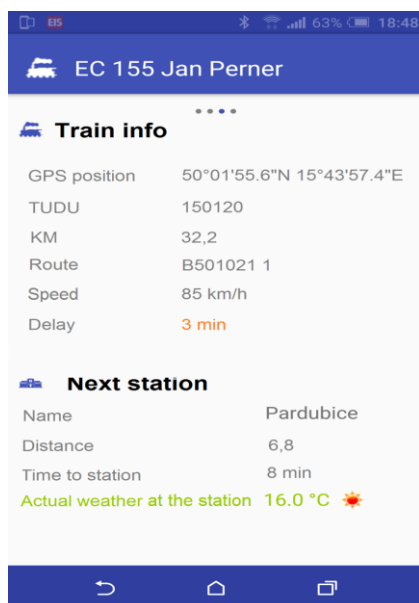


Figure 8: Android application

CONCLUSION

Attention was paid to the possible use of GNSS technology for utilization of web services for the localization of rolling stock. Oracle Spatial technology was employed. A multi-layered model (based on undirected graph) of the railway network infrastructure was designed. Further, algorithm for identification of the position of rolling stock was implemented in the railway network. This algorithm included the search of the previous or next railway station. For exploiting the localization position of rolling stock by other

applications, the web services REST/SOAP were designed. These web services provide the basic set of information about the position of rolling stock. For the testing, as mobile application based on the Android platform was designed. This testing application uses a designed web service, which is based on the input parameter/s (for example the train number), provides a complex set of information about the selected train position.

ACKNOWLEDGMENTS

This work has been supported by the project “SGS_2016_018 Models of infrastructure and operation of land transport systems” (financed by the University of Pardubice).

REFERENCES

- Becker, U. and J. Poliak. DemoOrt repositions trains with satellite. In: EURAILmag Business & Technology. 18. BLUE LINE & Bro, France, 2008,s. 216-219.
- Chudaček, V. and L. Lochman. Vlakový zabezpečovací systém ERTMS/ETCS. In: Vědeckotechnický sborník ČD, č. 5/1998
- Dorazil, P. Základní vlastnosti kolejových obvodů bez izolovaných styků. Pardubice, 2008. Bachelor thesis. University of Pardubice. Supervisor: Milan Kunhart.
- Fikejz, J. and A. Kavička. Modelling and simulation of train positioning within the railway network. In: KLUMPP, Matthias. ESM'2012. The European simulation and modelling conference. Ostende: EUROSIS - ETI, 2012, s. 366 -376. ISBN 978-9077381-73-1.
- Fikejz, J. and A. Kavička. Utilisation of computer simulation for testing additional support for dispatching rail traffic. In: European Simulation and Modelling Conference, 2011. Ostende: EUROSIS - ETI, 2011. p. 225-231. ISBN 978-90-77381-66-3.
- Fikejz, J. and E. Řezanina, Utilization of computer simulation for detection non-standard situations within the new data layer of railway network model. In: The 26th European Modeling & Simulation Symposium. Bordeaux, 2014 s. 371-377, ISBN 978-88-97999-32-4
- Ghazel, M. Formalizing a subset of ERTMS/ETCS specifications for verification purposes. In:Transportation Research Part C: Emerging Technologies. Elsevier Limited, 2014, pp. 60-75 ISSN: 0968-090X
- Kothuri, R. et al. Pro Oracle Spatial for Oracle database 11g. New York, NY: Distributed to the book trade worldwideby Springer-Verlag New York, c2007, xxxiv, 787 p. ISBN 15-905-9899-7.

- Lieskovský, A. and I. Myslivec. ETCS a AVV - poprvé společně. In: EuroŽel, Žilina, 2010
- Lieskovský, A. Automatické vedení vlaků Českých drah. In: Automatizace. Praha: Automatizace, 2004, roč. 10. ISSN 0005-125x.
- Murray, Ch, et al. ORACLE® Fusion Middleware : User's Guide for ORACLE MapViewer 11g Release 1 (11.1.1) [online]. 2010 [cit. 2012-07-09]. Available from : http://docs.ORACLE.com/cd/E14571_01/web.1111/e10145.pdf
- O'connor, M. L. Carrier-phase differential gps for automatic control of land vehicles, In: Dissertation Abstracts International, Volume: 59-06, Section: B, page: 2876.; 158 p. 1997, Stanford University, ISBN: 9780591909272
- Senesi, F. Satellite application for train control systems, In: The Test Site in Sardinia, Journal of Rail Transport Planning and Managemt. Elsevier BV, 2012, s. 73-78, ISSN:2210-9706
- Stadlmann, B. Automation of operational train control on regional branch lines by a basic train control. In: Proc. IEEE Intelligent Transportation Systems Conference, Toronto, Canada, September, 2006, s17-20.

EMULATION AS PART OF AN INTEGRATED LOGISTICS LEARNING ENVIRONMENT

Sebastian Langer^(a), Wladimir Hofmann^(b),
Sebastian Lang^(c), Tobias Reggelin^(d), David Weigert^(e)

^{(a), (b), (c), (d), (e)} Otto-von-Guericke University, Magdeburg

^(a) sebastian.langer@st.ovgu.de, ^(b) wladimir.hofmann@st.ovgu.de,
^(c) sebastian.lang@st.ovgu.de, ^(d) tobias.reggelin@ovgu.de, ^(e) david.weigert@ovgu.de

ABSTRACT

To meet the expectations towards future logistics education, an integrated learning environment (ILE) will be implemented at German-Kazakh University Almaty. Developed by Otto-von-Guericke University Magdeburg, the ILE will enable students to design, implement and verify logistics processes along a supply chain. The ILE includes a physical, miniaturized, automated material flow system with PLC and RFID, a self-developed Excel-based Control Centre and material flow simulation. Main emphases of this paper are the learning approach of process design with predefined process modules and the development of a discrete event emulation model as virtual commissioning tool for process control validations.

Keywords: Logistics Education, Material Flow Laboratory, Virtual Commissioning, Emulation

1. INTRODUCTION

Customization and regionalization of products are trends in today's industry. Together with service quality and schedule flexibility requirements, the resulting product diversity is overstressing the capabilities of classical production and logistics systems (Bauernhansel, ten Hompel, and Vogel-Heuser 2014). That is especially true for the all-connecting material flow systems and its controls. The concept "Industry 4.0"/"Internet of Things" is addressing these challenges by digitalizing and connecting all elements in the material flow (Schenk 2014). These decentralized control systems need clear identification before a decision process can take place. RFID is an approach to solve this dilemma.

Beside technological changes, the shifting of worldwide production volumes towards East Asia will establish new hubs in the global logistics networks. One of these hubs could be Kazakhstan as it is located on the shortest overland connection between Europe and East Asia. In its long-term development plan "Kazakhstan 2030" the country aims at building a powerful transport infrastructure. The strategy paper recognizes that beside the investments into infrastructure and technical solutions the education of logistics specialists is a critical success factor (Nazarbayev 1997).

The world bank is ranking the individual logistics competence of relevant employees in Kazakhstan only on the 81st place (Arvis, Saslavsky, Ojala, Shepherd, Busch, and Raj 2014).

In order to exploit this fallow potential, the LogCentre project was established in 2014. As a cooperation between leading German logistics research/education institutions (e.g. Otto-von-Guericke-University Magdeburg, Fraunhofer IFF Magdeburg) and the German-Kazakh University (DKU) in Almaty, it is a nucleus for up-to-date logistics education, consulting and industry-related research in Kazakhstan. The LogCentre is sponsored by the German Federal Ministry of Education and Research. The Centre's education services are aiming at logistics and telematics students of DKU as well as industry practitioners.

A practical and accessible didactics concept and an appropriate learning environment can address the challenges in modern material handling systems and can include Industry 4.0 solutions in logistics education. Main objectives of the teaching should be a deep understanding of intralogistics processes (design, implementation and verification) and the key technologies used (Material flow automation, RFID, simulation and ERP integration).

Today's industrial engineering education is addressing the linking role of logistics with a broad training program. Engineering basics, first insights into automation technology, the usage of ERP systems and simulation software are mostly done with the help of abstract tasks. Especially the context and the interaction between the different elements (e.g. the key technologies identified above) of a logistics systems are lost in the class structure. While all of these separated classes help the student to understand how wide and complicated the field is, it is a challenge to impart practical experience and understand the possible synergies between process and technology (Baumgarten and Hildebrandt 2008).

The approach described in the following is to integrate computer simulation and automation technology to form an innovative learning environment which enables

students to design, implement and verify logistics processes along the whole supply chain.

2. CREATING AN INTEGRATED LEARNING ENVIRONMENT

To understand complex problems and to develop future-oriented solutions, students (and other users of the LogCentre education services) need to think beyond the borders of the curriculum. Bringing together normally separated fields of technology can help to trigger this cross-border thinking.

Beside the class content, the new complexity is also challenging the classical, lecture oriented teaching methods. The needed depth as well as the required breadth of study can only be archived with a mixture from theory-based lectures, involvement-provoking seminars and tutorials, and, most important, excursions and project works (Baumgarten and Hildebrandt 2008).

As logistics is understood as an application-oriented science with the primary task to provide convincing recommendations for action (Delfmann, Dangelmaier, Günthner, Klaus, Overmeyer, Rothengatter, Weber, and Zentes 2011), hands-on experience is just as important as the theoretical foundation.

In this context, the combination of future technologies (e.g. RFID) and a physical material flow system with different methods from the world of computer simulation can provide an environment able to fulfill the posed standards. This environment will be called "Integrated Learning Environment" (ILE) in the following.

2.1. Learning with Material Flow Laboratories

The central and most striking element of the ILE is a material flow laboratory. While the history of laboratory work in logistics research dates back to the 1950's, there have been no studies on the effectiveness of use in (higher) logistics education (Uckelmann 2012).

With the results from other fields (e.g. engineering) in mind, it can be assumed that laboratory work is substantially different to the predominant theoretical teaching (Raghunath 2001):

"The biggest teacher is the failed experiment." Raghunath argues that not only the subject-related work is a valuable experience, but also the open way of working. Instead of repeating memorized definitions, laboratory work is about finding working hypothesis and making intelligent predictions about an (at least partly) unknown experiment outcome. Another fruit for students could be the realization that exam results do not always correlate with the results of the actual knowledge application (Raghunath 2001).

The other way around, practical work ("active learning") is improving the chances of undergraduate students to perform well in courses. Freeman, Eddy, McDonough, Smith, Okoroafor, Jordt, and Wenderoth (2014) show that students in traditional, exclusively theoretical lectures are 55 percent more likely to fail an exam.

2.2. Learning with Computer Simulation

While the above-mentioned showed that a physical material flow laboratory has a reason for existence, the tight budgets of university teaching do not allow the providing of sufficient laboratory capacity for long, individual implementation and testing procedures (Uckelmann 2012).

The use of simulation in logistics laboratories can be a remedial action. The literature provides many examples for successful integration of simulation into education. Bruzzone, Bocca, Longo, and Massei (2007) are for example highlighting the contribution of simulation to education: increased interaction between student and system and the possibility of direct experiments (Montevechi, Leal, Miranda, and Pereira 2013) emphasize the positive motivation through interaction and are also stating improved results in the student's further project works (Rangel, Rangel, and Nascimento 2014). show the application of discrete event simulation to demonstrate operational sequences in computer network technology archiving "favorable results".

Closer to the case at hand is the education-related paper of (Cardoso, Rangel, Nascimento, Laurindo, and Camacho, 2014), describing the connection of a discrete event simulation model to a PLC. This use case of steering a simulation model with an external control system is known as emulation. Emulation can be used to find functional errors in control algorithms and to verify developed solutions (Boer and Saanen 2008).

To find error sources not included in the emulation model, it still needs a physical laboratory on site (Henke, Ostendorff, and Wuttke 2012).

Another possible risk of simulation/emulation usage in education is the misunderstanding of the reality due to the simplified modelling. In addition, familiarization with the simulation tool can be time consuming. Also, the availability of simulation tools is mentioned as a restricting factor (Bruzzone, Bocca, Longo, and Massei 2007).

Nevertheless, the usage of simulation is promising great results in education as students are showing enthusiasm when dealing with virtual, interactive learning environments (Leão, Soares, Rodrigues, Seabra, Machado, Farinha, and Costa 2012).

The benefits of a virtual over a physical laboratory are the better cost effectiveness, the higher capacity of the learning environment and the higher number of training scenarios, while the combination with a physical laboratory can eliminate most of the risks connected with simulation use.

2.3. Specification of Laboratory Scope

In logistics, it is common sense not only to think across the borders of technologies and curriculums, but also across the borders between different companies. That leads to the premise that a model for a comprehensive logistics education needs to picture a whole supply chain (Ballou 2006), including inter- and intralogistics processes from supplier to point of sale, from SCM/ERP level down to automation layer.

The following process layout was created by analyzing problem fields of today's process design, especially in the context of RFID usage. All flow objects (goods, pallets and trucks) are RFID tagged and tracked/traced along the whole supply chain. The stock levels in the different supply chain stages are recorded and can be used to control the system, e.g. automatically reorder products.

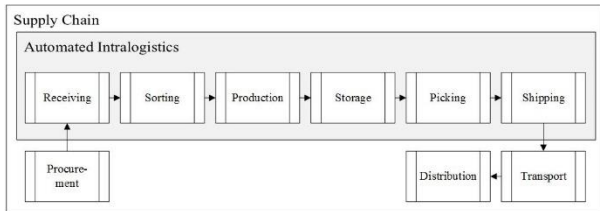


Figure 1: Process Structure of the ILE Supply Chain

2.4. Automated Intralogistics

The main part of the model is the automated intralogistics technology depicted below (Higher quality pictures and other project related media can be obtained from <http://www.logcentre.ovgu.de>). It shows all the main processes inside a production system in a miniaturized manner: goods receiving, transport, sorting, automated storage, production, goods dispatch. The processes are automated. The user is only needed to feed palletized products into the system and to sort goods on the picking stations.

The conveying mechanics are made by the company Fischertechnik, which is producing a construction kit under the same name. Coming with a high individualization potential, the kit can be complemented by different supplementary materials (e.g. steel bars) which positively influence the durability in the tough environment of a learning laboratory. Several vendors are offering pre-assembled industrial modules, including numerous intra-logistics relevant elements like the ones shown above.

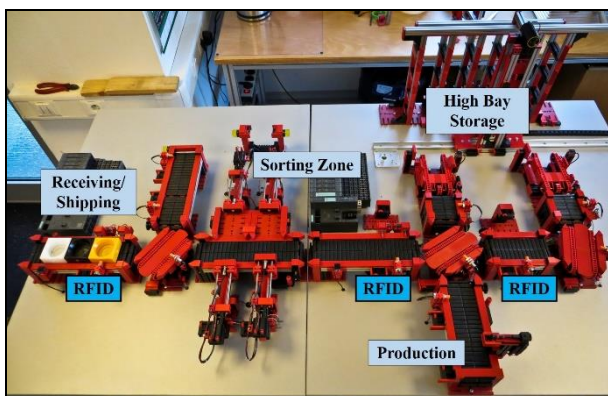


Figure 2: Automated Intralogistics in the ILE

2.5. Supply Chain Periphery

After leaving the automated material flow system, goods are transported down the supply chain to the distribution part represented by an outlet. The store has

a two-step layout, consisting of a storage area and a sales area. The transition points transport to storage area, storage area to sales area and sales area to the point of sale are equipped with RFID gates.

The main RFID processes in commerce can be demonstrated here: Automated bulk checkout, first-in-first-out assurance and the implementation of refined reorder procedures.

3. CONTROL SYSTEMS

The material flow along the modeled supply chain is managed by a number of tools which are also controlling the automated intralogistics.

Figure 3 shows the line-up of the software and hardware components installed in the ILE. Intersystem communication is done via the Control Centre (CC). Read from the top to the bottom, the structure shows a hierarchy similar to the automation pyramid (Heinrich, Glöckler, and Linke 2015).

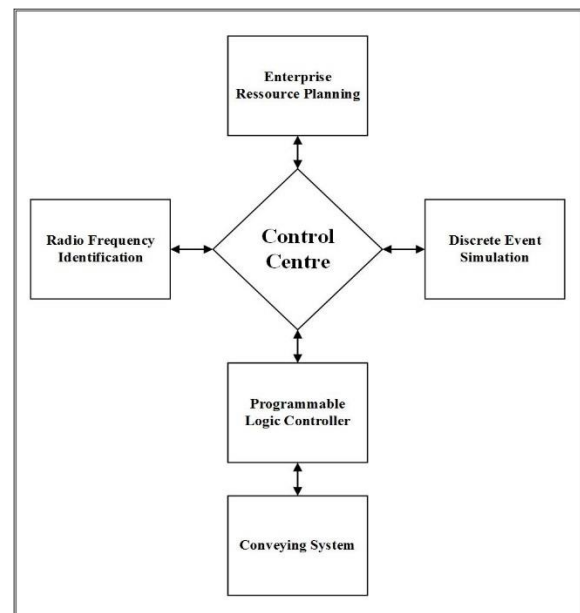


Figure 3: The Technical Components of the ILE

3.1. Control Centre

The Control Centre is the all-connecting knot in the material flow control. It receives RFID readings, manages the data base, processes user inputs and communicates the resulting process commands to the PLC.

The CC is integrated in an .xlsm file, consisting of several sheets and VBA code. As Excel is a standard tool used extensively for the most different kinds of business tasks, most students have been using the software before. In connection with VBA as programming language, which is close to natural language and a good starting point if the student was programming never before, it is easily accessible.

Excel seems to be a unique choice for a material flow control software but communicates very well with all other software elements of the ILE. That simplifies the

development of interfaces to the existing elements and ensures the future upgradability.

3.1.1. Usage of the CC to Demonstrate Processes

The CC is the user interface of the system. All main system routines are triggered by purchase orders which can be placed through user forms on the main control sheet. The user is guided through the order placing with the help of self-explaining input boxes. A short user guide describing all processes and how to trigger them is provided.



Figure 4: New Order User Form Screenshot

If the ILE is used for demonstration purposes only, nothing about the detailed functioning of the system needs to be known. In this way, teachers not directly involved in the project can use the material flow laboratory in their courses, too.

3.1.2. Usage of the CC to Modify Existing and Develop New Processes

However, for the main field of application the manipulation of the process sequences in the CC is elementary. It was necessary to choose an easily accessible programming approach that allows students to change the existing process structure and to implement new functionalities. While parts of the target group, especially the telematics students, are expected to have a solid IT knowledge, some of the logistics students as well as other users might not have a lot of experience in programming.

To face this challenge, an approach using graphical programming was developed. The students only need to combine elementary, pre-defined process modules to process groups, which in turn are linked to process chains through VBA-coded decisions.

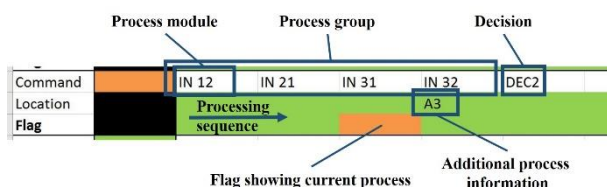


Figure 5: Screenshot Graphical Programming

Figure 5 shows an example line from the graphical programming sheet. The process module identifier above the flag is read by a superordinate VBA method and communicated to the PLC. Once the PLC confirms the successful execution of the process module, the flag jumps to the right follower. If additional information is needed to carry out a process (e.g. a location in the high bay storage) it is communicated in the field below the process module name. At the end of the process group, a decision method realized in VBA is called. This decision method chooses the next process group depending on the system status, e.g. RFID readings and the stock situation.

The single process module is clearly defined. It has a starting point and an ending point. The process module is reported as successfully executed once the good was transported to and detected by the physical ending point sensor.

Every possible process chain in the automated intralogistics part can be described as a sum of process modules.

3.2. PLC

The process modules described above are intended to simplify the design of process groups and process chains as well as to serve as commands from the CC to the PLC.

The PLC code is structured into different function blocks. These blocks are called by the communicated process module identifiers. The cyclic program sequence of the PLC is first checking for the active process module, then checking the system status represented by sensor readings and finally setting the actuators needed to complete the process module execution.

For almost all changes and process developments in the material flow laboratory, the PLC code does not need to be reconfigured. However, if the lecturer/the student wants to include PLC programming into his/her project, the system provides a state-of-the-art development environment. A Siemens Simatic S7 hardware PLC was chosen to control the automated intralogistics. This is the current industry standard for material flow automation. The used PLC programming environment is Siemens TIA Portal. It is supporting all commonly used PLC programming languages.

The communication between the Excel CC and the S7 was implemented by using an Industrial-Ethernet connection and the LGPL-rated communication library "libnodave.dll".

3.3. RFID Integration

RFID is one of the trending topics in the last decade's logistics research. Advantages are contactless and fast bulk reading without handling the goods (Gartner 2012, Kortmann 2006). Automated systems need this fast and reliable identification.

13.56 MHz USB readers with a suitable read range of 5-7 cm are in use in the ILE. The readers are positioned on seven points. Three are part of the intralogistics

automation, while the other four are keeping track of stock movements along the supply chain. All flow objects are equipped with passive tags for identification. The tag ID refers to the object's details which are saved in a central database. The reading data is written into the Control Centre. The data is used for tracking/tracing goods along the supply chain and, more important, for picturing the system status and deciding for the next process group.

4. INTEGRATING EMULATION INTO THE ILE

As mentioned above, the described physical laboratory does not offer sufficient capacity for all students as only one student after the other can implement and verify the developed process design. On the other hand, interaction with the system during the control development was identified as one of the central success factors for the laboratory.

Cardoso, Rangel, Nascimento, Laurindo, and Camacho (2014) showed the use and the usefulness of emulation models in automation engineering classes. Transferring this approach to the higher level of process control development solves the mentioned problem.

To test the developed process logic, an emulation model of the automated intralogistics and the supply chain periphery was created. This model is connected to and controlled by the Excel CC. This approach is known as high-level emulation and is typically used for virtual commissioning (Zhang, Le, Johnston, Nahavandi, and Creighton 2012).

The commissioning model emulates the physical part of the material flow model including PLC, conveying elements and RFID reader, emulating its input/output behavior towards the CC.

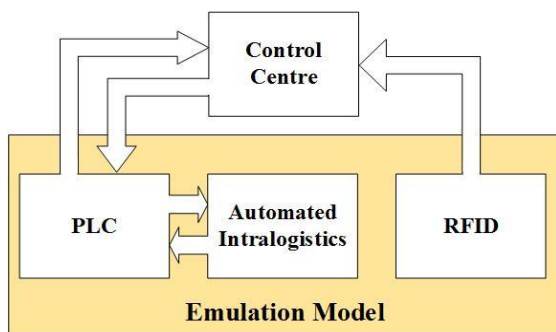


Figure 6: Structure of the Emulation Model

The implemented model does not emulate on sensor-actuator level (as not the PLC is connected) but the process level control done by the CC. Base for the emulation model were the planning outcomes: Flowcharts, conveyor layouts, the pre-defined interface CC to PLC and the pre-defined interface RFID to CC.

As part of the LogCentre, workstations with the simulation software ExtendSim are available for the students. The ExtendSim implementation of the model has the main objective to visualize/animate all

necessary information for the verification of the process sequence programmed in the CC. For this purpose, a special ExtendSim object library was created and custom made blocks were provided with a special animation behavior. In the simulation model, all of the physical components were modeled in a simple way in 2D, including material handling technology with conveyor belts, rotary tables and the production facility as well as goods and pallets with their attributes Tag ID and type. The high-bay storage is visualized by animating the information about put away goods that is written into a database table. The stock levels of the virtual HBS as well as the stock in other supply chain steps can be initialized with the currently defined stock levels from the CC.

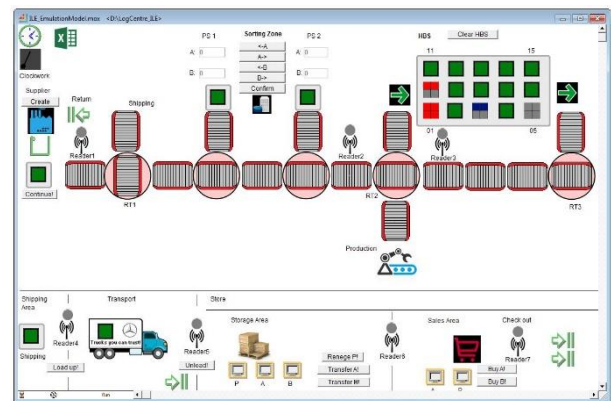


Figure 7: ExtendSim Emulation Model

The emulation model is controlled analogously to the control of the physical model's PLC: The CC transmits the ID of the process that has to be executed and further needed process information, e.g. location names for picking and put away. After successful execution of the process, the emulation model confirms the command. In addition, when passing the reader points the corresponding RFID tag IDs of the flow objects are communicated. The embedded software interfaces based on COM are used to write directly into corresponding CC database fields.

The writing of a process module execution confirmation message is triggering the next process module or a decision. In the emulation model, the execution – which means the simulation and animation of the communicated process – is released by opening the respective gates for the corresponding flow object.

Just as the physical system, the emulation needs human interaction. User inputs into ExtendSim can be done during runtime with the help of form fields and buttons. It is possible to create flow objects, to insert them into the automated intralogistics, to sort the goods on the picking stations and to move the goods down the supply chain.

CC and ExtendSim both run locally on the student's workstation. Test scenarios can be easily defined in the CC in the presented graphical way and transferred to the emulation model at the push of a button. As the solution can be tested virtually before, the work on the physical

model can be minimized to the actual implementation and presentation. That increases the number of students that can work in the laboratory.

5. USING THE LEARNING ENVIRONMENT IN LOGISTICS CLASSES

The concept presented in the preceding offers a wide range of teaching and learning possibilities. Some of these possibilities are presented to the teachers in form of learning materials, collected in an online knowledge management system. Using different kinds of media can improve the teaching results (Jensen, Wood, and Wood 2003), especially since it is respecting differences in learning habits, curiosity and experience (Mandal, Wong, and Love 2000).

The prepared tasks are aiming at different aspiration levels following the Bloom's taxonomy (Bloom and Engelhart 1976).

The easiest level is to demonstrate lecture contents with the help of the ILE. Example: The lecture "High Bay Storage" is accompanied with a practical demonstration of put away and picking processes in the automated intralogistics model. The students can *identify* and the main processes and can *compare* them to manual storage strategies.

On a medium aspiration level, the students can program a new process chain in the Control Centre. The given programming technique has to *be used* and the solution needs to *be structured* according to the material flow system layout.

The most complicated tasks are openly formulated ones. Example: A small group of students has to *develop* a RFID-stock take process for the high bay storage. The process needs to be *tested* and *evaluated*.

During the concept development, the ILE was always thought of as a living environment with constant growth and enhancement. The development of new functions through student project works and final papers is desired and makes it exceedingly attractive.

In the everyday classes at DKU, students of telematics and logistics hardly ever meet. To cross the border between the two curricula, the ILE can be used for project work in interdisciplinary teams consisting of students from both study programs. Especially for the logisticians a knowledge transfer within these projects could be essential – Industry 4.0 is asking for a firm knowledge of IT and integrated systems.

6. FURTHER RESEARCH

In later development stages an integration of virtual reality (VR) technology is considered. With the automated intralogistics part, students have a role model for constructing their first VR model. An interface between CC/PLC and VR could be used for the digital status control.

The SAP ERP learning module that is part of the LogCentre project will be integrated into the ILE. It could function as a superordinate user interface to control the material flow laboratory.

In order to measure the actual impact of the used methods, the learning progress of the students in the ILE will be tracked. Results from this study might answer the question about the usefulness of learning laboratories in logistics education.

ACKNOWLEDGMENTS

The LogCentre project is financed by the German Federal Ministry of Education and Research (BMBF).

REFERENCES

- Arvis J.-F., Saslavsky D., Ojala L., Shepherd B., Busch C., Raj A., 2014. Trade Logistics in the Global Economy 2014. Washington, D.C.: The International Bank for Reconstruction and Development/The World Bank.
- Ballou R., 2006. The Evolution and Future of Logistics and Supply Chain Management. *Produção*, 2: 375-386.
- Bauernhansel T., ten Hompel M., Vogel-Heuser B., 2014. *Industrie 4.0 in Produktion, Automatisierung und Logistik*. Berlin: Springer.
- Baumgarten H., Hildebrandt W.-C., 2008. *Studium Logistik: Akademische Ausbildung und Führungskräftenachwuchs in der Zukunftsbranche Logistik*. Berlin: Baumgarten/Hildebrandt.
- Bloom B., Engelhart M., 1976. *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz-Studienbuch.
- Boer C., Saanen Y., 2008. Controls: Emulation to Improve the Performance of Container Terminals. Proceedings of the 2008 Winter Simulation Conference, 2639-2647. December 7-10, Miami (Florida, USA).
- Bruzzone A., Bocca E., Longo F., Massei M., 2007. Training and Recruitment in Logistic Node Design by Using Web-Based Simulation. *International Journal on Internet Manufacturing and Services*, 1 (1), 32-50.
- Cardoso L. d., Rangel J. J., Nascimento A. C., Laurindo Q. M., Camacho J. C., 2014. Discrete Event Software for Teaching in Control Systems. Proceedings of the 2014 Winter Simulation Conference, 3608-3617. December 7-10, Savannah (Georgia, USA).
- Delfmann W., Dangelmaier W., Günthner W., Klaus P., Overmeyer L., Rothengatter W., Weber J., Zentes J., 2011. Positionspapier zum Grundverständnis der Logistik als wissenschaftliche Disziplin. In: Wimmer T., Grosche T., eds. *Flexibel – sicher – nachhaltig*. Hamburg: DVV Media Group, 262-274.
- Freeman S., Eddy S., McDonough M., Smith M., Okoroafor N., Jordt H., Wenderoth M., 2014. Active learning increases student performance in science, engineering, and mathematics. *PNAS - Proceedings of the National Academy*

- of Sciences of the United States of America, 111 (23), 8410-8415.
- Gartner Inc., 2012. Gartner's Hype Cycle for Supply Chain Management 2012. Stamford, Connecticut: Gartner, Inc.
- Heinrich B., Glöckler M., Linke P., 2015. Grundlagen Automatisierung. Wiesbaden: Springer Vieweg.
- Henke K., Ostendorff S., and Wuttke H.-D, 2012. A Concept for a Flexible and Scalable Infrastructure for Remote Laboratories - Robustness in Remote Engineering Laboratories. In: Uckelmann D., Scholz-Reiter B., Rügge I., Hong B., Rizzi A., eds. The Impact of Virtual, Remote, and Real Logistics Labs. Heidelberg: Springer, 13-24.
- Jensen D., Wood J., Wood K., 2003. Hands-on Activities, Interactive Multimedia and Improved Team Dynamics for Enhancing Mechanical Engineering Curricula. International Journal on Engineering Education, 19 (6), 874-884.
- Kortmann M., 2006. RFID - Einsatz in der Produktionslogistik. Saarbrücken: VDM Verlag Dr. Müller.
- Leão C. P., Soares F., Rodrigues H., Seabra E., Machado J., Farinha P., Costa S., 2012. Web-Assisted Laboratory for Control Education: Remote and Virtual Environments. In: Uckelmann D., Scholz-Reiter B., Rügge I., Hong B., Rizzi A., eds. The Impact of Virtual, Remote, and Real Logistics Labs. Heidelberg: Springer, 62-72.
- Mandal P., Wong K. K., and Love P. E., 2000. Internet-Supported Flexible Learning Environment for Teaching System Dynamics to Engineering Students. Computer Applications in Engineering Education 8 (1), 1-10.
- Montevecchi J. A., Leal F., Miranda R. d., Pereira T. F., 2013. Interactive Learning of Modeling and Discrete-Events Simulation through Lego® Parts. Proceedings of the 2013 Winter Simulation Conference, 3556-3567. December 08-11, Washington, D.C. (USA).
- Nazarbayev N., 1997. Prosperity, Security and Ever Growing Welfare of all the Kazakhstanis. Almaty: Government of the Republic of Kazakhstan.
- Raghunath M., 2001. Lab-based research is the best education. National University of Singapore, Singapore. Available from: <http://www.bioeng.nus.edu.sg/stu/Research-Metholody-Lectures/PART%208%20Lab-Based%20Research%20is%20the%20Best%20Education.pdf> [accessed 30 March 2016]
- Rangel C. d., Rangel J. J., Nascimento J. R., 2014. Discrete Event Simulation for Didactic Support Ressource. Proceedings of the 2014 Winter Simulation Conference, 3596-3607. December 7-10, Savannah (Georgia, USA).
- Schenk M., 2014. Fabrikplanung und Fabrikbetrieb: Methoden für die wandlungsfähige, vernetzte und ressourceneffiziente Fabrik. Berlin: Springer.
- Uckelmann D., 2012. The Role of Logistics Labs in Research and Higher Education. In: Uckelmann D., Scholz-Reiter B., Rügge I., Hong B., Rizzi A., eds. The Impact of Virtual, Remote, and Real Logistics Labs. Heidelberg: Springer, 1-12.
- Zhang J., Le V., Johnston M., Nahavandi S., Creighton D., 2012. Discrete Event Simulation Enabled High Level Emulation of a Distribution Centre. 14th International Conference on Modelling and Simulation, pp. 470-475. March 28-30, Cambridge (UK).

AUTHORS BIOGRAPHIE

Sebastian Langer is a M.Sc. student at the Institute for Logistics and Material Handling Systems (ILM) at Otto-von-Guericke University Magdeburg (OvGU). His main field of interest lies in logistics planning. He developed the concept for the physical material flow laboratory. Feel free to contact him via sebastian.langer@st.ovgu.de.

Wladimir Hofmann is a M.Sc. student at ILM OvGU. Since he is interested in material flow simulation in the context of IoT and Industry 4.0, he created the LogCentre virtual commissioning tool. His email address is wladimir.hofmann@st.ovgu.de.

Sebastian Lang is a M.Sc. student at ILM OvGU. His research interests include modeling and simulation in the field of production and logistics. His email address sebastian.lang@st.ovgu.de.

Dr.-Ing. Tobias Reggelin, M. Sc. is a project manager at the Fraunhofer IFF and research assistant at Otto-von-Guericke University Magdeburg (OVGU). In 2003 he graduated with a diploma in industrial engineering at the OVGU Magdeburg. His current teaching and research area is the modeling and simulation of logistic systems. In 2011 he received his doctorate in the field on "Mesoscopic modeling and simulation of logistic flow systems" at the OVGU rated summa cum laude.

David Weigert (M.Sc.) studied Industrial Engineering with specialization in Logistics at the Otto-von-Guericke-University Magdeburg. His areas of competence are the analysis and optimization of logistics processes, as well as modelling, simulation and optimization of logistics systems. Currently David Weigert is working on his doctoral thesis on the mesoscopic real-time simulation of logistic systems.

AN ADVANCED APPROACH TO SIMULATION OF SUPER-WIDE BANDWIDTH INFORMATION AND COMMUNICATION SYSTEMS COMBINING MICROWAVE AND PHOTONIC INDUSTRIAL TECHNOLOGIES

M.E. Belkin^(a), V. Iakovlev^(b), A. Sigov^(a), Y. Tyschuk^(c), V. Golovin^(c)

^(a) Moscow State Technological University (MIREA), Scientific and Technological Center “Integrated Microwave Photonics”, Moscow, Russian Federation, ^(b) RTI Research S.A., Yverdon-les-Bains, Switzerland, ^(c) Sevastopol State University (SevSU), Sevastopol, Russian Federation,

^(a) belkin@mirea.ru, ^(b) v.iakovlev@rti-research.com

ABSTRACT

A novel approach to simulation and design of combined information and communication systems based on microwave and photonic industrial technologies using high-power commercial microwave electronic CAD tool is proposed. A detailed validation of this approach based on the results of E-CAD simulation and experimental verification of a photonic beam former is presented as well.

Keywords: phased array antenna, photonic beamformer, electronic computer-aided design tool.

Microwave photonics (MWP) is an interdisciplinary scientific and technological field that combines the domains of microwave and RF engineering and photonics (Urlick Jr., McKinney, and Williams 2015). This field in the last 30 years has attracted immense interest and generated many new R&Ds from both the scientific community and the commercial sector. Emerging applications for radio-over-fiber (RoF) hybrid communication systems, sub-terahertz wireless systems, radar, and electronic warfare systems indicate that MWP is set to be a subject of increasing importance. Generally, MWP devices are the examples of an intimate integration of photonics, microwave electronics, and planar antenna technologies for producing a complicated functional module in a multichannel analog environment. In particular, MWP technology opens the way to super-wide bandwidth transmitting characteristics at lower size, weight, and power as compared with traditional electronic information and communication systems (Paoella, DeSalvo, Middleton, and Logan 2015). For example, in a typical arrangement of MWP-based software defined RF receiver, a photonic circuit is inserted between two microwave electronic chains (Fig. 1). For direct and inverse transformations of microwave and optical signals there are two interface units at its bounds: electrical-to-optical and optical-to-electrical converters. Between the interfaces there are various photonics processing units for transmission, switching, distribution, filtration, and frequency conversion of microwave signals in optical domain.

1. THE KEY MICROWAVE PHOTONICS ELEMENTS

There are 5 basic MWP types of active optoelectronic devices which are depicted in Fig. 2:

- optical-electric converter (Fig. 2, a), for example, photodiode;
- electric-optical converter (Fig. 2, b), for example, semiconductor laser;
- optically controlled microwave sensor (Fig. 2, c), for example, microwave generator which parameters (frequency, output power) depend on optical signal;
- converter of optical signal (Fig. 2, d), for example, optical modulator, laser amplifier;
- converter of microwave signal (Fig. 2, e), for example, microwave amplifier which gain is controlled by an optical signal, optoelectronic delay line.

Common distinctive feature of all above-mentioned devices and functional elements is their C- (1530...1565 nm) and/or L- (1565...1625 nm) operational spectral ranges, as specified in ITU-T. The main reasons for such a choice are the lowest losses in silica fiber, the widest operation spectral range and availability of low cost and high performance fiber amplifiers for compensation of losses.

In addition to above listed 5 types of optoelectronic devices, there are 2 types of devices based on all-optical interaction, that can be effectively applied in microwave photonics equipment:

- optically pumped converter of optical signals (Fig. 2, e), for example, Erbium, Raman and Brillouin fiber amplifiers;
- optically pumped sensor of optical signal (Fig. 2, g), for example, the Erbium fiber oscillator.
-

These elements and devices represent the principal building blocks for creation of essentially newer

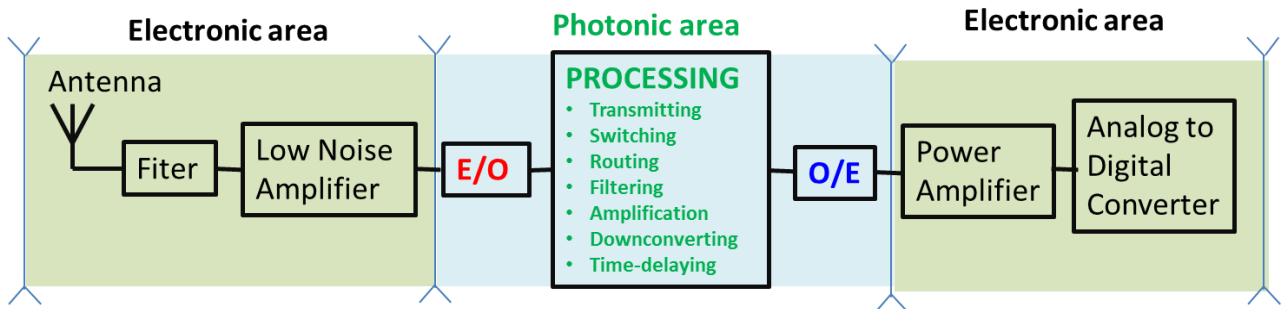


Figure 1: A typical arrangement of MWP-based software defined RF receiver

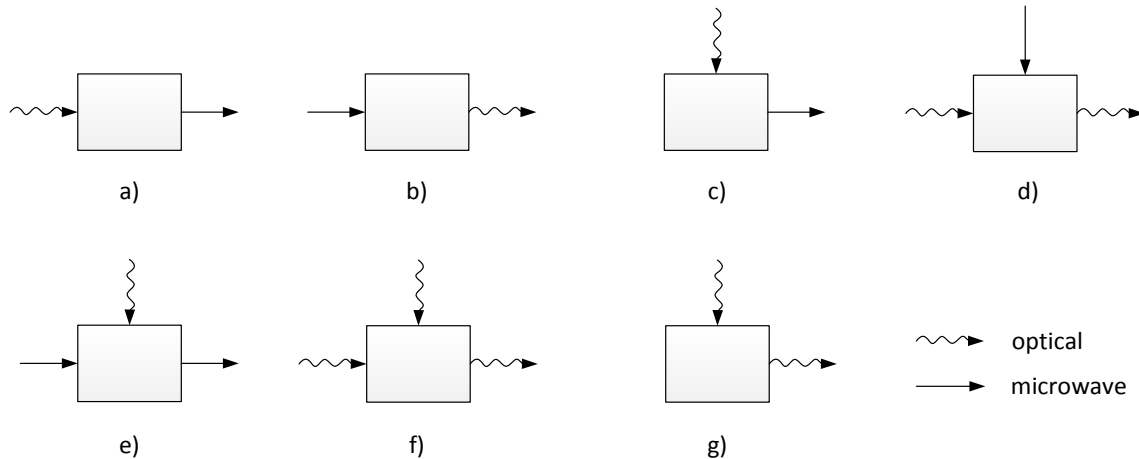


Figure 2: The classification of active optoelectronic devices and functional elements for microwave photonics technology

devices, which can significantly improve the key technical and economic characteristics of modern information and communication systems such as speed, operating frequency band, dynamic range, electromagnetic compatibility, etc.

2. SELECTION OF AN OPTIMAL PRODUCT FOR COMPUTER DESIGN

The developer of new MWP devices is facing a problem of choosing an appropriate computer tool for their modeling and design. As of today, a system designer is forced to use means of several computer-aided design (CAD) tools because the existing optical and optoelectronic CAD tools (OE-CAD) are not as developed as compared with the CAD tools intended for modeling of microwave devices (E-CAD) (Leijtens, Lourec, and Smit 1996). Increasing the accuracy of photonic circuits taking into account influence of their parasitic parameters in microwave band requires the use of the high-power microwave E-CAD tool (Leijtens, Lourec, and Smit 1996) working at symbolical level.

Table I lists detailed comparison of typical modern OE-CAD tool VPI Transmission Maker of VPI Photonics and well-known E-CAD tool AWRDE of National Instruments. Recently, we have developed by NI AWRDE E-CAD tool a number of models for active optoelectronic elements, such as semiconductor laser (Belkin, et al 2011; Belkin and Iakovlev 2015), p-i-n photodetector (Belkin and Sigov 2015) and for rather simple microwave photonics devices, such as optoelectronic microwave oscillator (Belkin and

Golovin 2015)(Belkin and Golovin 2015) and optoelectronic microwave frequency mixer (Belkin and Tyschuk 2015). Below we present the results of E-CAD simulation and experimental verification for a complex device such as photonic beam former of a super-wide bandwidth microwave phased array antenna (PAA).

3. MODELING OF SUPER-WIDE BANDWIDTH MICROWAVE PHOTONICS-BASED BEAM FORMER

So far, there are two approaches for the beam steering in modern microwave-band PAAs: using phase shifters or delay lines. The first one has simplest architecture but it has difficulties to achieve a wide scanning angle in super-wide bandwidth and suffers from so-called beam squint distortion. In order to achieve wide angle scanning of PAAs, the developer must implement real time delay lines based on so-called true time delay (TTD) approach (Mailloux 2005) to substitute the traditional phase shifters. Conventional TTD is composed of waveguide or coaxial cable, which has high loss of signal and engineering complexity. Application of microwave photonics and fiber technologies to PAA can reduce transit time and aperture effect, and resolve the problem of wide angle and wide bandwidth scanning (Yao 2009).

Let us compare the above beam former architectures for microwave photonics-based PAA. To be specific, the calculations were performed for the 16-element PAA.

Table I: Comparison of modern E-CAD and OE-CAD tools

#	Feature	Realization		Comments
		By E-CAD (NI AWRDE)	By OE-CAD (VPI Photonics)	
1	Analysis approach	Building Blocks, 3D electromagnetic analysis	Building Blocks,	
2	Simulation methods			
	- Linear circuits	S-,Y- matrix, equivalent circuits	S-matrix	
	- Nonlinear circuits	Harmonic Balance Engine ALPAC, 3D planar electromagnetic simulator AXIEM modeling	S-matrix, combination of time-and-frequency domain modeling	
3	Element representation			
	-active microwave elements	Multirate harmonic balance, HSPICE, Volterra, based on measured characteristics models	Ideal or based on measured characteristics models	
	- active MWP elements	Absent	Rate equation-based, transmission line models	
	- passive elements	Lumped and distribution, microwave-band specialties	Lumped, ideal	Waveguides, couplers, resonators, resistors, capacitor, inductor
4	Possibility for calculating the key parameters of MWP circuits and links	By one-click operation	By user-created complicated schemes	Large-signal transmission gain, Noise figure, Phase noise, Intermodulation distortion, Intercept points, so on
5	IC Layout design and analysis	Yes	No	
6	Built-in design kits from the main foundries	Yes	No	
7	Parameter optimization	Yes	No	For more sophisticated investigations
8	Sensitivity analysis	Yes	No	
9	Design of tolerance	Yes	No	
10	Statistical design	Yes	No	
11	Yield optimization routine	Yes	No	
12	Built-in library of producer specific models	Yes	No	

3.1. Phase shifter-based photonic beam former

To evaluate phase shifter-based photonic beam former performances an equivalent circuit including the models of previously studied microwave photonics component and built-in AWRDE's library models was developed. The simulation was performed by calculating the PAA's normalized radiation patterns (NRP) in the frequency band of 6-18 GHz. Fig. 3 shows examples of the calculation of the NRPs for the case of discrete phase delay selected in the center of the frequency band. As can be seen from the Figure, when a discrete phase shift $\Delta\phi = 13^\circ$ at the extreme frequencies of the band, there

are significant errors $\delta\theta = 20^\circ$ at the frequency of 6 GHz and $\delta\theta = 7^\circ$ at a frequency of 18 GHz. From the Figure, one can make an unambiguous conclusion: to achieve acceptable error value the instantaneous band of RF signal must be narrower.

4. TTD-BASED PHOTONIC BEAM FORMER USING SWITCHABLE DELAY LINES

As described above, the main advantage of this arrangement is in the relative simplicity and straightforward implementation, as the scheme is a complete "optical" analogue of a standard microwave PAA scheme. To evaluate its characteristics an

equivalent circuit including the models of previously studied microwave photonics components and built-in AWRDE's library models was developed too. Modelling was also conducted by calculating the PAA's NRP in the frequency band 6-18 GHz.

Figure 4 shows an example of calculation of the NRP in the upper frequency of the band for the photonic beam

former based on switchable optical delay line with 3-ps increments. As one can see from the Figure, the photonic beam former investigated in this section is provided by a uniform spatial shift of the main lobe that is more than 40°. The same was observed on other frequencies of the band.

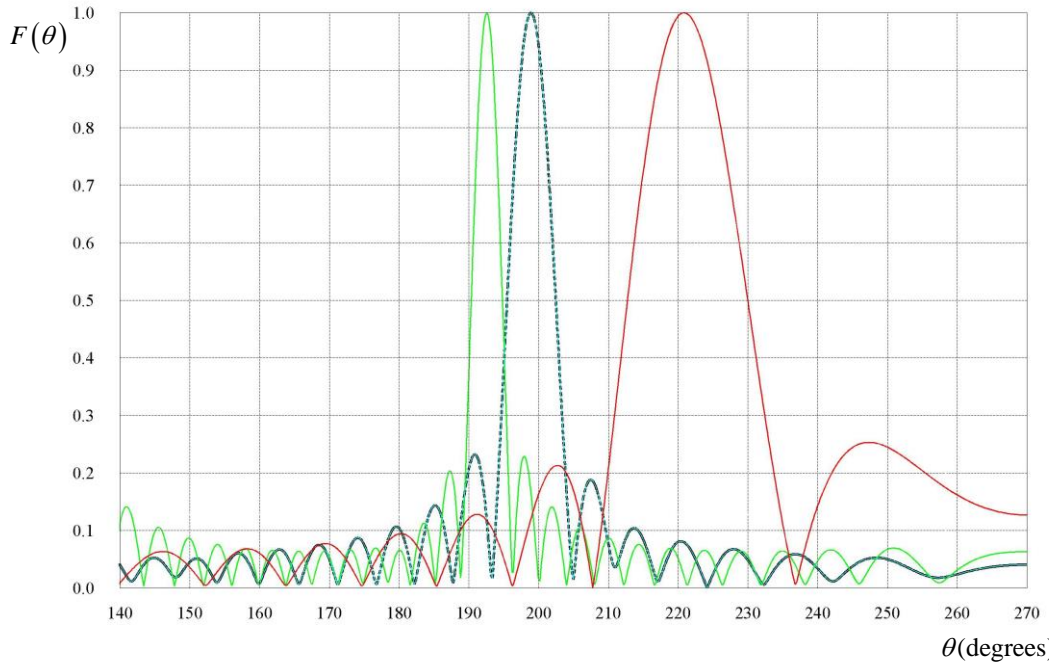


Fig. 3. Microwave photonics PAA's normalized radiation patterns at central frequencies of 6 GHz (right), 12 GHz (middle), and 18 GHz (left)

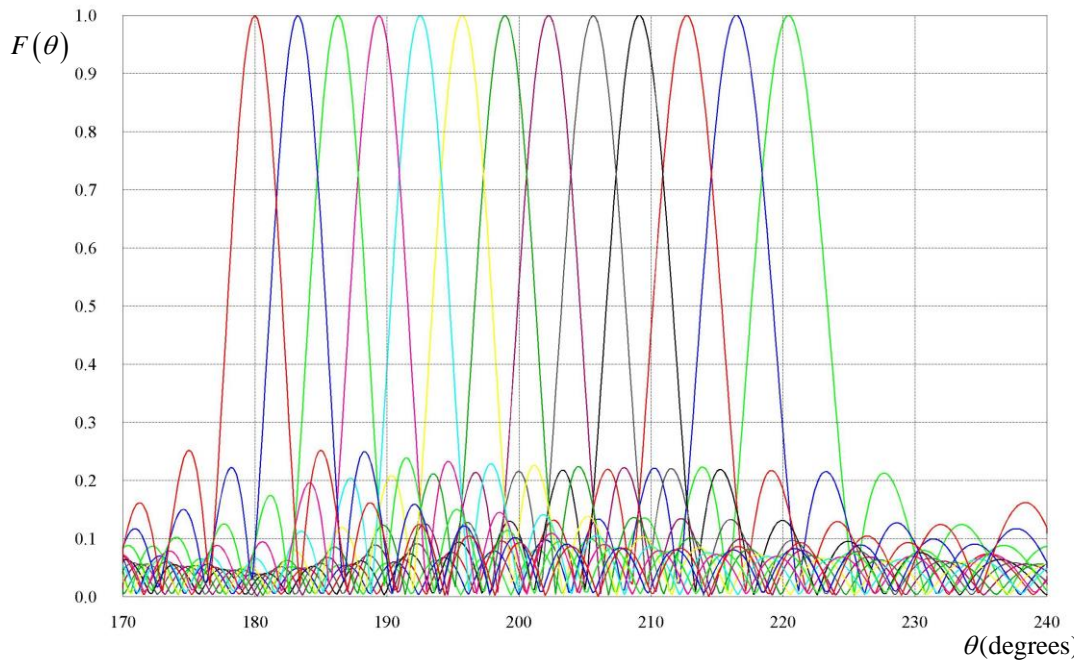


Fig. 4. Microwave photonics PAA's normalized radiation patterns at the frequency 18 GHz for photonic beam former using switchable delay lines

4.1. Super-wide bandwidth photonic beam former combining switchable delay lines and multichannel Bragg grating

The key advantage of the building principle proposed in (Pham, Arokiaswami, and Desmond 2008) is in efficiency when used in multi-element PAA. The results of its simulation and experimental verification are described in more details below.

The model of photonic beam former, implemented in AWRDE CAD's Schematic environment, which combining switchable optical delay lines and multichannel Bragg grating (Figure 5) includes:

- section of semiconductor lasers and spectral multiplexer;
- section of single sideband external modulation including two electro-optic Mach-Zehnder modulators;
- four-channel Bragg grating;
- section of the optical amplifier;
- optical distribution section (ODS), which forms 16 optical channels with preset values of the delay time;
- section of photodetector, which provides detection of 16 RF signals with the set time delays.

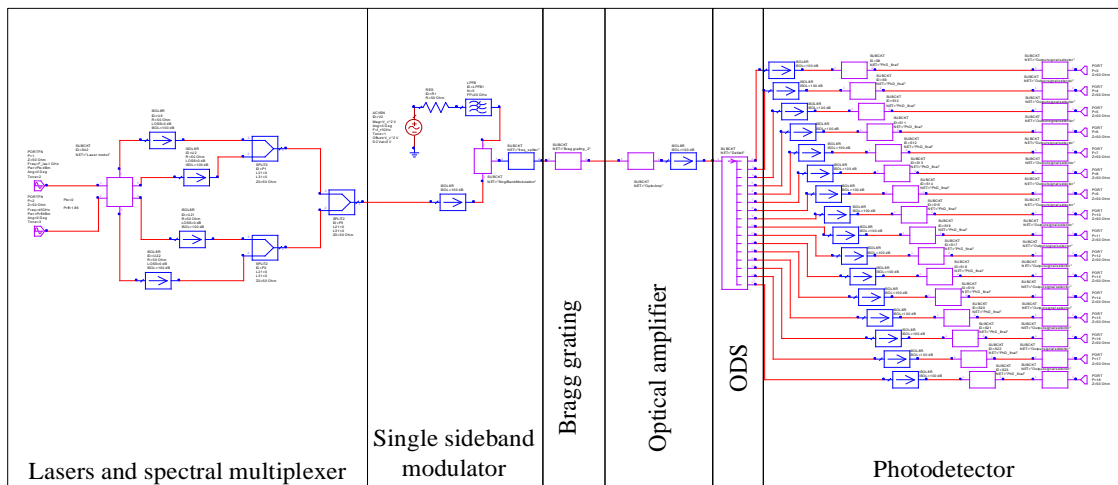
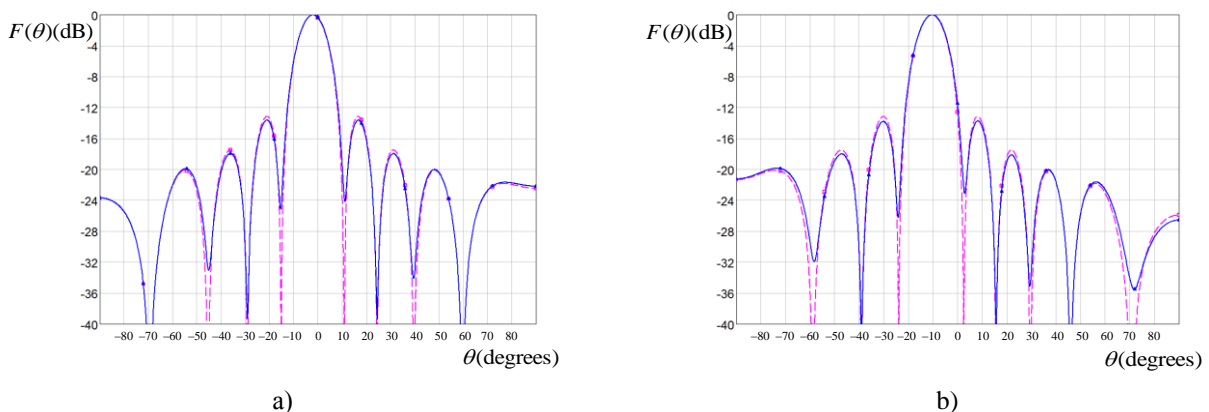


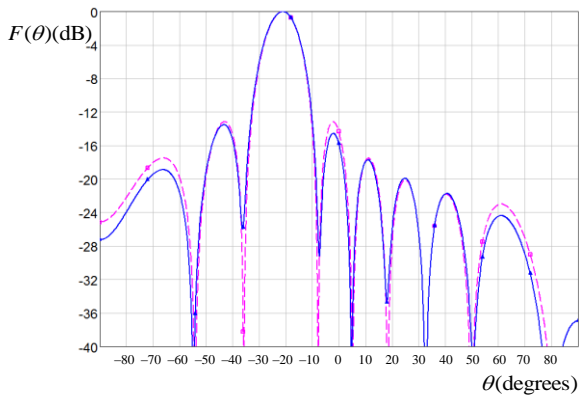
Fig. 5. Equivalent circuit of super-wide bandwidth photonic beam former combining switchable delay lines and multichannel Bragg grating



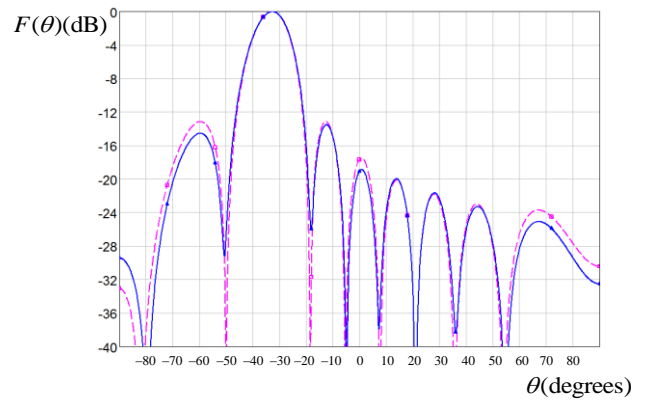
For simulation and experimental verification for the circuit of Fig. 5 the following reference data were taken: Operating frequency band of PAA is 1-18 GHz; The antenna array has 16 non-directional antennas spaced at $0,5\lambda_{\min}$ (corresponding to the maximum operating frequency of 18 GHz). Simulation and experiment are carried out for 2 cases:

- 1) at a fixed microwave frequency 10 GHz when the delay time $\Delta t = 1, 5, 10$ and 15 ps;
- 2) at a fixed time delay $\Delta t = 15$ ps when operating frequencies $F = 6, 12$ and 18 GHz.

Figure 6 depicts the results of calculations (dashed curves) and experiments (solid curve) of NRPs at various time delays. Besides, Figure 7 depicts the results of calculation (dashed curve) and experiment (solid curve) of NRPs at the operating frequencies of 6, 12 and 18 GHz when time delay is fixed on 15 ps. As follows from the Figures, there is a close agreement between the calculated and experimental data, due to the high accuracy of the built library models of AWRDE tool, as well as due to the fact that our models of optoelectronic components are based on their experimentally measured parameters.

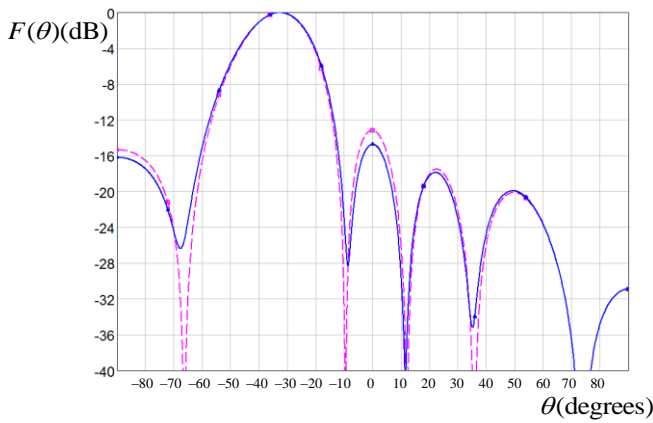


c)

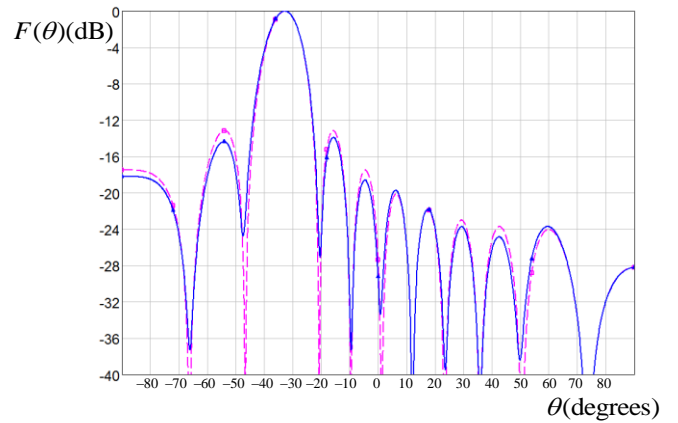


d)

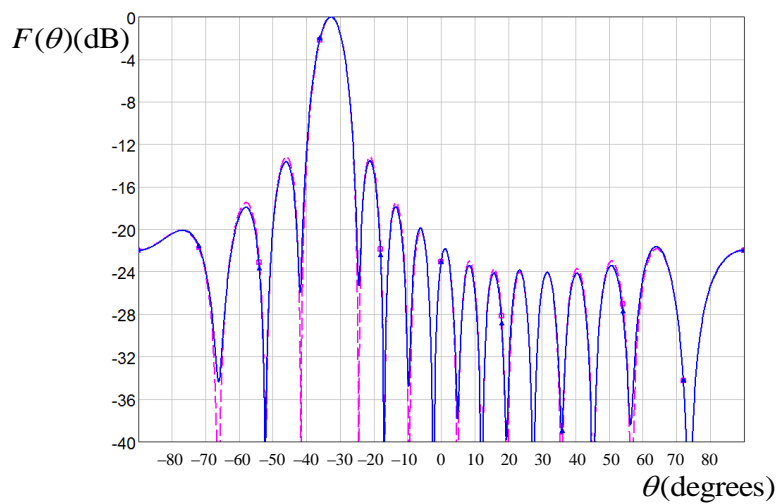
Fig. 6. Calculated (dashed curve) and measured (solid curve) NRPs of microwave photonics PAA at 10 GHz: a - $\Delta t = 1$ ps (offset of the main lobe $\Delta\theta = -2.2^\circ$); b - $\Delta t = 5$ ps (offset of the main lobe $\Delta\theta = -10.4^\circ$); c - $\Delta t = 10$ ps (offset of the main lobe $\Delta\theta = -21.3^\circ$); d - $\Delta t = 15$ ps (offset of the main lobe $\Delta\theta = -32.9^\circ$)



a)



b)



c)

Fig. 7. Calculated (dashed curve) and measured (solid curve) NRPs of microwave photonics PAA at time delay of 15 ps when operating frequency is 6 GHz (a), 12 GHz (b), and 18 GHz (c)

5. CONCLUSION

We have presented the results of design, simulation and testing of super-wide bandwidth photonic beam former for microwave photonics phased array antenna as a complex example of combined information and communication systems based on microwave and photonic industrial technologies using high-power commercial microwave electronic CAD tool. The good fit of the calculated and experimental data is achieved due to the high accuracy of the built library models of AWRDE tool, as well as due to the fact that our models of optoelectronic components are based on their experimentally measured parameters.

LIST OF REFERENCES

- Vincent J. Urick Jr., Jason D. McKinney, Keith J. Williams, 2015, by John Wiley & Sons, Fundamentals of microwave photonics.
- Paoella A.C., DeSalvo R., Middleton C. and Logan C., "Direction in radio frequency photonic systems," in WAMICON Conference Proceedings, p. 1-6, 2015.
- Leijtens X.J.M., Le Lourec P. and Smit M.K., "S-Matrix Oriented CAD-Tool for Simulating Complex Integrated Optical Circuits," IEEE Journal of Selected Topics in Quantum Electronics, 1996, v. 2, No 2, p. 257-262.
- Belkin M.E., Belkin L., Sigov A.S., Iakovlev V., Suruceanu G. and Kapon E., "Performances of Microwave-Band Analog Signal Transmission using Wafer-Fused Long Wavelength VCSELs," IEEE Phot. Technol. Lett., 2011, vol. 23, No 20, p. 1463-1465.
- Belkin M.E., Iakovlev V., "Microwave-band Circuit-level Semiconductor Laser Modeling," in Proc. 9th European Modelling Symposium on Mathematical Modelling and Computer Simulation EMS2015, p. 1-3, 6-8 October 2015, Madrid, Spain.
- Belkin M. E., Sigov, "Circuit-Level Large-Signal Modeling of Microwave Bandwidth Photodetector," in Proc. International Conference on Electromagnetics in Advanced Applications, ICEAA 2015, p. 1587-1589, 7-11 Sept. 2015, Torino, Italy.
- Belkin M.E., Golovin V., "Microwave Electronic CAD Modeling of Microwave-Band Optoelectronic Oscillator Based on Long Wavelength VCSEL," in Proc. International Conference on Microwaves, Communications, Antennas and Electronic Systems, COMCAS 2015, p. 1-3, 2-4 November 2015, Tel Aviv, Israel.
- Belkin M.E., Tyschuk Y., "Microwave Electronic CAD Modeling of Microwave Photonic Devices Based on LW-VCSEL Mixing," in Proc. II International Conference on Microwave and Photonics, ICMAP-2015, p. 1-2, December 11-13, 2015, Dhanbad, India.
- Mailloux R.J.. Phased Array Antenna Handbook. Artech House, Inc. – 496 pp. (2005).

Yao J. , 2009, "Microwave Photonics," IEEE Journal of Lightwave Technology, vol. 27, no. 3, pp. 314-335.

Pham Q.T., Arokiaswami A., and Desmond R.L., 2008, "A Novel Simplified Dual Beam-Former Using Multichannel Chirped Fiber Grating and Tunable Optical Delay Lines," IEEE Journal of Lightwave Technology, vol. 26, no. 15, pp. 2629-2634.

ACKNOWLEDGMENTS

This work is supported by Russian Fund of Basic Researches. The project number is obr-M No14-29-08141.

A SIMULATION MODEL OF HUMAN DECISION MAKING IN EMERGENCY CONDITIONS

Fabio De Felice^(a), Domenico Falcone^(a), Antonella Petrillo^(b), Agostino Bruzzone^(c), Francesco Longo^(d)

^(a) Department of Civil and Mechanical Engineering - University of Cassino and Southern Lazio, Italy

^(b) Department of Engineering - University of Naples "Parthenope", Italy

^(c) Simulation Team, DIME, University of Genoa, Italy

^(d) MSC-LES, University of Calabria, Italy

^(a) defelice@unicas.it, ^(a) falcone@unicas.it, ^(b) antonella.petrillo@uniparthenope.it, ^(c) agostino@itim.unige.it,
^(d) francesco.longo@unical.it

ABSTRACT

The optimization and automation processes taking place in industry to face globalization are significantly increasing the complexity of productive processes. The high complexity requires the adoption of sophisticated methods to manage any emergency situations that may arise in case of an accident. In this context the simulation approach is one of the most innovative tool to manage emergency conditions. Simulation tools seek to represent the different scenarios that may occur and its consequences. The objective of the present research is to develop a simulation model which analyzes emergency scenarios and the impact that the operation and man's decisions on the emergency management. The paper aims to develop a simulator model for approaching emergency management and human errors in industrial and critical infrastructure. A real case study is analyzed.

Keywords: simulation, human error, disaster management, nuclear power plant

1. INTRODUCTION

In recent decades, after the serious incidents developed in industrial plants that led to the loss of many lives, the analysis of emergency situations in critical infrastructure is becoming increasingly important (Yates and Paquette, 2010). The study of critical infrastructure is important, as the company is completely dependent on these infrastructures and emergency situations that may develop into a plant, could have disastrous consequences for the whole city and community (De Felice and Petrillo, 2011). Thus, the analysis of emergency management is carried out both for individual facilities, but also to entire neighborhoods or cities. It must first identify the causes that could lead to accidents and look before the event and find advanced features to reduce the risk of emergency situations (Hernandez and Serrano, 2001). The study of the vulnerability of critical infrastructure is of paramount importance when considering the sociological changes, technology and economics related to privatization and globalization. Obviously the analysis of emergencies is also closely linked to the

analysis of *human reliability and human behaviour*, as is the man the one who manages the emergency situation and sometimes it is the one that created it (Ra'ed and Keating, 2014).

Analysis of human reliability is a multidisciplinary problem, calling for knowledge and expertise from probabilistic safety analysis, plant design and operations, decision science, and the behavioral sciences (Hollnagel, 2002). HRA grew up in the 1960s, with the intention of modeling the likelihood and consequences of human errors (Sharit, 2012).

The researchers' great efforts to propose models of human behavior favoring numerical values of error probability in order to predict and prevent unsafe conduct are clearly evident (French *et al.*, 2008).

Human reliability is a crucial element to ensure industrial plant performance and to manage situations or activities where the stress factors including physiological and psychological stressors like fear, monotonous workload, overload, and so influence on the operator. Human reliability cannot be analyzed in the same manner as that of equipment/component. The main issue with human reliability analysis (HRA) is the *uncertainty* of the data concerning human factors, together with the difficulty in modeling the human behavior (Hollnagel, 2005).

According the above considerations, we can state that nowadays, the analysis of human factors constitute a *highly interdisciplinary field* of study. But it is not yet well defined, therefore, a complete and universally accepted taxonomy of different types of human errors and causes determining them, does not exist (D'Elia *et al.*, 2013).

For this purpose, the study of emergencies management must necessarily identify all the factors that influence human reliability within the accident scenario and its evolution. The emergency services are complex and dynamic, as in the workplace can result in accidents (Reznek *et al.*, 2003). This is even more evident as today industrial systems are becoming more complex, therefore more and more advanced tools for the analysis of incidental situations are needed. The most used tool in different areas is the simulation, as it can be on a computer a series of accident scenarios too complex and

verify their evolution. Using the simulation approach it is possible to evaluate the impact of any changes to the system by running different scenarios. Thus, a simulation approach could be a useful tool to manage emergency operations and to assess their impact (Kuo *et al.*, 2016). The specific purpose of the present study is to develop a *simulation model* for the analysis of accident and emergency management in a nuclear power plant. This is a hot topic. In fact, the use of nuclear power to satisfy energy demand is a controversy ever since, but it has recently gained momentum due to the Fukushima disaster. For the above considerations, in our opinion, it is necessary to analyze in deeper this topic in order to propose a useful tool to manage nuclear disaster. In this context, *emergency management simulation* has become one of the most popular method of preparing decision makers for various scenarios. In this paper, a simulation model which simulates the behavior of an operator controlling a complex system during the management of accidents is described. Particular attention is paid to the theoretical foundations of the model, to its computational implementation. The approach aims to build a structure for the various kinds of cognitive functions that are performed by an operator in complex environments. This paper deals with various aspects of human behavior that can influence operator reliability, considering the environment in which operator is working. The focus is on understanding the nature of human performance variability and eventually, how to describe and analyze it.

The rest of the paper is organized as follows. In Section 2 a literature review on emergency management and simulation approach is presented. Section 3 describes the nuclear power simulation model. In Section 4 a case study is analyzed. Section 5 presents discussion and results. Finally, in Section 6, conclusions are analyzed.

2. LITERATURE REVIEW

In the literature, there are different risk-based approaches reported, ranging from the purely qualitative to the quantitative (Longo and Ören, 2008). Many authors used probabilistic risk assessment. A vast majority of tools and techniques available for the HRA are meant for high risk sectors like nuclear, petrochemical industries, and so on, applied within the context of probabilistic safety assessment (Cacciabue, 1996). In this context, it is necessary to evaluate the alternative operations that can be done to manage the situation (Brady, 2003). The underlying causes of emergency situations may be human, natural or mechanical. Human causes are mainly due to human error, natural causes are due to natural disasters such as earthquakes, tsunamis, hurricanes, etc., while the mechanical causes are due to breakdowns of machines. There are various models for the management of emergencies. In particular it is possible to consider 2 main approaches, as follows:

a) *Operations Research*, which provides mathematical tools to support decision-making activities, coordinating activities with scarce

resources in order to maximize an objective function (Shannon *et al.*, 1980);

b) *Multicriteria analysis* that analyzes a number of important aspects and decision makers that make the difficult procedure due to a single goal (De Felice *et al.*, 2016).

The evaluation of the results obtained with these methods is achieved by *simulation*. The simulation is the true representation of real systems on a virtual platform. In the literature, there are many simulation models for the management of emergencies (Bruzzone *et al.*, 2014). Some examples of simulation model are defined below. For instance, Schafer *et al.* (2007) supports community emergency management planning through a geo-collaboration software architecture. It examines geo-spatial maps together and it develops emergency plans and procedures. The software architecture facilitates the development of geo-collaboration solutions. Geo-collaboration tools can be used for emergency management planning. Currión *et al.* (2007) present Sahana Eden, a simulation tool to manage coordination problems faced during an emergency, how to search for missing people, managing aid, and it maintains contacts in the fields of nongovernmental groups, civil society and the victims themselves. Rauschert *et al.* (2002) have a multimodal GIS Interface to Support Emergency Management. In this model the external environment plays a predominant role. With the recent spread of social networks, including in the area of emergency management are trying to develop similar tools. White *et al.* (2009) are studying how social networks can be used for emergency management. The literature presents several simulation models for the management of emergencies based on the behavior and human choices. A cognitive simulation represents the mental processes of the operators during the execution of their duties (1998). The simulation of human action can be qualitative and it represents the evolution of a mental process, or it can be accompanied by calculations and is a quantitative simulation. The output of a qualitative simulation can be a list of actions and errors committed by the operator during work. While in a quantitative simulation there are also the numerical values resulting from the system (Sasou *et al.*, 1995).

Some of the most well-known simulation models are follows:

1. Simulator for Human Error Probability Analysis (SHERPA): It is an integrated simulation model between the HEART method and SPAR-H method. The simulator evaluates the error probability. It can be used both in a preventive phase and in a phase retrospective. It is a quantitative model (Di Pasquale *et al.*, 2015).
2. Probabilistic Cognitive Simulator (PROCOS): It is a model that mimics the behavior of an operator. The model develops a simulator that analyzes the errors. It is a quantitative simulator (Trucco and Leva, 2007).

3. Simulation System for Behavior of an Operating group (SYBORG): The model simulates a group of workers at a nuclear plant. They show some possible combinations of operator errors that can lead to sequences of accidents. The model proposes several strategies to improve collaboration in the group. It is a qualitative simulator (Kirwan, 1998).
4. Cognitive Environment Simulation (CES): It simulates the behavior of a control operator in a nuclear power plant during an emergency situation. The model is developed using artificial intelligence (Woods *et al.*, 1987). It is a semi-qualitative simulator.

3. THE PROPOSED SIMULATION MODEL

The simulation model, presented in this research, aims to identify emergency situations that may occur in different scenarios and quantitatively analyzes the probability of human error during the handling of the emergency phase. The model is divided into several steps, as shown in Figure 1 and as defined follows:

- **Step#1. Activities choice:** It identifies the activities on which simulate the emergency condition, identifying the actions that must make the operator to handle the emergency.
- **Step#2. The emergency situation is simulated at time t.** They identify a number of possible accident scenarios.
- **Step#3. The Generic Tasks (GTTs)** related to the operator's activities are described. Each generic task is defined by a Weibull reliability function that decreases with the increase of working hours. The reliability is maximum in the first hour of work and minimum at the eighth hour of work.

- **Step#4. The nominal error probability** is calculated using the Weibull function related to the generic tasks considered.

$$\begin{cases} HEP_{nom}(t) = 1 - k * e^{-\alpha(1-t)^\beta} & \forall t \in [0; 1] \\ HEP_{nom}(t) = 1 - k * e^{-\alpha(t-1)^\beta} & \forall t \in]1; \infty[\end{cases}$$

This error probability is theoretical and does not take into account the influence of external factors.

- **Step#5. External influences** are inserted into the model through the use of Shaping Performance Indicators (PSF's) (Table 2).

Table 1: Definition of Performance Shape Factors

PSF	Scen. 1	Scen. 2	Scen. n
Decision Time			
Expertise			
Procedure			

The PSF's are indexes that allow you to manage the accident scenarios simulating an improvement or a worsening in the operator's working conditions.

$$PSF_{comp} = \prod_{i=1}^n PSF_i$$

- **Step#6.** Finally, by combining the nominal error probability and the values of the PSF's the simulator returns as output the **real error probability**, influenced by the external factors.

$$HEP_{cont} = \frac{HEP_{nom} * PSF_{comp}}{HEP_{nom} * (PSF_{comp} - 1) + 1}$$

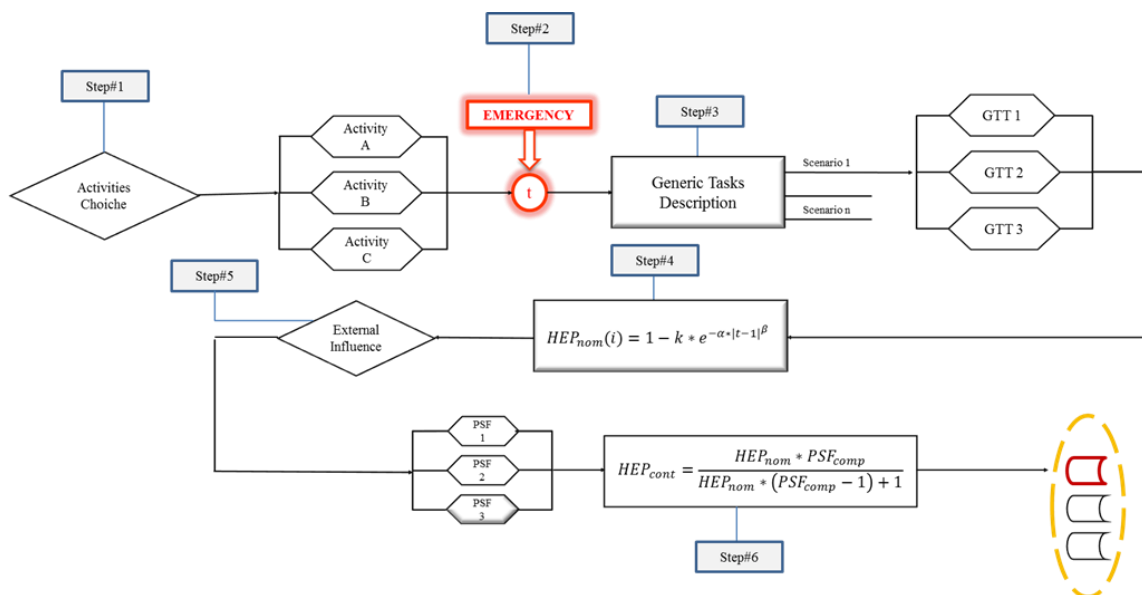


Figure 1: Simulator flowchart

4. THE EXPERIMENTAL SCENARIO

The simulation model replicated the working environment in a way that, overall, is perceived as realistic. The scenario under study analyzed a nuclear power plant. The reason to analyze this specific scenario is because worldwide there are several nuclear plants that could

cause high-known risks. Figure 2 shows the map of European nuclear plants. The analyzed scenario aimed to analyze a meltdown occurred at reactor. Figure 3. shows the immersive virtual environment of the Nuclear plant.

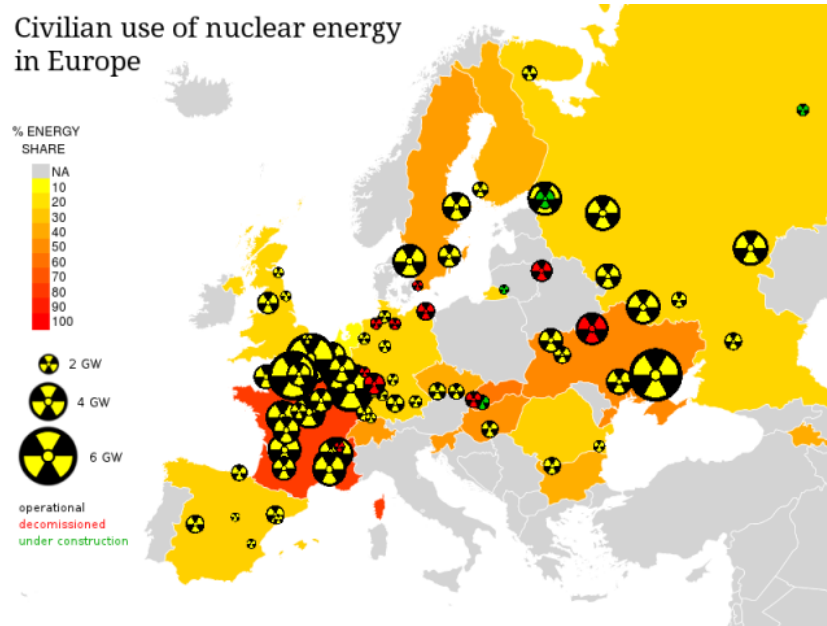


Figure 2: Nuclear energy in Europe, with country colors as nuclear energy share on total energy and symbols representing nuclear power plants with more than one Gigawatt. Source: IAEA - International Atomic Energy Agency



Figure 3: Immersive Virtual Environment of the Nuclear plant

While in Figure 4 is shown a typical replica of a Control Room. A schema of nuclear power plant is shown in Figure 5.



Figure 4: A typical replica of a Control Room

The aim of the nuclear power plant is to produce as much power as possible without causing a meltdown. But it is essential to run the plant up not beyond its safe operating limits. For this purpose it is important to set plant parameters and to check temperatures and power output day by day. Figure 6 shows an example of parameters setting. However, some problems could occur and cause a meltdown, as shown in Figure 7.

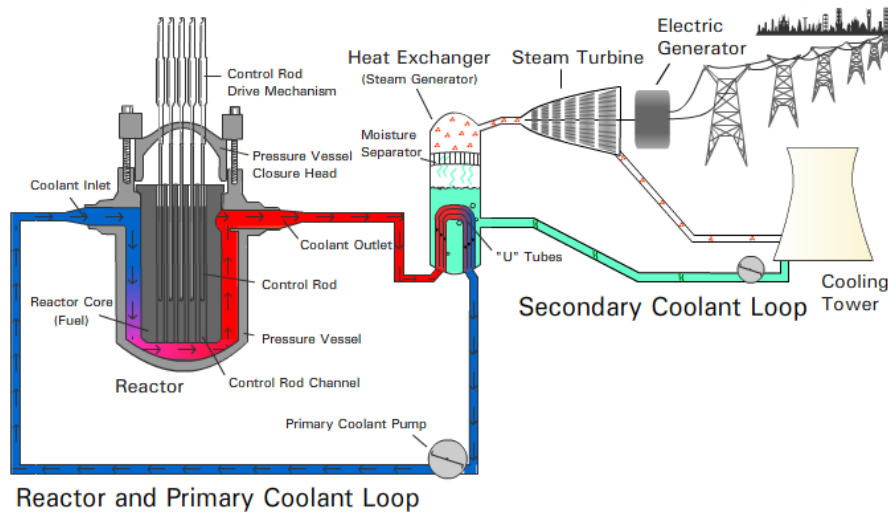


Figure 5: Nuclear power plant – schema

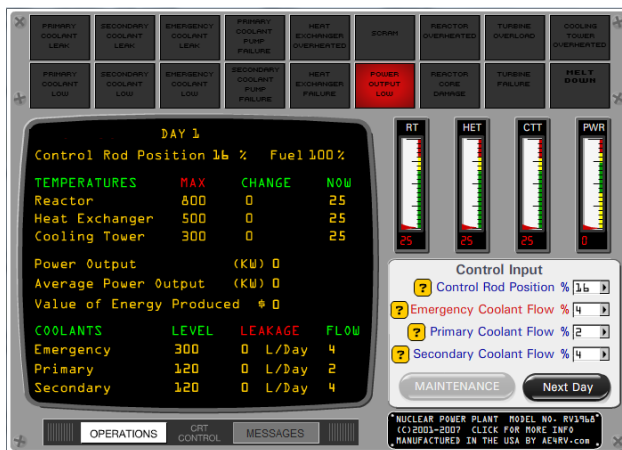


Figure 6: Example of initial parameters



Figure 7: Example of meltdown

The case study is focused on the management of an emergency by operator in the control room, during the meltdown of the reactor in the nuclear power plant.

The operator in the control room, ascertained and understood the magnitude of the emergency activates emergency plan that includes:

- Emergency signal activation;
- Activation cooling system;
- Staff evacuation;
- Alert external rescue system;
- Isolating damaged area;
- Emergency end.

The three simulation scenarios are:

1. **Low Hazard:** occurring while the emergency the decision maker has been monitoring the situation;
2. **Moderate hazard:** the occurrence of the emergency the decision maker can take wrong decisions;
3. **High Hazard:** concerns the decision maker can make a mistake with a good chance.

In the case study is considered the generic tasks in Table 3 that the best represents the operator’s activities.

Table 3: Generic Task Analyzed

Generic Task	Limitations of unreability
Routine, highly-practised, rapid task involving relatively low level of skill	0,007-0,045

Considering the three emergency conditions PSF’s are defined in Table 4 and the the overall PSF index is calculated for the three condition.

Table 4: Numerically Performance Shape Factors

PSF	Low	Moderate	High
Time	0,01	0,1	1
Expertise	0,1	1	3
Procedure	1	5	20
PSF comp	0,001	0,5	60

5. DISCUSSION AND RESULTS

The results of the simulation of the model defining the nominal error probability and the probability of the contextualized error. (Table 5).

The nominal probability of error is theoretical and it depends only on the considered generic tasks, while the contextualized probability of error depends on the external environment function and it is affected by the PSF values. The simulator returns three different error probability associated with the three different scenarios that can be considered:

- Low Hazard;
- Medium Hazard;
- High Hazard.

As the Table 5 shows the error values are greater for high hazard scenarios.

Table 5: Simulation Results

Generic Task	HEP _{nom} (t)		Low Hazard	Medium Hazard	High Hazard
	t	HEP _{nom}	HEP _{cont}	HEP _{cont}	HEP _{cont}
Routine highly practiced, rapid task involving relatively low level of skill	t=1	0,007	7,05E-06	3,51E-03	2,97E-01
	t=2	0,009	9,18E-06	4,57E-03	3,55E-01
	t=3	0,012	1,31E-05	6,49E-03	4,39E-01
	t=4	0,017	1,81E-05	8,98E-03	5,21E-01
	t=5	0,023	2,42E-05	1,19E-02	5,92E-01
	t=6	0,030	3,10E-05	1,53E-02	6,51E-01
	t=7	0,037	3,87E-05	1,90E-02	6,99E-01
	t=8	0,045	4,71E-05	2,30E-02	7,39E-01

Figure 8 shows the trend of the PSF as a function of the different scenarios considered.

The performance of the PSF has a fundamental role in the simulation, because it represents the environmental conditions under which the operator must work. Figure 9 shows the result of simulation for the worst case scenario concerning a high risk. The model results are affected by the Weibull reliability function and in fact the reliability of the operator decreases with time, and on the contrary it increases the probability of error. Moreover, the results of the graph are influenced by the working conditions presented with the PSF that worsen the error probability.

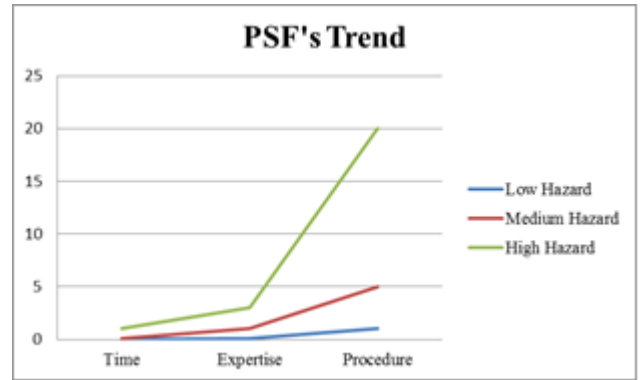


Figure 8: Trend of PSF's

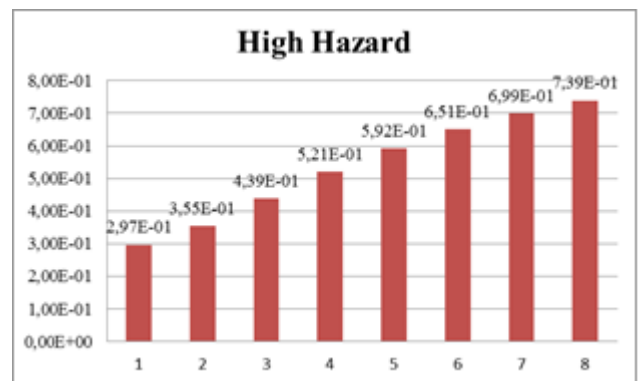


Figure 9: Simulation Results

6. CONCLUSION

Operations of safety-critical systems during emergency conditions, such as the nuclear plant, are certainly difficult to manage. Failing to identify the proper action to do in emergency could cause high risk. Thus, the proposed simulation model makes it possible to analyze how an action influences critical activity. The proposed approach is semi-probabilistic. Using the model it is possible to take into consideration different contexts by modifying the PSFs involved. A weakness of the proposed model is that in this early stage does not perform a PSFs time-dependent simulation process. For this reason, future research will consider this kind of dependence.

REFERENCES

Brady, T.F., 2003. Public health: emergency management: capability analysis of critical incident response. In Proceedings of the 35th conference on Winter simulation: driving innovation (pp. 1863-1867). Winter Simulation Conference, December 2003.

Bruzzone, A.G., Frascio, M., Longo, F., Chiurco, A., Zandoni, S., Zavanella, L., Fadda, P., Fancello, G., Falcone, D., De Felice, F., Petrillo, A., Carotenuto, P., 2014. Disaster and emergency management simulation in industrial plants. Proceedings of the International Conference on Modeling and

- Applied Simulation (EMSS 2014). September 10-12, Bordeaux, France, pp 649-656.
- Cacciabue P.C., 1996. Understanding and modelling man-machine interaction. *Nuclear Engineering and Design*, 165:351-358.
- Cacciabue, P.C., 1998. Modelling and simulation of human behaviour in system control. London. Springer & Verlag.
- Curron, P., Silva, C.D., Van de Walle, B., 2007. Open source software for disaster management. *Communications of the ACM*, 50(3), 61-65.
- D'Elia, C., De Felice, F., Mariano, P., Petrillo, A., Ruscino, S., 2013. A bayesian network analysis for safety management, 25th European Modeling and Simulation Symposium, EMSS 2013. Athens; Greece; 25 September 2013 through 27 September 2013; Pages 446-455.
- De Felice, F., Petrillo, A., 2011. Methodological Approach for Performing Human Reliability and Error Analysis in Railway Transportation System. *International Journal of Engineering and Technology*, 3, 341-353.
- De Felice, F., Petrillo, A., Zomparelli, F., 2016. A Disaster Risk Management Performance Index to Assess Safety and Security in Industrial Plants. *Applied Mechanics & Materials*, 84, 373-378.
- Di Pasquale, V., Miranda, S., Iannone, R., Riemma, S., 2015. A Simulator for Human Error Probability Analysis (SHERPA). *Reliability Engineering & System Safety*, 139, 17-32.
- French et al., 2009. Human Reliability Analysis: A Review and Critique. Manchester Business School Working Paper.
- Hernandez, J.Z., Serrano, J.M., 2001. Knowledge-based models for emergency management systems. *Expert Systems with Applications*, 20(2), 173-186.
- Hollnagel E., 2002. Understanding accidents – from root causes to performance variability. *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, 15-19 September, Scottsdale, Arizona, 2002; 1-6.
- Hollnagel E., 2005. Human reliability assessment in context. *Nuclear Engineering and Technology*, 37(2):159-166.
- Kirwan. B., 1998. Human error identification techniques for risk assessment of high risk systems—Part 1: review and evaluation of techniques. *Applied Ergonomics*. Volume 29, Issue 3, Pages 157-177.
- Kuo, Y.H., Rado, O., Lupia, B., Leung, J.M.Y., Graham, C.A., 2016. Improving the efficiency of a hospital emergency department: a simulation study with indirectly imputed service-time distributions. *Flexible Services and Manufacturing Journal*, 28(1-2), 120-147.
- Longo, F., Ören, T., 2008. Supply chain vulnerability and resilience: A state of the art overview. 20th European Modeling and Simulation Symposium, EMSS 2008, pp. 527-533.
- Sharit J. Human error and human reliability analysis. In *Handbook of Human Factors* (4th edn). Salvendy G (ed). John Wiley and Sons, Inc.: Hoboken, New Jersey, 2012; 737.
- Ra'ed, M.J., Keating, C.B., 2014. Fragility of oil as a critical infrastructure problem. *International Journal of Critical Infrastructure Protection*, 7(2), 86-99.
- Reznek, M., Smith-Coggins, R., Howard, S., Kiran, K., Harter, P., Krummel, T., 2003. Emergency Medicine Crisis Resource Management (EMCRM): Pilot study of a simulation-based crisis management course for emergency medicine. *Academic Emergency Medicine*, 10(4), 386-389.
- Rauschert, I., Agrawal, P., Sharma, R., Fuhrmann, S., Brewer, I., MacEachren, A., 2002. Designing a human-centered, multimodal GIS interface to support emergency management. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems* (pp. 119-124).
- Sasou, K., Takano, K., Yoshimura, S., Haroko, K., Kitamura, M., 1995. Modelling and simulation of operator team behaviour in nuclear power plants. In: *Proceedings of the HCI international '95*, Tokyo.
- Schafer, W.A., Ganoë, C.H., Carrol, J.M., 2007. Supporting community emergency management planning through a geocollaboration software architecture. *Computer Supported Cooperative Work (CSCW)*, 16(4-5), 501-537.
- Shannon, R.E., Long, S.S., Buckles, B.P., 1980. Operation research methodologies in industrial engineering: A survey. *AIIE Transactions*, 12(4), 364-367.
- Trucco, P., Leva, M.C., 2007. A probabilistic cognitive simulator for HRA studies (PROCOS). *Reliability Engineering & System Safety*, 92(8), 1117-1130.
- White, C., Plotnick, L., Kushma, J., Hiltz, S.R., Turoff, M., 2009. An online social network for emergency management. *International Journal of Emergency Management*, 6(3-4), 369-382.
- Woods, D.D., Roth, E.M., People, H.E., 1987. Cognitive environment simulation: an artificial intelligence system for human performance assessment. (Technical report NUREG-CR-4862, US Regulatory Commission, Washington DC, US.
- Yates, D., Paquette, S., 2011. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International journal of information management*, 31(1), 6-13.

PERFORMANCE OF FDS VERSIONS 5 AND 6 IN PASSENGER CAR FIRE COMPUTER SIMULATION

Weisenpacher, P.^(a), Glasa, J.^(b), Sipkova, V.^(c)

^{(a),(b),(c)}Institute of Informatics, Slovak Academy of Sciences, Dubravská cesta 9, 84507 Bratislava, Slovakia

^(a)upsyweis@savba.sk, ^(b)utrrglas@savba.sk, ^(c)upsyvvsip@savba.sk

ABSTRACT

In this paper, the performance of FDS versions 5 and 6 and their accuracy in high performance parallel simulations of passenger car fire is evaluated. The results of measurements obtained during a full-scale fire experiment conducted in Povazsky Chlmec (Slovakia) in 2009 have been used to test the geometrical representation and material properties of two passenger cars and their passenger compartments used in the model. The tested simulation scenario includes a fire in passenger compartment of one car and its spread to an adjacent car. A series of simulations is executed on a cluster of computers using the NIST FDS system, versions 5.5.3 and 6.1.2. The differences between the simulation results obtained by both FDS versions are described and discussed.

Keywords: passenger car fire, fire simulation, parallel simulation, high performance computing

1. INTRODUCTION

Advances in Computational Fluid Dynamics (CFD) as well as increasing performance of computers and clusters of computers have recently enabled to simulate the course of fire for safety purposes in unprecedented detail. CFD simulations dealing with smoke propagation in car parks and road tunnels have achieved significant level of reliability and practical applicability. Knowledge on car fire behaviour is an important factor contributing to efficiency of fire safety measures and to CFD simulations applicability. Extensive experimental research has been conducted in order to investigate the course of single car fire course as well as the spread of fire to adjacent vehicles (Mangs and Keski-Rahkonen 1994; Shipp and Spearpoint 1995; Okamoto et al. 2009; Zhao and Kruppa 2004). These experimental results determined the heat release rate (HRR), the mass loss rate, the heat flux, temperatures and other quantities inside and outside the car inevitable for fire description. CFD simulations of car fires focus mainly on smoke propagation (for example Deckers et al. 2013) as the smoke pose the main threat for people's safety.

Prevailing approach in the literature is to model the source of fire (usually burning cars) as a pool fire, i.e., with a constant or prescribed HRR. This approach is well justified by the need for a well-defined fire source

in order to compare the performance of different ventilation systems. However, in some cases a mutual interaction between the fire and the ventilation (or sprinklers) must be taken into account and the assumption of prescribed HRR not depending on air flows is of limited use. In (Weisenpacher, Glasa, and Halada 2016) we elaborated the model of passenger car with its interior modelled in detail and with specified parameters of corresponding materials for the purposes of computer simulation of passenger car fire by Fire Dynamics Simulator (FDS). The reliability of such car representation is based on comparison with the results of full-scale experiments carried out in open air in Povazsky Chlmec (Slovakia) in 2009. The experiments included a fire in car interior and its spread to an adjacent vehicle (Polednak 2010; Svetlik 2010). During this experiment, temperatures inside and outside passenger compartment were measured to be used for the calibration of computer simulation. Subsequent simulations confirmed credibility of this model.

The model enables to simulate the fire of passenger car located in a car park or a tunnel and to capture the response of the fire on ventilation performance. Similar approach was used in (Partanen and Heinisuo 2013) to investigate the effect of sprinklers on fire of three cars in a car park. Due to significant computational requirements caused by the combination of fine mesh resolution and large dimensions of compartments in which fire occurs, parallel computation is inevitable for practical purposes. However, the process of parallelization in FDS leads to a loss of accuracy. The performance and accuracy of FDS simulations, version 5.5.3 is evaluated in (Weisenpacher, Glasa, and Halada 2016). For practical applicability and validation of simulations performed by new versions of the simulation software it is also necessary the ability to evaluate and compare the results of simulations executed by different versions of simulation software, especially if new models of particular phenomena or improved computational algorithms are included in the new version. Such comparison may lead to an important adjustment of the used geometrical representation of burning objects or their material properties and to enable significant improvement of the simulation accuracy. In this paper, we compare the performance and accuracy of parallel computation performed by the

FDS versions 5 and 6 and discuss specific features of the simulated fire behaviour.

2. MANAGEMENT OF SIMULATION RUNS

Fire Dynamics Simulator is CFD simulation system capable to simulate fire-driven fluid flow, developed at National Institute of Standards and Technology (NIST), USA (McGratten et al. 2010; McGrattan et al. 2013).

FDS is a Fortran program solving conservation equations that describe the evolution of fire. It reads input parameters from a text file, computes a numerical solution to the governing equations, and writes user-specified output data to files.

FDS has been developed to run on a variety of hardware platforms under operating systems MS Windows, Mac OS X and Linux. The employed computer ought to have fast processors (CPUs) – this will determine how long the computation will take to finish, and at least 4 GB RAM per processor – this determines how many mesh cells can be held in memory. To save the output of the calculation a large hard disk is required, since the results can consume more than 1 GB of storage space. FDS supports the configuration of various programming models – a serial model, designed for running on a single computer, and several parallel models, designed for running on multiple computers. There are three ways to execute FDS in parallel:

- OpenMP (Open Multi-Processing), model exploits multiple cores on a single computer or a compute cluster node (OpenMP, Internet source).
- MPI (Message Passing Interface) model exploits multiple processors distributed over the computers on a network or cores embodied in a compute cluster. In this case the computational domain must be divided into multiple meshes and typically, each mesh is assigned its own MPI process (Open MPI, Internet source).
- Combined MPI and OpenMP model, where each MPI process can be supported by OpenMP threads. This approach enables a two-level parallelization: at first, the computational domain is break up into multiple meshes (MPI), and then within each mesh the multi-threading (OpenMP) on some selected code regions is applied.

Most of the speedup is achieved by the MPI which is the better choice for multiple mesh simulations. OpenMP works best when exploiting multiple cores associated with a single (physical) processor or a node socket. It can speed the calculation by an extra factor up to about 2.

Simulations were performed using the FDS package versions 5.5.3 and 6.1.2. Both have been installed on the HP compute cluster (SIVVP, Internet source) located at the Institute of Informatics, Slovak Academy of Sciences, Bratislava. The cluster consists of 54

compute nodes, each comprising of two 6-core processors Intel E5645 @2.4 GHz, 48 GB of RAM and 2x500 GB scratch disks. All nodes are connected by the Infiniband interconnection network with the bandwidth of 2x40 Gbit/s per link and direction. In addition, the cluster includes 10 nodes with GPU accelerators NVIDIA Tesla K20 interconnected via the 8 Gbps Ethernet network, two 72 TB disks, and several managing servers. The cluster is running the OS Scientific Linux.

For comparison purposes the source code of FDS 5.5.3 was compiled using two compilers: the GNU 4.4.7 (gfortran, gcc, OpenMP) and Intel 14.0.1 (ifort, icc, OpenMP). The source code of FDS 6.1.2 was compiled using the compiler Intel 14.0.1. Parallel models of FDS are built with Open MPI version 1.10.0.

The evaluation of simulations results requires carrying out a great number of experiments. In order to facilitate and automate the whole simulating process, for each model we have developed a couple of command line scripts, called *fds-manager* scripts, which enable the user in an efficient way to submit FDS jobs to the cluster. Scripts are created in the Shell language including submission commands of the underlying middleware. By default, the execution of jobs on a local cluster is performed through the PBS (Portable Batch System, Internet source) that represents a workload management system providing a unified batch queuing and job management interface to a set of computing resources. The *fds-manager* script is responsible for the accomplishment of the following actions:

1) Initially, it accepts and checks the given input parameters specifying the FDS input file and asked cluster configuration: the number of nodes and cores, and eventually, the number of MPI processes and number of OpenMP threads. Parameters values are used in the subsequent operations.

2) Based on given input parameters, it produces the description of the application, called *fds-submission* script, which serves as the input to the submission command “PBS *qsub*”. The *fds-submission* files are written in Shell including PBS commands.

3) Finally, it provides for the execution of the FDS simulation using the previously generated *fds-submission* script. Each simulation instance is running in its own working directory.

The complete execution of the FDS simulation is realized by the invocation of the appropriate *fds-manager* script accompanied with the input parameters.

3. CAR FIRE EXPERIMENT AND ITS FDS SIMULATION

In November 2009 we performed a full-scale experiment of a passenger car interior fire in open air and its spread to an adjacent vehicle (see Fig. 1). The experiment was conducted in the testing facilities of the Secondary School of Fire Protection of the Ministry of Interior of the Slovak Republic in Povazsky Chlmec (Polednak 2010, Svetlik 2010). The fire was initiated in a new functional car Kia Cee'd. The right front and left

rear side windows were broken in order to increase the oxygen supply. The second car was an older model of BMW, located lengthwise in the 50 cm distance. Gas temperatures inside and outside the cars were measured by thermocouples. The fire behaviour was observed and recorded by infra-red and digital cameras.



Figure 1: Fire experiment in Povazsky Chlmec

The fire was ignited by burning of a small amount of gasoline (about 10 ml) placed onto the back seat behind the Kia driver's seat. The fire grew progressively and after flashover at the 150 s into the fire the whole passenger compartment of the Kia was involved. During the next minutes remaining car windows were broken and the temperature inside the interior reached the value of 1000°C. After 7 minutes the rubber sealing of the nearest window of the BMW ignited. The fire was suppressed after 12 minutes of the experiment. The fire scenario was subsequently simulated by FDS.

The simulated scenario includes two cars in configuration corresponding to the experiment (see Figures 2, 3). The first car model includes the interior equipment consisting of seats, a dashboard with a steering wheel and interior lining. This equipment is modelled by two materials: 'UPHOLSTERY' for seats and 'PLASTIC' for other equipment. The fire source is represented by a 6 x 6 cm burning surface with the 1000 kW.m⁻² heat release rate per unit area (HRRPUA) placed on the back seat. The second car includes a window rubber sealing at the place where ignition was observed during the experiment. According to the experimental observation its interior does not burn in the simulation. Material properties and the conditions of windows breakage are discussed in detail in (Weisenpacher, Glasa, and Halada 2016) in which FDS 5.5.3 simulation of the described scenario is performed

and evaluated and the accuracy of the burning car representation is tested.

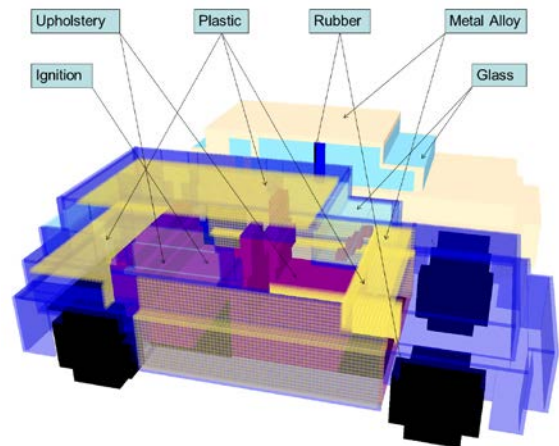


Figure 2: Car representation and used materials

In order to compare both FDS versions correctly, three modifications of the simulated scenario were made. First, we used more intensive initial fire to fulfil condition required by (McGratten et al. 2010) relating fire HRR and computational mesh resolution which is sufficient for accurate simulation (the condition is not fulfilled for less intensive experimental fire). Nine times more intensive fire produced by 18 x 18 cm burning surface with the same HRR prevents small random fluctuation of particular quantities, as discussed in (Weisenpacher, Glasa, and Halada 2016). Second, the duration of the simulated fire is 420 s to avoid extremely long CPU times. Finally, we explicitly determined chemical reaction 'ETHYLENE' (required in FDS 6) for both FDS versions. Due to these differences, simulated scenario outputs cannot be directly compared with the experimental values.

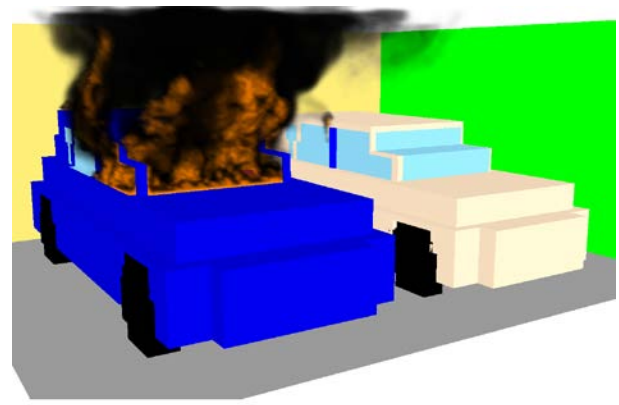


Figure 3: Fire behaviour after 400 s of fire

The computational domain size is 576 x 486 x 240 cm with the 3 cm mesh resolution. The total number of cells is 2,488,320. Sequential calculation (1M) and six variants of parallelisation are evaluated: 12M, 24M, 48M, 96M, 192M and 288M, in which the computational domain is divided into 12, 24, 48, 96,

192 and 288 computational meshes, respectively, each of which is then assigned to one CPU core.

4. SIMULATION RESULTS

4.1. Performance and accuracy of the simulations

The simulation results documenting the performance and accuracy of particular simulation variants executed by both FDS versions are shown in Tables 1 and 2 where t_{400} is the time at which the thermocouple located in the middle of passenger compartment front part (the same place as in the experiment) reached the value of 400°C for the first time, t_{br} is the time at which the windscreen of the first vehicle fell out, and t_{ign} is the time at which the window lining of the second car ignited. The parameter t_{400} indicates the time of flashover inside the passenger compartment.

Table 1: Simulation performance and accuracy, version FDS 5.5.3.

Simulation Results						
	CPU Time [hrs]	Speed up	Efficiency	t_{400} [s]	t_{br} [s]	t_{ign} [s]
1M	444.1	1.00	1.00	50	132	281
12M	58.9	7.54	0.63	49	131	288
24M	30.5	14.55	0.61	52	138	285
48M	15.9	27.95	0.58	51	135	289
96M	8.80	50.48	0.53	51	137	286
192M	4.32	102.9	0.54	79	175	-
288M	2.92	151.7	0.53	81	174	390

Table 2: Simulation performance and accuracy, version FDS 6.1.2.

Simulation Results						
	CPU Time [hrs]	Speed up	Efficiency	t_{400} [s]	t_{br} [s]	t_{ign} [s]
1M	975.1	1.00	1.00	50	129	-
12M	208.2	4.68	0.39	44	121	-
24M	111.3	8.76	0.37	43	129	-
48M	63.1	15.46	0.32	41	126	-
96M	36.2	26.92	0.28	42	126	-
192M	21.1	46.05	0.24	43	127	385
288M	14.78	65.97	0.23	43	124	-

CPU time of FDS 6 sequential calculations is about 2.2 times longer than in the case of FDS 5 simulations, which is consistent with (McGrattan et al. 2013). Moreover, the values of speed up of FDS 6 parallel calculations are lower and their efficiency decreases more considerably than for the corresponding values obtained by FDS 5. CPU time of FDS 6 288M calculations is about 5 times longer than the CPU time of the corresponding FDS 5 calculation. The probable reason is larger information transfer between mesh

boundaries (i.e., between nodes) which increases the simulation accuracy at increased cost of CPU time.

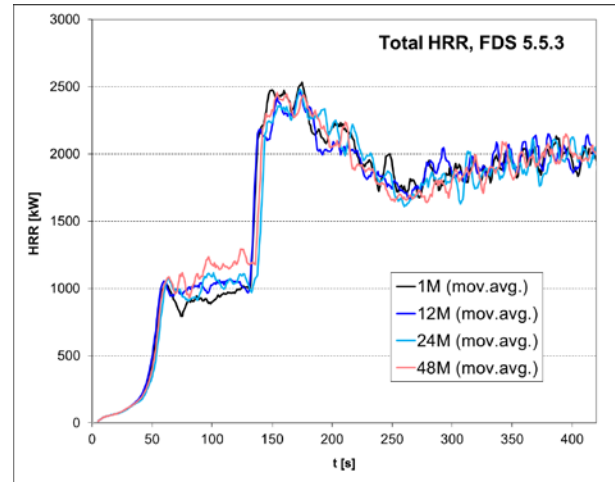


Figure 4: HRR simulated by FDS 5.5.3 for decompositions up to 48M

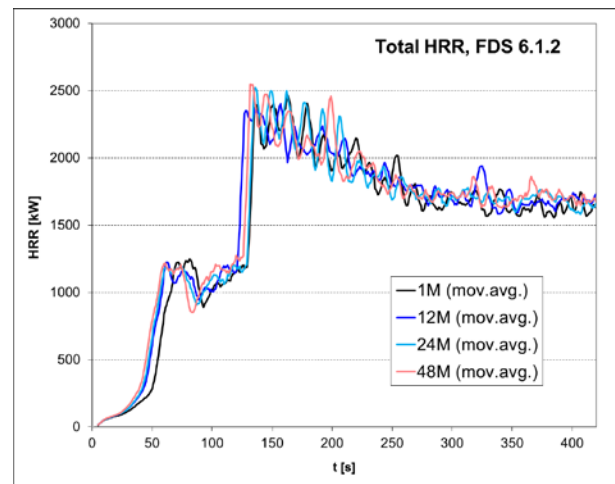


Figure 5: HRR simulated by FDS 6.1.2 for decompositions up to 48M

4.2. Heat Release Rate Behavior

Figures 4 – 7 show the HRR of simulations, which is the main quantity characterising the fire. The behaviour of 1M simulations during the first 250 s of fire is very similar for both FDS versions. The HRR increase at about 50 s into the fire is caused by the flashover in passenger compartment, while the HRR achieves the value of about 1MW. It is succeeded by another considerable HRR increase at about 130th s caused by a windscreen glass breakage and fall out which increases the HRR above 2.5 MW. Subsequently HRR decreases. After 250 s the fire behaviour starts to differ in both simulations and FDS 6 provides slightly lower values of HRR, probably due to the new turbulent combustion model leading to different results in under-ventilated compartment containing the objects which burn out. In the consequence of lower HRR the second car does not ignite in the simulation.

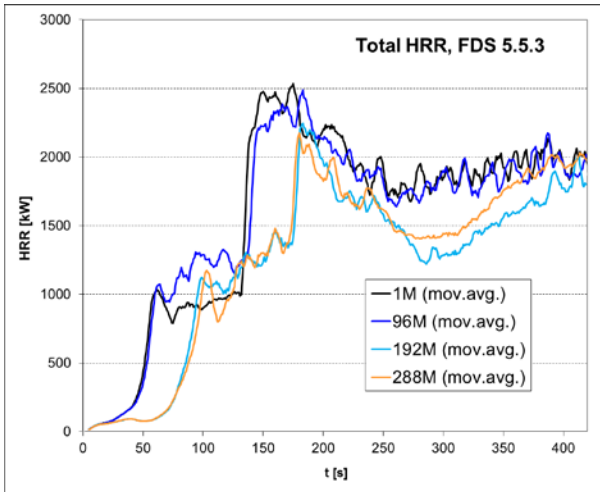


Figure 6: HRR simulated by FDS 5.5.3 for decompositions exceeding 48M

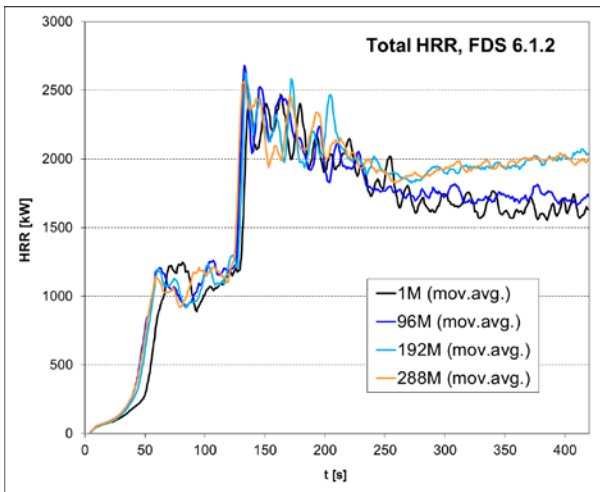


Figure 7: HRR simulated by FDS 6.1.2 for decompositions exceeding 48M

Parallel calculations of both FDS versions lead to very similar results as it is in the case of corresponding 1M calculation up to 96M (see Figures 4 - 7). The differences in the simulations 192M and 288M are more considerable and the simulation errors increase, although in FDS 6 they are significantly smaller (see Figures 6, 7). FDS 5 parallel calculations tend to delay fire behaviour (Weisenpacher, Glasa, Halada, Valasek, and Sipkova 2014) as can be seen in Table 1, columns 5 – 7. On the contrary, FDS 6 parallel calculations accelerate it slightly (Table 2, columns 5 – 6). While delays in FDS 5 increase relatively gradually with increasing mesh number, the differences between FDS 6 calculations look random. Even 12M simulations is distinguished by 6 s acceleration of the time of flashover in comparison with sequential simulation, which is a significant difference; however, in simulations containing more meshes this error is not considerably larger. Therefore, the impact of larger parallelisation on simulation accuracy is smaller in FDS 6. The values of energy released during the fire by particular simulation variants and their relative changes

in comparison with sequential calculation illustrate this tendency as well (see Tab. 3, column 5).

Table 3: Energy released during the fire by particular simulation variants and its relative change in comparison with sequential calculation

Released Energy				
	E_{FDS5} [MJ]	d_{FDS5}	E_{FDS6} [MJ]	d_{FDS6}
1M	663.0	0.00	624.9	0.00
12M	665.1	0.00	646.6	0.03
24M	653.3	-0.01	640.9	0.03
48M	662.5	0.00	649.9	0.04
96M	660.4	0.00	650.9	0.04
192M	515.7	-0.22	689.2	0.10
288M	537.0	-0.19	694.1	0.11

4.3. Temperatures Evolution

The evolution of temperatures inside the passenger compartment and the impact of parallelisation on it show similar patterns as can be observed in HRR behaviour (see Figures 8, 9).

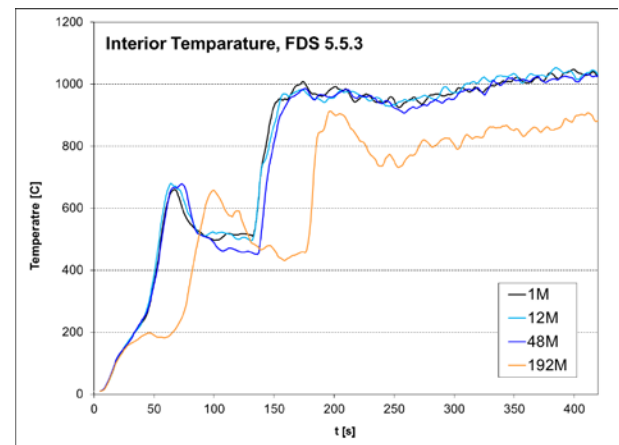


Figure 8: Interior temperature simulated by FDS 5.5.3

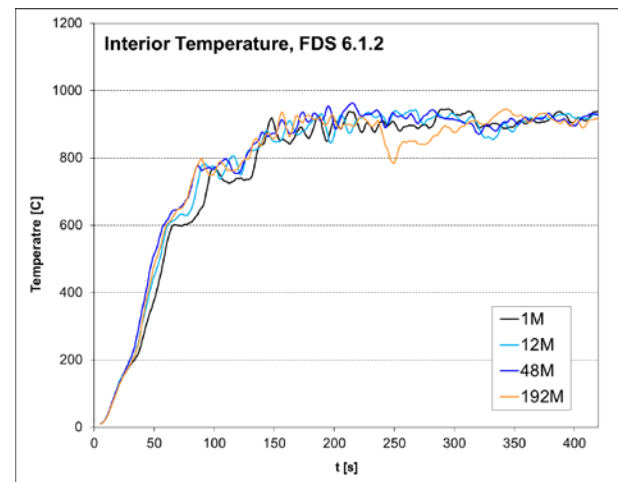


Figure 9: Interior temperature simulated by FDS 6.1.2

After initial gradual increase, temperature drops temporarily after flashover due to insufficient oxygen

concentration. In FDS 6 simulation the drop is less pronounced. Temperature increases again after windows fall out and finally reaches the values corresponding to the fully developed state of fire. The most important difference between both FDS version is significantly lower interior temperature of this state simulated by FDS 6. The temperature difference between both versions is about 100 °C. The errors influenced by parallelisation are almost negligible in FDS 6 simulation, while in FDS 5 simulation significant deviations can be observed in the case of the 192M simulation.

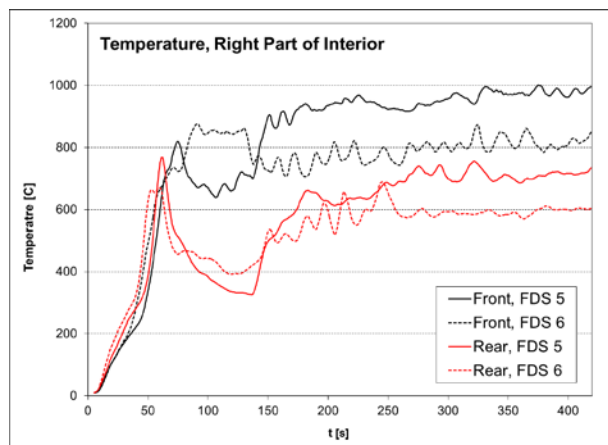


Figure 10: Temperature in the right part of car interior simulated by FDS 5.5.3 and 6.1.2

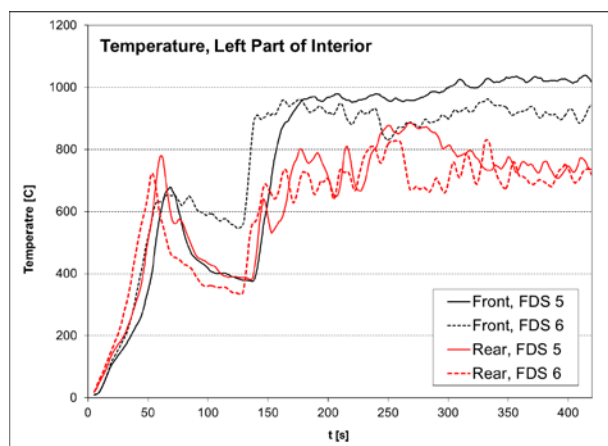


Figure 11: Temperature in the left part of car interior simulated by FDS 5.5.3 and 6.1.2

However, this behaviour does not capture differences between the evolutions of temperature in different parts of the passenger compartment. Figures 10 and 11 show the temperature behaviour in the front left, front right, rear left and left right part of the passenger compartment. Fig. 12 shows gas temperature in the horizontal cut of car interior (20 cm under the roof). It can be seen that the temperature differences between both versions in the right part of the interior are significantly higher than in the left part. Table 4 quantifies these differences.

Note that rear right window of the car was set “unbreakable” in the simulation, considering experimental observation that the window remained intact up to the 10th minute of the fire. Therefore, the right part of the compartment is more under-ventilated than the left part, which supports the hypothesis that FDS 5 overestimates the temperatures in under-ventilated compartment fires, while FSD 6 prevents such error.

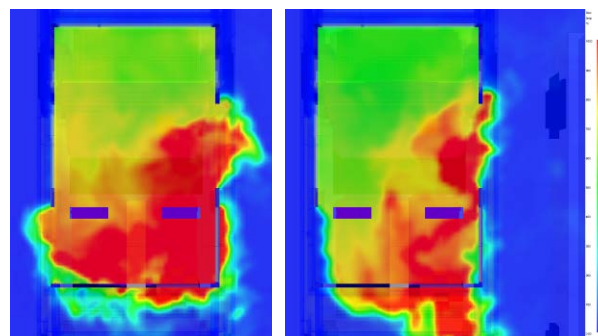


Figure 12: Gas temperature in the horizontal cut of car interior (20 cm under the roof) simulated by FDS 5.5.3 and 6.1.2 at the 350th s of fire.

Table 4: Average temperatures during the last 250s of the fire in different parts of the passenger compartment by both FDS versions and their differences D_{56} and relative differences d_{56} .

Average temperatures				
	FDS 5	FDS 6	D_{56}	d_{56}
Front, right	956	796	160	0,17
Front, left	993	914	79	0,08
Rear, right	689	591	98	0,14
Rear, left	774	723	51	0,07

CONCLUSION

Series of parallel simulations of passenger compartment fire have been performed by FDS versions 5.5.3 and 6.1.2. Simulations have confirmed and quantified several differences between the results obtained by both FDS versions. Due to a larger information transfer between computational mesh boundaries in parallel computations the FDS 6 simulation accuracy is preserved at the cost of lower parallel efficiency and CPU time. FDS 6 simulation are distinguished by considerable accuracy even when 288 meshes is used, while in FDS 5 simulations accuracy drops more significantly if more than 100 meshes is used. FDS 6 gives slightly lower values of HRR and temperature in some situations, especially if burning occurs in significantly under-ventilated compartments. These results suggest that the representations of the burning cars and their interiors should be adjusted for use in FDS 6 simulations.

ACKNOWLEDGMENTS

The authors would like to thank to P. Polednak, J. Svetlik, M. Simonova, and J. Flachbart for organizing the car fire experiments and to M. Dobrucky and J.

Astalos for technical support for SIVVP cluster computing. This work was supported by the National Science Foundation VEGA (project No. 2/0184/14) and by Slovak Research and Development Agency under the contract No. APVV-15-0340.

REFERENCES

- Deckers, X., Haga, S., Tilley, N., Merci, B., 2013. Smoke control in case of fire in a large car park: CFD simulations of full-scale configurations. *Fire Safety Journal*, 57, 22–34.
- Mangs J., Keski-Rahkonen O., 1994. Characterization of the fire behaviour of a burning passenger car. Part I: car fire experiments. *Fire Safety Journal*, 23(1), 17–35.
- McGratten, K., Baum, H., Rehm, R., Mell, W., McDermott, R., Hostikka, S. and Floyd, J.: *Fire Dynamics Simulator (Version 5)*, Technical Reference Guide, NIST Special Publication 1018-5, NIST, Gaithersburg, Maryland, USA, 2010.
- McGrattan, K., Hostikka, S., McDermott, R., Floyd, J., Weinschenk, C. and Overholt K.: *Fire Dynamics Simulator*, Technical Reference Guide. National Institute of Standards and Technology, Gaithersburg, Maryland, USA, and VTT Technical Research Centre of Finland, Espoo, Finland, sixth edition, September 2013
- Okamoto, K., Watanabe, N., Hagimoto, Y., Chigira, T., Masano, R., Miura, H., Ochiai, S., Satoh, H., Tamura, Y., Hayano, K., Maeda, Y. and Suzuki, J., 2009. Burning Behaviour of Sedan Passenger Cars. *Fire Safety Journal*, 44, 301–310.
- OpenMP (Open Multi-Processing), API Specification for Parallel Programming. Available from: <http://openmp.org/>
- Open MPI (Message Passing Interface), A High Performance Message Passing Library. Available from: <http://www.open-mpi.org/>
- Partanen, M. and Heinisuo, M., 2013. Car fires with sprinklers: A study on the eurocode for sprinklers. *Proceedings of International Conference Applications of Structural Fire Engineering*, pp. 23-28. April 19-20, Prague (Czech Republic).
- PBS (Portable Batch System). Available from: <http://www.adaptivecomputing.com/> <http://www.mcs.anl.gov/research/projects/openpbs>
- Polednak, P., 2010. Experimental Verification of Automobile Fires (in Slovak). *Proceedings of the 4th International Conference on Fire Safety and Rescue Services*. June 2-3, Zilina (Slovakia).
- Shipp M., Spearpoint M., 1995. Measurements of the severity of fires involving private motor vehicles. *Fire and Materials*, 19(3), 143–151.
- SIVVP – Slovak Infrastructure for High Performance Computing. Available from: <http://www.sivvp.sk/>, <http://hpc.ui.savba.sk/>
- Svetlik J., Flachbart J., 2012. Skip fire from vehicle to vehicle – experiment. *Proceedings of the 3rd International Scientific Conference on Safety Engineering and the 13th International Scientific*

Conference Fire and Explosion Safety, pp.113–118. October 18–19, The Higher Education Technical School of Professional Studies, Novi Sad (Serbia).

- Weisenpacher P., Glasa J., Halada L., Valasek L. and Sipkova V., 2014. Parallel computer simulation of fire in road tunnel and people evacuation. *Computing and Informatics*, 33(6), 1237–1268.
- Weisenpacher, P., Glasa, J. and Halada, L., 2016. Automobile interior fire and its spread to an adjacent vehicle: parallel simulation. *Journal of Fire Sciences*, 34(4), 305–322.
- Zhao B., Kruppa J., 2004. Structural behaviour of an open car park under real fire scenarios. *Fire and Materials*, 28, 269–280.

AUTHORS BIOGRAPHY

Peter Weisenpacher studied theoretical physics at Comenius University, Faculty of Mathematics and Physics, Department of Theoretical Physics, Bratislava, Slovakia and received PhD. in 2003. He works as research scientist at Slovak Academy of Sciences, Institute of Informatics. His current research interests include computational fluid dynamics, fire computer simulation and parallel computing. He participates in various research projects on fire simulation.

Jan Glasa graduated in numerical mathematics in 1986, received the RNDr. degree in numerical mathematics and optimization methods and algorithms in 1986 at the Faculty of Mathematics and Physics of Comenius University in Bratislava. He has received the C. Sc. degree (equivalent to Ph.D.) in computer science at the Slovak Academy of Sciences. He is with the Institute of Informatics of Slovak Academy of Sciences in Bratislava as senior scientist. He serves as head of scientific council of the institute. His current research interests include mathematical modelling and computer simulation of fires and parallel computing.

Viera Sipkova received the M. Sc. degree in mathematics and the RNDr. (Rerum Naturalium Doctor) degree in computer science at the Comenius University in Bratislava, Slovakia. Her research deals with parallel and distributed computing technologies. For several years she has officiated at the Vienna University, where her research efforts were focused on the high performance Fortran compiler and the automatic parallelization of scientific applications for distributed memory systems. At present she works at the Institute of Informatics of the Slovak Academy of Sciences and her main research interests include technologies for high performance computing (cluster, grid, cloud), preferring the development and porting complex scientific applications into a HPC environment.

TWO TIME STEPS PREDICTIVE CONTROL APPLICATION TO A BIOPROCESS SIMULATOR

S Altuntaş^(a), H Hapoğlu^(b)

^(a)Provincial Directorate of Environment and Urbanization, 55070 Samsun, Turkey

^(b)Department of Chemical Engineering, Ankara University, 06500 Ankara, Turkey

^(a)seminaltuntas@hotmail.com, ^(b)hapoglu@eng.ankara.edu.tr

ABSTRACT

The control of batch bioreactors to maintain the pH at a constant set point is quite different from the continuous system applications, because acid and base simultaneous addition has a strong effect on the closed loop performance which may never reach desired value. While it is most desirable to use approximate models based on the chemistry of the system, this may not always be possible for batch bioprocesses. The present paper discusses usage of a simulator with appropriate input and output data obtained from the real batch bioprocess without control. The real process pH behaviour was simulated by using 7M citric acid flowrate changes in the face of constant base addition. Two time steps predictive control was applied to the bioprocess simulator by coordinating 0.5M citric acid and 0.5M sodium bicarbonate flow rates. The best control weighting factor magnitude was determined. The constant set point following was achieved successfully.

Keywords: predictive control, bioprocess simulator, pH control, constant desired pH

1. INTRODUCTION

The pH control of batch bioprocesses is considered quite difficult. These processes may move from one operating point to another operation point or never reach a predetermined point. The batch system identification and a small amount addition of controller manipulating variables and determination of weighting factor magnitude are the most important and time consuming tasks in the proper implementation of control.

Bharathi, Sivakurnar and Shanmugam (2006) applied model based pH control algorithm to highly nonlinear systems such as fed-batch reactors. They noted that proportional integral derivative (PID) controllers are not adequate to obtain efficient pH control of such processes with extra nonlinear features. It is also proved with the simulation results in the same work that model based PI controller gives better performance and robustness. Meenakshipriya, Saravanan, Somasundaram and Kanthabhabha (2011) investigated a pH neutralization system control by using PI and P

algorithm based on the coefficient diagram method. Simulation results showed that the performance of these controllers was good when compared to other techniques. Pekel, Zeybek, Hapoglu and Alpbaz (2010) reported successful generalized predictive pH control application to textile wastewater treatment batch process. Camcioglu, Ozyurt and Hapoglu (2015) applied proportional integral derivative pH control to paper mill wastewater treatment batch process. The coordinated pH control application was achieved properly by using acid and base flow rates as manipulated variables.

For the food quality, the output variable measurements are compared to the desired values which can be converted into instrumental process set points. Paraskevopoulou, Athanasiadis, Blekas, Koutinas, Kanellaki and Kiosseoglou (2003) investigated the stability of kefir type drink production by adding kefir granules into cheese whey mixture containing fructose, black raisin extract and milk. A continuous low alcohol drink production by using whey mixture with kefir yeast was proposed by Kourkoutas, Psarianos, Koutinas, Kanellaki, Banat and Marchant (2002). The self-tuning proportional integral and derivative pH control and generalized predictive pH control algorithm were theoretically applied to discrete-time modelling of the cheese whey drink production respectively (Altuntaş, Ertunç, Hapoglu and Alpbaz 2011; Altuntaş, Ertunç, Hapoglu and Alpbaz 2012).

In this work, the appropriate citric acid flow rate changes introduced into the simulator from out of the system stands for the dynamic pH behaviour of the cheese whey batch kefir-type drink production in the face of constant 0.5M sodium bicarbonate addition. Two time steps predictive control was applied to the bioprocess simulator to hold the mixture pH value at a constant desired value. The pH value and the coordinated acid and base flow rates were chosen as the controlled and the manipulating variables respectively. One of the manipulating variables was kept at a small constant value, while the other was changed versus time by using control algorithm. The best control tuning parameter values were chosen according to performance criteria.

2. TWO TIME STEPS PREDICTIVE CONTROL

Two time steps predictive control is designed on the basis of generalized predictive control (Clarke, Mohtadi and Tuffs 1987) as below:

$$J(u, t) = E\{(y_{t+2} - r_{t+2})^2 + (y_{t+1} - r_{t+1})^2 + W(u_{t-1} - u_{t-2})^2\} \quad (1)$$

Where W is a weighting factor for manipulating variable. The expectation E presents the control value $u(t)$ that is estimated from data obtained up to and including time t and that a stochastic disturbance model is assumed. The quadratic minimization of equation (1) corresponds to a direct minimization problem.

The logic for applying incremental input Δu_t in the criterion is that Equation (1) does not allow zero static error in the case of a non-zero constant reference unless the open loop system algorithm involves an integrator, which would permit y_t to stay at a constant value when the control input is zero. One feature of using Δu_t in the control law term is that definite bounds on u_t may be difficult to achieve. This particular control design seems also to be a variant of the minimum variance strategy and the idea is to maintain weighting factor (W) as small as possible in order to achieve minimum output variation while still sustaining closed loop stability. The method can handle a variable dead time but not with a model which has an over specified order. It is necessary to maintain numerical robustness to keep control weighting close to zero (Hapoglu 2002; Ertuğ, Akay, Bursali, Hapoglu and Alpbaz 2003; Hiti, Boyacioglu, Ozyurt, Ertuğ, Hapoglu and Akay 2014). To establish the controller algorithm it is supposed that a model of the system is expressed in terms of the following controlled autoregressive integrated moving average (CARIMA) form:

$$\Delta y(t) = \frac{b_0 \Delta u(t-1)}{1 + a_1 z^{-1} + a_2 z^{-2}} \quad (2)$$

Altuntaş, Ertuğ, Hapoglu and Alpbaz (2012) obtained the coefficients of the model for the experimental open loop process of cheese whey drink production with kefir yeast using a pseudo random binary sequence as the forcing function and Bierman (1976) Algorithm. The magnitude of the coefficients were determined as $a_1 = -0.93$, $a_2 = -0.00013$ and $b_0 = 0.000065$. These values were also used in the control algorithm utilized in the present work.

The function of the Δ operator ($\Delta = 1 - z^{-1}$) is to guarantee integral action in the controller which eliminate offset, i.e. a steady-state output disturbance.

Definition of the vectors f and y ,

$$f = [y_{t+1|t} \ y_{t+2|t}]^T \quad (3)$$

Where $y_{t+j|t}$ is the free response prediction of y_{t+j} assuming that future control increments after time $t-1$ will be zero.

$$y = [y_{t+1} \ y_{t+2}]^T \quad (4)$$

$$G^T = [g_0 \ g_1] \quad (5)$$

$$y = G \Delta u_t + f \quad (6)$$

The future incremental control vector Δu_t is given by:

$$\Delta u_t = [G^T G + W I]^{-1} G^T (r - f) \quad (7)$$

In which r has been clearly described from the reference signal as:

$$r = [r_{t+1} \ r_{t+2}]^T \quad (8)$$

The control gain computed in equation (7) stays fixed and only the vectors r and f are update at the next sampling time.

In the present work, the steps used in the application of two time steps predictive control algorithm may be summarised as:

1. Apply a pseudo random binary sequence to the system as a forcing function and attain the plant output.
2. Estimate the coefficients from equation (2) using the Bierman U-D update algorithm.
3. Apply equation (7) to estimate the control signal utilizing the control procedure, and use this at every sampling time step.

3. EXPERIMENTAL PROCEDURE

The bioprocess simulator is given in Figure 1. The display screen shown in Figure 2 is designed for the simulator. The bioreactor is a 1 L glass jacketed cylindrical vessel and the mixture that is mixed by a stirrer. Water is used as coolant in the jacket. For all the experiments, the manipulated variable was acknowledged as a valve opening for the coordinated acid and base flow input rates while one of these was kept at a small constant value. It was accepted that enough agitation was provided by the mixer in the simulator. During cheese whey drink production by fermentation with kefir yeast, the open loop dynamic pH behaviour was used as a sample profile which is simulated by using the acid flow rate changes introduced into the simulator from out of the system. At the beginning of the experiment, the simulator mixture which contains cheese whey, glucose, grape juice and milk was carried to the desired temperature. Then the appropriate flow rate changes of citric acid were added to the simulator in the face of constant base addition and the mixture temperature was held at the desired value by means of on-off control.



Figure 1: The Experimental Bioprocess Simulator

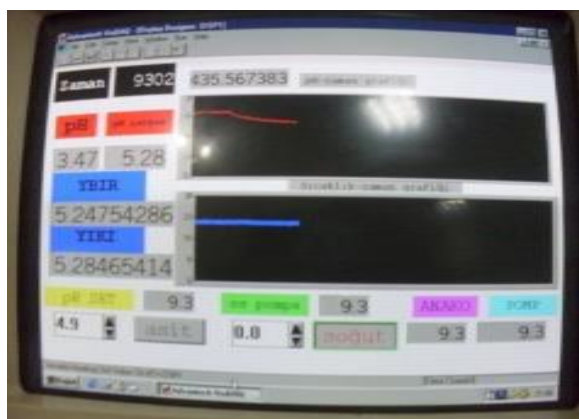


Figure 2: The display screen of the simulator

Internal pH of the simulator was measured and transferred to the computer. A computer with A/D and D/A converters was used for data acquisition and the advance pH control algorithm was applied to the simulator. The converter modules were connected to regulate the flow rate of the pump and the heat input given to the reactor. The converter module had two outputs (pump and heater) and three inputs (mixture pH, cooling water inlet temperature and outlet temperature). Digital signals from the system were transferred into the module and these signals were sent to a computer connected on-line to the system to convert the signals into analogue signals.

4. RESULTS AND DISCUSSIONS

An experimental work has been carried out to acquire the medium pH variation versus time during the open loop batch cheese whey drink production with kefir yeast. The pasteurized mixture of the reactor which contains 350ml cheese whey, 10g glucose, 70ml grape juice and 273 ml milk was used. Kefir yeast was added as 7ml inoculum. The actual pH behaviour without control was monitored throughout the fermentation process. VisiDAQ programming for data acquisition was advanced and applied during the experiments.

In order to simulate the open loop bioprocess pH disturbance, the pH effect of citric acid flow rate variations at different concentrations has been examined to stand for the experimental pH profile while a small amount of 0.5M sodium bicarbonate flow rate was consistently introduced. The results given in Figures 3, 4, 5 and 6 indicate that the citric acid concentration and very small flow rate variations have strong effect on the pH profile obtained for simulation. Under control free condition, pH responses were compared to each other and real pH change of medium with kefir yeast to see the performance of the citric acid flow rate change pattern with various concentrations. Figure 3 presents the pH in medium with kefir yeast which ranged from 8 to 4, and the pH simulation without kefir yeast over a 28000s period. The pH simulation tested by 1M citric acid flow rate changes in the face of constant base addition was ranged from 6.8 to 2.8. The pH decreased somewhat linearly to about 8000s. Above 8000s, the pH tended to level off around pH 2.8. This simulation did not support the real pH profile obtained with kefir yeast. Figure 4 presents that if the pH signal for comparison were obtained by adding 3M citric acid flow rate changes and a constant base flow rate, the desired pH simulation would not be achieved. Visual inspection of the pH profiles in Figure 5 indicates that the delayed pH changes are almost exactly opposite in about 6000s time domain. To gain insight into the patterns and mechanisms hidden in the bioprocess pH data with kefir yeast, it is shown how the simulation pH tends to under-predict the target profile by introducing 4M citric acid flow rate changes to the process without kefir yeast in the face of constant base flow. At similar operating conditions by using 7M citric acid flow rate changes, the pH profiles obtained with and without kefir yeast were in good agreement. 7M citric acid flow rate pattern was chosen as the simulator disturbance (see Figure 6). To determine the adequate concentration of sodium bicarbonate and citric acid for control purpose, 250ml medium of bioprocess without kefir yeast was tested by adding small amount of base or acid solution. The concentration of acid and base solutions was varied and the bioprocess pH values were measured for each case (see Table 1 and 2). A small amount of base flow rate with 0.5M concentration and 0.5M acid concentration were chosen as the controlled bioprocess constant inputs. The acid flow rate was chosen as variable input of the process.

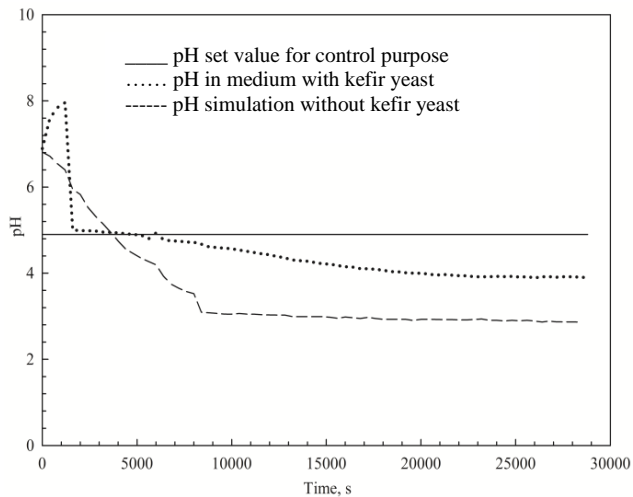


Figure 3: The Simulation Of The Real Bioprocess By 1M Citric Acid Flow Rate Changes

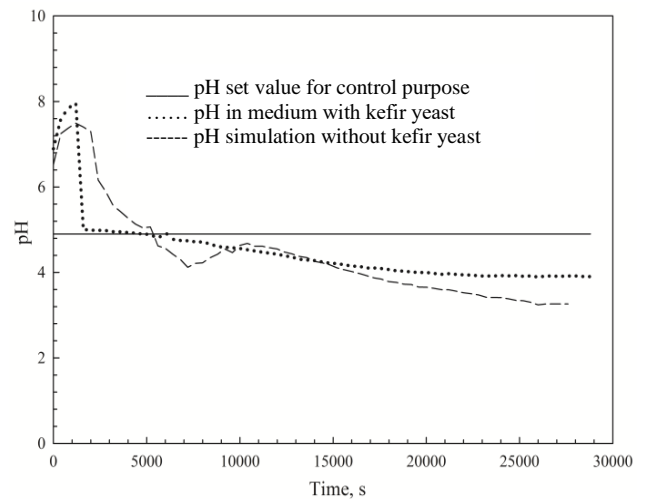


Figure 6: The Simulation Of The Real Bioprocess By 7M Citric Acid Flow Rate Changes

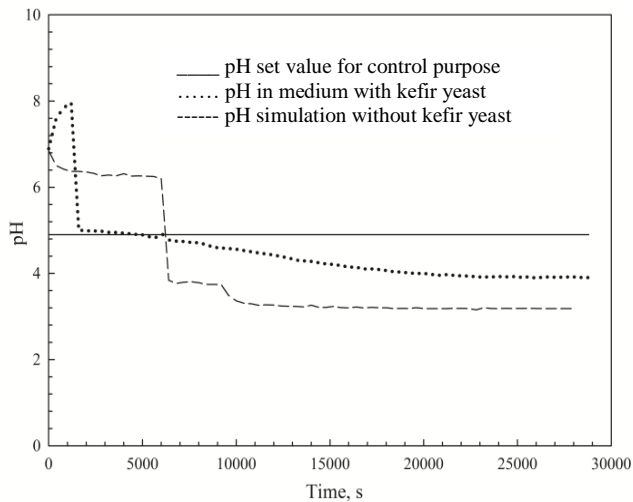


Figure 4: The Simulation Of The Real Bioprocess By 3M Citric Acid Flow Rate Changes

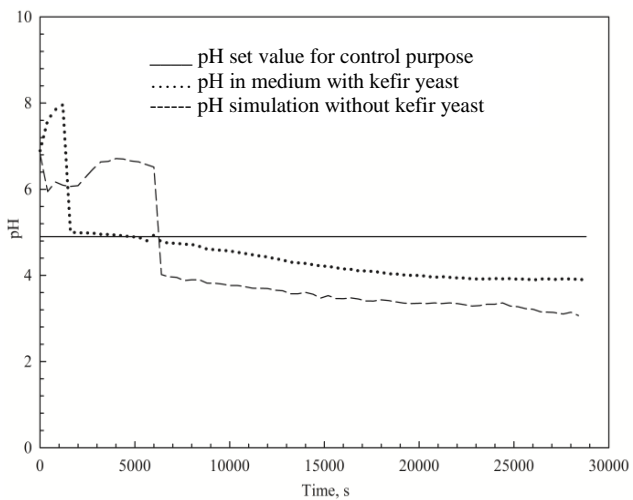


Figure 5: The Simulation Of The Real Bioprocess By 4M Citric Acid Flow Rate Changes

Table 1: Medium pH Changes Versus Volume And Concentration Of Base

	Volume of base solution, ml								
	0	5	10	15	20	25	30	35	
Base concentration	pH for sodium bicarbonate solution								
	0.01M	5.51	5.55	5.64	5.71	5.77	5.79	5.86	5.95
	0.05M	5.51	5.69	5.87	5.99	6.12	6.21	6.33	6.45
	0.1M	5.40	5.95	6.16	6.36	6.44	6.54	6.67	6.78
	0.5M	5.51	6.56	6.86	6.99	7.13	7.24	7.32	7.41
	1M	5.51	6.77	7.06	7.26	7.42	7.55	7.62	7.76

Table 2: Medium pH Changes Versus Volume And Concentration Of Acid

	Volume of acid solution, ml								
	0	5	10	15	20	25	30	35	
Acid concentration	pH for citric acid solution								
	0.01M	7.15	3.82	3.47	3.35	3.09	3.05	2.98	2.89
	0.05M	5.45	4.91	4.56	4.28	4.07	3.93	3.82	3.69
	0.1M	5.32	4.64	4.15	3.86	3.66	3.49	3.48	3.33
	0.5M	5.45	3.60	3.01	2.69	2.46	2.29	2.18	2.08
	1M	5.53	3.14	2.61	2.26	2.11	1.91	1.79	1.76

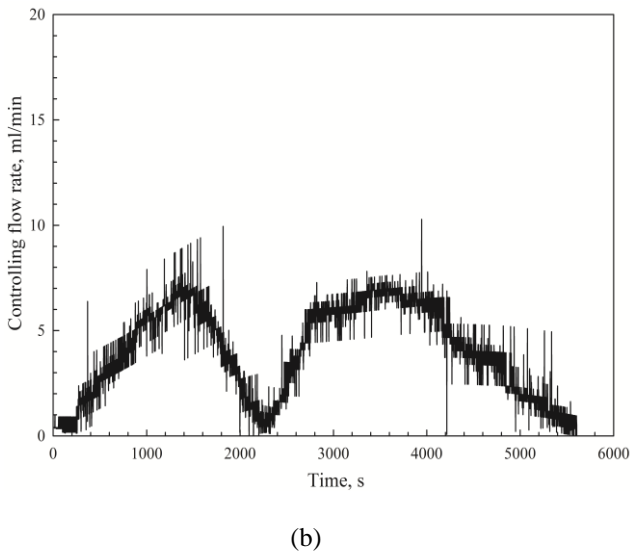
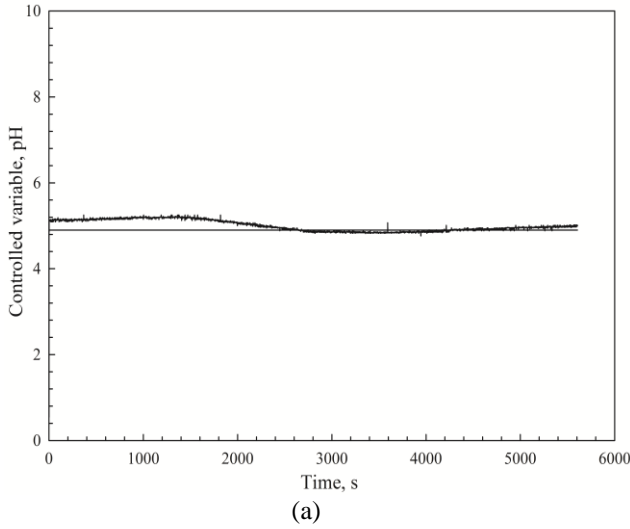


Figure 7: The Control Of The pH Under The Effect Of 7M Citric Acid Flow Rate Changes. The Controller Tuning Factor Is $W=0.00001$

Control of bioprocess medium pH was achieved by utilizing a predictive algorithm. The weighting factor (W) was varied in order to achieve appropriate performance. The integral square of the error (ISE) and the integral of absolute value of error (IAE) criteria are computed by comparing the performances of two time steps predictive controller with two different tuning weight factors in Table 3. From this table, it is clearly seen that the controller with 0.00001 tuning factor magnitude exhibited better results. This closed-loop pH changes versus time was shown in Figure 7a. The controller brings the medium pH back to set point after 2400 seconds but controlling flow rate continues oscillation until the end of the experiment (Figure 7b). An overshoot is observed at the beginning of the simulated bioprocess during the control in the simulator while acid flow rate is manipulating and base flow rate is held constant.

The performance criteria used in this work are given as follows:

$$ISE = \sum (y - r)^2 \quad (9)$$

$$IAE = \sum |y - r| \quad (10)$$

Table 3: ISE And IAE Values Obtained For Two Time Steps Predictive Control Of The Bioprocess Simulator Medium pH

Controller	ISE	IAE	Controller tuning factor, W
Two time steps predictive control	2	5.2	0.0001
	0.35	3.9	0.00001

5. CONCLUSION

The proposed strategy was performed experimentally to control the pH at a constant desired value during the pre-determined pH condition created by 7M citric acid flow rate changes as disturbance in a batch bioprocess simulator. The performance comparison of the controller was compared by means of two different feasible small magnitudes of weighting factor. The results showed that the magnitude of 0.00001 was adequate for the bioprocess pH controller as the tuning factor, and that the batch process control optimization was effectively implemented for two future output values taking into account as an even number of time step range. In addition, it is noted that tolerable manipulation of coordinated acid and base flow rates can be used for more nonlinear batch bioprocesses to maintain process pH at desired constant value.

ACKNOWLEDGMENTS

Financial support from Ankara University Scientific Research Projects Coordination Unit is gratefully acknowledged.

REFERENCES

- Altuntaş S., Ertunç S., Hapoglu H. and Albaz M., 2011. Discrete-time modelling of the cheese whey drink production and servo control of pH. Ordu University Journal of Science Technology, 1 (1): 15-21.
- Altuntaş S., Ertunç S., Hapoglu H. and Albaz M., 2012. Servo Control Application of Predictive Algorithm to the Cheese Whey Drink Production. Journal of Engineering and Architecture Faculty of Eskişehir Osmangazi University, 25 (2): 31-42.
- Bierman G.J., 1976. Measurement updating using the U-D factorization. Automatica, 12 (4): 375-382.
- Bharathi N., Sivakurnar E. and Shanmugam, J., 2006. Control of pH in fed-batch neutralization processes. IEEE International Conference on Industrial Technology, 1 (6): 1821-1825.
- Camcioglu S., Özyurt B. and Hapoglu H., 2015. pH control of paper wastewater treatment with

- electrocoagulation method. *Anadolu University Journal of Science and Technology- A Applied Sciences and Engineering*, 16 (2): 107-115.
- Clarke D.W., Mohtadi C. and Tuffs P.S., 1987. Generalized predictive control-part I: the basic algorithm. *Automatica*, 23 (2), 137-148.
- Ertunç S., Akay B., Bursali N., Hapoglu H. and Alpbaz M., 2003. Generalized minimum variance control of growth medium temperature of baker's yeast production. *Food and Bioproducts Processing*, 81: 327-335.
- Hapoglu H., 2002. Nonlinear long range predictive control of an openloop unstable reactor. *Computers and Chemical Engineering*, 26: 1427-1436.
- Hitit Z.Y., Boyacioglu H., Ozyurt B., Ertunc S., Hapoglu H. and Akay B., 2014. Self-tuning GMV control of glucose concentration in fed-batch baker's yeast production. *Applied Biochemistry and Biotechnology*, 172: 3761-3775.
- Kourkoutas Y., Psarianos C., Koutinas A.A., Kanellaki M., Banat I.M. and Marchant R., 2002. Continuous whey fermentation using kefir yeast immobilized on de-lignified cellulosic material. *Journal of Agricultural and Food Chemistry*, 50 (9): 2543–2547.
- Meenakshipriya B., Saravanan K., Somasundaram S. and Kanthabhabha P., 2011. CDM-based PI-P control strategy in pH neutralization system. *Instrumentation Sciences and Technology*, 39: 273-287.
- Paraskevopoulou A., Athanasiadis I., Blekas G., Koutinas A.A., Kanellaki M. and Kiosseoglou V., 2003. Influence of polysaccharide addition on stability of a cheese whey kefir-milk mixture. *Food Hydrocolloids*, 17 (5): 615-620.
- Pekel L.C., Zeybek Z., Hapoglu H. and Alpbaz M., 2010. Textile wastewater treatment with coagulation and generalized predictive control. *Chemical Engineering Transactions*, 21 (21): 817-822.

AUTHORS BIOGRAPHY

Semin Altuntaş is currently a senior expert of the Ministry of Environment and Urbanization. She holds a B. Sc., M. Sc., and Ph.D. from the Chemical Engineering Department of Ankara University, Turkey. She has written a number of articles on simulation and process control.

Hale Hapoglu is a professor in the Chemical Engineering Department, Ankara University, Turkey. She holds the B.Sc. and M.Sc. from the same department and a Ph. D. from the Chemical Engineering Department of Wales University, U.K. She has written over hundred articles on modelling, simulation, and process control.

SIMULATION AND OPTIMIZATION OF THE MANUFACTURING PROCESS OF E- BICYCLES: THE CASE OF INNOVAUNAM

Pedro González-Hernández^(a), Aida Huerta-Barrientos^(b)

^{(a),(b)}Department of Systems Engineering
National Autonomous Mexico University

^(a)black-pedro@hotmail.com, ^(b)aida.huerta@comunidad.unam.mx

ABSTRACT

The electric bikes are a new concept of bikes designed to promote exercise and at the same time the reduction of road congestion and pollution, spending on fuel and autonomous mobility. In Mexican context, the entrepreneurial sector was the first producing electric bicycles, but its assembly system was not optimized. In fact, due to the high costs of production the lifetime of their enterprises is limited to less than three years. The aim of this paper is to develop a simulation model of electric bicycles manufacturing process maximizing the productivity as a tool for supporting the decision-making of Mexican entrepreneurs of this sector. We develop the conceptual model of manufacturing process based on requirements of real demand of e-bicycles. Then, we implement the simulation model using AnyLogic software. The simulation model is validated via a sensitivity analysis. Finally, we maximize the productivity of the manufacturing process using the OptQuestTM solver.

Keywords: assembly process, electric bicycles, sustainability, simulation-optimization.

1. INTRODUCTION

Globally we suffer from high levels of pollution from the use of internal combustion vehicles, that creates serious effects on our life such as health problems, respiratory rate, saturation of transport routes and overweight, as a result of putting aside the physical effort to move from side to side. Perhaps the biggest problem in the world is pollution, before this have been raised solutions from government authorities such as traffic restrictions, encouraging the use of public transport, the use of hybrid and electric vehicles, the reduced speed limits and the encourage in the use of bicycles, creating pedestrian streets, some initiatives to share business vehicles and the supports to purchase electric vehicles and bicycles. In this direction, electric bicycles (e-bikes) are a cost effective alternative for industry and consumers, since these do not generate a high rate of pollution. E-bikes can be used anywhere, allowing the physical effort and even movement from the classic 13 km/hr offered by normal bike to 23-26 km/hr. The e-bike is very versatile and is adaptable to the needs of users, but perhaps its most important characteristic is the possibility of mobility free fuel costs. It requires a small parking space and even in the

folding version could be saved within an office. Pike Research (2016) forecast that the worldwide market for e-bikes will grow at an annual rate of 7.4% between 2012 and 2018, resulting in global sales of 47.6 million vehicles at the end of 2018 (see Fig. 1). Almost 90% of the total world market is concentrated in China, followed by Western Europe.

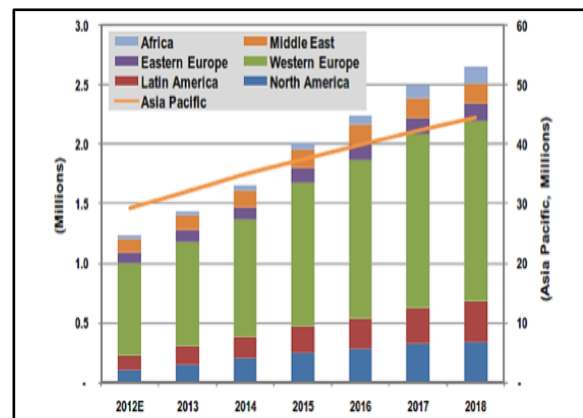


Figure 1: Electric bicycle sales by Region, World Markets: 2012-2018 (Pike Research, 2016)

In Mexico, in terms of business the entrepreneurial sector was the first producing e-bikes. At the beginning, the e-bikes in Mexico were neither good nor attracted attention and were on par with some attempts to work with internal combustion engines such as motorcycles. Just a decade ago, improvements were made, also experience was achieved and adequate profitability as well as an elegant, functional and highly efficient design were attained. Although the market is still small there is a future for e-bikes in Mexico but the number of suppliers is still limited by a few players. In this direction, Mexican entrepreneurs of e-bikes industry face a new challenge, the lack of theoretical tools for supporting the design and implementation of their production system, in order to increase the lifetime of their business which actually is limited to less than three years. The aim of this paper is to develop a simulation model of e-bicycles manufacturing process maximizing the productivity as a tool for supporting the decision-making of Mexican entrepreneurs of this sector.

In this paper, we present the case of InnovaUNAM whose entrepreneurs designed a new e-bike in order to commercialize it (see Fig. 2) based on the technical specifications presented in Table 1.

This paper is prepared as follows: the conceptual model of the manufacturing process based on real demand requirements of e-bicycles by InnovaUNAM entrepreneurs is described in Section 2. The simulation model is developed considering discrete-event simulation approach, then it is implemented using AnyLogic™ software, and it is validated via a sensitivity analysis in Section 3. The productivity of the manufacturing process is maximized in Section 4. Concluding remarks are drawn in Section 5.

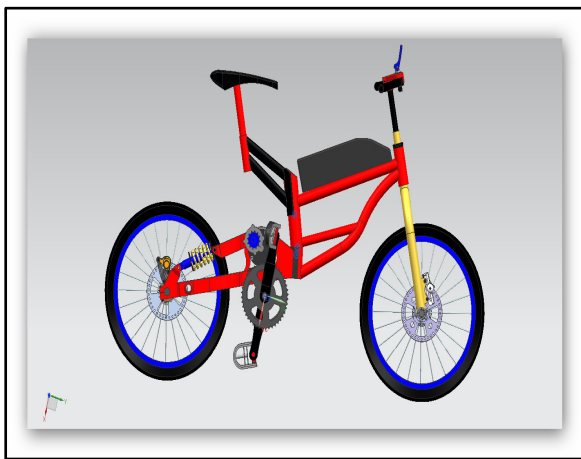


Figure 2: The e-bicycle designed by InnovaUNAM entrepreneurs

Table 1: Technical specifications of the e-bike designed by InnovaUNAM entrepreneurs

Technical specifications	
E bike	Folding
Distance	30 km
Battery	Li-ion
Time for recharging the battery	3 hours
Maximum velocity	25 km/hour
Weight allowance	90 kg
Frame material	Aluminum
Rolled rim	0.66 meters
Brake	Disc

2. THE CONCEPTUAL MODEL OF THE MANUFACTURING PROCESS OF E-BIKES

We propose a general methodology for designing the manufacturing process of e-bikes to support the decision making of InnovaUNAM entrepreneurs about the layout. The methodology starts with the definition

of the product, which in this case is an e-bike. Then, the required supplies, resources and mediums are listed. Also, the legal aspects need to be verified and the entrepreneur vision needs to be identified. The work zone is selected as well as the suppliers. After that, the processes are designed and the layout is created (see Fig. 3).

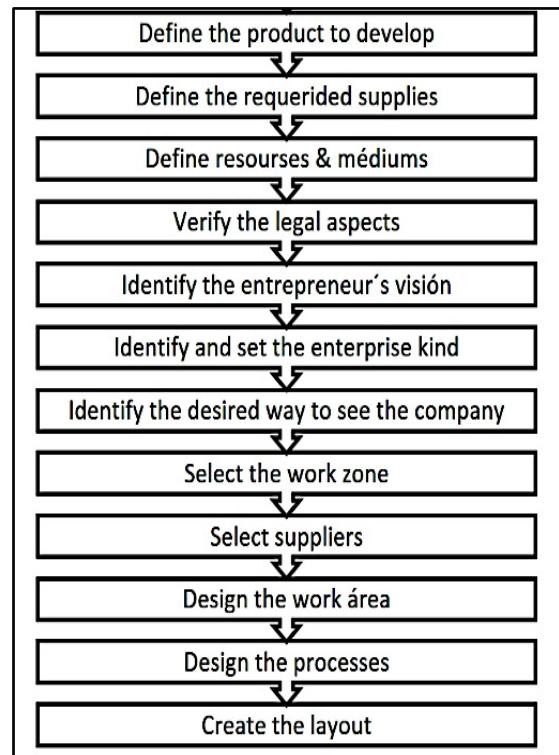


Figure 3: A proposed methodology for designing manufacturing process

In this study, the manufacturing process is modeled as a system that consists on processes interrelated. In general, the flow between the processes is a flow of mechanical and electric parts. As is observed in Fig. 4, the cross-functional diagram consists on five processes: cutting, bending, scuttled, welding, and assembling. The manufacturing of e-bikes starts with the general cutting of materials. Then, the load and force application continue as part of the bending process. After that, the tubes are drilled, as well as the bars and the mass. The drilled bars pass to the square belding process. While the mass drilled passes to the assembling process that starts with the wheels and finished with the electric and storage assembling (see Fig. 4). The e-bike manufactured is approved only if it passes the quality check, and then is carried on to the store for sale. If the quality required is not achieved, the e-bike is repaired, entering to the manufacturing process again.

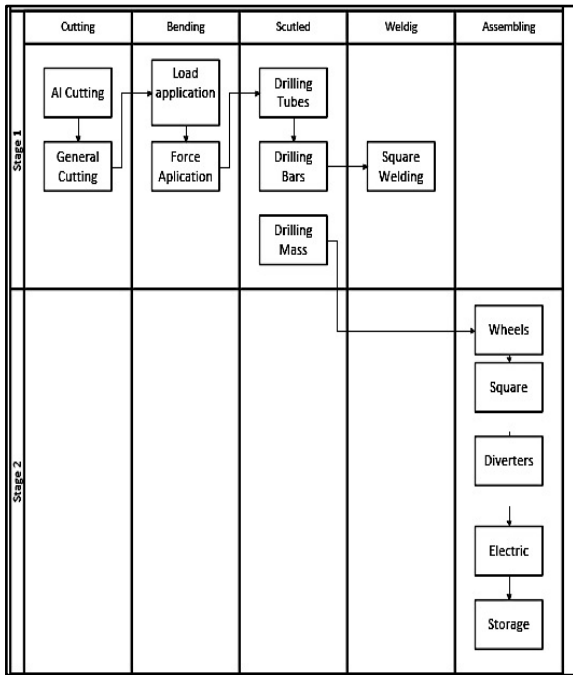


Figure 4: Cross - functional diagram of e-bikes manufacturing process

Based on the methodology proposed in Fig. 3 and considering the cross-functional diagram in Fig. 4, we design the layout (see Fig. 5). Each process is mapped in a physical area and the flow is sequential. We propose two storage areas, one of them for the materials and supplies and the other one for the e-bikes finished.

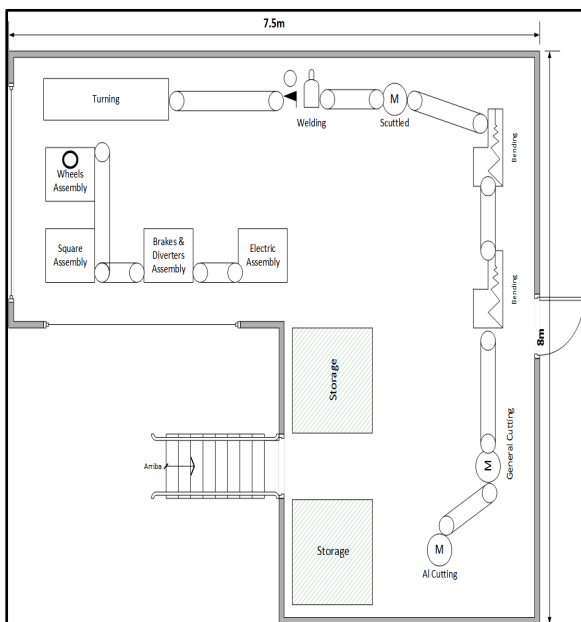


Figure 5: The layout

3. THE SIMULATION MODEL BASED ON DES APPROACH

We develop the simulation model of the manufacturing process using discrete-event simulation (DES) approach that is based on the top-down perspective (Robinson

2004; Banks 1998). In general, a DES model attempts to represent the components of a system and their interactions so the state variables change only at those discrete points in time at which events occurs, considering an event as an occurrence that changes the state of the system (Banks 1998). In this study, the events are generated by the internal changes in the five processes: cutting, bending, scuttled, welding, and assembling. The conceptual model of the manufacturing process of e-bikes is implemented in computer using AnyLogic™ software release 7.3. It is a Java-based and general-purpose simulation tool that takes advantage of the power of Java in any part of the model or in any library. These features make Anylogic™ a suitable tool for simulating complex systems such as manufacturing systems. Furthermore it allows the simulation of various domains using different approaches, and also provides animation, which can be useful in supporting the decision-making processes (Longo, Huerta-Barrientos and Nicoletti 2013). We use the Process Modeling Library whose blocks allow users to use combinations of agents, resources, and processes to create process-centric models of real-world systems (Grigoryev 2015). We use the source, sink, time measure, delay, service, assembler, and conveyor blocks (see Fig. 6).

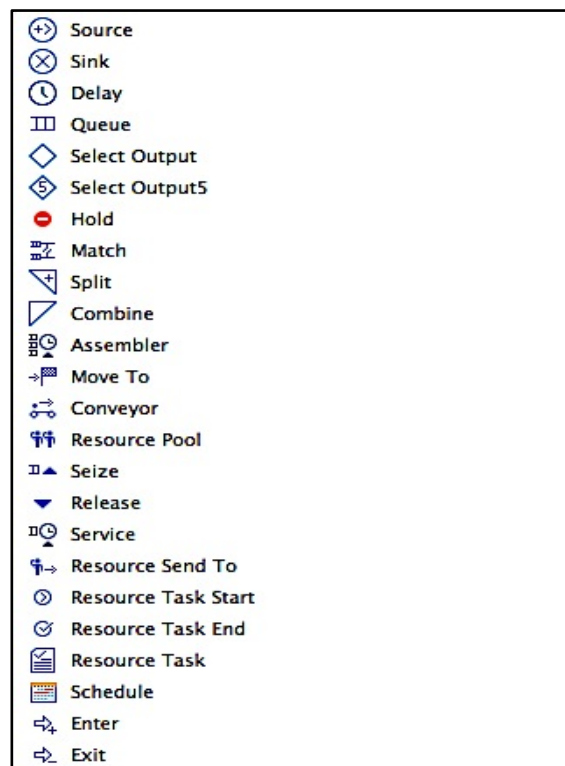


Figure 6: Block of Anylogic's Process Modeling library used in the simulation model implemented

The layout in the simulation model was built considering the real restrictions in space. The size of machinery was adjusted also to the real scale. The layout in the simulation model includes industrial safety signs that were allocated on the walls.



Figure 7: The simulation model of the manufacturing process of e-bikes: the welding, bending and scuttled processes



Figure 8: The simulation model of the manufacturing process of e-bikes: the general cutting and aluminum cutting processes

3.1. Simulation model verification

Simulation model verification deals with building the model right, but also with the accuracy in the conversion of a model representation in a micro flowchart into an executable computer program (Banks 1998). In this direction, we use a tool used rely heavily on human reasoning called *face validation*. During a technical visit, the potential users of the model, that means the entrepreneurs from InnovaUNAM-Engineering Unit, compared the simulation model and the manufacturing system behavior under identical operational conditions and they judged that results were reasonable. In this direction, it can be said that the simulation model represents the real system.

3.2. Simulation model validation via a sensitivity analysis

The model validation deals with building the right model within its domain of applicability (Banks 1998). In this direction, to validate the simulation model we use the dynamic technique called *sensitivity analysis* intended for evaluating the model based on its execution behavior. According to Banks (1998), through the sensitivity analysis values of input variables and parameters are systematically changed over some range of interest and the model behavior is observed. Using sensitivity analysis input variables and parameters to which model behavior is very sensitive are identified. We consider the turning process time as the input variable with two extreme values: 0.1 and 1 hour. The run time was during one month but considering just 40 working hours in the week.

Table 2: Input variables for designing the sensitivity analysis

Input variables		
	Low value	High value
Turning process time	0.1 hour	1 hour

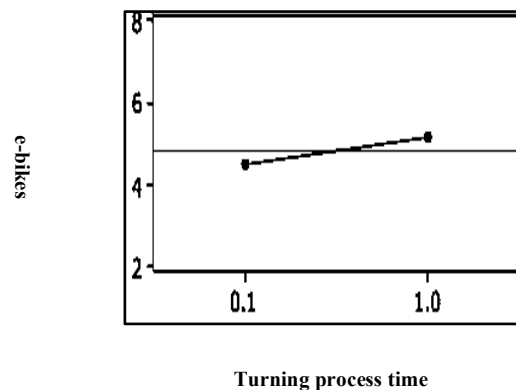


Figure 9: The sensitivity analysis results

As is observed from Fig. 9, the effect of the increment in the turning process is reflected on the increment in the number of e-bikes productivity, as it is in the real system.

3.3. Design of simulation experiments

We conduct a 2^4 factorial design (Montgomery, 2005) of simulation experiments. It means that we consider four factors (see Table 3): material flow, welding, turning, and electric assembly processes time in hours, with two different values. The simulation run time is one month but considering just 40 working hours per week.

Table 3: Time of different processes (hours)

Input variables and parameters			
Parameter	Input variable	Low value	High value
A	Material flow	0.05	0.1
B	Welding	0.5	1.5
C	Turning	0.1	1
D	Electric assembly	0.5	1.5

In Fig. 10, the individual effects of each factor (or the main effects) on e-bikes productivity are shown. The results of this factorial experiment indicate that material flow process time effect is larger than either the welding time or the turning time. We also obtain the metamodel (Eq. 1) that represents the e-bikes productivity considering the four factors with two values. From Fig. 11 to Fig. 15, the effects of the four different parameters are showed.

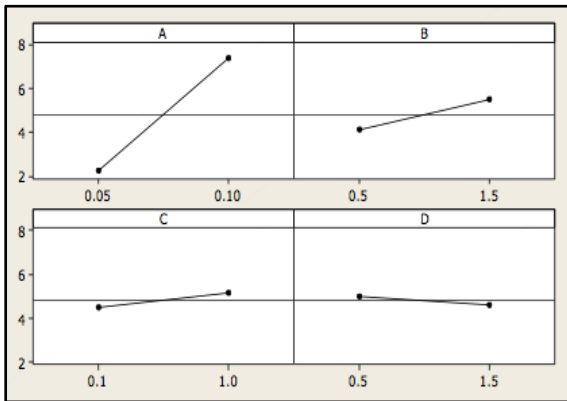


Figure 10: The main effects graphic of parameters A, B, C, and D on e-bikes productivity

$$Y = 169.44A + 6.11B - 2.50C + 3D - 56.67AB + 5.56AC - 58.89AD + 3.89BC - 2.22BD + 5CD - 33.33ABC + 33.33ABD - 11.11ACD - 7.78BCD + 66.67ABCD - 9 \quad (1)$$

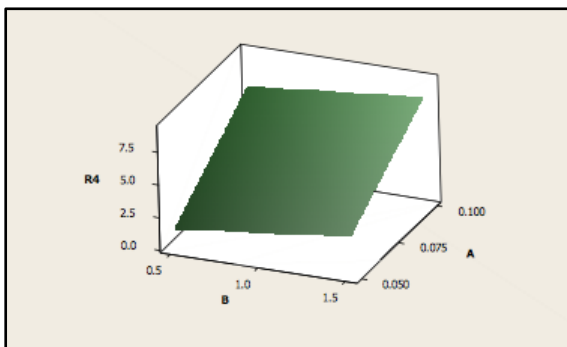


Figure 11: The effect of parameters A and B on e-bikes productivity

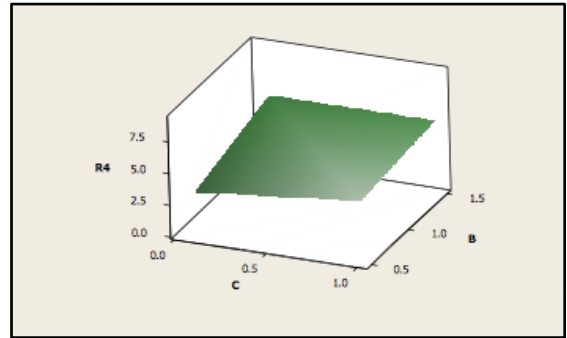


Figure 12: The effect of parameters B and C on e-bikes productivity

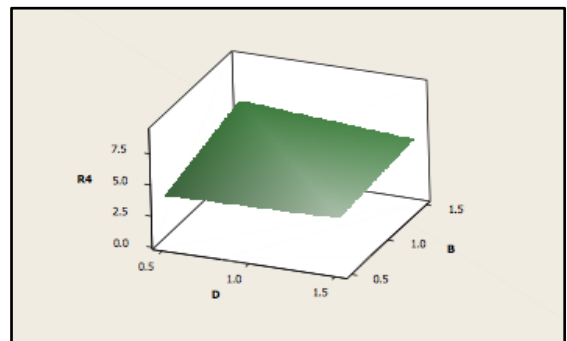


Figure 13: The effect of parameters B and D on e-bikes productivity

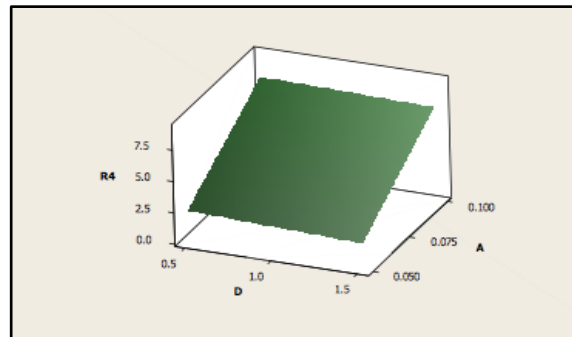


Figure 14: The effect of parameters A and D on e-bikes productivity

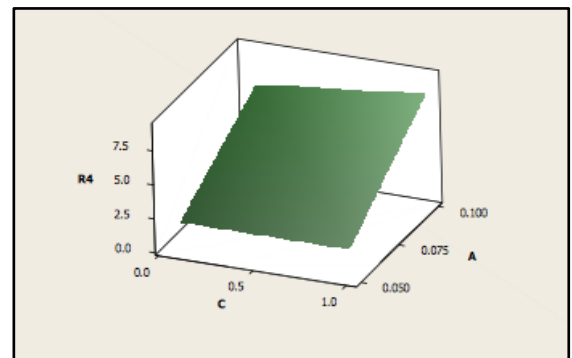


Figure 15: The effect of parameters A and C on e-bikes productivity

4. OPTIMIZING THE ASSEMBLY PROCESS OF ELECTRIC BICYCLES

In this study, we maximize the productivity, in terms of e-bikes assembled using the OptQuest™ solver that is included in AnyLogic™ software. OptQuest™ (by OptTek Systems, Inc.) is a general purpose commercial optimization software that operates by treating the objective function evaluation as a black box allowing users to represent solutions as mixture of continuous, discrete, integer, binary, permutation and other variables (Laguna 2011). As suggested by Fu (2002), OptQuest™ is an add-on to the underlying simulation engine. Also, OptQuest’s main optimization engine is based on the scatter search methodology, tabu search strategies, and neural networks. As is observed in Fig. 16, the metamodel of Eq. (1) is used to maximize the e-bikes productivity. Also, the parameters: material flow, welding, turning, and electric assembly processes time in hours, with two different values, were considered.

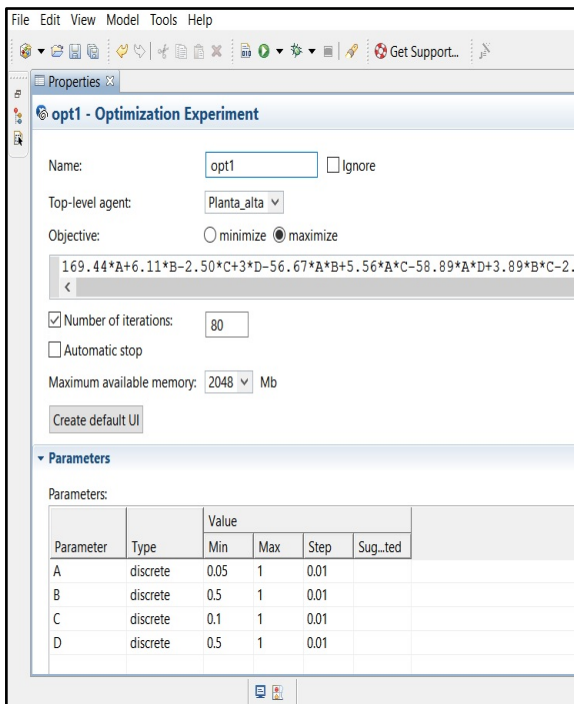
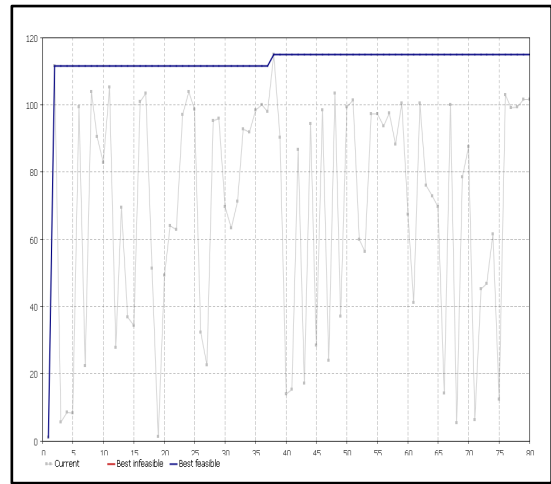


Figure 16: Maximizing the productivity of assembly process using OptQuest™

The optimization results of the e-bikes productivity are presented in Fig. 17 and Table 4. The best value for the productivity is 117 e-bikes during one month considering 40 working hours per week. The best time for the material flow process is set up in 1 hour, while for the welding process is set up in 0.94 hour and for the turning process in 0.79 hour. Finally, the electric assembly process time is allocated at its minimum value in 0.53 hour.



(a)

	Current	Best
Iteration:	80	37
Objective:	↓ 101.642	115
Parameters		
A	0.95	1
B	0.69	0.94
C	0.92	0.79
D	0.94	0.53
Copy the best solution to the clipboard		<input type="button" value="copy"/>

(b)

Figure 17: Productivity optimization results, a) the graphic of evolution in time, b) the best values found

Table 4: Optimized time of different processes (hours)

Input variables and parameters		
Parameter	Input variable	Best value
A	Material flow	1
B	Welding	0.94
C	Turning	0.79
D	Electric assembly	0.53

5. CONCLUDING REMARKS

We developed a simulation model of electric bicycles manufacturing process to maximize its productivity, as a tool for supporting the decision-making of Mexican entrepreneurs. The conceptual model of the manufacturing process of e-bikes was implemented in computer using AnyLogic™ software release 7.3. The layout in the simulation model was built considering the real restrictions in space. The results of a 2⁴ factorial design of simulation experiments indicate that material

flow process time effect was larger than either the welding time or the turning time. The optimum values for the productivity was on 117 e-bikes during one month considering 40 working hours in the week. To reach this productivity, the time of the processes such as material flow, welding, turning, and electric assembly needs to be less than 1 hour per e-bike. Finally, we consider that the simulation model 3D animation can be used as a tool for supporting decision making of entrepreneurs because it is useful to visualize the entire manufacturing e-bike process and recognize where improvements could be made.

ACKNOWLEDGMENTS

The authors appreciate the partial support by National Council for Science and Technology of Mexico (CONACyT) (SNI number 65477). We are also grateful to InnovaUNAM-Engineering Unit for sharing empirical data about a Mexican entrepreneur.

REFERENCES

- Banks, J., 1998. Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Georgia: John Wiley & Sons, Inc.
- Fu, M., 2002. Optimization for Simulation: Theory vs. Practice. Journal on Computing, 14 (3), 192-215.
- Grigoryev, I., 2015. Anylogic 7 in three days. A quick course in simulation modeling.
- Laguna, M., 2011. OptQuest. Optimization of Complex Systems. OPTTEK SYSTEMS, INC.
- Longo, F., Huerta-Barrientos, A., and Nicoletti, L., 2013. Performance analysis of a Southern Mediterranean seaport via discrete-event simulation. Strojinski Vestnik/Journal of Mechanical Engineering, 59, 517-525.
- Pike Research, 2016. Available from: [<http://www.greencarcongress.com/2012/03/ebike-20120327.html>]
- Robinson, S., 2004. Simulation: the practice of model development and use. Chichester: Wiley.

AUTHORS BIOGRAPHY

PEDRO GONZÁLEZ HERNÁNDEZ currently is Master student of the Department of Systems Engineering at the School of Engineering, UNAM. His research interests are industrial and production systems, simulation and optimization of complex systems.

AIDA HUERTA-BARRIENTOS received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the School of Engineering, UNAM and she is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

SIMULATION OPTIMIZATION OF PEDESTRIAN EVACUATION IN MEXICAN SUBWAY: THE CASE OF PINO SUAREZ STATION

Alma Vera-Morales^(a), Aida Huerta-Barrientos^(b)

^{(a),(b)}Department of Systems Engineering
National Autonomous Mexico University

^(a)alma.vera.m@outlook.com, ^(b)aida.huerta@comunidad.unam.mx

ABSTRACT

Mexico City subway started operating in 1961 and actually it is considered the public transport preferred by citizens with an annual ridership of 1.6 billion in 2015. The subway network is geographically located in the seismic areas of high risk in Mexico City. In fact, the subway Line 1 transports users to strategic points from East to West such as workplaces, touristic areas and the Mexico City downtown. The aim of this paper is to develop a simulation model of PINO SUAREZ subway station to optimize the number of passengers evacuating PINO SUAREZ subway station. We implement the simulation model using AnyLogic™ software and it is validated via a sensitivity analysis. We maximize the number of passengers evacuating PINO SUAREZ subway station at the busy hours in a seismic situation. From our results, we conclude that is important for subway authorities to strong the security measures and open new emergency exits.

Keywords: hybrid modeling, Mexican subway, seismic events, pedestrian evacuation.

1. INTRODUCTION

Mexico is located in one of the world's most seismically active regions. Plate tectonics and volcanic eruptions are the main causes of earthquakes in this country. Given the geographical location and the historical earthquakes that have occurred in the last decades, the governmental authorities have generated official documents as the official map or seismic regionalization to identify regions vulnerable to these natural phenomena in order to protect the life of Mexicans. In this direction, the official map of seismic regionalization classify four seismic zones called A, B, C and D, (see Fig. 1). For instance, Class A represents the geographical region without records of earthquakes in the last 80 years. While, Class B and Class C represent geographical areas where earthquakes are recorded not as frequently as other and Class D represents geographical areas where the largest number of earthquakes has been recorded. Over the last years, the Mexican area that has been most affected by earthquakes is Mexico City. In this case, the soil composition is characterized by clay, which tend to increase the waves when a seismic occurs.

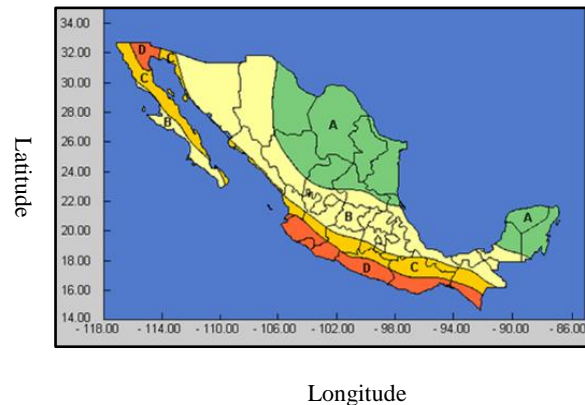


Figure 1: Seismic regionalization of Mexico, location of class A, B, C, and D areas

The earthquake occurred in 1985 was considered a catastrophe that collapsed Mexico City because the amount of material and human losses. After that, the governmental authorities published the Building Regulations of Mexico City that includes a map with the three types of soil in order to prevent the construction of tall buildings in susceptibles zones. On the other hand, Mexico City's population has increased over the last years exponentially, generating a big problem of urban mobility. Mexico City subway started operating in 1961 and there are currently 12 lines in service. (see Fig.2.) It is considered the public transport preferred by citizens with an annual ridership of 1.6 billion in 2015. From the geographical perspective, the Mexico City's subway network is located over the seismic areas.

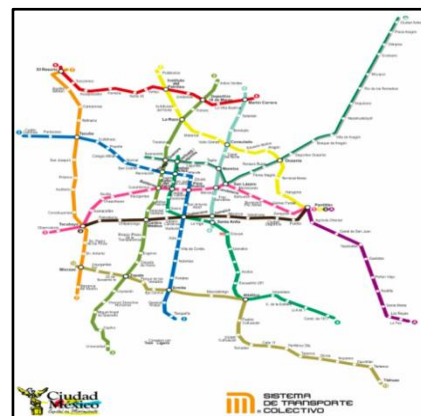


Figure 2: Subway network in Mexico City

We can distinguish three geographical areas based on their seismic risk in Mexico City: high (red), medium high (orange) and medium (yellow) (see Fig. 3). Many subway stations are geographically located over the risk area so this condition represents a warning for citizens and government. For example, work zones are located in the downtown of the city, so a lot of subway's passengers are in imminent danger if an earthquake occurs when they are in the subway.

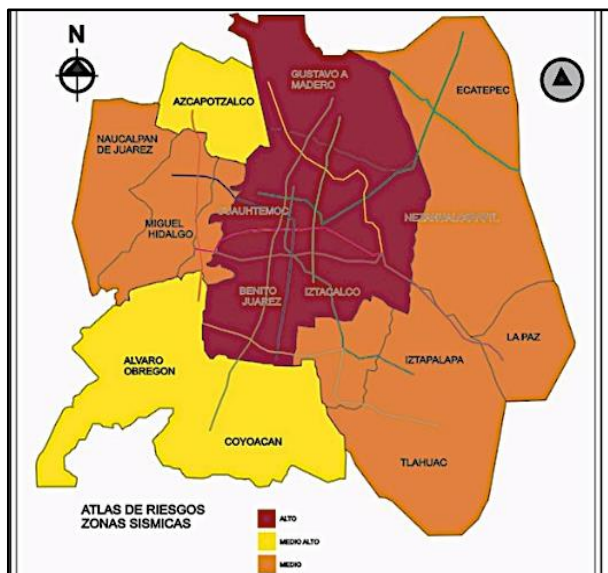


Figure 3: The three geographical areas based on their seismic risk in Mexico City

The subway Line 1 was the first to come into operation. It started in 1969 and consists on 20 subway stations with a total length of 18,828 kilometers. It transports passengers to strategic points running in a East to West direction and vice versa, such as workplaces, touristic areas and the Mexico City downtown. Also, the subway Line 1 has eight intersections with Lines 2, 3, 4, 5, 7, 8, 9, A, and B. Actually, subway Line 1 is considered the second busiest subway line, with an annual ridership of 263, 708, 660 in 2014.

PINO SUAREZ subway station, is located in Mexico City downtown, one of the areas of high seismic risk, and it is the connection between subway Line 1 and subway Line 2. PINO SUAREZ subway station has on average 25,986 riderships per day with two busy periods of time. The first starts from 5:30 am to 10 a.m. and the second starts from 5:30 to 10:30 pm. At busy hours, the subway platforms are completely full, indeed the ascent and descent of passengers is very complicated (see Fig. 4). So, if an earthquake occurs during busy hours, passengers would be seriously affected in the absence of space or better conditions for evacuating. The aim of this paper is to develop a simulation model of PINO SUAREZ subway station to optimize the number of passengers evacuating PINO SUAREZ subway station. We consider our study can help to Mexico City governmental authorities to reinforce the civil protection measurements to preserve the life of subway's passengers.



Figure 4: PINO SUAREZ subway station Line 1 at busy hour

This paper is prepared as follows: the literature review about simulation models of pedestrian evacuation and simulation software is presented in Section 2. A conceptual model of PINO SUAREZ subway station is developed in Section 3. A hybrid simulation model of PINO SUAREZ station is implemented using AnyLogic™ software and validated via a sensitivity analysis in Section 4. Then we maximize the number of passengers evacuating PINO SUAREZ subway station at the busy hours in a seismic situation and some recommendations for evacuating Mexico City subway in seismic situations are made in Section 5. Concluding remarks are drawn in Section 6.

2. PEDESTRIAN MOVEMENT MODELING & SIMULATION

From the quantitative perspective, an increasing amount of literature is devoted to the pedestrian movement and mass evacuation in both indoor and outdoor using simulation. For example, Banos and Charpentier (2007) developed an agent-based prototype, MAGE to simulate pedestrian movements in very constrained and limited spaces, like subway stations. Through “what if scenarios” they explored some fundamental aspects of complex spatial systems. Their investigations demonstrated the major role played by individual interactions and their influence on the overall behavior of the system.

On the other hand, Yang, Li and Zhao (2014) optimized the opening number of the entrance ticket windows at the Hangzhou subway station. They developed a simulation model based on Anylogic software and study the impact of parameters as the pedestrian arrival rate and the opening number of ticket windows in peak and off-peak periods. From their simulation results, it was observed that during the off-peak period, opening two ticket windows would achieve their best utilization rate.

While in the mass evacuation context, Chen et al. (2016) analyzed the related factors which included the number and width of passage, channelization setting and the number of pedestrians with the evacuation time in the passages of one subway station in Beijing. They recorded the pedestrian evacuation conditions and from their analysis it was revealed that when the number of pedestrians gets reach to more than 200, the evacuation time increased significantly. Mustafa, Rahman, and Bachok (2013) used SimWalk to plan for an evacuation during a panic situation. It was revealed that when the bottleneck occurred, pedestrians would collide with each other, increasing the friction between them.

There is also a variety of commercial software to simulate the dynamic of pedestrians (see Table 1). For the purpose of this study, we use Anylogic software, a simulation tool that supports discrete event, agent-based and system dynamics simulation (Borshchev and Filippov 2004). The libraries that we use in this study are described in the Section 4.

Table 1: Commercial software used for simulating pedestrian evacuation

Simulation software	
Name	Provider
Egress Section in FPETool	H.E. Nelson National Bureau of Standards, U.S.
EVACNET 4	University of Florida, U.S.
TIMTEX	University of Maryland, U.S.
WAYOUT	Fire Modelling & Computing, AU
STEPS	Mott MacDonald, UK
PedGo	TraffGo
Pedroute and Paxport	Halcrow Fox Associates, UK
SIMULEX	Integrated Environmental Systems, UK
GRIDFLOW	D. Purser & M. Bensilum, UK
ASERI	Integrierte Sicherheits-Germany

3. THE CONCEPTUAL MODEL OF PINO SUAREZ SUBWAY STATION

We develop the conceptual model of the processes involved when a passenger enters to PINO SUAREZ subway station to board the subway. We visited PINO SUAREZ station at busy hours in order to observe the dynamic of daily operations. We identify the access points to understand the behavior of users within the station. In that sense, we identify three areas: the admission ticket, the subway platform and the route through passengers walk to arrive to the subway platform. The first point we focus is the access by turnstile, we find two types of passengers: the first type had ticket and the second one did not. So, once they entered the station, they choose between different directions to reach their final destinations (see Fig. 5).

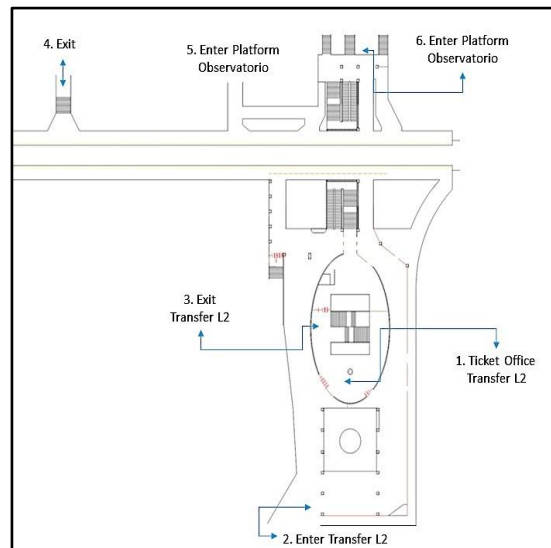


Figure 5: PINO SUAREZ subway station layout

As is observed in Figure 6, passengers sometimes have to wait if the station is too congested or they wait to another user to make the trip or get out of the system together. A lot of users enter the station at busy hour, this situation makes complicated the traffic on it. Sometimes containment measures are applied to make more agile the transit of pedestrians. Another point of interest is the connection between line 1 and line 2 of the PINO SUAREZ station. We describe feasible options that users have to move around the station, sometimes user use to use transfer just to get out of the station, perhaps their destination is closer by that exit or they decide to change the line, to reach their final destination.

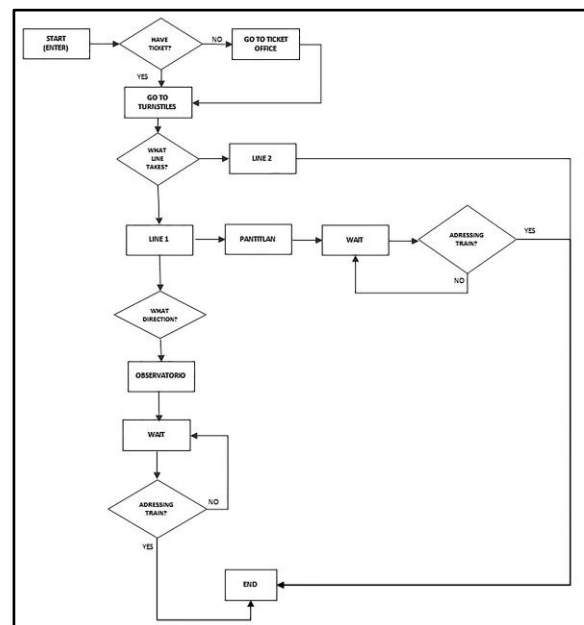


Figure 6: Flow chart of user activities inside the subway station

Finally, we design a flowchart based on Chen (2016) where it can observe how passenger behavior is when an earthquake reaches the station, shares which normally have passengers in a panic situation are: to look around and seek to safety place or to get out of dangerous situation. The station has four emergency exits, in which we optimize the number of passengers evacuating PINO SUAREZ subway station (see Fig. 7).

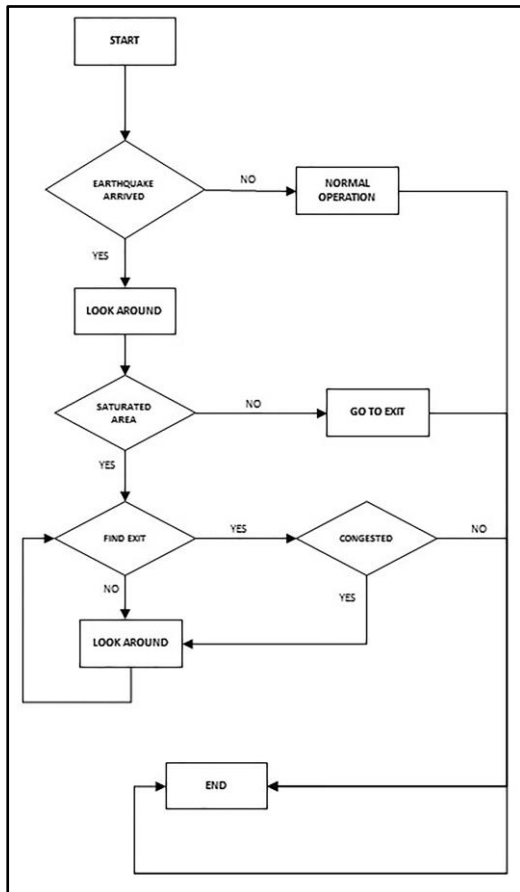


Figure 7: Flow chart of user evacuation inside the subway station during a seismic situation

4. A HYBRID SIMULATION MODEL BASED ON ABM AND DES APPROACHES

We develop the simulation model of pedestrian and subway activities in the PINO SUAREZ subway station using discrete-event simulation (DES) approach (Balci 1994, Banks 1998, Robinson 2004) and agent-based modeling and simulation (ABMS) approach (Grimm and Railsback 2005, Wilensky and Rand 2015). The DES approach uses the top-down perspective, to develop simulation models. In this case, collective phenomena are described at global level. In contrast, ABMS approach uses bottom-up perspective. In this last case, it is possible to understand the interactions rules between elements at local level. Using DES it is not fundamental the definition of a space scale as it is in ABMS. It is important to note that the time scale in DES and ABMS varies at discrete steps.

The hybrid simulation model is implemented using AnyLogic™ Personal Learning Edition software. We used the Pedestrian and Rail libraries. The Anylogic™ Rail Library allows users to efficiently model, simulate and visualize the dynamic of the subway along the PINO SUAREZ subway station. This library integrates well with Anylogic™ Enterprise and Pedestrian libraries (Borshchev 2013). In this simulation model we use the Train Source, Train Move To, and Train Dispose blocks (see Fig. 8).

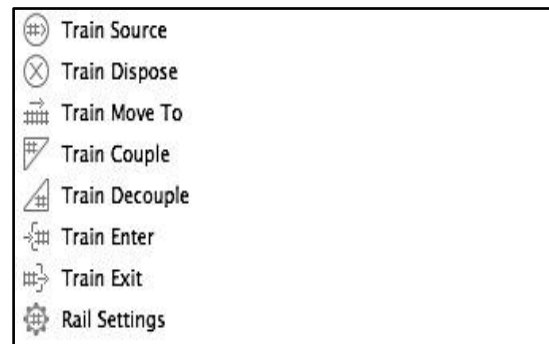


Figure 8: Block of Anylogic's Rail Library used in the simulation model implemented

On the other hand, the Anylogic™ Pedestrian library simulates pedestrian flows in “physical” environments by allowing to create models of buildings and areas with large numbers of pedestrians such as subway stations (Grigoryev 2015). In this case, we use Ped Source, Ped Sink, Ped Move To, Ped Escalator, Ped Select Output, Ped Area Descriptor, Ped Enter, and Ped Service blocks (see Fig. 9).

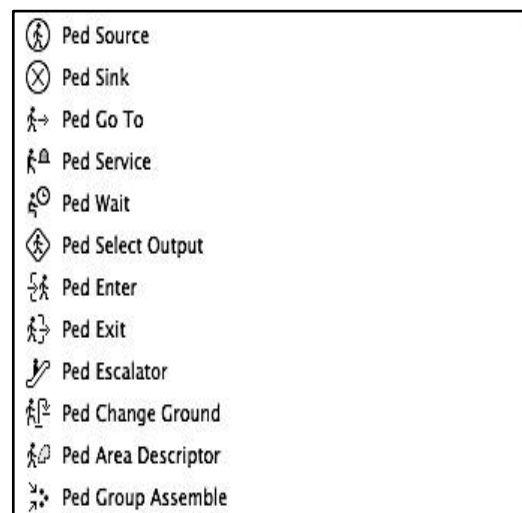
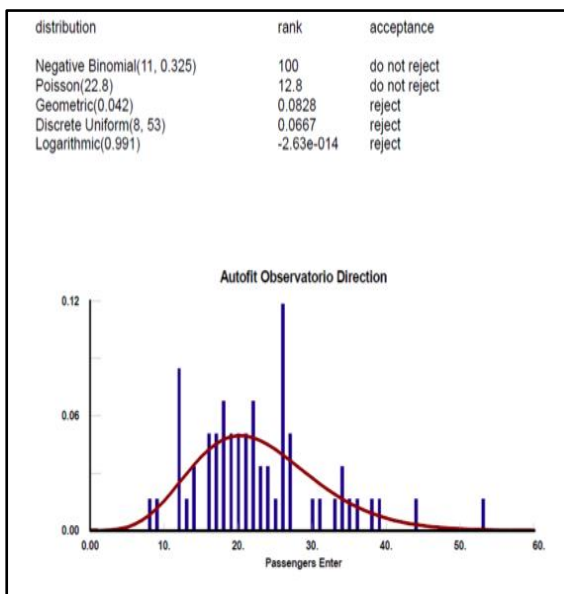


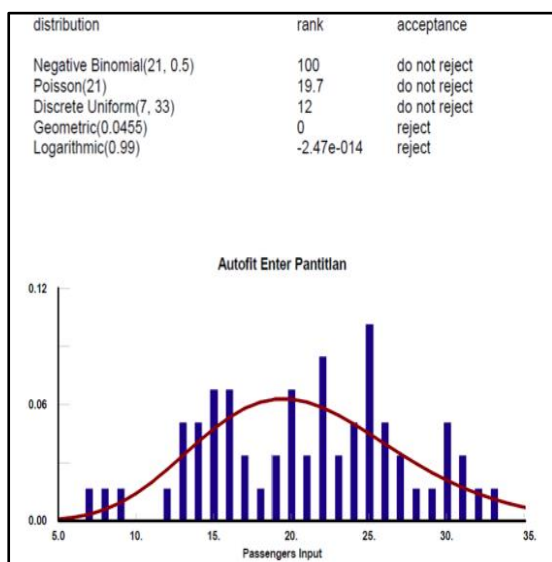
Figure 9: Block of Anylogic's Pedestrian Library used in the simulation model implemented

4.1. Input data analysis

During the technical visits to PINO SUAREZ subway station we collected the data about passengers entering to the subway station during one hour on Wednesdays. Then, we used Stat::Fit® version 3.0.1.0 Standard to fit probabilistic distributions to our empirical data. Stat::Fit® is a probability distribution fitting software package designed to help users more easily test the fit of hypothesized statistical models to empirical data to identify the best candidate distribution for a given scenario (Benneyan 1998). The flow of passengers is fitted to Negative Binomial and Poisson distributions as can be observed in Fig. 10a and Fig 10b.



(a)



(b)

Figure 10. Autofit of empirical data of flow of passengers a) Observatorio Enter and b) Pantitlan Enter

The layout of PINO SUAREZ subway station is implemented in the simulation model considering real dimensions. We built a 3D animation of the simulation model to visualize in the most realistic way the dynamic of passengers, for example on the platform (see Fig. 11) and at the entrance (see Fig. 12).



Figure 11: The 3D animation of passenger's activities in PINO SUAREZ subway platform

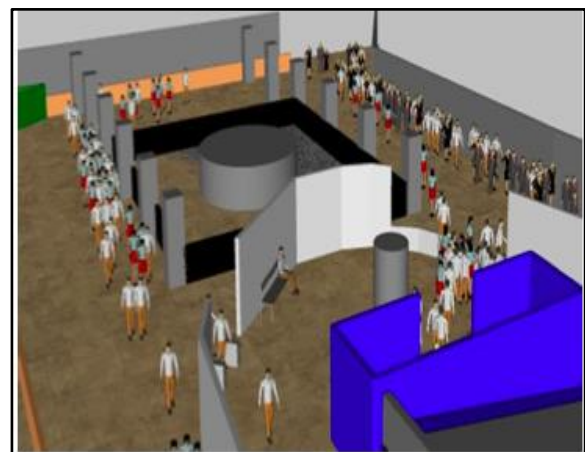


Figure 12: The 3D animation of passenger's activities in PINO SUAREZ subway station entrance

4.2. Simulation model verification

The process of verification has the goal of elimination of “bugs” from the code, that means, the model needs to be correctly implemented (Wilensky and Rand 2015). In this case, we carry out a check on the Java StackTrace Console included in Anylogic about the warnings. Additionally, as suggested by Banks (1998), we used the 3D animation to detect some logical errors of passengers and convoys. The most common error detected is when passengers look for their final destination and it is unreachable. It was due mainly when the passengers are not adjusted in scale to the subway station dimensions.

4.3. Simulation model validation

As is suggested by Wilensky and Rand (2015), if a model is to be useful for answering real-world questions, it is important that the model provides outputs that address the relevant issues and that the outputs are accurate. In this direction, we are interested in the study of the dynamic of passengers in the PINO SUAREZ subway station, so in order to validate the simulation model, we determine whether the implemented model in Anylogic™ explains such dynamic. We conduct a sensitivity analysis where passenger's speed is allocated in two extreme values based on a Uniform probability distribution (see Table 2).

Table 2: Input variables for designing the sensitivity analysis

Input variables		
	Low	High
Passenger's speed	Uniform(0.0, 0.5) m/s	Uniform (0.5, 1.5) m/s

In the case of passenger's speed distributed as Uniform (0.0, 0.5) meters per second (see Fig. 13), we observe three zones of maximum passenger's density, that were really crowded, marked in red color. This is due to the bottleneck effect when passenger's walk slowly. In the bottleneck situation, in both the real and the simulated system, passengers are too close to each other suffering inter personal friction. Contrary, in the case of passenger's speed distributed as Uniform (0.5, 1.5) meters per second (see Fig. 14), we observe a weak density in the subway station. This preliminary results can suggest that to avoid bottleneck in the subway station, passengers need to move really fast.



Figure 13: Passenger's speed at low value, strong density of passengers in red color



Figure 14: Passenger's speed at high value, weak density of passengers in green color

5. MINIMIZING THE EVACUATION TIME IN PINO SUAREZ SUBWAY STATION IN SEISMIC SITUATIONS

We maximize the number of passengers evacuating PINO SUAREZ subway station at the busy hours in a seismic situation. We used the OptQuest™ tool that is included in AnyLogic™ software. OptQuest™ is a general purpose optimizer that works with an optimization problem defined outside of the system (Glover, Kelly, and Laguna 1999), which is represented in our case by the implemented model of the PINO SUAREZ subway station (see Fig. 15). In this direction, our simulation model can change while the optimization routines remain the same. The optimization procedure uses the outputs from the simulation model which evaluate the outcomes of the inputs that were fed into the model (Glover, Kelly, and Laguna 1999). OptQuest™ is based on scatter search, tabu search, and neural networks algorithms.

In this study, the outputs from the simulation model are obtained via a design of simulation experiments. To do that, we consider two factors: a) the speed passengers (meters per second) and b) the number of passengers that enter to PINO SUAREZ subway station. From the simulation experiments we obtain the metamodel (Eq. 1) that represents the passengers that evacuate the station when the two factors changes.

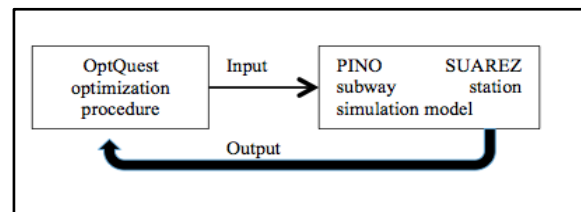


Figure 15: Interaction between OptQuest™ and the simulation model

$$Y = -11.05a + 0.0017b - 0.0033ab + 41.8 \quad (1)$$

We use the metamodel for optimizing the number of passengers that evacuate the subway station during a seismic situation (see Fig. 16). During the optimization execution, agents walk towards the nearest exit area. The path followed for evacuating is not necessarily the fastest route. From the optimization results, we observe that the optimum number of simulated agents that evacuate is 44 and the evacuation time was 10 minutes (see Fig. 17). Most agents remain inside the station. So it is important for subway authorities to strong the security measures and open new emergency exits.

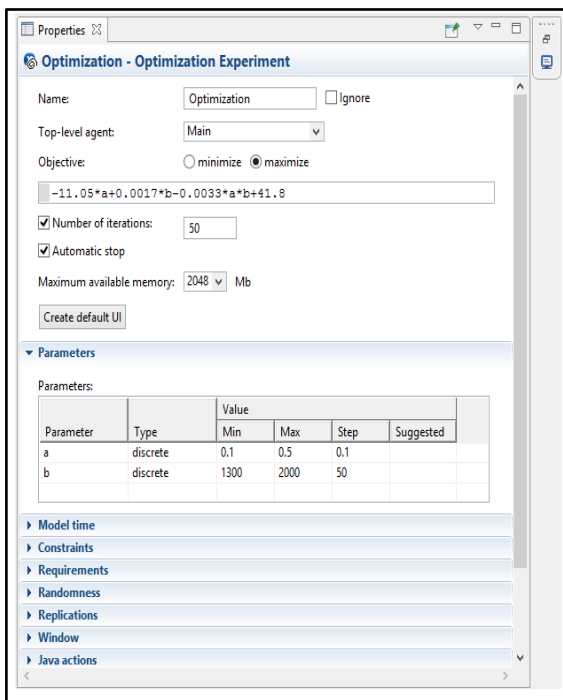
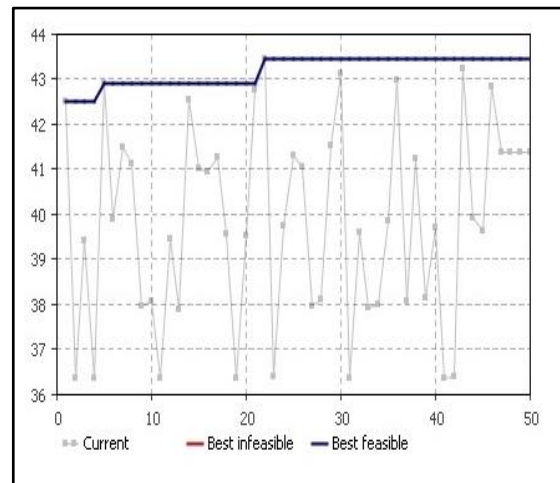


Figure 16: OptQuest™ interface in Anylogic



a)



b)

Figure 17: Passengers that evacuate PINO SUAREZ subway station during a seismic situation a) the best values found, b) the graphical representation

6. CONCLUDING REMARKS

We developed the conceptual model of the processes involved when a passenger enters to PINO SUAREZ subway station to board the subway. The simulation model was implemented based on discrete-event and agent-based modeling and simulation approach. During the technical visits to PINO SUAREZ subway station we collected the data about passengers entering to the subway station during one hour on Wednesdays. The flow of passengers was fitted to Negative Binomial and Poisson distributions using Stat::fit software. We maximized the number of passengers evacuating PINO SUAREZ subway station at the busy hours in a seismic situation. We used the OptQuest™ tool that is included in AnyLogic™ software. From the optimization results, we observed that the optimum number of simulated agents that evacuate is 44 and the evacuation time was 10 minutes. Most agents remain inside the station. In conclusion, this study is useful to optimize and evaluate the layout of PINO SUAREZ subway station for evacuating situation.

ACKNOWLEDGMENTS

The authors appreciate the partial support by National Council for Science and Technology of Mexico (CONACyT) (SNI number 65477). We are also grateful to Sistema de Transporte Colectivo Metro for the facilities in the data collection.

REFERENCES

Balci, O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research*, 53, 121-174.

Banks, J., 1998. Simulation. New York: John Wiley & Sons.

Banos, A. And Charpentier, A., 2007. Simulating pedestrian behavior in subway stations with agents. Proceedings of the 4th European Social simulation Association. September 10-14, Toulouse.

Benneyan, J.C., 1998. Software Review: Stat::Fit. OR/MS Today. Available from: <http://lionhrtpub.com/orms/orms-2-98/swr.html> [accessed 5 July 2016]

Borshchev, A. And Filippov, A., 2004. From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools.

Borshchev, A., 2013. The big book of simulation modeling. Multimethod modeling with Anylogic 6.

Chen, Y., Yang, R., and Shen, S., 2016. Impact study on mass evacuation in urban underground passages. IACSIT International Journal of Engineering and Technology, 8 (3), 222-226.

Glover, F., Kelly, J. and Laguna, M., 1999. New advances for wedding optimization and simulation. Proceedings of Winter Simulation Conference, pp. 255-260. December 5-8, Phoenix.

Grigoryev, I., 2015. Anylogic 7 in three days. A quick course in simulation modeling.

Grimm, V. and Railsback, S. F., 2005. Individual-based modeling and ecology. New Jersey: Princeton University Press.

Mustafa, M., Rahman, Z., and Bachok, S., 2013. Agent-based simulation for pedestrian movement during panic situation. Proceedings of 20th ITS World Congress TOKYO 2013. October 15-18, Tokyo.

Robinson, S., 2004. Simulation: the practice of model development and use. Chichester: Wiley.

Yang, Y., Li, J., and Zhao, Q., 2014. Study on passenger flow simulation in urban subway station based on Anylogic. Journal of Software, 9 (1), 140-146.

Wilensky, U. and Rand, W., 2015. An introduction to agent-based modeling. Cambridge: The MIT Press.

AUTHORS BIOGRAPHY

ALMA VERA MORALES currently is Master student of the Department of Systems Engineering at the School of Engineering, UNAM. Her research interests are planning and project management, simulation and optimization of complex systems.

AIDA HUERTA-BARRIENTOS received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the School of Engineering, UNAM and she is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

PRESERVING AND MANAGING HERITAGE BUILDINGS IN A SUSTAINABLE WAY WITH INTEGRATION OF BIM – A REVIEW

Lina Seduikyte^(a), Angeliki Kylili^(b), Jolanta Sadauskiene^(c) Paris Fokaides^(d)

^{(a), (c)} Kaunas University of Technology, Faculty of Civil Engineering and Architecture, Department of Building Energy System, Kaunas, Lithuania

^{(b), (d)} Frederick University, School of Engineering and Applied Sciences, Cyprus

^(a)lina.seduikyte@ktu.lt, ^(b)res.ka@frederick.ac.cy, ^(c)jolanta.sadauskiene@ktu.lt, ^(d)eng.fp@frederick.ac.cy

ABSTRACT

Heritage buildings through retrofitting are able to minimize CO₂ emission, the use of energy and “work” as sustainable buildings.

For the renovation process and preservation of the heritage building new digital tools such as Building Information Model (BIM) should be used. BIM is also useful for the decision making and choosing the right strategies and materials.

The paper, through literature review, is analysing sustainable implementations during the retrofit process of heritage buildings and analyses the use of progressive BIM tool for preservation and management of heritage buildings.

Keywords: heritage, sustainability, BIM

1. INTRODUCTION

Each country should preserve and investigate the past in order to create the future.

With nowadays technologies and knowledge it can be done in a sustainable and digital way. Figure 1 presents the main components of sustainability: environment, economic and social. This diagram was first launched at the 2005 World Summit on Social Development and continues to be a valid attempt to describe sustainability. Smart retrofitting of existing, equally heritage, buildings can play a significant role in creating a sustainable future.

The increasing use of Building Information Model (BIM) in construction sector because of many benefits and resource savings during design, planning, and construction of new buildings is noticed in the past decades. Moreover, in the last years BIM is applied for conservation and management of heritage buildings as well.

The aim of this paper, through literature review, (1) to analyses sustainable implementations during the retrofit process of heritage buildings and (2) to analyse the use of progressive BIM tool for preservation and management of heritage buildings.

The conclusions and recommendations of this research will be useful for heritage professionals, researchers and decision makers in both governmental and industry level who are involved in preservation of heritage buildings.



Figure 1: Main components of sustainability

2. HERITAGE AND SUSTAINABILITY

Buildings of European countries use almost 40% of the global energy, by generating the 40–50% of the total output of greenhouse gases. The heating requirement is responsible for almost 80% of the energy demand in houses. Demand for cooling is growing as well.

Since historic buildings constitute 25% of the European built environment they have a role to play in delivering CO₂ emissions reduction targets along with the rest of the domestic stock (Moran et al. 2014).

Sustainability of energy systems are related to sustainable energy supply systems, integration of renewable energy, management of energy demand, and sustainable energy consumption. Building sector,

including heritage buildings, has the largest potential for energy saving.

Heritage buildings generally use more energy compared to the buildings which were built recently. However, retrofit of heritage buildings should be done in a different way, with respect to historical, sociocultural and architectural values.

The implementation of sustainable retrofit of historic buildings represents a huge challenge, which gives a significant impact on the global energy consumption.

Heritage buildings, such as churches, should also meet required indoor microclimate conditions. First of all to save artworks and secondly to insure good thermal comfort for people staying indoors. Thermal comfort and the preservation of artworks are often in conflict with each other, so a balance between these two is needed.

Table 1 presents several investigation of heritage buildings and strategies implemented for energy saving (1-7) and preservation of art works (8-9).

Lassandro et al. (2015) made a survey on heritage school building and developed methodology based on a holistic approach that correlates students post occupancy evaluation with instrumental survey and software simulations. The results showed, that the energy audit through virtual tour is an innovative tool that, due to the replicability of the energy retrofiting solutions categories, can be transferred to other school buildings especially with the same construction type and implemented with further evaluations about acoustic comfort and indoor air quality.

The study presented by Ben and Steemers (2014) demonstrates that balanced approaches can be developed to retrofit heritage and meet the requirement of both energy efficiency and heritage conservation.

Pankhurst and Harris (2013) presented “eco” renovation example of Victorian stable yard (Modern Hall Park, south-west London). Building is saving and generating energy. In the presented project it was specified that, where possible, materials must be of low embodied energy, from sustainable resources or recycled materials, locally sourced and have a low environmental impact.

Grytli et al. (2012) examined different short and long term impacts from various energy efficiency measures on a model building by combining life cycle assessment, energy calculations and a self-developed heritage value assessment system. By combining the results from the different analyses in an integrated decision-making tool, it was possible to discuss optimal solutions for energy improvement, taking both environmental and heritage aspects into consideration.

Thermal-energy, environmental, and economic assessment of retrofit solutions and optimized scenarios for energy saving, could present a key answer to strike down the energy requirements in heritage buildings. When smart methodologies are integrated in heritage buildings, these buildings are more “friendly” for public use.

3. HERITAGE AND BIM

BIM is the process of generating, storing, managing, exchanging and sharing building information in an interoperable and reusable way (Vanlande et al. 2008).

BIM can be used for buildings of different size and purpose (residential, commercial, healthcare, educational, heritage and other types).

The use of BIM differs between countries. The increase of digitalisation, sustainability requirements, cloud computing will increase and extend the use of BIM in new and existing buildings.

Volk et al. (2014) presented literature review of BIM used for existing buildings with focus on maintenance and deconstruction life cycle stages.

Khosrowshahi and Arayici (2012) presented roadmap for implementation of BIM in the UK construction industry. The findings suggested three structured patterns - organisational culture, education and training, and information management - to tackle technology, process and people issues in BIM implementation.

Wong and Zhou (2015) made literature review and concluded that green BIM has emerged as a popular energy performance analysis tool during buildings design stage. BIM was applied for emission estimation and monitoring of the carbon footprints of construction projects.

New research integrates BIM with Life Cycle Assessment (LCA) for the maximization of its impact in building project and the demonstration of the benefits that arise from the synergy of the two (Kylili et al. 2015, Ajayi et al. 2015).

The need of retrofit of heritage buildings is becoming as much important as construction of new buildings. New approach, such as implementation of BIM, will improve the way of working with saving of time and money.

Mezzino (2014) discusses digitalization of cultural heritage and connection of BIM and Geographic Information System (GIS), which could provide a platform for creation of new meanings stemming from the interaction of different users and stakeholders engaged in its preservation and enhancement.

Del Giudice and Osello (2013) introduced national InnovANCE project (Italy), which aims to develop the first national database able to share information among professionals through the help of BIM.

Hichri et al. (2013) presents critical analysis of the application of BIM in the field of cultural heritage.

Table 2 present nine case studies on historic building and BIM. In analysed cases implementation of BIM resulted in representation of historical and archaeological objects, creating of BIM models of historic buildings, representation of urban centres, and automation of engineering drawings for conservation process.

Table 1: Investigation of heritage buildings and strategies implemented for energy saving and preservation of art works

Nr.	Authors	Heritage buildings	Tested strategies or implementations	Results
1.	Pisello et al. (2016)	Palazzo Gallenga Stuart, a four story university building, Italy	Building energy efficiency has been pursued through two strategies: - innovative cool tiles with the same appearance of traditional historic tiles; - geothermal heat pump system with water storage tanks positioned in the under-ground unoccupied areas of the building.	Application of the innovative cool tiles lead to a maximum cooling energy saving of 14.0% and 3.8% in the classrooms of the top floor and in the whole building, respectively. The installation of a more effective energy plant leads to an average energy saving of 64.3% and 67.0% in terms of heating and cooling demand, respectively. The combination of the two effects leads to an average energy saving of 64.0% for heating and 69.2% for cooling.
2.	Todorovic et al. (2015)	Museum of Aviation in Belgrade, Serbia	This study encompasses six versions of BIPV (Building Integrated Photovoltaic), including non-transparent opaque type and two types of semi-transparent (with 20% and 30% visible light transmittance). Defined models have been studied by BPS (Building Performance Simulation) and co-simulation coupling BES (Building Energy Simulation) and two domains CFD simulation (outdoor and indoor).	The results of the present study confirm that the total electricity demand of heat pump operation in heating and cooling regimes, as well as the whole HVAC system's power demand, and in addition lighting demand are smaller than potential PV electricity power and produced energy. The Museum's deep RES (renewable energy sources) integrated refurbishment could result in reaching the Museum's Zero CO2 emission status, cost-effectively transforming this historic building in sustainable Energy Plus building.
3.	Burattini et al. (2015)	Historic center of Tivoli, a town close to Rome, Italy	Four passive strategies were examined and applied based on the location of the building and the non-alteration of the structure and the landscape. Energy analyses of historical buildings were done according to recommendations of the UNI TS 11300-Part1. The calculations were performed only on the building envelope, based on passive solutions and alternatives.	Results showed that it is possible to improve the energy performance of an existing building achieving a significant energy saving with the respect of the building architecture, shape, function and the surrounding landscape. The annual energy savings reached a maximum value of 25%.
4.	Sahin et al. (2015)	Basmane Neighborhood Centre, Turkey	Three energy efficient retrofit packages (which included: air tightening of the buildings, indoor air temperature control, changing type of the heating and/or fuel, insulation of the attic floor, additional roof insulation, insulation of walls, additional insulation of the ground floor) were evaluated with respect to their effect on the energy consumption.	Results showed, that energy savings of more than 34% could be obtained without damaging the heritage values.
5.	Ascione et al. (2015)	Educational building of the University of Sannio, Benevento, Italy	Three main steps were conducted: - the building performance assessment and energy diagnosis, - numerical studies by means of hourly energy simulations with a deepening about the calibration of the simulation model, - investigation of potential energy savings and	In the presented research, the most suitable combination of energy efficiency measures has resulted with the replacement of windows with low emissive ones, the application of thermal plaster and the new thermal insulation of the roof slab. Annual primary energy demand for the microclimatic control and the greenhouse emissions was reduced by 38%. The total cost of the package was around 123,610 €

			environmental benefits, by evaluating their economical profitability.	
6.	Harrestrup and Svendsen (2015)	A multi-storey building Copenhagen, Denmark	Investigation of possibility to carry out moisture safe energy renovations and saving 50% of the building's energy consumption by use of existing technologies.	In the analysed project moisture safety was put prior to energy savings. Internal insulation and mechanical ventilation systems were installed and analysed, and the windows were replaced with new ones. It was found that the measured annual energy consumption was reduced to 47%.
7.	Bellia et al. (2015)	Palazzo Fuga, Naples, Italy	Energy efficiency interventions on the building envelope and its plants.	Results showed, that: - an opaque PV roof results to be more convenient than a semitransparent PV roof, for aesthetic, energy production and the-mal comfort aims; - given the windows' geometry and location, no significant variation in daylight access can be obtained without modifying the envelope's aesthetic; -the presence of thick and heavy external walls requires no intervention of thermal insulation; - the estimated PV energy production with the opaque solution is about 107.9 MWh/year; - interventions on windows and on the floor at the VII level determine estimated energy savings of about 27.1% for the entire heating season; - the estimated annual reduction of carbon dioxide emissions is 32.1 t.
8.	Camuffo et al. (2010)	Church of Santa Maria Maddalena in Rocca Pietore, Italy	The proposed novel heating strategy is based on low-temperature radiant emitters mounted in a pew to provide a desirable distribution of heat to the feet, legs and hands of people occupying it; while leaving the conditions in the church, as a whole, undisturbed.	Due to little heat dispersion, this novel system significantly reduces the risk of mechanical stress in wooden artworks and panel or canvas paintings, fresco soiling and cyclic dissolution-recrystallization of soluble salts in the masonry, but is also energy-efficient.
9.	Samek et al. (2007)	Wooden church of St. Michael Archangel in Szalowa, Poland	Investigation in detail the effect of electric overhead radiators on the indoor environment of historic churches.	The heating system was not found to increase the concentration of SPM indoors; no re-suspension of particles already present in the church was observed. This work has demonstrated that the overhead radiant heaters are capable of providing localised heat to the areas where people congregate without adversely affecting painted walls and the works of arts displayed in churches.

Table 2: Case studies on historic building and BIM

Nr.	Authors	The purpose of use of BIM	Results
1.	Simeone et al. (2014)	To enhance information management during the investigation and restoration activities.	The core of the presented model is the integration of a BIM-based modelling environment and a knowledge base developed by means of ontologies, in order to represent all the semantics needed for a comprehensive representation of the historical artefact. Archaeological investigation process of the Castor and Pollux temple at Cori, Italy was made.
2.	Titomanlio et al. (2014)	BIM is the concrete method allowing the peculiarities of the architectural/restoration design and the structural counterpart to be simultaneously contemplated. In the BIM methodology for structures all the features of the virtual model are suitably contemplated.	In this paper, the BIM for cultural heritage is dealt with, outlining its main steps (inspections, measurements and surveys, non-destructive investigations, integrated structural modeling and s.o.) and configuring the new B.I.M.4S&D methodology. The latter is shown through the application to a historic masonry building.
3.	Rua et al. (2013)	This work is presenting an urban application developed into a GeoBIM tool, ESRI City Engine Software (CE), that integrates GIS (Geographic Information Systems) and BIM concepts.	Spatial analyses were conducted with the use digitalization - interactive representation of urban centres the downtown Lisbon city engine model.
4.	Dore et al. (2013)	For generating digital heritage models from laser scan or photogrammetric data using HBIM.	A library of parametric architectural objects has been designed from historic manuscripts and architectural pattern books. These parametric objects were built using an embedded programming language within the ArchiCAD BIM software called Geometric Description Language (GDL). Procedural modelling techniques have been implemented to create a parametric building facade which automatically combines library objects based on architectural rules and proportions.
5.	Garagnani et al. (2013)	To introduce a methodology destined to process point cloud data in a BIM environment with high accuracy.	Results of experiences on monumental sites documentation, generated through a plug-in written for Autodesk Revit and codenamed GreenSpider after its capability to layout points in space as if they were nodes of an ideal cobweb.
6.	Ludwig et al. (2013)	To counteract the process of disintegration and to get three dimensional building model of the Church of St. Catherine's.	The Gothic Church of St. Catherine was largely destroyed in a devastating bombing raid on January 2nd 1945. For the preparation of parametric architectural mode a heterogeneous set of data was used. The prepared BIM model was perfect for delivering impression of the interior and exterior of the church to present observers.
7.	Apollonio et al. (2013)	To prepare virtual reconstruction of buildings just documented by partial sketches, or partially built, or no more existing.	Conceptual similarity between the treaties and BIM were presented; a new and more robust solutions to the 3D modeling from 2D drawings for cultural heritage artifacts, able to allow the verification of the assumptions used during the reconstruction pipeline were presented.
8.	Baik et al. (2013)	BIM was used as a method to document and manage buildings of historical city of Jeddah.	New approach for heritage recording by using introduced Jeddah Historical Building Information Modelling (JHBIM).
9.	Murphy et al. (2013)	Development of conservation documentation; to automate conservation documentation in the form of engineering drawing and schedules from laser scan and image based surveys of historic structures.	A historic framework for building a parametric library of architectural elements was proposed. Several solutions for the development of parametric and shape rules to reproduce classical elements detailed using geometric descriptive language (GDL) were presented. A design scenario for end users was proposed to assess the suitability of HBIM as a tool for automation of engineering drawings for conservation process.

4. DISCUSSION AND CONCLUSIONS

The review and analyses presented in this paper showed that heritage buildings could be considered as high energy performance buildings with innovative retrofit strategies. Such buildings cannot have typical retrofit strategies. However, innovative methods for the energy retrofit of heritage buildings should be implemented in all heritage projects.

Heritage buildings should also ensure required indoor microclimate conditions to save artworks and to insure good thermal comfort for people staying indoors.

As a conclusion of this paper, recommendations could be given to officers working at different levels (continental, EU, country, town) and dealing with preservation of heritage buildings. In order to save, monitor and retrofit these buildings in sustainable and digital way, some actions should be taken. Figure 2 presents fields where BIM could be used to ensure sustainable preservation and implementation of progressive actions during inventory and monitoring, retrofitting and strategical planning.

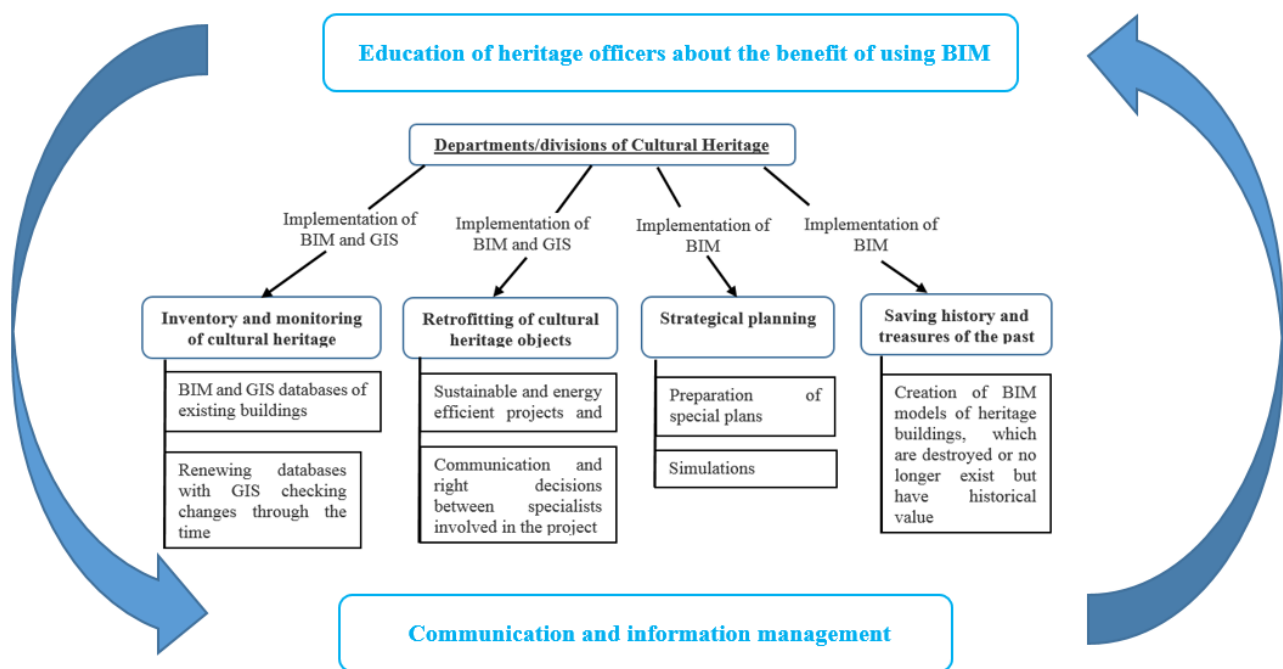


Figure 2: Digitalization of decision making in cultural heritage

Officers working in ministries and municipalities should invite BIM professional in order to ensure sustainable and advance work and decision making. Visualizations of the projects or objects always helps to ensure smooth communication between different specialists.

Using BIM and GIS for inventory and monitoring of heritage buildings will give possibilities to register changes through the time and take actions for preservation of cultural heritage.

BIM could be used in cases when heritage buildings no longer exist or there are only ruins. If the object has high historical value it could be recreated in digital way and presented through web platforms or Aps applications.

BIM could be used for strategical planning, when special plans are prepared. With BIM different simulation scenarios could be prepared for revival of heritage buildings and areas around them. With the right

solutions, the functions of heritage buildings could be changed and the area could get “authority”.

Each heritage building is unique and all actions should be taken case by case, taking into account all aspects.

The success of BIM depends on the size of the project, skills of team members in BIM, communication within the project team members.

In general, adaptation of BIM would be higher if there would be more experience within the project team and external organisations.

ACKNOWLEDGMENTS

This work was supported by COST Action TD1406 “Innovation in Intelligent Management of Heritage Buildings (i2MHB)”.

The author would like to thank colleagues from Faculty of Civil Engineering and Architecture (KTU): prof. K.Zaleckis, assoc. prof. dr. V. Karvelytė-Balbierienė and assoc. prof. dr. D.Rekus for consultancy.

REFERENCES

- Ajayi S.O., Oyedele L.O., Ceranic B., Gallanagh M., Kadiri K.O., 2015. Life cycle environmental performance of material specification: a BIM-enhanced comparative assessment. *International Journal of Sustainable Building Technology and Urban Development* 6(1): 14-24.
- Apollonio F. I., Gaiani, M., Sun, Z., 2013. 3D modeling and data enrichment in digital reconstruction of architectural heritage. *Proceedings of 24th International CIPA Symposium*, pp. 43-48. September 2-6, Strasbourg, France.
- Ascione, F., Bianco, N., De Masi R.F., De'Rossi F., Vanoli G.P., 2015. Energy retrofit of an educational building in the ancient center of Benevento. Feasibility study of energy savings and respect of the historical value. *Energy and Buildings* 95: 172-183.
- Baik A., Boehm J., Robson S., 2013. Jeddah historical building information modelling "JHBIM" old Jeddah – Saudi Arabia. *Proceedings of 24th International CIPA Symposium*, pp. 73-78. September 2-6, Strasbourg, France.
- Bellia L., Alfano F.R.D., Giordano J., Ianniello E., Riccio G., 2015. Energy requalification of a historical building: A case study. *Energy and Buildings* 95: 184-189.
- Ben H., Steemers K., 2014. Energy retrofit and occupant behaviour in protected housing: A case study of the Brunswick Centre in London. *Energy and Buildings* 80: 120-130.
- Burattini C., Nardecchia F., Bisegna F., Cellucci L., Gugliermetti F., Vollaro A., Salata F., Golasi I., 2015. Methodological approach to the energy analysis of unconstrained historical buildings. *Sustainability* 7(8): 10428-10444.
- Camuffo D., Pagan E., Rissanen S., Bratasz L., Kozłowski R., Camuffo M., Valle A., 2010. An advanced church heating system favourable to artworks: A contribution to European standardisation. *Journal of Cultural Heritage* 11(2): 205-219.
- Del Giudice M., Osello A., 2013. BIM for cultural Heritage. *Proceedings of 24th International CIPA Symposium*, pp. 225-229. September 2-6, Strasbourg, France.
- Dore C., Murphy M., 2013. Semi-automatic modelling of building facades with shape grammars using historic building information modeling. *Proceedings of Conference on 3D Virtual Reconstruction and Visualization of Complex Architectures (3D-ARCH)*, pp. 57-64. February 26-26, Trento, Italy.
- Garagnani S., Manferdini A.M., 2013. Parametric accuracy: building information modelling process applied to cultural heritage preservation. *Proceedings of Conference on 3D Virtual Reconstruction and Visualization of Complex Architectures (3D-ARCH)*, pp. 87-92. February 26-26, Trento, Italy.
- Grytli E., Kvaerness L., Rokseth L.S., Ygre K.F., 2012. The Impact of Energy Improvement Measures on Heritage Buildings. *Journal of Architectural Conservation* 18(3): 89-106.
- Harrestrup M., Svendsen S., 2015. Full-scale test of an old heritage multi-storey building undergoing energy retrofitting with focus on internal insulation and moisture. *Building and Environment* 85: 123-133.
- Hichri N., Stefani C., De Luca L., Veron P., Hamon G., 2013. From point cloud to BIM: a survey of existing approaches. *Proceedings of 24th International CIPA Symposium*, pp. 343-348. September 2-6, Strasbourg, France.
- Khosrowshahi F., Arayici Y., 2012. Roadmap for implementation of BIM in the UK construction industry. *Engineering, Construction and Architecture Management* 19(6): 610-635.
- Kylili A., Fokaides P.A., Vaiciunas J., Seduikyte L., 2015. Integration of Building Information Modelling (BIM) and Life Cycle Assessment (LCA) for sustainable constructions. *Journal of Sustainable Architecture and Civil Engineering*, 13(4), 28 – 38.
- Lassandro P., Cosola T., Tundo A., 2015. School building heritage: energy efficiency, thermal and lighting comfort evaluation via virtual tour. *Proceedings of 6th International Building Physics conference*, pp. 3168-3173. June 14-17, Torino, Italy.
- Ludwig M., Herbst G., Rieke-Zapp D., Rosenbauer R., Rutishauser S., Zellweger A., 2013. The advantages of parametric modeling for the reconstruction of historic building. The example of the in war destroyed church of St. Catherine (Katharinenkirche) in numbering. *Proceedings of Conference on 3D Virtual Reconstruction and Visualization of Complex Architectures (3D-ARCH)*, pp. 161-165. February 26-26, Trento, Italy.
- Mezzino D. 2014. The digitalization of Cultural Heritage's tangible & intangible dimensions. *Proceedings of 12th International Forum of Studies -The Paths of the Merchants*, pp. 115-122. June 12-14, Aversa, Italy.
- Moran F., Blight T., Natarajan S., Shea A., 2014. The use of Passive House Planning Package to reduce energy use and CO2 emissions in historic dwellings. *Energy and Buildings* 75: 216-227.
- Murphy M., McGovern E., Pavia S., 2013. Historic building information modelling – adding intelligence to laser and image based surveys of European classical architecture. *ISPRS Journal of Photogrammetry and remote Sensing* 76: 89-102.
- Pankhurst C., Harris A., 2013. Conservation and Innovation - The Challenge of 'Eco' Renovation in Heritage Buildings. *Journal of Architectural Conservation* 19(2): 18-34.

- Pisello A.L., Petrozzi A., Castaldo V.L., Cortana F., 2016. On an innovative integrated technique for energy refurbishment of historical buildings: Thermal-energy, economic and environmental analysis of a case study. *Applied Energy* 162: 1313–1322.
- Rua, H., Falcao, A.P., Roxo, A.F., 2013. Digital Models - Proposal for the Interactive Representation of Urban Centres The downtown Lisbon City Engine model. *Proceedings of 31st International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe)*, pp. 265-273. September 18-20, Delft, Netherlands.
- Sahin C. D., Arsan Z.P., Tuncoku S.S., Brostrom T., Akkurt G.G., 2015. A transdisciplinary approach on the energy efficient retrofitting of a historic building in the Aegean Region of Turkey. *Energy and Buildings* 96: 128-139.
- Samek L., Maeyer-Worobiec A., Spolnik Z., Bencs L., Kontozova V., Bratasz L., Kozłowski R., Grieken R., 2007. The impact of electric overhead radiant heating on the indoor environment of historic churches. *Cultural Heritage* 8: 361-369.
- Simeone, D., Cursi, S., Toldo, I., Carrara, G., 2014. B(H) IM - Built Heritage Information Modelling Extending BIM approach to historical and archaeological heritage representation. *Proceedings of 32nd International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe)*, pp. 613-622. September 10-12, Newcastle upon Tyne, England.
- Titomanlio, I., Faella, G. BIM and Structural Modeling: Application on Cultural Heritage. *Proceedings of 12th International Forum of Studies -The Paths of the Merchants*, pp. 973-979. June 12-14, Aversa, Italy.
- Todorovic M.S., Ecim-Duric O., Nikolic S., Ristic S., Polic-Radovanovic S., 2015. Historic building's holistic and sustainable deep energy refurbishment via BPS, energy efficiency and renewable energy—A case study. *Energy and Buildings* 95: 130-137.
- Vanlande R., Nicolle C., Cruz C., 2008. IFC and building lifecycle management. *Automation in Construction* 18(1): 70-78.
- Volk R., Stengel J., Schultmann F., 2014. Building Information Modelling (BIM) for existing buildings – Literature review and future needs. *Automation in Construction* 38: 109-127.
- Wong J.K.W., Zhou J., 2015. Enhancing environmental sustainability over building life cycles through green BIM: a review. *Automation in Construction* 57: 156-165.

DISCRETE EVENT SIMULATION TO IMPROVE CUTTING-WELDING WORKSHOP IN A SHIPYARD

Marcos Rouco-Couzo^(a), Diego Crespo-Pereira^(b), Alejandro Garcia-del-Valle^(c), Mar Cebral-Fernandez^(d), Rosa Rios-Prado^(e)

^(a) ^(b) ^(c) ^(d) ^(e)University of A Coruña UMI Navantia-UDC

^(a)marcos.rouco@udc.es, ^(b)diego.crespo@udc.es, ^(d)alejandrogarcia.delvalle@udc.es, ^(c)mar.cebralf@udc.es,
^(e)rosa.rios@udc.es,

ABSTRACT

The Cutting-Welding workshop is the first step of a shipbuilding process where the basic structural elements of the ship are fabricated. It is a fairly automated process where the main operations can be automated using robots. However, the geometrical diversity of the elements produced cause challenges from an operations management point of view since the throughput rates change and depend on the ship and the block types. Modelling and Simulation provides a practical methodology for analyzing and optimizing this workshop. This paper describes the development, validation and initial results of a discrete events simulation model of a Cutting-Welding workshop. It is intended to assist in the planning phase of a new ship, particularly focused on the capacity design of this workshop. Its applicability has been tested on a realistic scenario based on an existing shipyard with the aim of evaluating the accuracy of the model. Preliminary results show that the proposed simulation framework can be successfully applied to process planning in shipbuilding providing improvements in terms of time and costs.

Keywords: Workshop Design, Discrete Event Simulation, Shipbuilding, Shipyard, Cutting-Welding.

1. INTRODUCTION

Shipbuilding is a complicated process where collaboration of many actors both intra and inter-organizational is necessary (Tann & Shaw, 2007). Nowadays, much time, efforts and resources are spent by all parties, especially by ship-owners and shipyard engineers in coordinating all design details of the ship under construction (Solesvik et al 2012). This ambitious goal needs efficient coordination activities in order to provide the complex solutions for the ship-owners fluctuating workload and strong competition in terms of price, time to market and product quality. Its relevance is reflected by the fact that up to 70 % and more of a

shipyard's value creation is based on purchased equipment and services (Bronstert, Gau, Luckau, & Sucharowski, 2005).

Moreover, a large number of shipbuilding-related firms have great concern for offering complete vessels, designed and outfitted by their customers, within the previously established schedule. Therefore, in order to make shipbuilding projects successful, the optimization of quality, time and cost must be taken into account as the most important aspects to consider in the evaluation of any project.

This environment has led to the emergence of the Industry 4.0 which, as Kolberg & Zühlke (2015) state, is based on machines and devices that have their own intelligence and, as a result, a factory becomes "smart". According to this phenomena, this new Industry gives paramount importance to the optimization processes by implementing a dynamic production. Besides, it also can be integrated in Lean Production providing outstanding improvements in production processes.

In line with these current developments, some of these ship-owners have decided to invest in innovative software which aims to support design and coordination activities along the shipbuilding process. Shin et al., (2004), analyzed, using the DELMIA QUEST software, a subassembly line in a large shipyard and the different processes. They were able to simulate the material flow in the line in order to show the bottlenecks and the interferences between objects.

There are multiple difficulties when a complex system like a shipyard has to be described realistically. Mathematical analyses, most of the times, do not reflect in the best possible way all the casuistic that may occur in a huge structure like this one. Simulation enables not only to replicate the whole process (with all the particularities) but also to recognize and solve problems in the system (bottlenecks, interruptions, etc.) (Ljubenkovic, Dukic, & Kuzmanic, 2008).

As a result of the importance of simulation methodologies, many efforts are being made in this field. However, these approaches are not always

successful due to the requirement of much time and effort for developing a new system. In addition, it is very difficult to adapt a commercial simulation software to the process planning (CAD information, scheduling information). However, it is possible to set up accurate process planning in advance if a consistent simulation framework is developed (Cha & Roh, 2010).

Burnett et al., (2008) developed, by means of the 3D software Flexim, a discrete event model simulation with a methodology that provided a rapid generation of models from the enterprise database. This Flexim model reads data from PLM software using the standard Excel format.

Using a virtual shipyard provides a perfect testing framework to find the best solution to apply in the real world. Unfortunately, due to the efforts required to develop an efficient simulation model, many shipyards are reticent to use these techniques. But, there are some shipyards like Jos, L. Meyer and Aker Ostsee, in Germany, that already use a simulation tool successfully (Krause, Roland, Steinhauer, & Heinemann, 2004).

Simulation is also a tool for enterprises in the project planning activities. In manufacturing industry, such as ship assembling, these kind of simulation models are used successfully to improve production lines. Related to this, König, et al., (2007), proposed by means of the Plant Simulation software, a simulation framework to model outfitting processes in shipbuilding and building engineering. Using a constraint-based simulation approach, different practical outfitting schedules were created and evaluated in terms of work and material flow organization, utilization of space and worker's efficiency as well as process costs.

Thanks to these achievements, nowadays discrete events simulation has become an essential part in order to improve the overall process in a shipyard.

This paper focuses in the simulation model of a Cutting-Welding Workshop of a shipyard. This is the first stage of the manufacturing process which is required to produce in time all the components that will be assembled to build the structure of the ship. The case study will highlight the two main operations performed in this workshop: Cutting and Welding.

The products to be processed will also be of two kinds: steel pieces and structural profiles in order to assembly them to create panels and webs (longitudinal girders, web frames, double bottom girders, soled floors, bulkheads, beams etc.). The study will reflect these processes in this shipyard workshop by means of a discrete event simulation model created by a Discrete Event Simulation Software called ExtendSim.

It should be noted that a Cutting-Welding Workshop can be considered as the first one among all the others in the shipyard. The products (panels and webs) generated in this facility will be the required materials to create the rest of the blocks in all other workshops of the shipyard.

The identification of bottlenecks, the adjustment of manufacturing time and minimizing the makespan are the main objectives of this model.

The paper is organized as follows. Section 2 presents issues related to the manufacturing process in this shipbuilding. Section 3 presents the validation of the model. Then, section 4 states the results of the proposed simulation model. Finally, the main conclusions of the paper are summarized.

2. PROCESS DESCRIPTION

In order to describe the whole process in this workshop a simplified explanation will be presented in the following lines.

Two kinds of products are manipulated: structural profiles (stiffeners) and steel pieces. These products will suffer different processes depending on the element they will be part of: a web or a panel, (the intermediate and structural products that will be the base of the future block).

Cutting and Welding processes can be taken as the principal ones. First of all, the cutting process starts with the reception of the profiles and steel sheets (usually with dimensions of 12 x 3 meters). Both products are split in several pieces in order to create the intermediate ones (panels and webs).

The profiles suffer a blasting process before they are cut by an oxi-cut machine. The original steel sheet is cut by means of two plasma stations, depending on if the resultant pieces will be destined to webs or panels.

The second part consists in the welding process: the base of the merger pieces. Once the pieces are cut, there are two main destinations to where the pieces can be moved: to the Panel Line or to the Web Line. All welding processes take place in these two Lines.

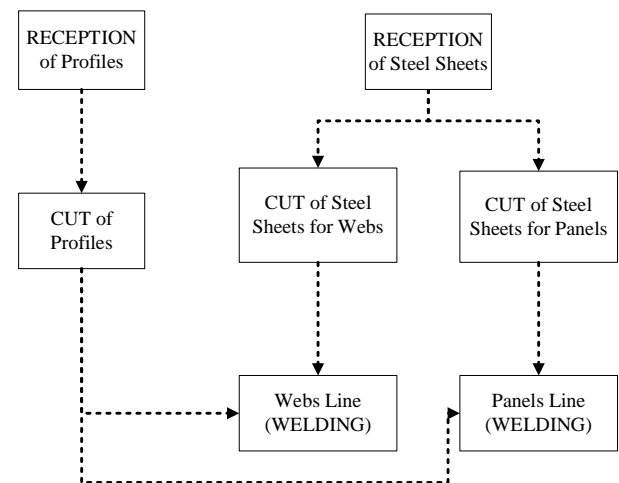


Figure 1: Main processes in cutting-welding workshop

The Panel Line can be described as a completely automated area where, firstly steel pieces (large plates) and secondly large profiles, are processed to manufacture the Panel. In the first part of the Line, the large plates are assembled in the One Side Welding station in order to create the base of the panel. In the second part, many stiffeners are welded to provide the structural strength to the panel.

The Web Line is described as an area where steel pieces and profiles go through the manufacturing process, that is, they are assembled, in order to create the structural component (web). The differences in shape between webs is a handicap in this process. Most webs are unique and vary depending on the structural function they will be destined for (longitudinal girders, web frames, double bottom girders, soled floors, bulkheads, beams etc.). The Figure 1 shows the simplified summary of the process.

2.1. Model development

The simulation of the cutting-welding workshop has been made with the Discrete Event Simulation software ExtendSim. The high level of resolution that this model requires makes ExtendSim a perfect solution to represent this workshop.

A discrete event simulation model has been created from the information stored in an Excel file. This file contains the information related to the items represented in the model. The ExtendSim model “reads” the data stored in the Excel file with the aim of creating the attributes of the items.

Figure 2 shows a small section of the model, specifically the arrival of the steel sheets and the destination to the appropriate plasma station. After that, the cutting process takes place and multiple pieces are created.

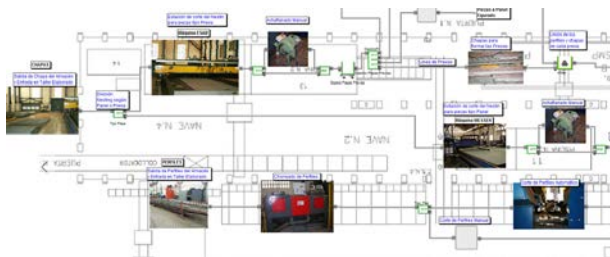


Figure 2: Section of the ExtendSim model

The data for the simulation model was obtained from historical records and visits to the shipyard. The times of the cutting and welding operations were estimated from cutting and welding lengths in the registries and engineering planes. An extensive effort in data analysis was carried out in order to fit the adequate statistical distributions for each block type. A set of correction factors were estimated to extrapolate the results from a specific ship to other ships that have different structural designs.

3. MODEL VALIDATION

Once the initial version of the model was build according to the information provided by the workshop managers and the information gathered during visits to the plant, a validation phase was carried out. The validation consisted of two related activities:

- A data based validation in which the real data from two workloads was compared to the simulated results for the two workloads.
- An experts based validation in which the results of the model were shown to a set of experts and their reviews and comments taken into account for improving the models.

The simulation experiment for the validation consisted of compering the flow times of the products of two selected workloads through different parts of the process. The flow times estimated by the model were compared to the historical data and represented in a Value Stream Mapping (VSM) type of diagram. The results for the workload 1 are shown in Figure 3. In this chart, the triangles symbolize delays between processes and the total flow time of each process is represented in the white squares. The reason for using this type of diagram was establishing the difference between the durations of the real processing steps and the delays in the historical records caused by the assignment of other orders to this process. The time units of the diagram are hidden due to confidentiality reasons and the real values of the times have been scaled by a confidential value.

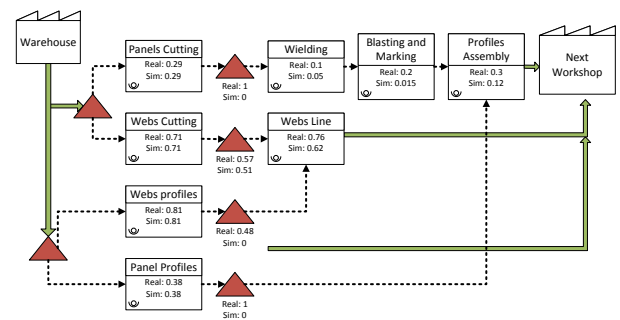


Figure 3: VSM of workload 1.

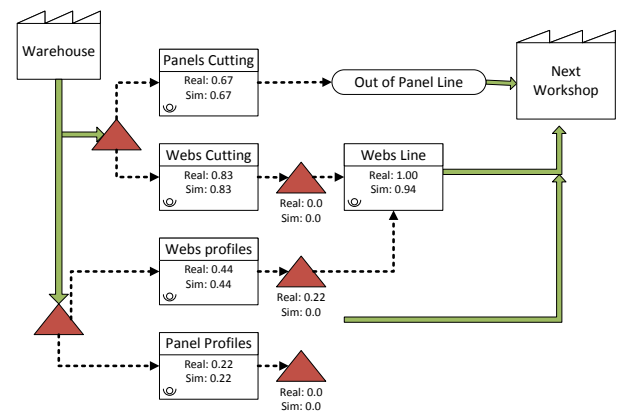


Figure 4: VSM of workload 2.

The results of this validation experiment showed differences in some of the process stages. However, once the causes of this differences were checked with the process experts, it was found out that they were caused by the assignment of higher priority orders which were not part of the experiment. Thus, the

workload 1 scenario was only used valid for checking some of the times.

The Figure 4 represents the same comparison for the workload 2 scenario. This second scenario corresponds to another case with more controlled conditions in the plant and the results show a good match between the model predictions and the historical data. The model was thus accepted as a valid representation of this workshop.

4. EXPERIMENTATION AND RESULTS

The experimentation phase presented in this paper is a preliminary analysis of the workshop evaluated with a realistic production plan. The scenario does not correspond to a real ship but to a representative mix of block types that are common in most of ships.

The main two challenges face for defining a scenario for experimentation are the following.

The goal of this workshop is to produce panels and webs in time for the next steps of the process. The composition in panels and webs of a ship's blocks is heterogeneous. Thus, the workshop never reaches a "stationary state" as in mass production processes. Hence, the simulation scenario must reproduce the workload of a ship and the response variables must be the total makespan for producing each type of product.

The objective of the simulation experiment is to provide a capacity design of this workshop to satisfy the makespan requirements at minimum cost. In mass production systems, the throughput of the line is given by the bottleneck station so the only way to improve the system is to act on it. However, in this system the requirement is to satisfy a makespan requirement and this can be achieved by acting on different stages of the process. Thus, for instance, it could be more convenient to allocate overcapacity in a low cost workstation than in a costlier one even if the second would be the stationary state bottleneck.

4.1. Capacity Design

The factors initially considered in the simulation experiment were:

- The number of cutting machines.
- The number of parallel web welding stations and shifts.
- The number of parallel panel welding lines, profiles assembly and shifts.
- The number of cranes used for internal transport.

An initial screening process was made to assess which were the limiting resources. It was found that the webs line was the slowest one. Figure 5 shows a graph comparing the number of panel and web units completed along the simulation time (which is expressed in generic time units). An important feature that can be observed in the graph is that the slope of the graph is not constant. This is caused by the variation in the types and dimensions of the webs that happen

between different blocks. It shows that in this process is not possible to define a maximum throughput capacity in units of items per time because it depends of the geometry of the webs being produced.

Table 1 shows the utilization results for the webs fabrication line. It can be noticed that the Cutting process is less busy than the welding operation. Also, adding additional capacity for this step is less costly since there is the option of manual welding webs. Thus, in this case the best strategy to increase the capacity of this workshop would be to act on the welding webs station. However, the decision on whether this increase is really necessary would depend on whether the total makespan for the simulated scenario is enough or not.

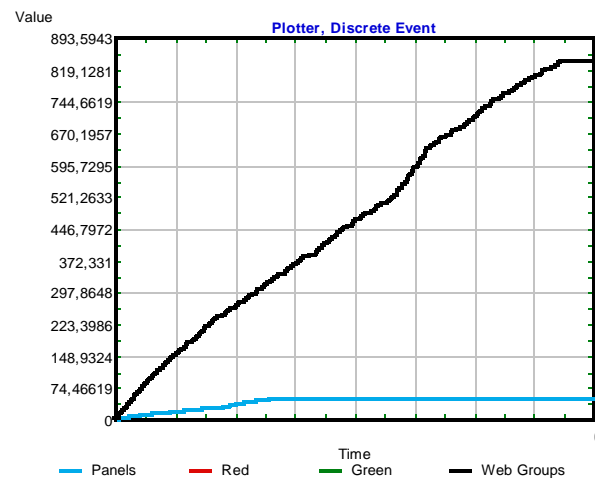


Figure 5: Number of web and panel units completed.

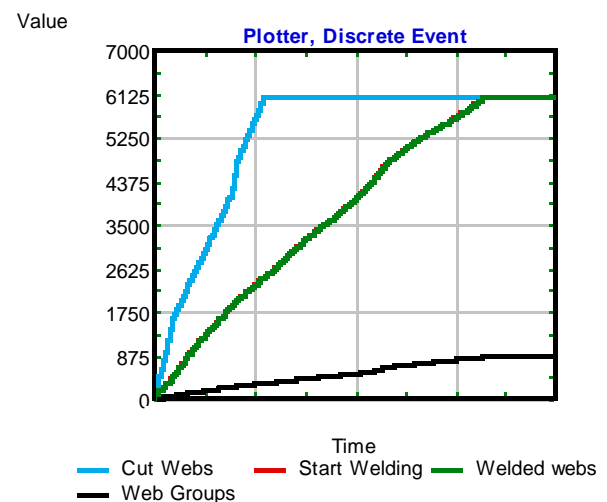


Figure 6: Number of web units that complete each step of the process.

The figure 6 displays the number of webs that have completed the cutting process, welding and which have been grouped in batches for transport to the next step of the process. From this plot, it can be concluded that the

bottleneck of this process is the welding station. Table 1 supports this conclusion showing the utilization rate.

Table 1: Utilization of the webs line.

Utilization (5 runs)	
Cutting Webs	29.0%
Welding Webs	97.2%

The following heuristic procedure for capacity design was proposed for future application to this case:

1. Define a workload scenario and the maximum admissible makespan.
2. Run the simulation model and identify which are the resources or stations that delay the fabrication the most. The previously shown graphs are a useful tool for doing so.
3. Make a sorted list of the resources that could be increased or reduced. Estimate the cost of increasing it or the saving from reducing it.
4. If the slowest station is the same as the one with the lowest cost of increasing it, run the model after making this change and return to step 2.
5. If they are different, simulate the scenarios of increasing the capacity of the less costly resource and the slowest one. If the makespan constraint is satisfied by acting on the less expensive resource, select this option. If not, select the option that reduces the makespan the most. Return to step 2.
6. If no more changes are required to satisfy the makespan constraint, then reduce the capacity of the least used resources that yield the highest savings in a similar way, stopping when the makespan requirement is no longer satisfied.

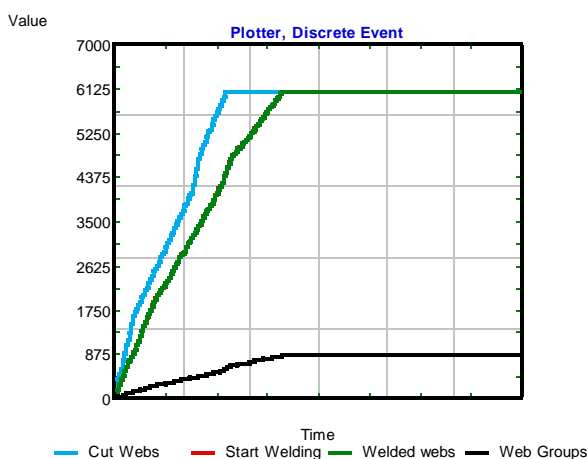


Figure 7: Number of web and panel units completed.

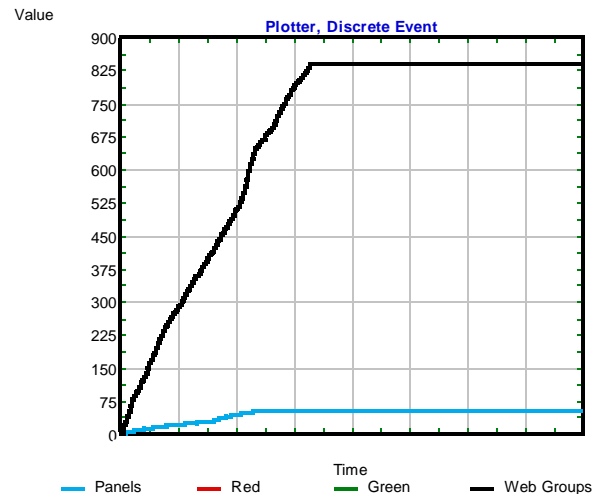


Figure 8: Number of web and panel units completed.

The figures 7 and 8 show the results after applying the procedure to increase the capacity of the welding stations. The total makespan for the cutting and welding operations could be reduced to half by acting on the critical activities of the welding process.

Thus, the throughput of panels and webs can be better synchronized. It must be noted that this results do not correspond to a real workload scenario. It is an scenario arranged for demonstrating the capabilities of the simulation model and the methodology for designing the capacity of the real workshop.

5. CONCLUSION

The simulation model of a Cutting-Welding workshop has been presented. The methods employed for validating the model have been described. They show a reasonable degree of similarity between the real data and the simulation results. Some differences were observed in some delays, but further analysis showed that they were caused by other pending workload that was blocking the process during the period analyzed.

The preliminary experimentation was focused in the capacity design. The simulation model proved to be a useful tool to select the number of required cells and how to meet the time requirements with the minimum number of resources. The plots generated showing the increase in completed elements in each stage of the process was an useful tool for identifying the bottleneck resources and selecting the proper course of action to improve them.

ACKNOWLEDGMENTS

The authors are thankful to Unidad Mixta de Investigación (UMI) Navantia-UDC for its valuable support.

REFERENCES

Bronart, R., Gau, S., Luckau, D., & Sucharowski, W. (2005). Enabling Distributed Ship Design and Production Processes by an Information

- Integration Platform. In *2005 International Conference on Control, Automation and Systems* (pp. 1–10).
- Burnett, G. A., Medeiros, D. J., Finke, D. A., & Traband, M. T. (2008). Automating the development of shipyards manufacturing models. In *Winter Simulation Conference* (pp. 1761–1767).
- Cha, J. H., & Roh, M. I. (2010). Combined discrete event and discrete time simulation framework and its application to the block erection process in shipbuilding. *Advances in Engineering Software*, *41*(4), 656–665.
- Kolberg, D., & Zühlke, D. (2015). Lean Automation enabled by Industry 4.0 Technologies. In *Proceedings of 2015 International Federation of Automatic Control* (Vol. 48, pp. 1870–1875).
- König, M., Beißert, U., Steinhauer, D., & Bargstädt, H.-J. (2007). Constraint-based Simulation of Outfitting Processes in Shipbuilding and Civil Engineering. *Proceedings of the 6th EUROSIM Congress on Modeling and Simulation*.
- Krause, M. ., Roland, F. ., Steinhauer, D. ., & Heinemann, M. . (2004). Discrete event simulation: An efficient tool to assist shipyard investment and production planning. *Journal of Ship Production*, *20*(3), 176–182.
- Ljubenkov, B., Dukic, G., & Kuzmanic, M. (2008). Simulation Methods in Shipbuilding Process Design I, *54*, 131–139.
- Shin, J., Lee, K., Woo, J., Kim, W., Lee, J., Kim, S., ... Yim, H. (2004). A Modeling and Simulation of Production Process in Subassembly Lines at a Shipyard. *Journal of Ship Production*, *20*(2), 79–83.
- Tann, W., & Shaw, H. J. (2007). The collaboration modelling framework for ship structural design. *Ocean Engineering*, *34*(5-6), 917–929.

IMPROVING PLANNING AND RESOURCE UTILIZATION OF A SHIPBUILDING PROCESS BASED ON SIMULATION

Mar Cebal-Fernandez^(a), Diego Crespo-Pereira^(b), Alejandro Garcia-del-Valle^(c), Marcos Rouco-Couzo^(d)

^{(a), (b), (c), (d)} University of A Coruña and UMI Navantia-UDC

^(a)mar.cebralf@udc.es, ^(b)d.crespo@udc, ^(c)alejandrogarcia.delvalle@udc.es,
^(d)marcos.rouco@udc.es

ABSTRACT

This paper presents the experience obtained from the initial steps of the development of a simulation model of a shipyard. Shipbuilding is one of the most complex manufacturing processes due to the high number and diversity of elements that must be assembled. The process can only be partially automated and it requires the coordination of many resources and the resolution of complex scheduling algorithms. Modelling and Simulation M&S is a methodology suited for studying complex systems and as such, it offers a great potential for improving operations in this sector. The paper is focused in describing the general steps of the process and the difficulties found during the model development along with the solutions adopted. The preliminary results obtained emphasize the need for studying and applying scheduling methods in order to maximize the utilization of the resources and reduce the duration of a project.

Keywords: discrete events simulation, modelling and simulation, shipbuilding, scheduling.

1. INTRODUCTION

Shipbuilding of medium and large size vessels is a complex manufacturing process which is usually managed in a project-oriented approach. Each individual ship (specially for military applications) has some degree of customization and there are only few units based on the same design. Thus, from an operations management point of view, shipbuilding has probably more common features with the construction sector than with mass production processes.

Still, efforts for applying Lean principles and standardizing processes have led to a modular approach in which each ship is subdivided in fairly standard blocks which are assembled in the so-called block erection process. Under this approach, common in the sector already for many decades, the unit of production is the "block" for most of the steps of the process.

The competitiveness of a Shipyard is determined by its ability to make offers minimizing costs and the project duration. This requires to optimize the production process of the blocks as well as ensuring that they are

available for the erection process in time and in the proper order.

The processes involved to produce a block span cutting and welding of steels plates and profiles to form structural units as well as outfitting operations and painting. Ideally, the blocks assembled should contain all the structural elements as well as all the systems and equipment installed on them. In practice however, not all the elements can be effectively mounted before the block erection process due to physical constraints, availability of materials and time requirements. For instance, the tubes located in the union of two blocks are often mounted after the block assembly process since the tolerance control required to ensure that they match could be too high and require reworking.

The large number of operations required to produce the blocks of a ship, the necessity to synchronize multiple workflows and numerous resources make the management of such a production system very challenging. Modelling and Simulation M&S provides a powerful tool for analyzing decisions in such systems. This paper describes the main elements required to build a simulation model of a shipyard which can be applied for evaluating the effect of different blocks characteristics, planning methods and resources. The

2. M&S IN SHIPBUILDING

Process simulation began to develop in the 60s, when languages of discrete event systems appeared on the market. Since then, there have been significant advances in this field that have transformed the simulation in an indispensable tool for optimizing production processes. The simulation as a tool for improving systems, can be included within the practices or thought "Lean".

The shipbuilding industry has seen a mixed degree of extension of M&S tools. Some of the best shipyards adopted this technology at least since the 90s. However, many shipyards are not still aware of these tools and their potential. For instance, in the Spanish shipbuilding sector, in despite of being one of the countries with highest market share, M&S is in general an unknown technology for process optimization.

M&S technologies help to evaluate, decide and make production plans while enabling a continuous and

transparent review of the performance. This useful tool improves the project definition and the evaluation of the production of each type of vessel (Kasemaker, 2006).

In the shipbuilding industry, simulation models can be used as a tool to analyze the impact of programing a new workload, evaluate different scenarios and identify the resource constraints. Likewise, it can be used to analyze the expected results of the integration of new technologies or equipment in a shipyard, with their impact over the operating costs and on the planning processes (McLean, 2001). In fact, “the National Research Council (NRC) has repeatedly identified M&S as a high priority area... one of two breakthrough technologies that will accelerate progress in addressing the grand challenges facing manufacturing in 2020” (Leong, S., Y.T. Lee, 2006).

A peculiarity of shipbuilding is that manufacturing is not repetitive and hundreds if not thousands of different operations, labors and resources distributed along different workshops need to be synchronized. This causes high costs for developing simulation models which is a factor that makes difficult the application of M&S in this sector.

Due to the high costs of development and maintenance of the simulation models, cooperation among companies, universities and research centers is convenient. The group SimCoMar (Simulation Cooperation in Maritime Industries) is an example. The Flensburger Nordseewerke Emden shipyard, the universities TUHH (Technische Universität Hamburg-Hamburg), DUT (Delft University of Technology), ANAST (University of Liege), and the Center of Maritime Technology (CMT) in Germany are participating at this initiative. Besides SimCoMar, other partnerships have been established between shipyards and universities such as the University of Seoul South Korea, Japan's Kinki, Michigan University, and Federal University of Brazil (LABSEN laboratory) (Caprace, Jean-David Moreira Freire, Assiss, Martin Pires, & Rigo, 2011).

In 2006 the SIMoFIT is founded (Simulation resourcing in shipbuilding and Civil Engineering) as a professional collaboration between shipbuilding and civil engineering (Steinhauer, 2007). The shipyards Flensburger, the Bauhaus-Universität Weimar, the Ruhr University of Bochum and SimPlan AG as a consultant simulation joined this cooperation (Steinhauer, 2011). Other examples are the Meyer and Flensburger shipyards (Caprace, Jean-David Moreira Freire et al., 2011).

In addition to the high costs, model generation is “knowledge intensive, time consuming, and error-prone” (Lee & Kang, 1996). The time required to build accurate models is a barrier to the use of discrete event simulation in the shipyards (Medeiros & Williams, 2000). One way to face these challenges is using automatic programming techniques. According to Madden & Neill, 2005, the use of auto-generated models allows the simulation tool to find more widespread usage among the project managers.

Another problem is how to store and read the simulation data. Although there is not any standard format to store simulation data, some formats have been proposed such as SDX. Microsoft Excel spreadsheets are often used because they provide a simple and convenient interface to store and represent the system data (Burnett, Medeiros, Finke, & Traband, 2008).

In this paper, the ExtendSim software was chosen to focus only in the analysis of the process and not in the way of automatically processing process' data. Microsoft Excel is used to store all the data that configures a simulation scenario.

Some of the aspects that have been more extensively researched are the block erection process and the capacity design of the previous block production workshops.

The block erection problem can be seen as an assembly jobshop scheduling problem with additional constraints imposed by the ship geometry and the stability requirements of the erection process. To optimize the blocks schedule, one of the earliest works proposed the use of an algorithm called the constraint-directed graph search (CDGS) (Jae Kyu Lee, 1995). Other authors have opted for genetic algorithms (Okumoto, Yasuhisa, 2002 and Bao Jinsong, Hu Xiaofeng, 2009). Although the initial works focused on this problem separated from the blocks scheduling in the rest of the proces, in the recent years some authors have adopted a global approach employing different techniques of optimization (Jean-David Caprace, Clarice Trevisani Da Silva, Philippe Rigo, 2011).

The determination of the optimal number of cells in the previous workshops has been approached as a task scheduling problem. Scheduling problems with the jobs grouped in batches and spatial constraints are generally the most studied in the literature. The main research work about the spatial scheduling problem for shipyards was done in the context of the DAS project (Kyoung Jun Lee and Jae Kyu, 1996).

3. PROCESS DESCRIPTION

The manufacturing process in the shipyard studied includes the following steps: cutting-welding, assembly, outfitting 1, blasting and painting, outfitting 2, block erection and finally the ship is launched where the out fitting works are finished. Figure1, shows the shipbuilding process.

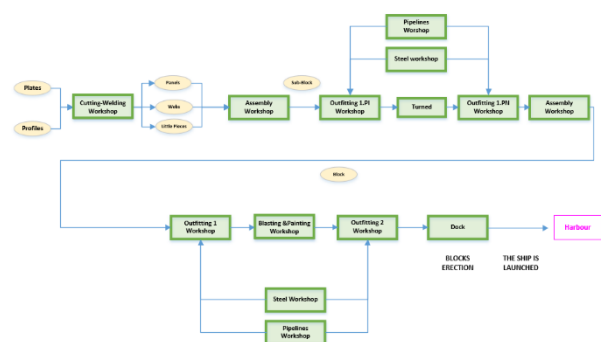


Figure1: Shipbuilding process

Firstly, steel plates are cut into small parts in accordance with the specified design in the cutting-welding workshop. The plates are then used along with metallic profiles to form the smallest assembled units that will constitute the structural elements of the blocks. Two main types of structural components are defined: panels and webs. The panels are large steel plates that can be curved in some cases and make up the hull, the deck and the different levels of the ship. The webs are smaller parts that provide resistance to the main structure and also divide spaces or perform other various functions.

The next step of the process is the sub-blocks assembly in the “assembly workshop” followed by the first outfitting process (Outfit 1). The assembly process mainly consists of welding operations among the panels and the webs, according to the constructive strategy designed for the sub-block. The outfitting process mainly consists of installing pipes, brackets, equipment and other auxiliary components inside the sub-block.

Then, contiguous sub-blocks are positioned together and welded to form the blocks. Other outfitting operations are performed thereafter. The next operations are blasting and painting the blocks in the painting stations. The high costs associated with this facilities and the long painting times mean that this is often a bottleneck.

After painting, the second outfitting process takes place (Outfitting 2). It involves installing wires and diverse equipment. In general, all the components that could be damaged in the blasting and painting process are assembled in this step. Finally, the blocks are transported to the dry dock to build the ship. The ship is then launched and the process is finished at the harbor.

Critical planning stages to be performed are splitting the ship in blocks and defining the optimal erection sequence. The erection process is a very complicated operation that involves decision-making taking into account a lot of structural items and the different erection strategies. On the other hand, it is not possible to optimize the assembly sequence in the blocks erection if the principal workshops do not have the appropriate number of cells in order to have the different blocks that are necessary on time and in the correct order.

4. MODEL DEVELOPMENT

The established goals for the development of this simulation model were:

- To provide an analysis tool capable of estimating the total makespan of building a ship from the beginning of the cutting operations to the launching and the lead times of each block at each workshop.
- To provide a tool capable of reproducing the effect of different numbers of workshop cells, internal transports vehicles, schedules, machines operation rates, welding rates,

workshops productivity and erection sequence and constraints.

- To design a flexible and parameterized model that is capable of simulating the building process of different ships.

Thus, a shipyard employing this model would get an important analysis tool to optimize for planning and capacity design. The model has been designed to provide an overall view of the shipyard, without entering into the details of the workshops. Thus, the capacity parameters are the number of cells in the main workshops of the shipyard (assembly, outfitting 1 and outfitting 2). Planning of all the shipbuilding processes is strongly linked to these workshops. Starting date and ending date of each block are imposed by these steps.

As pointed out before, the simulation software employed was ExtendSim. A screenshot of a part of the model is shown Figure 2. Although ExtendSim provides some 3D visualization features, they are not used at this stage of model development. The purpose is to obtain a model valid for conducting preliminary analysis of the shipyard which should later be refined by more detailed models of the workshops.

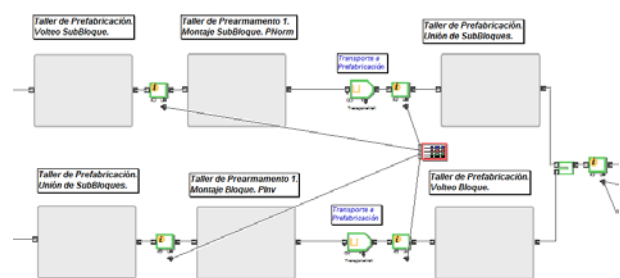


Figure 2: Model Simulation Shipbuilding Process

The main components of the simulation model are shown in Figure 3. The model starts its execution by reading data from an Excel file where the list of blocks along with their characteristics to be produced. The blue arrows and blocks indicate the steps of the process included in the simulation model. The green blocks and arrows are used for the scheduling and planning elements. The data in the Excel file also includes a schedule that determines the times from which the different blocks are allowed to initiate each step. When two blocks are available, the priority is given to the earliest one. The red blocks and arrows show the main resources which are shared among the different steps of the process and which are included in the model.

During the model development phase several challenges were faced. The first one is the high number of components types and diversity of elements that must be manufactured for building a ship. Just building the structure requires several types of steel components with diverse geometries. The most common types of processes are welding operations whose times depend on the welding lengths. Characterizing the welding lengths of the numerous elements is a time consuming

process which requires detailed engineering plans of the ship. However, this plans are not usually available at the early stages of planning for which this aggregated model is most useful. Due to this reason, during the model development a simplified approach was taken which consisted of employing different equations that estimate the total duration of each step as a function of some block features. The block's features used are:

- The block size in the three axis.
- The block type (according to a classification defined by the shipyard).
- The number of structural elements that form each block (such as the number of panels).
- The quantity of elements assembled in the outfitting process.
- The block's weight.
- Other specific attributes.

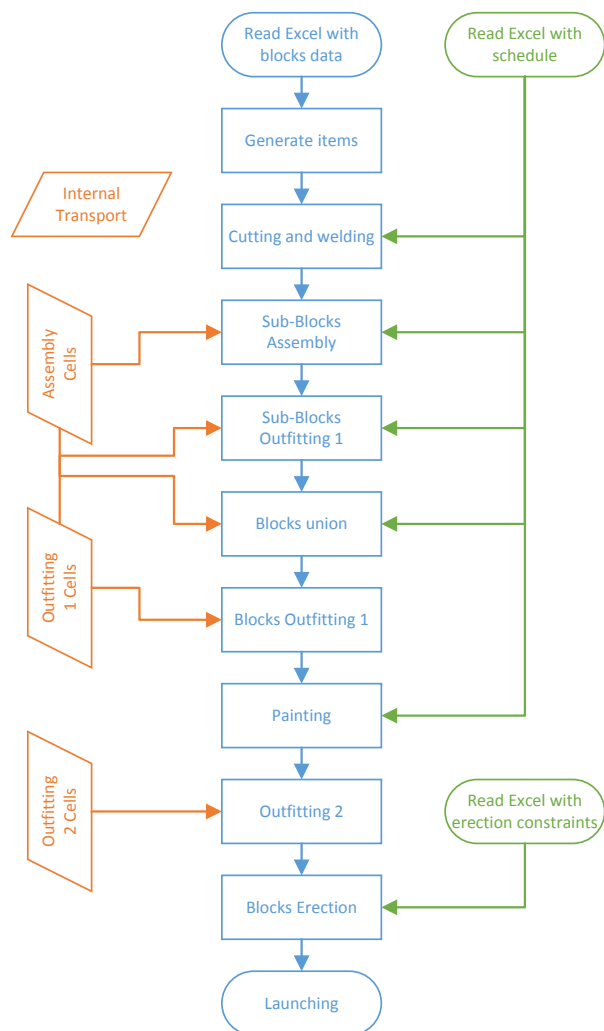


Figure 3: Simulation Model Components

Thus, in general, the model used for estimating the duration of the process step i is:

$$\bar{t}_i = f(\text{Size, Type, Elements, Weight, ...}) \quad (1)$$

The uncertainty in the duration estimated by the previous equation and the variability due to the process and the manual tasks is taken into account by a statistical distribution. In this model, Lognormal distributions are used. The reasons for this decision (supported by the experience of the team that developed this model) are the following.

- The lognormal distribution is a simple distribution with two parameters that allow to specify the mean and the variance, which are the two main parameters that affect the performance of a queuing system.
- The lognormal distribution is one the distributions that often fits well real data.
- The lognormal distribution arises from a random process of multiplicative variations of a variable. Thus, it reflects well the variability in production times whose variance grows with the mean values. For instance, the standard deviation of the duration of long tasks is usually larger than that of short tasks.

Thus, the model used for simulating the times of the processes in the model is:

$$t_i \sim \ln N(\bar{t}_i, CV \cdot \bar{t}_i) \quad (2)$$

Where CV stands for the coefficient of variance of the process.

Another critical challenge faced in the model development was the need to generate all the elements that will form the blocks' structure and that will be outfitted. A simple solution could be to define a single item in the model that represents all the elements that will be matched to form a block. However, this approach would reduce the capabilities of the model since it would not reflect properly the capacity constraints of some processes and the effects of batching. For instance, the assembly process of a block could start even if not all the panels and webs that will form it are available yet. Thus, it is necessary that the items that flow through the initial steps of the process are the single entities that represent the webs and the panels.

However, the level of detail in data available at early stages of the project planning does not allow to know the specific parts that will be necessary to manufacture in order to build the ship. Thus, the solution adopted consisted of designing a first component that generates in a random but realistic way all the individual elements that will need to be produced to assemble each block. The model starts with information at an aggregate level which is read from an Excel file and then it disaggregates this information splitting the items that represent the blocks into increasingly smaller units that correspond to the all the parts that need to be manufactured for building the ship.

The requirements for designing the disaggregation process were:

- To generate the number of items of each type randomly with a realistic distribution that depends on the block type and dimensions. To do so, different statistical distributions were fitted for different block types and a procedure was designed to correct the estimations based on the block's length. To do so, a naval architect was in charge of defining the rules on how the size attributes affect the number of elements.
- To automatically generate unique ID's for each part and for each part within each group and count them. This is necessary in the ExtendSim model in order to know in later steps how many items need to be assembled and exactly which parts need to be matched to each block.
- To group the parts in the production batches that are used for production planning in a realistic way. This step matters since the lead times of the products in the different workshops are mainly affected by the batch sizes. The designed model uses a heuristic algorithm to generate batches of similar parts that are usually grouped to specify which model items will be matched together. However, since the times of the assembly steps are not defined at the group level but at the block level, a procedure was devised to distribute randomly the total assembly time of a block among the groups of parts that will form it proportionally to the workload that each of them requires.

Another great challenge which has only partially solved at this stage of the model development was the production planning. The simulation model developed represents the real processes of a shipyard and thus optimizing the schedule is a complex combinatorial problem. As in the real shipyard, the synchronization of the flow of items, the assignment of priorities and the definition of the sequences of blocks and batches is a big challenge that affects the results of the model. Actually, there are two main issues involved in this subject.

The first issue is how to reproduce the real dynamic planning that happens in the shipyard. The construction of a ship may take several months or even years so deviations from the original plan are common. The rescheduling decisions are made by the directors of the workshops or by the project managers depending on the issue and thus are subject to their personal criteria as experts which is not easy to implement in a mathematical model. Thus, it is impossible that the simulation model fully represents the reality of how the system is managed.

The second issue is that no optimal efficient algorithms are known to schedule the production in the simulation model. Thus, the results of the model are only as good as the methods implemented for scheduling. This means that if the model is intended for planning purposes but

the scheduling methods used are inefficient, the model will forecast longer durations than actually should be achieved.

So far the planning methods implemented in the model are simple:

- There is an initial schedule that defines the start times for cutting and assembling each block. This schedule is defined outside the simulation model and imported through an Excel file.
- The workshops follow FIFO rules for managing all the queues. This means that whenever two blocks are available to enter the next step of the process, the priorities will be assigned based on the earliest arrival. This is an aspect that must be improved in further refinements.
- The shared resources (such as assembly cells or the painting cabins) are assigned in following a FIFO order.

5. EXPERIMENTATION

This paper presents the preliminary experiments conducted with this simulation model. The scenario designed does not correspond to a real ship but it contains a representative mix of block types. The real time values and time units are omitted due to confidentiality reasons. The results displayed have been modified in such a way that are valid for supporting the conclusions but do not allow to know the real duration of the process.

The analysis has been focused on inspecting the model results in order to obtain a list of measures for future improvement.

The response variables used are:

- The makespan of each stage of the process.
- The time that the blocks take to complete each processing step.
- Graphs that show the number of blocks that have completed each step along time.

The utilization rate of the different resources was inspected to check for possible lacks of capacity. However, its average value is not a useful indicator in this process because the utilization of some resources are very high in some specific stages that cause delays in the production but very low during the rest of time. Thus, it is not a sufficient indicator of where to act for improving the system.

The figure 4 shows an example of the average utilization of the assembly cells during the simulation time. It can be noticed an initial phase in which the blocks are mainly going through the cutting and welding operations and this workshop is idle. Then, occupation grows as new blocks become available until it reaches a maximum. Then, there is a period in which the last blocks leave this process and the average utilization reaches a final value of 70%.

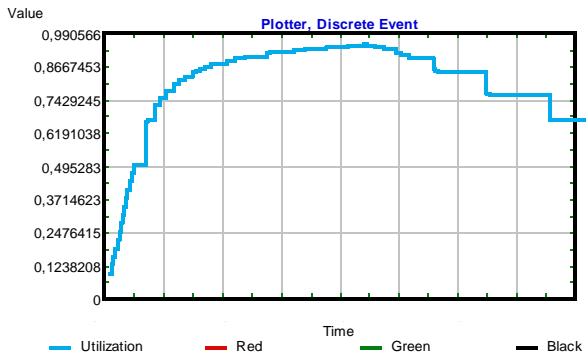


Figure 4: Utilization graph of the assembly cells.

The Figure 5 shows the evolution of the number of completed blocks on the main stages of the process. The blue line corresponds to assembled sub-blocks, the red line to assembled blocks, the green one to painted blocks and the black one to erected blocks. The gaps between the lines indicate the time passing since a block completes a processing step and the next. Thus, it is an important tool for detecting where the problems occur.

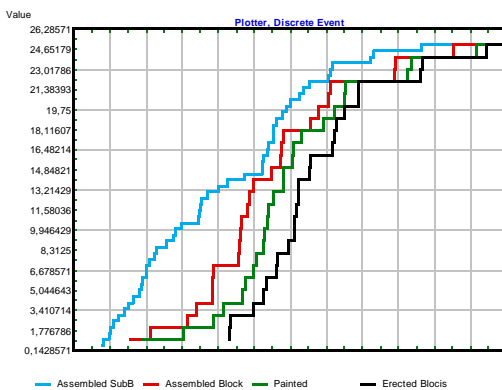


Figure 5: Simulation Model Components

This graph corresponds to the preliminary results obtained before applying optimization techniques and optimizing the schedule. As noted in the previous section, it is not a realistic representation of the planning procedures at the shipyard and thus the conclusions from the analysis cannot be directly extended to the real plant. The conclusions withdrawn must only be circumscribed to the simulation model developed.

The following measures were proposed to improve the model in further simulation studies:

1. Long delays between the sub-blocks and the blocks assembly times were observed in the initial 50% of the blocks and also in the last 20%. This indicates that a production plan that equilibrates better the workload in this process could help to smooth the workload and reduce the total timespan.
2. A 30% of the total makespan was caused by the last blocks. This suggests that optimizing

the schedule could lead to a great improvement in the duration of the project.

3. Even if the painting step was initially regarded as the system bottleneck, the results suggest that the initial processes are not capable of reducing the makespan under the current scheduling methods implemented in the model.

6. CONCLUSION

Shipbuilding is a complex manufacturing process that requires big efforts in project planning and scheduling. Project planning is even more difficult at the initial stages of a project (when engineering plans may not yet be fully available) because of the high number and variety of parts that need to be assembled to build a ship.

Modelling and Simulation M&S offers a practical tool to cope with this complexity. However, great challenges are faced in the development of these kind of simulation models. This paper has described the experience gathered in the initial steps of the development of a simulation models and offers some solutions to deal with them.

The preliminary results demonstrate the necessity for implementing appropriate planning methods and scheduling algorithms when building the model. Without them, a non-optimized sequence causes delays and inefficiencies in the system. Although this represents a challenge for applying M&S, it offers the opportunity of learning about the behavior of the real system and designing measures of improvement.

ACKNOWLEDGMENTS

The authors are thankful to Unidad Mixta de Investigación (UMI) Navantia-UDC for its valuable support.

REFERENCES

- Bao Jinsong, Hu Xiaofeng, J. Y. (2009). A genetic algorithm for minimizing makespan of block erection in shipbuilding. *Journal of Manufacturing Technology Management*, Vol. 20 Iss 4 Pp. 500 - 512.
- Burnett, G. A., Medeiros, D. J., Finke, D. A., & Traband, M. T. (2008). Automating the development of shipyard manufacturing models. *Proceedings - Winter Simulation Conference*, 1761–1767.
- Caprace, Jean-David Moreira Freire, R., Assiss, L. F., Martin Pires, C., & Rigo, P. (2011). Discrete Event Production Simulation in Shipyard Workshops. *Proceeding of the 21th Copinaval*, 20.
- Jae Kyu Lee, Kyoung Jun Lee, June Seok Hong, Wooju Kim, Eun Young Kim, Soo Yeoul Choi, Ho Dong Kim, Ok Ryul Yang, and H. R. C. (1995). DAS : Intelligent Scheduling Systems for Shipbuilding. *The AL Magazine*, Vol. 16, no4, pp.78-94., 16(4), 78–94.
- Jean-David Caprace, Clarice Trevisani Da Silva,

- Philippe Rigo, F. C. M. P. (2011). Discrete Event Production Simulation and Optimisation of Ship Block Erection Process. 10th International Conference on Computer Applications and Information Technology in the Maritime Industries, 271–282.
- Kyoung Jun Lee and Jae Kyu. (1996). A Spatial Scheduling System and its Application to Shipbuilding: DAS-CURVE. Expert System With Applications, Vol. 10, No. 3, pp.311-324. 1992., 10(3).
- Lee, J.-W., & Kang, S. (1996). Efficient Simulation Model Generation Using Automatic Programming Techniques. Proceedings of the 1996 Winter Simulation Conference Ed., 708–713.
- Leong, S., Y.T. Lee, and F. R. (2006). A Core Manufacturing Simulation Data Information Model for Manufacturing Applications. In Proceedings of the Systems Interoperability Standards Organization 2006. Fall Simulation Interoperability Workshop., 1–7.
- Madden, M. G., & Neill, D. E. O. (2005). A Strategy for Autogeneration of Space Shuttle Ground Processing Simulation Models for Project Makespan Estimation, 1251–1259.
- McLean, C. (2001). Simulation of Shipbuilding Operations, 870–876.
- Medeiros, D. J., & Williams, D. (2000). Simulation Based Design for a Shipyard Manufacturing Process, 1411–1414.
- Steinhauer, D. (2011). The Simulation Toolkit Shipbuilding (STS) –10 Years of Cooperative Development and Interbranch Applications. *Compit11*, 453–465.

A SIMULATION-BASED TOOL TO SUPPORT DECISION-MAKING IN LOGISTICS DESIGN OF A CANS PACKAGING LINE

Victoria G. Achkar^(a), Valentina Bär^(b), Franco Cornú^(c), Carlos A. Méndez^(d)

^{(a),(d)} INTEC (UNL – CONICET), Güemes 3450, 3000 Santa Fe, Argentina

^{(b),(c)} Facultad de Ingeniería Química, UNL, Santiago del Estero 2829, 3000 Santa Fe, Argentina

^(a)vachkar@intec.unl.edu.ar, ^(b)valentinabar.26@gmail.com, ^(c)cornufranco@hotmail.com.ar, ^(d)cmendez@intec.unl.edu.ar

ABSTRACT

This paper proposes an advanced discrete-event simulation-based tool developed in order to support decision-making in the design of internal logistics associated to a packaging line of a multinational brewery company. The selected software, SIMIO, allows emulating, advising and predicting the behavior of complex real-world systems. It also provides a modern 3D interface which facilitates the verification and validation of the model. In this work, it is used to understand the dynamic interactions between multiple performance measures (including both material-handling and inventory system performances) to help defining necessary quantities and capacities associated to a future cans packaging line. Based on the proposed model, a what-if analysis is performed to determine thresholds values and critical variables in order to optimize the current system.

Keywords: discrete-events simulation, packaging line, logistics, design

1. INTRODUCTION

Due to increased logistic costs associated to returnable beer bottles, in the last years, an important brewery company decided to introduce cans package into the local market in order to reduce the consume of the mentioned bottle format. The main goal is to make savings related to inverse logistic costs and to optimize transport carrying a bigger amount of beer in the same vehicle, inducing a change in customer habits. This strategic decision conduces to necessarily increase production capacities in its factories or indeed to install new packaging lines. Strategic decisions in manufacturing systems typically concern design problems and resources allocation in the medium/long period. Usually, problems at this level may involve contrasting objectives therefore requiring a strong experience (for people involved in the decision process) as well as advanced decision support tools. (Bruzzone and Longo 2013). They also have a strong impact concerning financial issues. This work focuses on the design of a new cans packaging line. It aims to support decision-making to ensure the best configuration of

internal logistics, including storage and materials-handling, to avoid incurring in additional costs.

To address this kind of issues (representation and optimization design and operation processes), there are two solution strategies typically used: (i) mathematical analysis and (ii) simulation. There are many reasons for using the first strategy opposed to the second one (Seila, Ceric and Tadikamalla 2003). For example, the system under study may yield a model that is so complex that it cannot easily be described using this method. Using mathematical analysis may leave the options of either further simplifying the model and perhaps making it unrealistic or using simulation.

Manufacturing and material-handling systems provide one of the most important applications of simulation. It has been used successfully as an aid in the design of new production facilities, warehouses, and distribution centers. Engineers and analysts using simulation have found it valuable for evaluating impact of capital investments in equipment and physical facility of proposed changes to material handling and layout. Managers have found it useful in providing “test drive” before making capital investments, without disrupting the existing systems with untried changes (Banks, Carson, Nelson and Nicol 2005).

In consequence, discrete events simulation methods are adopted to represent the whole real-world process as an integrated form. The simulation mode, developed with SIMIO software, is used to accurately represent future process operation in order to analyze its behavior with its critical variables, mainly those associated to storage and material-handling. The goal is to evaluate the feasibility of getting the desired production capacity and at the same time trying to optimize internal logistics resources. Therefore, different scenarios are created varying production schedules, to also aid verification and validation processes. Other works has recently used this modern tool to solve similar situations in decision-making processes (Achkar, Picech and Méndez 2015; Aguirre, Müller, Seffino and Méndez 2008; Basán, Cóccola and Méndez 2015).

The work continues with the following structure: (ii)methodology, (iii)internal logistics system (iv)simulation model, (v)verification and validation, (vi)sceneries and experimentation (vii)conclusions.

2. METHODOLOGY

To address the problem by modeling and simulating the system, several steps must be taken to create an accurately model that reflects the real system. The following steps were taken: (i) information collection, (ii) data analysis, (iii) conceptual model definition, (iv) simulation model developing, (v) verification and validation, (vi) scenarios definition, (vii) results analysis. The first step consists in collecting accurate information from the system to start comprehension process and to take into account every component that could influence. Different techniques were applied on this stage: (i) collection of historical data and daily records, (ii) “in situ” observation and time keeping, (iii) iterative interviews with personal involved.

One part of the information obtained is used to fully understand the problem and generate the conceptual model. The conceptual model is used to obtain a sufficient abstraction level of the problem and to define assumptions. The other part of the information is filtered and analyzed to become input data. Truck types and their arrivals, materials rejection rates, packaging speed, resources efficiency, forklift characteristics, warehouses operation rules and capacities, among other factors, are studied in this step.

The fourth step is to create the simulation model on the basis on the information gathered and the conceptual model. SIMIO, the software chosen, is a simulation modeling framework based on intelligent objects. A model is built by combining objects that represent the physical components of the system (Thiesing, Watson, Kirby and Sturrock 2015). These objects could be forklifts, pallet of materials, warehouses, packaging lines, trucks, racks, etc. We choose SIMIO for its power to represent the system in three dimensions and to model realistic spatial relationships of layout. This results in a user friendly interface that facilitates model verification and validation.

Once the simulation model is completed, verified and validated, it is used to analyze critical performance variables and represent scenarios with desired operational policies, evaluating system feasibility and defining necessary level resources such as quantity of forklift or stock area capacities.

3. INTERNAL LOGISTICS SYSTEM

The packaging sector has relationship with different areas of the company, such as logistics, warehouses, maintenance, quality control, etc. These sectors perform simultaneously different tasks and depend on each other for proper operation. This work features on two of them: warehouses and internal logistics. The last one manages raw materials and final products through the plant, using forklift belonging to different sectors. High season is a critical period due to intense material flows between all sectors. Final product dispatches and truck arrivals are also in charge of this sector.

Warehouses are used to stock final products, raw materials and inputs. Packaging is an internal client of

them, requesting for bringing raw material and inputs depending on its needs, and carrying final products. The following picture presents the main sectors involved in this problem with the flow of materials.

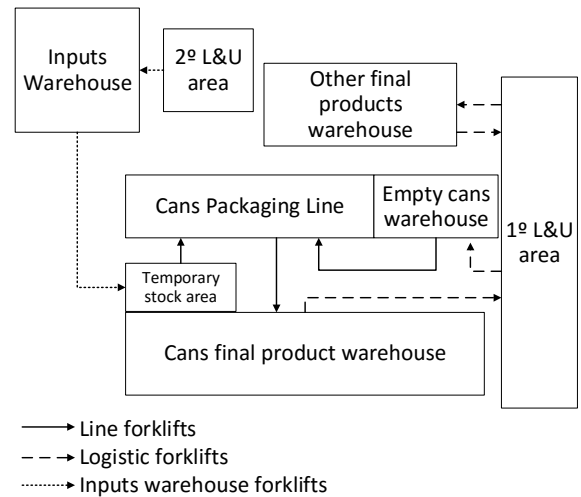


Figure 1: Sectors and material flow involved in internal logistics of the cans packaging line

As shown, the system includes (i) a packaging line, (ii) four warehouses, (iii) one temporal stock area and (iv) two loading and unloading (L&U) areas. Components of the system are described below.

3.1. Forklifts

In the plant there already exist five forklifts that are related to the system under study: four logistic forklifts and one exclusive forklift for the inputs warehouse. There are still no line forklifts (neither packaging line). All of them have the same speed, loading and unloading times, but not same carrying capacities. They also have different work schedules, breaks and assigned tasks depending on the sector each one belongs. They perform different tasks carrying pallets through the system as it is shown in the previous picture. In Table 1, some characteristics of every type are listed.

Table 1: Forklifts characteristics

Type	Quantity	Capacity	Work Schedules
Logistic	4	2	Sunday 20hs to Saturday 5hs
Input warehouse	1	1	Monday 0hs to Friday 17hs
Line	0	1	24hs

3.2. Materials

There is a wide variety of material moving through the system. To produce a pallet of final product, ten different inputs are needed, such as empty cans, can lids, different types of cartons, wraps, stickers, ribbons,

etc. Most of them are packaged in pallets. There are also intermediate material pallets and scrap traveling through the system.

3.3. Trucks

On the one hand, trucks enter to the principal L&U area to load final products and to unload empty cans. They also make loads and unloads related to other packaging lines. On the other hand, trucks entering to the secondary area unload inputs to stock them in its exclusive warehouse. Most of them can transport 24 pallets, except for those transporting empty cans that transport 22 pallets. Depending on the type of truck, there could be different restrictions to enter into the factory.

3.4. Packaging Line

The future packaging line is wanted to work 24 hours a day, 3 to 7 days a week depending on the season.

The expected line speed rounds 80,000cans/h and its efficiency is about 75%. Input consume rates are given by a bill of materials and there is a temporary stock area nearby for immediate supply. This temporary stock area will have defined replenishment frequency and a fixed capacity to define. The output of the line also has a maximum capacity. Forklifts are expected to store final products, feed the line and perform secondary tasks associated to residual input pallets that must be returned to suppliers.

3.5. Warehouses

There are three principal warehouses of interest in the system: input, empty cans and final product warehouses. Each of them have desired security stock levels. Inputs warehouse has the same work schedule as its exclusive forklift previously mentioned. On days off, it must ensure that the temporary stock area has enough quantity of material to avoid interrupting the production. The remaining warehouses are directly related to de packaging line, so they must be able all the time.

3.6. Loading and unloading areas

The principal L&U area has four parking places. There, trucks bringing and retiring materials of all packaging lines share this parking places. Focus on tasks related to packaging line, empty cans pallets are unloaded to be stored in its correspondent warehouse. Only one truck of this type can enter at a time. An analyst examines the lot of empty cans and decide of the truck is rejected or not. About 5% of these trucks are rejected (empty cans is a very fragile product). Trucks that arrive to load cans final product has no entrance restrictions.

The secondary L&U has a single parking place and most of the tasks performed are unloading input pallets for all the packaging lines. Inputs warehouse forklift is in charge of this area.

3.7. Problem definition

Once all components and their interactions are fully understood, different conflicts can be detected. There must be a coordinate relation between every sector. Variations on the line speed or on its work schedule will directly impact on warehouses occupation rates. The line could need to be replenished and to retire final product with variable frequency and will consequently impact on line forklifts utilization. Furthermore, final product warehouse will tend to get full during operational days and empty when the line is not working, and in a similar way empty cans warehouse and input temporary stock area occupation rates will vary. To deal with the situation, the number of daily trucks must increase, both to bring inputs and to dispatch final product. Consequently, logistic and input warehouse forklifts are immediately affected. Moreover, security stock levels should be reconsidered, taking into account capacity limitations.

This work analyzes the possibility of getting the desired speed and work schedules without collapsing the system. Trucks arrivals must be designed according to final product desired sales volume and to guarantee certain stability in warehouses occupation. Security levels will be set according to line speed and to support a desired number of days of production. Temporary stock area must be dimensioned according to desired replenishment frequency. Line forklifts optimal quantity must be proposed and also if any forklift must be added to other sectors. Parking places will be evaluated to determine if they can deal with the amount of trucks entering into the system. The simulation model must give enough information to make this set of decisions.

4. SIMULATION MODEL

Once the system is completely understood, the next step is to develop the simulation model. SIMIO makes modeling dramatically easier by providing a new object-based approach. Objects represent the physical components in your system such as workstations, conveyors, and forklift trucks in a manufacturing facility. Object-based modeling is a very natural and simple approach to simulation modeling (Pedgen 2009). The developed model uses a factory scale plan with representative distances. The following subsections describe the major components of the model.

4.1. Model assumptions

The major assumptions made in the model are listed below:

- The efficiency of the line is traduced in a daily productive period of time with constant production speed.
- A generic cans product was defined pondering both existing sizes according to its sells. A pallet of this generic product contains 2280 cans.

- Final product pallets are immediately stored in its warehouse and the output of the line has a limited capacity.
- Temporary stock area can store all pallets that are brought to it, and it has defined replenishment frequency.
- Every sector and forklift has a defined work schedule.
- Every truck transports exclusively one type of material.
- Only trucks transporting empty cans have entrance restriction: one at a time.
- Trucks have defined capacities according the material it transports.
- Forklifts do not fail and drivers do not absence (in real system there is an additional forklift ready to supply any of them).
- Trucks are loaded/unloaded using at maximum three forklifts.
- Line forklift does never stop working (drivers take turns to replace).

4.2. Input variables

Some variables are entered in the model as input data, and others are decision variables.

On the one hand, the main variables are:

- Truck characteristics (capacities, rejection rates, entering restrictions, etc.).
- Forklift characteristics (capacities, speed, work schedules, L&U times).
- Packaging line speed.
- Replenishment frequency and quantities of the temporary stock area.
- Bill of materials.
- Quantity of pallets that can be accumulated at the output of the line.
- Secondary task characteristics.
- Line efficiency (traded in productive and unproductive daily periods of time).

On the other hand, decision variables are:

- Line work schedule.
- Quantity of forklifts.
- Arrivals of trucks.
- Security and initial stock levels.
- Restocking frequency to temporary stock area.

4.3. Output variables

The main variables used to measure the performance of the system are:

- Total production.
- Quantity of line interruptions because of the lack of inputs.
- Average, maximum and minimum stock levels of each material in warehouses.

- Average, maximum and minimum stock levels of each material in the temporary line stock area.
- Total number of trucks arrived and total number of trucks attended.
- Forklifts and parking places utilizations.
- Average and maximum stock levels in the output of the line.

Analyzing the values of these variables allows detecting if any restriction is not been accomplished or if the model collapsed.

4.4. SIMIO Model

The computer model was made using different objects provided by the software and setting its properties to adapt them. For further customization, internal logic processes were created and associated to different objects. These processes allow, using events, states, monitors, timers and other definitions of SIMIO, modeling every detail necessary to create an accurately system representation. For example, entrance restrictions, rejection rates, material routes and destinations and requests of materials were modeled with internal logic processes. Appendix A has a small glossary including the major objects of SIMIO used in the model.

How the major components of the system were modeled is described in the following subsections:

- Pallets of materials and trucks: were represented using entities, with different priorities to differentiate them.
- Forklifts: using the vehicle object, they were modeled introducing its characteristics such as speed, capacity, L&U times, etc. Figure 2 shows trucks, pallets and forklifts of the model. Objects animation facilitates to identify each one.

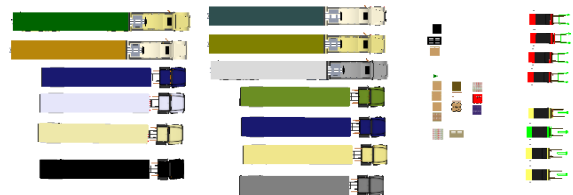


Figure 2: Trucks, pallets and forklifts objects in 2D SIMIO model

- Cans packaging line: it was represented using several objects. The main one is a Workstation object. It contemplates consumption and production rates defining a BOM (Bill of materials) matrix. It has monitors with threshold values defined to request input pallets when they are crossed. Figure 3 shows the packaging line at right and the temporary stock area at left.

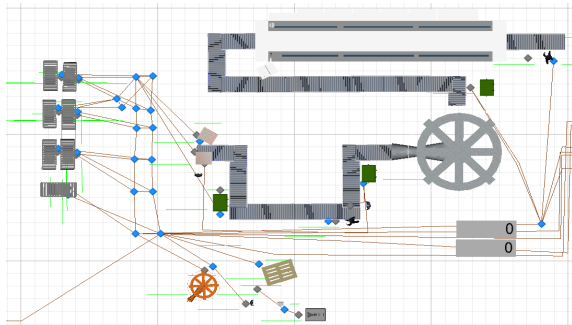


Figure 2: Packaging line and temporary stock area in 2D SIMIO model

- Warehouses: were modeled with server objects for each kind of material, customized using internal logic processes to retain pallets until they are requested from another sector and liberate the exact requested quantity. They also have monitors that control stock levels. Figure 3 presents empty line (left) and cans final products (right) warehouses.

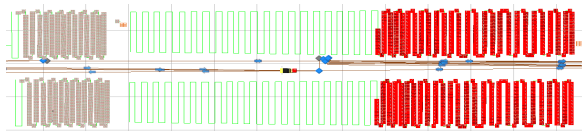


Figure 3: empty cans and cans final product warehouses in 2D SIMIO model

- Temporary stock area: modeled with server objects, they work similar to warehouses, with the difference that it has not a defined capacity (it must be determined as a model output).
- L&U areas: they were also represented with server objects, one for each parking place. These sectors uses a big quantity of internal processes, to model every restriction, rejection rate, setting entities destination or requesting warehouses to bring them depending on the type of truck. In Figure 4 there is the input warehouses conformed by several server objects and at left there is the secondary L&U area.

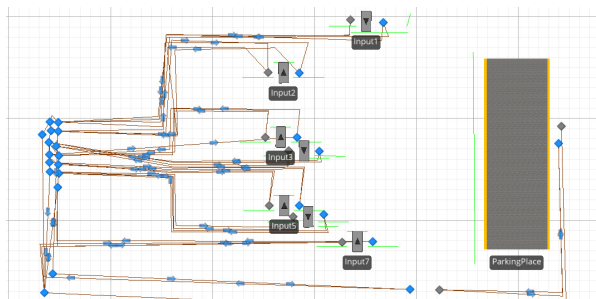


Figure 4: inputs warehouse and secondary L&U area in 2D SIMIO model

The figure below shows a global view of the animated model. Each sector with its elements can be easily identified, and their operation and interactions can be simultaneously watched.

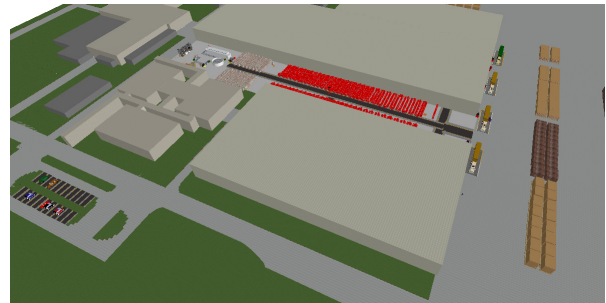


Figure 5: Global view of the 3D SIMIO model

5. VERIFICATION AND VALIDATION

Verification is concerned with determining if the conceptual model with its specifications and assumptions were correctly translated in computerized representation (Law Averil 2007). During the conceptual model development, several requirements of the different elements were determined, concerning expected values and system behaviors, such as production rates and arrival rates, restrictions of entry and operation, etc. The expected values were compared with output variables thrown by the simulator. Figure 6 shows a comparison between empty cans trucks required to satisfy the production in the conceptual model and those arrived on the simulated system. Then, Figure 7 presents estimated monthly production rates versus the output from the model. Similar values prove that the behavior of the computer model is in accordance with estimated values.

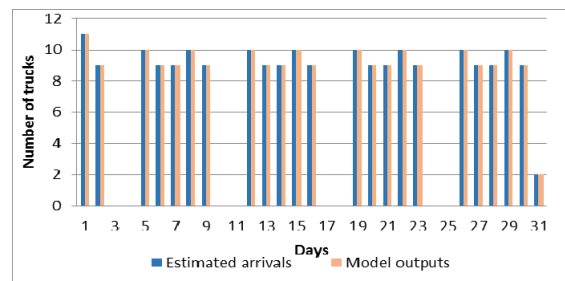


Figure 6: Empty can trucks arrivals verification

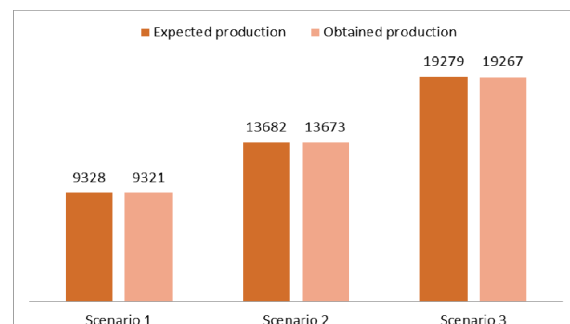


Figure 7: Monthly production verification

Therefore, validation was carried out, which is concerned with determining how closely the simulation model represents the real system (Law Averil 2007). In this step, several iterative comparisons between output variables obtained by the simulator and information gathered from the company are performed. The most difficult aspect in this step consist in validating a line which does not exists yet. Thus, there are many aspects that cannot be compared with a real model. To carry out validation, some aspects were taken from resources that exists in other lines in the same plant. Form example: forklifts times to perform tasks and permanence times of trucks in parking places. Distances that must be covered by forklifts and trucks in the simulation model were compared with the real scale plant layout provided by the company. Other aspects were validated using as reference another cans packaging line belonging to the same company, such as times related to secondary tasks and inputs consumption rates. Finally, all aspects were discussed with experienced staff from the company. In each iteration, changes and necessary adjustments are made on the model programming in order to achieve the desired values.

6. SCENERIES AND EXPERIMENTATION

As a result of the seasonal demand characteristic on this kind of enterprises, it was imperative to analyze three types of operational configuration for the can packaging line. The scenarios evaluated were:

- *Type 1*: 4 operational days during the first week, 4 operational days in the second and 3 operational days in the third.
- *Type 2*: 5 operational days of the line during the week.
- *Type 3*: 7 operational days of the line during the week.

Due to each scenario has a different work schedule but packaging speed keeps constant, it was necessary to define for each of them an exclusive policy of operation. The main variables affected were: (i)packaging line work schedule, (ii)daily trucks arrivals, (iii) forklifts quantities, (iv)restocking frequency to temporary stock area and (v)maximum quantity admissible of other truck arrivals. The simulation period is a month. Table 2 shows a description of every scenario, with major decision variables involved. Each scenario has defined the same quantity of forklifts: 4 for logistics, 3 for the line and 1 for the inputs warehouse. Then, Tables 3 and 4 show the results of principal output variables. Maintaining the same quantity of resources available, the company wanted to know the number left of arrivals that they can use per day to attend other lines. The maximum daily arrivals of other trucks admissible keeping the same quantities of forklifts for each scenario are: 117, 113 and 80 for types 1, 2 and 3 respectively.

Table 2: Scenarios decision variables

Scenarios: Decision variables				
Sceneries	Packaging line work schedule	Empty cans trucks	Final product trucks	Inputs trucks
1	Type 1	141	389	28
2	Type 2	207	572	38
3	Type 3	291	804	53

Table 3: Maximum stock levels to measure warehouses

Sceneries	Max. stock level in empty cans WH	Max.stock level in final product WH	Max. stock level in temporary stock area	Max. stock level in Inputs WH
1	2538	1105	113	1216
2	2560	1168	165	1305
3	2702	1623	282	2149

Table 4: Forklifts Utilizations

Sceneries	Monthly production (pallets)	Logistic Forklifts Utilizacion	Line Forklifts Utilizacion	Inputs WH Forklifts Utilizacion
1	9321	81,77%	84,25%	64,38%
2	13673	93,70%	80,45%	69,34%
3	19267	84,80%	76,20%	69,43%

As shown in the table, variation in the work schedule of the packaging line, can strongly impact on the quantity of final product that must be shipped and on empty cans and raw materials requirements. This causes a great increase in the quantity of trucks that must arrive to unload empty cans and load final product. The variation also affects warehouses capacities, but does not affect forklifts required quantities. As there are no can line forklifts yet, three of them are recommended to satisfy production requirements without collapsing the system. Results obtained help to determinate which is the ideal production alternative to be chosen depending on the season, having into account resources and capacities required for proper operation of the packaging line. Other operational policies can be easily simulated and enable the customer to select those policies according to his requirements.

7. CONCLUSIONS

In the present work, a discrete event simulation modeling tool is used to support decision-making in the design of internal logistics of a can packaging line. The proposed simulation model allows evaluating and defining the best strategy for minimizing costs related to the impact of introducing a new packaging line in logistic operation.

The main attraction of this tool for the client is the possibility to emulate future operation policies and analyze the behavior of a system which is currently on its design step. It also allows experimenting new plant layouts, transport systems, hardware systems, etc. with no need of investing money on their acquisition or interrupting the normal operation of other lines that will interact with the new one. It results a strong tool to perform a what-if analysis which is particularly useful in the design of new systems.

ACKNOWLEDGMENTS

This paper was partially founded by CONICET under Grant PIP-2221 and from UNL under Grant PI-81-481.

APPENDIX

A. GLOSSARY

Path (Figure 6): used to define a pathway between two node locations where the travel time is determined by the path length and a traveler's speed. Entities or vehicles can go through it. Some of its properties are speed, capacity and length.

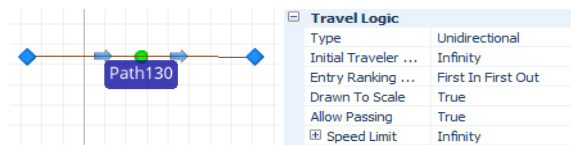


Figure 6: Path module and characteristics

Server (Figure 7): represents a processing activity in the model. Between its properties they must be set: processing time, resources needed failures, internal process and events associated.

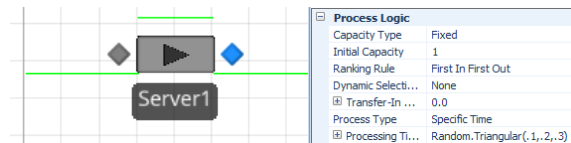


Figure 7: Server module and characteristics

Sink: represents a final point in the model where entities go to be eliminated.

Vehicle (Figure 8): transports entities from one point to another. It has assigned a pick up and drop off point. Other properties are speed, loading and unloading time and capacity.

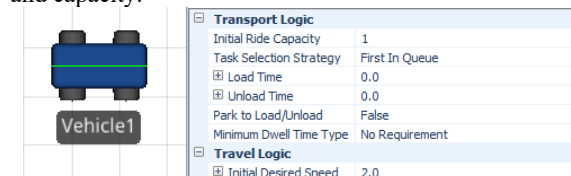


Figure 8: Vehicle module and characteristics

Workstation: represents a more complex server. It has properties such as setup time and it consider consumption and production of materials based on a BOM matrix.

Internal logic process (Figure 9): A sequence of commands that dictate the behavior of an object. It allows including inside standard modules some tasks to custom them such as seizing or releasing resources, assigning variables and firing events.(Achkar, Picech and Méndez 2015).

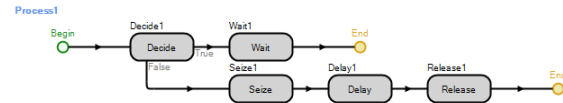


Figure 9: Path module and characteristics

REFERENCES

- Achkar V.G., Picech L.S., Méndez C.A., 2015. Modeling, simulation and optimization of logistics management of a cans packaging line. Proceedings of the 27th European Modeling & Simulation Symposium, 395-402. September 2015, Italy.
- Aguirre A.M., Müller E.J., Seffino S.E., Méndez C.A., 2008. Using an advanced discrete-event simulation framework to productive capacity management of a car-parts factory. Proceedings of the 20th European Modeling & Simulation Symposium, 159-168. September 2008, Italy.
- Banks J., Carson J.S., Nelson B.L., Nicol D.M., 2005. Discrete-Event System Simulation. U.S.A.: Pearson Prentice Hall.
- Basán N.P., Cóccola M.E., Méndez C.A., 2015. Conducting experimental design and optimization of the system configuration and operation of an innovative car rental business. Proceedings of the 27th European Modeling & Simulation Symposium, 166-171. September 2015, Italy.
- Bruzzone A.G., Longo F., 2013. An Advanced Modeling & Simulation Tool for Investigating the Behavior of a Manufacturing System in the Hazelnuts Industry Sector. International Journal of Food Engineering. 9(3): 241-257.
- Pedgen, C.D., 2009. An Introduction to SIMIO for Beginners. Available from: <http://www.simio.com/resources/white-papers>
- Seila A.F., Ceric V., Tadikamalla P., 2003. Applied Simulation Modeling. USA: Brooks/Cole Thomson Learning.
- Thiesing R., Watson C., Kirby J., Sturrock D., 2015. SIMIO Reference Guide, Version 8. Available from: SIMIO software support.

AUTHORS BIOGRAPHY

Victoria G. Achkar is an Industrial Engineer and PhD student at National Scientific and Technical Research Council (CONICET). Her research interests include hybrid simulation & optimization tools for logistic management of complex production and distribution processes of industrial interest.

Valentina Bär is an advanced Industrial Engineering student conducting research in optimization and

simulation tools for production planning and scheduling of automated production systems.

Franco Cornú is an advanced Industrial Engineering student conducting research in optimization and simulation tools for production planning and scheduling of automated production systems.

Dr. Carlos A. Méndez is a Titular Professor of Industrial Engineering at Universidad Nacional del Litoral (UNL) in Argentina as well as a Senior Researcher of the National Scientific and Technical Research Council (CONICET) in the area of Process Systems Engineering. He has published over 200 refereed journal articles, book chapters, and conference papers. His research and teaching interests include modeling, simulation and optimization tools for production planning and scheduling, vehicle routing and logistics. His group's web page is: <http://servicios.intec.santafe-conicet.gob.ar/grupos/capse/>.

SIMULATION OF LONGWALL COAL MINING TECHNOLOGIES

Victor V. Okolnishnikov^(a), Sergey V. Rudometov^(b), Sergey S. Zhuravlev^(c), Vasily V. Sinoviev^(d)

^{(a),(b),(c)}Design Technological Institute of Digital Techniques of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

^(d)Federal State Budget Scientific Centre «The Federal Research Center of Coal and Coal-Chemistry of Siberian Branch of the Russian Academy of Sciences», Federal State Budget Educational Institution of Higher Professional Education Kuzbass State Technical University named after T.F. Gorbachev, Kemerovo, Russia

^(a)okoln@mail.ru, ^(b)rsw@academ.org, ^(c)s-zhur@yandex.ru, ^(d)zv150671@gmail.com

ABSTRACT

This paper describes our own visual interactive discrete simulation system specialized in simulation of technological processes in coal mines. The simulation system allows creating complex simulation models fast, interactively, as well as using them for all kinds of simulation tasks: “what-if” ones, presentation, training, hardware development with all “in-the-loop” manners. This is ensured by the presence in the simulation system visually interactive interface and a set of libraries. There are libraries for simulation the following subsystems of coal mine: belt conveyor subsystem, power supply subsystem, ventilation subsystem, and pumping subsystem. A new library contains models of coal seam, and mining machines. Models of longwall technologies for coal mining in various types of coal seams were developed with the help of the new library.

Keywords: visual interactive simulation, coal mining, longwall mining

1. INTRODUCTION

Different investigations in various areas of underground mining that require automatic control (e.g. coal mining) specify that resources are becoming less available, and harder to be extracted.

The technologies of coal mining are well known. Today mine uses mining machinery for mining, transportation, roof support, etc. How exactly will these machines function? What additional machines will be required? What is the cost of machines installed in the mine? Will big universal machines be more appropriate rather than many small, specialized mechanisms? What will the cost of a final product be? The particular mining installation is effective in terms of minimum costs and maximum productivity only if it is correctly planned. Usually big layout has many components from different vendors interconnected. That makes it hard or, even, impossible to predict the exact effectiveness. The situation gets worse if there is also a requirement to create new components for management of a part of such layout. In all these cases, computer-based simulation can be used to solve these problems.

Because of the importance of these problems, there are a large number of papers on the use of simulation in the development and optimization of coal mining systems (Sturgul 2001; Kizil et al. 2011; Cai et al. 2012).

There are also a great deal of simulation tools both universal simulation systems and specialized systems and packages for simulation of coal mining systems.

A number of models for various technologies of coal mining were developed with the help of our own simulation system. Brief description of this simulation system is presented in section 2. Section 3 contains description of a new library and examples of the models of longwall coal mining technologies.

These models are used for developing of process control systems for underground coal mines in Kuznetsk Coal Basin (Russia, Western Siberia).

2. THE SIMULATION SYSTEM

A visual interactive Manufacturing and Transportation Simulation System (MTSS) is developed and used at Design Technological Institute of Digital Techniques (DTIDT). It is a process-oriented discrete simulation system intended to development and execution of models of technological processes (Okolnishnikov et al. 2010, 2011, 2014). MTSS is a set of program interfaces for creating *elementary* models as well as forming complex models from them. This corresponds to bottom-to-top simulation style. The elementary model is verified according to its specifications, ready-to-use simulation model of an equipment unit with capability of low-level control for it.

The elementary model consists of the following parts:

- Two-dimensional and three-dimensional animated graphic images.
- Input and output parameters.
- Functionality algorithm describing dependence between parameters.
- States that the elementary model can reach during the simulation process.
- Control commands defining switching process between elementary model states.

A model in MTSS is created by graphical connection of images of elementary models.

MTSS is also a tool for running of complex models built from elementary models. The running model performs the movement of the model time and visualization. Statistics is collected as well. Statistics are available as a short overview when model runs, and more statistics are available after model completion.

Elementary models can be then combined in a libraries forming simulation goal-specific tool, used by field specialists.

Such tool can be an application for rapid creation of correct simulation model by mining engineers. Usually engineers have not enough qualification to create simulation models in details, but they know how to connect correctly elementary models to create the required topology. MTSS uses 2D as a graphical editor and 2D, 3D for visualization of model running. Such approach seems more natural for mining engineers, when all installations and machines appear first on 2D plans. 3D is more useful for visualizing complex vertical movement.

Process control systems often have two levels: the low level of equipment and simple control logic in it and the upper level of complex control of production. One of the distinguishing features of MTSS is a division of the logic of a simulation model into two parts: a low-level logic and an upper level logic.

Such division allows us not only to correlate to the usual structure of the process control systems but also to use such models for embedding them into actual process control systems in the following ways: to emulate equipment, to simulate upper level logic, and to send commands to actual process control system for debugging and testing. The division into upper and lower logics let us also organize a switch between various implementations of decomposition. It makes possible coexisting simulation of upper level logic and a proxy that allows communicating with the upper level logic of actual process control system.

The model of coal mining can communicate with a new process control system developed in DTIDT, to be a source of input signals, emulate equipment, test actual control program with simultaneous visualization of overall process of mining. This allows debugging and tuning of a new process control system in accordance with behavior of simulated system, even simulating various accidents. This permits to minimize time and costs on site for commissioning.

There are libraries for simulation the following subsystems of coal mine (Okolnishnikov et al. 2013a, 2013b, 2015) like:

- Belt conveyor subsystem.
- Power supply subsystem.
- Ventilation subsystem.
- Pumping subsystem.

A new library of simulation models of mining machines for coal mining was developed. This library is based on

MTSS and its prime goal is to simulate interactively and visualize various aspects of coal mining in coal seams of various types and by various coal mining technologies, hiding simulation specifics from the end-user (field engineers).

The library consists of new elementary models of:

- Highwall mining system.
- Longwall mining system.
- Coal seam.
- Mining machines.
- Self-moving coal wagon.
- Storage area.

3. SIMULATION OF LONGWALL MINING

Longwall mining system is a highly automated, very powerful and productive way to mine coal. It is most-widely applied around the world. Its main advantage is that it leaves almost no coal inside mines. However, the depth of the mine (measured from the surface) limits it. In addition, it is applied in relatively flat areas of coal, from 0.8 up to 10 meters high, from 150 to 450 meters face, and up to 4 kilometers in depth. The depth of mine itself is relatively small – up to 200 meters.

Longwall system consists at least of an armored face conveyor (AFC), a shearer, and roof support sections. The AFC is connected to outbound belt conveyor. The shearer cuts the coal from coal seam face, in a series of passes along the AFC. The AFC delivers coal to the belt conveyor connected. The roof support moves itself and the AFC, pushing it (and itself) forward with hydraulics.

One of the problems of longwall mining is roof caving, which can lead to serious environmental problems. Another problem is that longwall requires significant amount of work to be done before its massive equipment is installed in production.

For detailed simulation of longwall mining system, we closed the following decomposition: an armored face conveyor, a shearer, and roof support sections.

Simulation model for longwall mining system can function only if it is connected (in terms of MTSS) with belt conveyor simulation model.

The simulation model built from the content of the library simulates movement of all mobile objects of the simulated system. Both 2D (top view) and 3D visualization are available. Statistical data is also collected.

Figure 1 presents a sample layout in a simulation model of longwall coalface.

Main window consists of at least 4 areas:

1. Main simulation window. The origin is at the left top corner of the view.
2. Parameters of a simulation model.
3. Specialized view for fast navigation in simulation model.
4. Settings for the time start, time end and current model time.

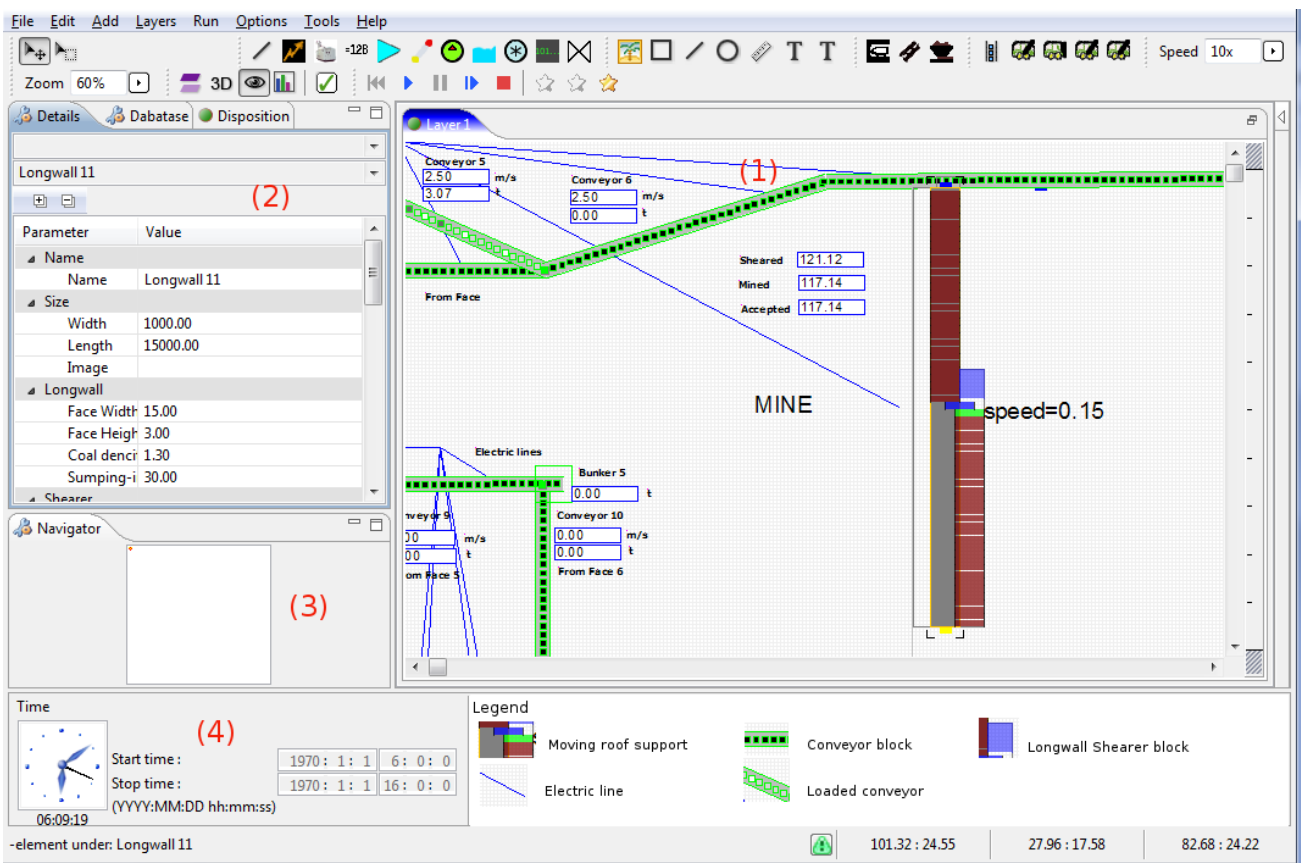


Figure 1: Main window of MTSS system running simulation model of longwall coal mining

The main task for the simulation model of longwall coal mining is to show how longwall automation produces the coal for the rest of the underground factory.

In the other words, it is a model of mine face. In contrast with “face” model, which is primarily statistical, this model will show “pulse” of coal. Figure 2 presents it.

Changing the algorithm the longwall model functions, it produces different loading for the rest of the simulation model.

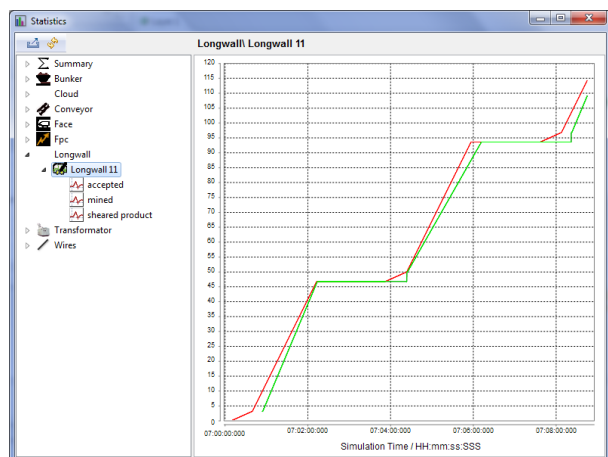


Figure 2: "Pulsing" coal by longwall

The algorithm for longwall mining system simulation is closer to the control algorithms of a real longwall mining system. There are some variations of the shearing task (Mitchell, 2009).

Longwall is managed by tasks, which are state machines in its nature. Tasks are:

- To shear (dig in and shear), with detailed animation. This is the main task.
- To deliver coal outside the longwall simulation.
- To move roof support when it is possible.

The main task will define the amount of time for the shearer to move from one end to the other. Task for roof support is required to simulate roof support advance. The amount of coal will be moved to the AFC that will deliver it to the belt conveyor. Tasks will repeat these steps until the “done” or “postpone” conditions will be achieved.

Main task is done when shearer reaches the end of an AFC line. Task is postponed when belt conveyor is overloaded and cannot accept the next portion of coal or gas level is not safe.

During any of these steps, the simulation model simulates roof fall (behind the roof support) and gas level increase. In addition, ventilation is simulated (remove gas from working area).

In most cases, there is no need to simulate in details any technology like longwall or flat-lying coal mining, if it is used just as a source of a coal for belt conveyor system, for example. Everything really needed in such cases is:

- To define that longwall is a source of coal for a big mining system like a conveyor.
- To know the performance of the longwall coal mining during some time (working day, 8-hour time interval, 1-hour time interval). Note that the emergency stops (like gas or coal dust) are already included in this statistical data.

However, if the goals of simulation are:

- To predict how longwall mining automation will behave in details (i.e. movement of its parts depending on various situations in mines).
- To make a detailed visualization of mining process.
- To define how this mining will influence on the overall performance of mines.
- To define scenarios of broken or temporarily inaccessible parts interactively -

then achieving these goals will require detailed decomposition of common longwall mining system and a detailed visualization of all its parts.

In our work, both detailed and statistical approaches are presented.

4. FUTURE WORK

Detailed simulation of longwall system, connected with detailed simulation of flat-lying coal seam (or multiple flat-lying coal seams), will allow creating of simulation that will not only predict the behavior of big underground mining system, but also simulation of land subsidence while using longwall, especially with very heavy longwall systems that can cut 10-meters-high coal seams.

Simulation system MTSS can be used not only for simulation of existing coal mining techniques but also for perspective robotized techniques (Sinoviev et al. 2015, 2016).

ACKNOWLEDGMENTS

This research was financially supported by the Russian Foundation for Basic Research (project 16-47-420490 p_a, and project 16-07-01179 A).

REFERENCES

Cai D., Baafi E., Porter I., 2012. Modelling a longwall production system using flexsim 3D simulation software. Proceedings of 21st International Symposium on Mine Planning and Equipment Selection (MPES 2012), pp. 107–114. November 28-30, New Delhi, India.

Kizil M.S., McAllister A., Pascoe R., 2011. Simulation of Development in Longwall Coal Mines. Proceedings of the 11th Underground Coal Operators' Conference, pp. 91–98. February 10-11, Wollongong, NSW, Australia.

Mitchell G.W., 2009. Longwall Mining. In: R. J. Kininmonth and E. Y. Baafi, eds. Australian Coal Mining Practice. Carlton, Victoria, Australia, Australian Institute of Mining and Metallurgy: 340–375.

Okolnishnikov V., Rudometov S., and Zhuravlev S., 2010. Simulation environment for industrial and transportation systems. Proceedings of the International Conference on Modelling and Simulation, pp. 161–165. June 22-25, Prague, Czech Republic.

Okolnishnikov V., 2011. Development of Process Control Systems with the Use of Emulation Models. International Journal of Mathematics and Computers in Simulation, 6 (5): 553–560.

Okolnishnikov V., Rudometov S., and Zhuravlev S., 2013a. Simulation Environment for Development of Automated Process Control System in Coal Mining. Proceedings of the 2013 International Conference on Systems, Control, Signal Processing and Informatics, pp. 285–288. July 16-19, Rhodes Island, Greece.

Okolnishnikov V., Rudometov S., and Zhuravlev S., 2013b. Simulation Environment for Development of Automated Process Control System in Coal Mining. International Journal of Systems Applications, Engineering & Development, 6 (7): 255–262.

Okolnishnikov V.V., Rudometov S.V., 2014. A System for Computer Simulation of Technological Processes. St. Petersburg State Polytechnic University Journal. Computer Science. Telecommunications and Control Systems, 1 (181): 62–68.

Okolnishnikov V., Rudometov S., and Zhuravlev S., 2015. Simulation as a Tool for Debugging, and Testing of Control Programs for Process Control Systems in Coal Mining. International Journal of Systems Applications, Engineering & Development, 9: 1–6.

Sinoviev V.V., Starodubov A.N., Dorofeev M.U., Okolnishnikov V.V., 2015. Discrete Event Simulation Robotic Technology of Mining. In: Mathematics and Computers in Science and Industry. Mathematics and Computers in Science and Engineering Series – 50: 75–77.

Sinoviev V.V., Okolnishnikov V.V., Starodubov A.N., Dorofeev M.U., 2016. Approach to Effectiveness Evaluation of Robotics Technology in Mining Using Discrete Event Simulation. International Journal of Mathematics and Computers in Simulation, 10: 123–128.

Sturgul J.R., 2001. Modelling and Simulation in Mining - Its Time Has Finally Arrived. Simulation, 76 (5): 286–288.

Simulation-optimization of schedules with sequence dependent setup times based on Petri net models

Gašper Mušič^(a)

^(a)University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

^(a)gasper.music@fe.uni-lj.si

ABSTRACT

The paper deals with simulation-optimization of schedules with sequence dependent setup times. Such scheduling problems are commonly found in industry where machine setups or cleaning procedures are required among different jobs. The problems can be effectively modelled by simple Coloured Petri nets (CPNs). CPN modelling is combined with new type of simulation based on predefined transition set sequence conflict resolution strategy and standard permutation representation of job schedules. This enables generation of neighbouring solutions that are always feasible and standard local search optimization algorithms can be effectively applied to derived models. Modelling approach and neighbourhood construction procedure are explained in detail. Results of tests on sample scheduling problems by simulated annealing and genetic algorithms are provided.

Keywords: Petri nets, simulation, optimization, scheduling, local search

1. INTRODUCTION

Scheduling problems have been one of the classical research topics within Operations research for decades. Recently, manufacturing operations scheduling is gaining importance in line with wide informatization of industrial processes that should facilitate the optimisation of various operational aspects of such systems. Among those, efficient operations scheduling can contribute to better production responsiveness by more efficient processing of released manufacturing jobs.

Petri nets (PNs) are a modelling technique that is well suited to description of manufacturing systems. In contrast to most system description languages, PNs are state and action oriented at the same time providing an explicit description of both the states and the actions. When modelling manufacturing systems PNs have several advantages, such as efficient modelling of concurrency and synchronization, ability of capturing functional, temporal and resource constraints within a single formalism (Recalde et al. 2004), and simple representation of production systems' specific properties (Tuncel and Bayhan 2007). These can be supplemented by various extensions that enable quantitative performance analysis of modelled systems. A

rich variety of extensions of the original PN formalism include the time concept, considering both deterministic and stochastic time (Viswanadham and Narahari 1992). In this way, PNs can be applied for simulation-based operations scheduling, that is, timed simulation can be performed to evaluate the quality of the schedules under certain operating conditions (Piera and Mušič 2011). If the schedule is modelled properly, the job completion times can be read from the timestamps of tokens in the final marking obtained by PN simulation. As the objective to be minimized is always a function of the completion times of the jobs (Pinedo 2008), this enables to implement scheduling systems with various objectives.

In our previous work a simulation based optimization approach applying PNs was intensively studied, as well as other, more classical approaches, such as dispatching rules and reachability tree based heuristic search (Gradišar and Mušič 2007, Löscher et al. 2007). In particular, the investigations focused on combination of PN modelling approach and local search methods (Löscher et al. 2007, Mušič 2012b). Other authors focus on the use of PN models for deadlock-free design, resource allocation and scheduling optimization (Sindičić et al. 2012, Huang et al. 2013, Xing et al. 2012), and use of PNs with other metaheuristics, e.g. genetic algorithms (Mejía et al. 2012, Han et al. 2014).

This paper focuses on PN models of scheduling problems with sequence dependent setup times (Kurz and Askin 2004) and explores the possibilities of PN models use in conjunction with state-of-the-art local search algorithms. The previous works on standard flow-shop and job-shop problems are generalized, so that also less restricted problem classes of larger practical relevance can be addressed and PN based scheduling methods are brought closer to the practical needs in manufacturing and services.

2. PETRI NET MODELLING OF SCHEDULES WITH SEQUENCE DEPENDENT SETUP TIMES

Petri net models of scheduling problems have been described in our previous works (Mušič 2011, 2012b, 2015). Both Place/Transition and Coloured Petri net classes have been used for that purpose. Yet, schedules with sequence dependent setup times have not been addressed so far.

Such problems are of great practical relevance as

complex flow lines exhibiting this property can be found in many industries. Ríos-Mercado and Bard (1998) state that sequence-dependent setup times are found in the container manufacturing industry as well as the printed circuit board industry. Also semiconductor and automobile industries encounter sequence dependent setup times, which result in difficult scheduling problems (Kurz and Askin 2004).

Here we follow the formulation of Ríos-Mercado and Bard (1998), who assume that setup can only be performed after the machine is no longer processing any job and the job for which setup is being performed is ready. The basic property that has to be captured within the model of such a schedule with sequence dependent setup times is the dependence between setup durations and job order. In the proposed model this is achieved by explicitly representing recently processed job by an additional control place. When using Place/Transition PN a set of places is required for that purpose to distinguish among different jobs. A sample resulting model for a single machine with three jobs is shown in Figure 1.

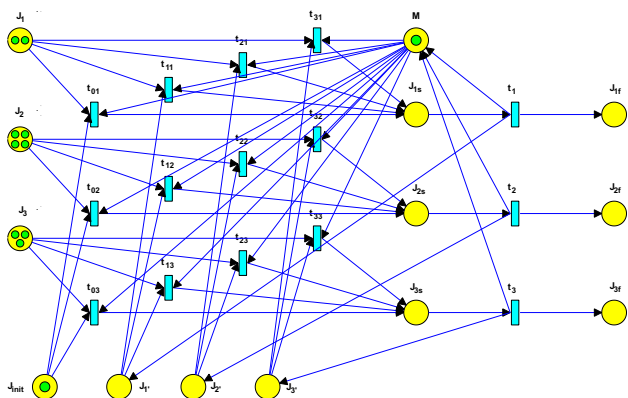


Figure 1: PN model of a sample scheduling problem with sequence dependent setup times

Here setup times are assigned to transitions t_{ij} , where i and j denote the previous and the current job in the sequence, respectively. Marking of J_i denotes i -th job requirements, while $J_{i'}$ is used to memorize the previous job and choose correct transition with proper setup time for the next firing. As initially all $J_{i'}$ are empty, J_{init} is used to enable an initial transition, the choice of which depends on the schedule. Similarly, also more than one t_{ij} can be enabled in subsequent firing steps, but all the transitions compete for a single shared resource M . The choice on which of them will fire and occupy the resource depends on the schedule. After the setup is complete the job operation is processed, which is modelled by transitions t_i with corresponding operation duration. Tokens collected in J_{if} indicate processed jobs. If more stages are required, these tokens are the input to a subsequent stage, so the model can be easily expanded to an arbitrary number of stages.

The number of nodes in the model is polynomial with the number of different jobs, so a graphical representation of the Place/Transition PN model quickly becomes unmanageable. On the other hand it possesses large degree of self-similarity, which indicates that representation can be

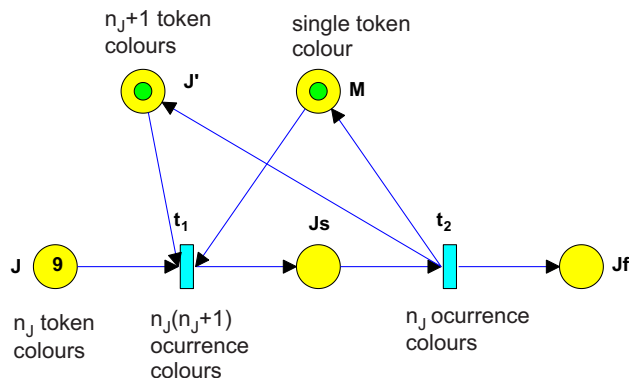


Figure 2: CPN model of a scheduling problem with sequence dependent setup times

simplified by employing Coloured Petri nets (CPNs). Figure 2 shows the same model in CPN framework.

Here all job types can be represented by a single node, containing tokens of various colours. Similarly, different setup times or processing times are represented by a single transition with multiple occurrence colours.

3. PETRI NET SIMULATION OF SCHEDULES WITH SEQUENCE DEPENDENT SETUP TIMES

To use the derived model within simulation-optimization approach, a simulation based evaluation of schedule objective function needs to be implemented. The simulation run must depend on externally changing influential variables, whose changes are driven by optimization algorithm.

To formalize the simulation algorithm, the modelling framework is first presented in a formal manner.

3.1. Coloured Petri nets

For clarity of presentation, a simplified version of Coloured Petri nets without arc expressions and guards will be considered in the following. Compared with standard CPNs (Jensen 1997), no arc inscriptions other than weight are used, and transition inscriptions are limited to specification of time delay only. This allows a more focused presentation of time aspects while the mechanism of time inclusion is kept as close as possible to the one of standard CPNs. This enables a straightforward extension of presented concepts to standard CPNs.

A formal definition of such a CPN is given as follows. Note that the definition closely follows one of the representations used in Basile et al. (2007) with an important difference: a different interpretation of transition delays is used, which is closer to that of Jensen (1997).

A $CPN = (\mathcal{N}, M_0)$ is a Coloured Petri net system, where: $\mathcal{N} = (P, T, Pre, Post, Cl, Co)$ is a Coloured Petri net structure:

- $P = \{p_1, p_2, \dots, p_k\}, k > 0$ is a finite set of places.
- $T = \{t_1, t_2, \dots, t_l\}, l > 0$ is a finite set of transitions (with $P \cup T \neq \emptyset$ and $P \cap T = \emptyset$).
- Cl is a set of colours.

- $Co : P \cup T \rightarrow Cl$ is a colour function defining place marking colours and transition occurrence colours. $\forall p \in P, Co(p) = \{a_{p,1}, a_{p,2}, \dots, a_{p,u_p}\} \subseteq Cl$ is the set of u_p possible colours of tokens in p , and $\forall t \in T, Co(t) = \{b_{t,1}, b_{t,2}, \dots, b_{t,v_t}\} \subseteq Cl$ is the set of v_t possible occurrence colours of t .
- $Pre(p, t) : Co(t) \rightarrow Co(p)_{MS}$ is an element of the pre-incidence function and is a mapping from the set of occurrence colours of t to a multiset over the set of colours of p , $\forall p \in P, \forall t \in T$. It can be represented by a matrix whose generic element $Pre(p, t)(i, j)$ is equal to the weight of the arc from p w.r.t colour $a_{p,i}$ to t w.r.t colour $b_{t,j}$. When there is no arc with respect to the given pair of nodes and colours, the element is 0.
- $Post(p, t) : Co(t) \rightarrow Co(p)_{MS}$ is an element of the post-incidence function, which defines weights of arcs from transitions to places with respect to colours.

$M(p) : Co(p) \rightarrow \mathbb{N}$ is the marking of place $p \in P$ and defines the number of tokens of a specified colour in the place for each possible token colour in p . Place marking can be represented as a multiset $M(p) \in Co(p)_{MS}$ and the net marking M can be represented as a $k \times 1$ vector of multisets $M(p)$. M_0 is the initial marking of a CPN.

Timed models

As described in Bowden (2000), there are three basic ways of representing time in PNs: firing durations (FD), holding durations (HD) and enabling durations (ED). When using HD principle, a firing has no duration but a created token is considered unavailable for the time assigned to transition that created the token.

HD principle is somewhat restrictive, as can be seen in a PN where a conflict appears. In this case more transitions are enabled by one marking, but not all can fire simultaneously. A choice has to be made on which of them will actually fire. With HD principle the firing occurs immediately after the choice is made and the conflict is resolved, and the delay is only active afterwards. This indicates that HD concept can only model non-preemptive tasks in the system. A more general concept is presented and formally described in Lakos and Petrucci (2007). When modelling several performance optimization problems, e.g. scheduling problems, a general framework is not needed. It is natural to use HD when modelling most scheduling processes as operations are considered non-preemptive. HD principle is also used in the timed version of CPNs defined by Jensen (1997), where the assignment of delays both to transitions and to output arcs is allowed. We further simplify this by allowing time delay inscriptions to transitions only. This is sufficient for the type of examples investigated here, and can be generalized if necessary.

To include a time attribute of the marking tokens, which implicitly defines their availability and unavailability, the notation of Jensen (1997) will be adopted. A group of coloured tokens is accompanied with a time stamp, which defines the time instant at which the tokens become available. The time refers to the observation time and is measured by a global clock. A timestamp is assigned to

every token at the moment of token creation; this can be at the definition of the initial marking or at a transition firing. A group of tokens is available and can participate in transition enabling if and only if the assigned timestamp values are lower or equal to the global clock value. In the opposite case the tokens are declared unavailable. As the global clock is increasing a momentarily unavailable token can become available at some future point in time.

The timestamp is written next to the token number and colour and separated from the number by @. E.g., two c-coloured tokens with time stamp 10 are denoted $2'c@10$. A collection of tokens with different colours and/or time stamps is defined as a multiset, and written as a sum (union) of sets of timestamped tokens. E.g., two c-coloured tokens with timestamp 10 and three d-coloured tokens with timestamp 12 are written as $2'c@10+3'd@12$.

Time stamps are elements of a time set TS , which is defined as a set of numeric values. In many software implementations the time values are integer, i.e. $TS = \mathbb{N}$, the same assumption will be used in the sequel. The timestamps could also be admitted to take any positive real value including 0, i.e. $TS = \mathbb{R}_0^+$, the restriction to integers merely simplifies the presentation of examples. Timed markings are represented as collections of time stamps and are multisets over TS : TS_{MS} . By using HD principle the formal representation of a Coloured Timed Petri net (CTPN) is defined as follows.

$CTPN = (\mathcal{N}, M_0)$ is a Coloured Timed Petri net system, where:

- $\mathcal{N} = (P, T, Pre, Post, Cl, Co, f)$ is a Coloured Time Petri net structure with $(P, T, Pre, Post, Cl, Co)$ as defined above.
- $f : Co(t) \rightarrow TS$ is the time function that assigns a non-negative deterministic time delay to every occurrence colour of transition $t \in T$.
- $M(p) : Co(p) \rightarrow TS_{MS}$ is the timed marking, M_0 is the initial marking of a timed PN.

Coloured Petri net firing rule

The firing rule of a CTPN is defined in accordance with the HD timing principle. Functions Pre and $Post$ define the weights of directed arcs, which are represented by arc inscriptions in the matrix form. In the case when the all the weights in the matrix are 0, the arc is omitted. Let $\bullet t_b \subseteq P \times Cl$ denote the set of places and colours which are inputs to occurrence colour $b \in Co(t)$ of transition $t \in T$, i.e., there exists an arc from every $(p, a) \in \bullet t$ to t with respect to colours $a \in Co(p)$ and $b \in Co(t)$.

To determine the availability and unavailability of tokens, two functions on the set of markings are defined. The set of markings is denoted by \mathbb{M} . Given a marking and a time $\tau_i \in TS$, $m : P \times \mathbb{M} \times TS \rightarrow Co(p)_{MS}$ defines the number of available coloured tokens, and $n : P \times \mathbb{M} \times TS \rightarrow Co(p)_{MS}$ defines the number of unavailable coloured tokens for each place of a CTPN at a given model time $\tau_i \in TS$. As mentioned before, the model time is represented by a global clock.

Two timed markings can be added (denoted $+\tau$) in a similar way as multisets, i.e. by making a union of the corresponding multisets. The definition of subtraction is somewhat more problematic. To start with, a comparison operator is defined. Let M_1 and M_2 be markings of a place $p \in P$. By definition, $M_1 \geq_\tau M_2$ iff $m(p, M_1, \tau_i) \geq m(p, M_2, \tau_i), \forall \tau_i \in TS, \forall a \in Co(p)$.

Similarly, the subtraction is defined by the number of available tokens, and the subtrahend should not contain any unavailable tokens. Let M_1, M_2 and M_3 be markings of a place $p \in P, M_1 \geq_\tau M_2$, and $m(p, M_1, \tau_i), m(p, M_2, \tau_i)$, and $m(p, M_3, \tau_i)$, be the corresponding numbers of available tokens at time τ_i , and $n(p, M_2, \tau_i) = 0$. The difference $M_3 = M_1 -_\tau M_2$ is then defined as any $M_3 \in \mathbb{M}$ having $m(p, M_3, \tau_i) = m(p, M_1, \tau_i) - m(p, M_2, \tau_i)$.

As the HD principle is used it is not important which of the available tokens is removed. All the subsequent transition firings only deal with present or future points in time, and all currently available tokens remain available for the eventual enablement of future firings. The implementation can be simplified if the new markings are always generated in the same way, e.g., by always removing the token with the most recent timestamp first.

Using the above definitions, the HD-based firing rule of a CTPN can be defined. Given a marked CTPN $= (\mathcal{N}, M)$, a transition t is *time enabled* at time $\tau_i \in TS$ w.r.t occurrence colour $b \in Co(t)$, denoted $M[t_b]_{\tau_i}$ iff $m(p, M, \tau_i) \geq Pre(p, t)(b), \forall p \in \bullet t_b$. An enabled transition can fire, and as a result removes tokens from input places and creates tokens in output places.

If transition t fires w.r.t occurrence colour b , then the new marking is given by

$$M'(p) = M(p) -_\tau Pre(p, t)(b)@_{\tau_i} +_\tau Post(p, t)(b)@_{(\tau_i + f(t, b))}, \forall p \in P \quad (1)$$

If marking M_2 is reached from M_1 by firing t_b at time τ_i , this is denoted by $M_1[t_b]_{\tau_i} M_2$. The set of markings of CTPN \mathcal{N} reachable from M is denoted by $R(\mathcal{N}, M)$.

3.2. Simulation with predefined transition set sequences

Petri nets can be used either to constructively build a schedule, employing a given set of scheduling rules, or to evaluate the objective function for a schedule, given in advance. Within the simulation-optimization approach used in this paper we follow the second approach.

The basic problem that has to be resolved when evaluating PN represented schedules is to translate a particular schedule into on-the-fly conflict resolution strategy that will ensure a proper ordering of conflicting transitions. The conflicting transitions represent jobs competing for a shared resource at a given moment in simulation time.

The conflict resolution strategy is here embedded within the PN simulation algorithm and parameterized by a prescribed sequence of transition sets.

Given a set $T_{be} \subseteq T \times Cl$ of time enabled transition t occurrence colours (t, b) , shortly denoted as t_b

$$T_{be} = \{t_b : t \in T, b \in Co(t), M[t_b]_{\tau_i}\}$$

and a set of schedule relevant transition occurrence colours

$$T_{sch} = \{t_b : \exists b \in Co(t), Pre(r, t)(b) > 0\}$$

where $r \in P$ is a resource place, a *transition set sequence* S is considered

$$S = \{T_k : T_k = \{t_{b,a} : t \in T, b \in Co(t), a \in Co(J), Pre(J, t)(b, a) > 0\}\}$$

where $J \in P$ is the job place (see Fig.2), a is the token colour of k -th job in the scheduled sequence, and $k = 1 \dots k_{max}$ is a sequence state.

We say that transition t is *sequence enabled* at sequence state s w.r.t occurrence colour $b \in Co(t)$, denoted $M[t_b]_{T_k}$ iff $t_b \in T_k \vee t_b \notin T_{sch}$. Note that transitions $T \notin T_{sch}$ are always sequence enabled.

During simulation run, at the stage where firing transitions are determined, the transition enabling check within the simulation algorithm implements an additional filter, which filters out the transitions that are not sequence enabled. Only transitions that are both marking- and sequence-enabled are permitted to fire and thus change the PN marking. The new marking is computed according to Eq. (1). Additionally sequence state k is increased, whenever a schedule relevant transition is fired. Such a simulation model will be denoted $PN(S)$.

This way a link is established between the scheduled job sequence and PN transition firing sequence. Compared to our previous work, e.g. (Löscher et al. 2007, Mušič 2012b), the approach is generalized to permit more than one schedule relevant transition occurrence to fire at particular sequence state, which is achieved by specifying set sequence instead of simple firing sequence. A broader class of scheduling models can be addressed this way, such as the case of sequence dependent setup times, where the particular firing is chosen based on the previous system state.

Repeating the above described enablement checking step and marking calculation step evolves the PN model from initial marking, indicating the required number of individual jobs' repetitions to the final marking, indicating the number of finished jobs. Timestamps adjoined to tokens in the final marking indicate the job completion times, which permits to calculate an arbitrary schedule objective (Mušič 2012a, Napalkova et al. 2014).

4. PETRI NET BASED OPTIMIZATION OF SCHEDULES WITH SEQUENCE DEPENDENT SETUP TIMES

4.1. Local search

The proposed optimization approach is to use local search methods. Basic optimization algorithm is shown in Algorithm 1.

One of the central issue within the approach is the neighbour solution construction procedure. In Mušič (2011) a procedure has been proposed, which constructs a new scheduling solution by permuting sequence vectors related to transitions in the PN model that are linked to shared resources. The resulting solution is always feasible but the approach only works for standard job-shop and

Algorithm 1 Local Search

```
generate initial solution  $s$ ;  
 $i := 0$ ;  
repeat  
  generate neighbour solution  $s' \in N(s)$ ;  
  perform acceptance test of  $s'$ ;  
  if test positive then  
     $s := s'$ ;  
  end if  
   $i := i + 1$ ;  
until termination criterion
```

flow-shop scheduling problems. A more general construction procedure is presented in Mušič (2015), which can also be used for flexible job shop problems.

Alternatively, standard permutation representations of schedules can be employed (Cheng et al. 1996, Werner 2013). Among others, the *operation-based representation* is used, where job sequence is coded by a vector of integer indices. Job operations are represented by repetition of indices within the vector, where i -th repetition indicates the execution of i -th operation within the job. Such a representation is used in Xing et al. (2012) where schedules are evaluated by a custom algorithm that takes into account the transition precedence relations within the PN model and corresponding operation durations. Although the work of Xing et al. (2012) is focussed on maintaining neighbour solution feasibility, in particular in the case of deadlock-prone schedule models, the schedule evaluation algorithm only works effectively for particular problem classes. As such the approach is not directly applicable to problems with sequence dependent setup times.

Using the simulation approach with predefined transition set sequences, presented in the previous section, a new schedule evaluation procedure within the local search strategy is proposed here in this paper.

The schedule is represented by a permutation vector

$$F = \{\pi(1), \pi(2), \dots, \pi(k_{max})\}$$

$f_k = \pi(k)$ denotes an integer obtained by permutation of $\{1, \dots, k_{max}\}$ and k_{max} corresponds to

$$k_{max} = n_{stages} \cdot \sum_a M(J)$$

i.e., it represents the number of all required job operations.

At every optimization step, a change in the permutation vector is imposed, e.g. two randomly chosen elements are swapped:

$$N : F \rightarrow F'$$

$$F' = \{f_1, \dots, f_{j-1}, f_k, \dots, f_{k-1}, f_j, \dots, f_{k_{max}}\}$$

Next the resulting vector is decoded into a sequence of transition sets. As the setup times are sequence dependent, no exact calculation of the setup duration is performed at this stage. A set of corresponding transitions that can trigger the setup is specified instead. This substantially simplifies the decoding step.

Next, the simulation is run with the determined transition set sequence as a parameter. During each simulation

step, the compliance with the set sequence is maintained and in case of more than one transition in a sequence element, triggering of one of participating transitions is allowed, as shown above. This way, the actual setup time that applies in a given moment is implicitly defined by a combination of the sequence element and the schedule model. It is determined automatically by checking marking- and sequence- enablement of transitions within the PN model.

Finally, the simulation result is used to evaluate the objective function and the new schedule is kept or discarded, depending on the detailed implementation of the local search strategy.

The optimization algorithm is summarized as follows

Algorithm 2 Schedule optimization

```
generate PN model;  
calculate the number of job operations;  
generate permutation vector  $F$ ;  
decode initial set sequence  $S$ ;  
simulate PN( $S$ );  
calculate initial schedule and related objective function value;  
 $i := 0$ ;  
repeat  
  generate neighbour solution  $F' \in N(F)$ ;  
  decode set sequence  $S'$ ;  
  simulate PN( $S'$ );  
  calculate new schedule and related objective function value;  
  perform acceptance test of  $F'$ ;  
  if test positive then  
     $F := F'$ ;  
  end if  
   $i := i + 1$ ;  
until termination criterion
```

By introducing eventual restrictions within neighbour solution construction procedure and by using different acceptance tests, various local search algorithms can be implemented, e.g., Simulated Annealing (SA) or Tabu search (Brucker 2007).

4.2. Genetic algorithms

The local search procedure can be easily adapted to the use of genetic algorithms (Cheng et al. 1996, Eiben and Smith 2015). The above described permutation vector is used as a chromosome, and the described vector change, which represents the neighbourhood operator in local search strategy, is used in mutation step of the genetic algorithm.

What remains is to define a crossover operator. Standard crossover operators for the permutation representations can be used, such as order crossover (OX), linear order crossover (LOX) or position-based crossover (PBX) (Cheng et al. 1999, Werner 2013, Eiben and Smith 2015).

Simulation approach with predefined transition set sequences is then used in the fitness evaluation stage of the genetic algorithm.

5. EXPERIMENTAL RESULTS

5.1. Single machine problem

The introductory example of Fig. 1 is used to show the feasibility of the approach. The setup times are integers selected from a uniform distribution between 2 and 4, and integer processing times are selected from a uniform distribution between 8 and 12.

Initial optimization was attempted by SA algorithm. Three problem instances were generated and 10 optimization runs were performed by each instance. Objective function was the makespan, i.e. the finish time of the last job. Figure 3 shows how objective function of the current solution changes with the iterations of the SA algorithm for the three problem instances. Figure 4 shows the convergence of the solutions for all optimization runs. It can be seen that although the convergence rates are different, all the runs converge to the same values for the same problem instances.

One of the solutions is illustrated by Gantt chart in Figure 5, where processing times for jobs J_1 , J_2 and J_3

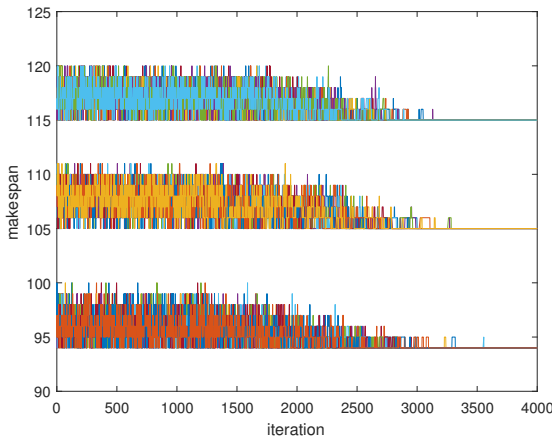


Figure 3: Solution changes during the run of SA algorithm

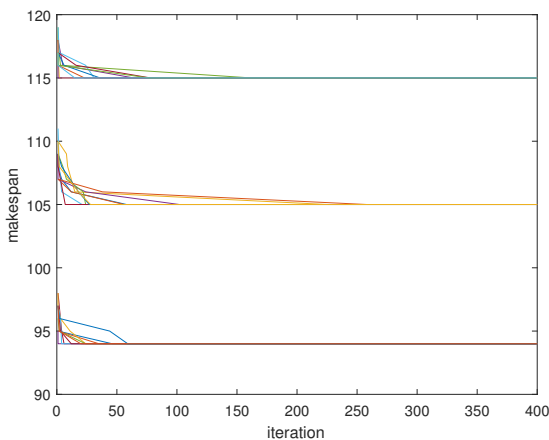


Figure 4: Convergence of solutions for three problem instances

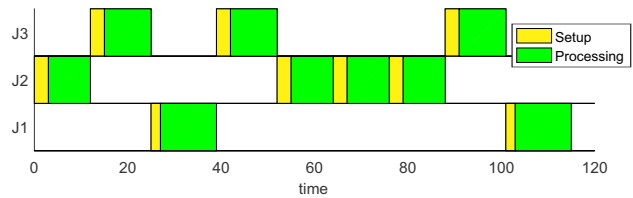


Figure 5: Sample Gantt chart for a problem with sequence dependent setup times

are 12, 9 and 10, respectively, initial setup times are 3, 3 and 4, and sequence depended setup times are given by matrix:

$$S_{ij} = \begin{bmatrix} 3 & 4 & 3 \\ 2 & 3 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

where S_{ij} denotes the setup time when job J_i is followed by J_j . The optimal makespan is 115 time units.

5.2. Two stage flow line

The two stage flow line example is inspired by the set of sample problems in Kurz and Askin (2004). Basically, the model of the first example is extended to 6 different jobs, and then repeated twice in sequel, where final job places of stage 1 are used as initial job places of stage 2. Integer setup times are selected from a uniform distribution between 5 and 90, and integer processing times are selected from a uniform distribution between 50 and 70.

Figure 6 illustrates a sample optimization run for the second example.

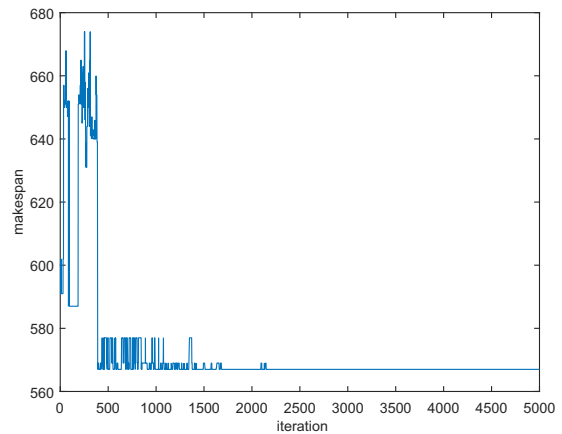


Figure 6: Solution changes during the run of SA algorithm for the two stage example

Additionally, the second example was optimized by genetic algorithm. The Matlab implementation of genetic algorithm within Global Optimization Toolbox was used, with custom defined fitness function, which implements the proposed set sequence simulation based evaluation of solutions. Additionally, custom crossover and mutation functions were defined, which are able to process permutation based solutions.

Although good solutions were obtained in general, significant dependence on algorithm parameter settings was observed, e.g., dependence on Fitness Scaling Options and Selection Options. In general, the settings that maintain greater population diversity lead to better overall results. Figures 7 and 8 show the convergence for one of the problem instances and two different sets of algorithm settings. The same crossover and mutation operators were used, but different fitness scaling (rank vs. proportional) and different selection settings (stochastic uniform vs. uniform).

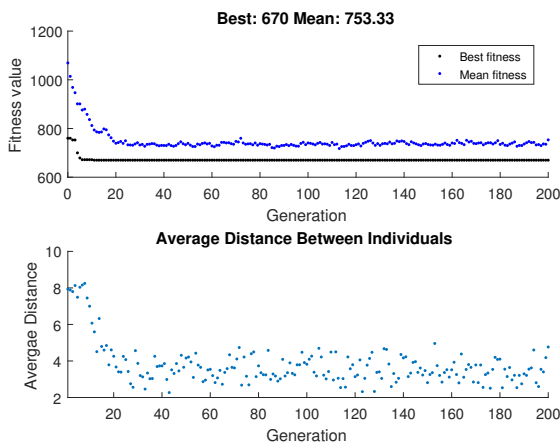


Figure 7: Optimization of the two stage example by genetic algorithm - fast convergence, poor solution

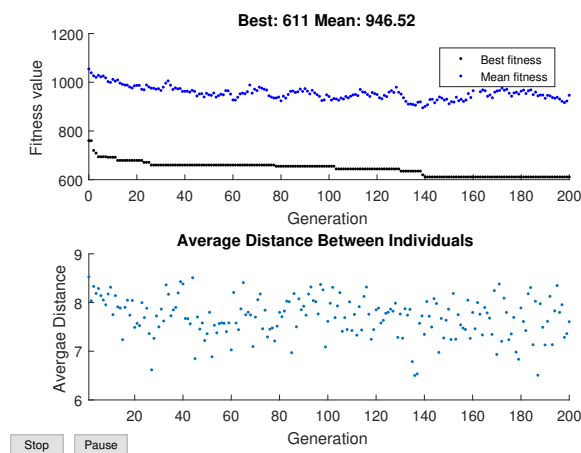


Figure 8: Optimization of the two stage example by genetic algorithm - slow convergence, good solution

Figures 9 and 10 illustrate the optimal solution by Gantt charts. Note that setup and processing stages of a job are represented by the same colour.

Finally, a comparison of different crossover operators was performed. No significant impact of the crossover operator choice to the solution quality is observed, which is illustrated in Figure 11 by showing average resulting

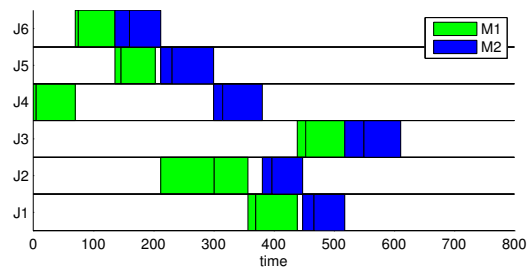


Figure 9: Gantt chart for the two stage example - job ordering

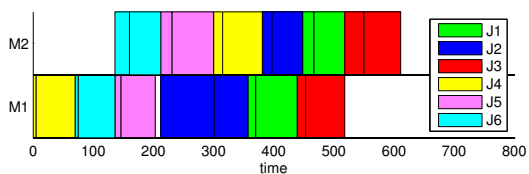


Figure 10: Gantt chart for the two stage example - machine usage

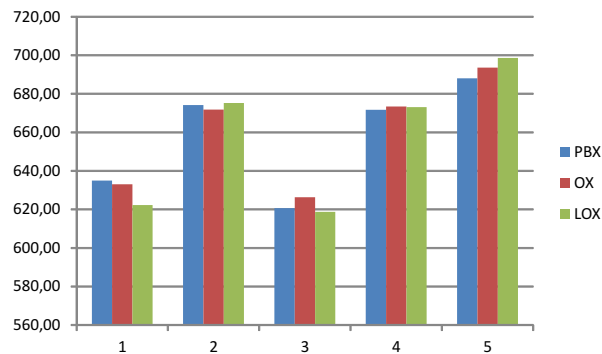


Figure 11: Average results of 10 optimization runs for five problem instances using different crossover operators

makespan of 10 optimization runs for five problem instances using operators PBX, OX, and LOX, respectively.

6. CONCLUSIONS

The paper shows how existing simulation-optimization framework using Petri net models and specific, sequence dependent PN simulation can be extended to the class of scheduling problems with sequence dependent setup times. Related PN modelling approach is presented, the advantage of Coloured Petri nets use is illustrated, and sequence dependent simulation algorithm is extended by allowing set sequence specification. This way an additional degree of freedom is introduced in specifying firing sequences, which permits to partially determine the actual sequence on-the-fly during simulation execution. Some illustrating examples show the feasibility of the approach in combination with local search techniques and genetic algorithms.

REFERENCES

- Basile, F., Carbone, C. and Chiacchio, P., 2007. Simulation and analysis of discrete-event control systems based on Petri nets using PNetLab, *Control Engineering Practice*, 15, 241–259.
- Bowden, F. D. J., 2000. A brief survey and synthesis of the roles of time in Petri nets, *Mathematical & Computer Modelling*, 31, 55–68.
- Brucker, P., 2007. *Scheduling Algorithms*, 5th edn, Springer.
- Cheng, R., Gen, M. and Tsujimura, Y., 1996. A tutorial survey of job-shop scheduling problems using genetic algorithms—I. representation, *Computers & Industrial Engineering*, 30 (4), 983–997.
- Cheng, R., Gen, M. and Tsujimura, Y., 1999. A tutorial survey of job-shop scheduling problems using genetic algorithms, part II: hybrid genetic search strategies, *Computers & Ind. Engineering*, 36 (2), 343–364.
- Eiben, A. E. and Smith, J. E., 2015. *Introduction to Evolutionary Computing*, Natural computing series, 2 edn, Springer-Verlag, Berlin Heidelberg.
- Gradišar, D. and Mušič, G., 2007. Production-process modelling based on production-management data: a Petri-net approach, *International Journal of Computer Integrated Manufacturing*, 20 (8), 794–810.
- Han, L., Xing, K., Chen, X., Lei, H. and Wang, F., 2014. Deadlock-free genetic scheduling for flexible manufacturing systems using Petri nets and deadlock controllers, *International Journal of Production Research*, 52 (5), 1557–1572.
- Huang, H., Du, H. and Ahmad, F., 2013. Job shop scheduling with Petri nets, in P. Pardalos (ed.), *Handbook of Combinatorial Optimization*, Springer, pp. 1667–1711.
- Jensen, K., 1997. *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use*, Vol. 1, 2 edn, Springer-Verlag, Berlin.
- Kurz, M. E. and Askin, R. G., 2004. Scheduling flexible flow lines with sequence-dependent setup times, *Eur. Journal of Operational Research*, 159 (1), 66–82.
- Lakos, C. and Petrucci, L., 2007. Modular state space exploration for timed Petri nets, *International Journal on Software Tools for Technology Transfer*, 9, 393–411.
- Löscher, T., Mušič, G. and Breiteneker, F., 2007. Optimisation of scheduling problems based on timed Petri nets, *6th EUROSIM Congress on Modelling and Simulation*, Vol. II, Ljubljana, Slovenia.
- Mejía, G., Montoya, C., Cardona, J. and Castro, A. L., 2012. Petri nets and genetic algorithms for complex manufacturing systems scheduling, *International Journal of Production Research*, 50 (3), 791–803.
- Mušič, G., 2011. Efficient exploration of coloured Petri net based scheduling problem solutions, *23th European Modeling & Simulation Symposium*, Rome, Italy, pp. 681–689.
- Mušič, G., 2012a. Efficient exploration and evaluation of Coloured Petri net based scheduling problem solutions, *14th IFAC Symposium on Information Control Problems in Manufacturing, INCOM'12*, IFAC, Bucharest, pp. 150–155.
- Mušič, G., 2012b. Schedule optimization based on Coloured Petri nets and local search, *7th Vienna Conference on Mathematical Modelling, full paper preprint volume*, ARGESIM, Vienna.
- Mušič, G., 2015. Generation of feasible Petri net based scheduling problem solutions, *IFAC-PapersOnLine*, 48 (1), 856 – 861. 8th Vienna International Conference on Mathematical Modelling MATHMOD 2015.
- Napalkova, L., Piera, M. A. and Mušič, G., 2014. Performance evaluation of flexible manufacturing systems by coloured timed Petri nets and timed state space generation, in J. Campos, C. Seatzu and X. Xie (eds), *Formal Methods in Manufacturing*, CRC Press, pp. 381–408.
- Piera, M. A. and Mušič, G., 2011. Coloured Petri net scheduling models: Timed state space exploration shortages, *Math. Comput. Simul.*, 82, 428–441.
- Pinedo, M. L., 2008. *Scheduling: Theory, Algorithms, and Systems*, 3rd edn, Springer Publishing Company.
- Recalde, L., Silva, M., Ezpeleta, J. and Teruel, E., 2004. Petri nets and manufacturing systems: An examples-driven tour, in J. Desel, W. Reisig and G. Rozenberg (eds), *Lectures on Concurrency and Petri Nets*, Vol. 3098 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 71–89.
- Ríos-Mercado, R. Z. and Bard, J. F., 1998. Computational experience with a branch-and-cut algorithm for flow-shop scheduling with setups, *Computers & Operations Research*, 25 (5), 351 – 366.
- Sindičić, I., Petrović, T. and Bogdan, S., 2012. Machine-job incidence matrix based analysis of manufacturing systems in time domain, *IEEE International Conference on Control Applications (CCA)*, pp. 234–239.
- Tuncel, G. and Bayhan, G. M., 2007. Applications of Petri nets in production scheduling: a review, *International Journal of Advanced Manufacturing Technology*, 34, 762–773.
- Viswanadham, N. and Narahari, Y., 1992. *Performance Modeling of Automated Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, USA.
- Werner, F., 2013. A survey of genetic algorithms for shop scheduling problems, in P. Siarry (ed.), *Heuristics: Theory and Applications*, Nova Science Publishers, pp. 161–222.
- Xing, K., Han, L., Zhou, M. and Wang, F., 2012. Deadlock-free genetic scheduling algorithm for automated manufacturing systems based on deadlock control policy, *IEEE Trans. Systems, Man, and Cybernetics, Part B*, 42 (3), 603–615.

AUTHOR BIOGRAPHY

GAŠPER MUŠIČ received B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Slovenia in 1992, 1995, and 1998, respectively. He is Full Professor at the Faculty of Electrical Engineering, University of Ljubljana. His research interests are in discrete event and hybrid dynamical systems, supervisory control, planning, scheduling, and industrial informatics. His Web page can be found at <http://msc.fe.uni-lj.si/Staff.asp>.

MINIMIZING THE IMPACT OF ESCALATOR FAILURES IN METRO TACUBAYA SUBWAY STATION ON USER'S MOBILITY

Hector Omar Israel Bautista-Martínez^(a), Aida Huerta-Barrientos^(b), Gerardo de Jesús Portillo-Villasana^(c)

^{(a),(b),(c)} Department of Systems Engineering
National Autonomous Mexico University

^(a)omarbautista07@gmail.com, ^(b)aida.huerta@comunidad.unam.mx, ^(c)gerardojpv@hotmail.com

ABSTRACT

TACUBAYA subway station interconnects subway Lines 1,7, and 9 of Mexico City subway system. In TACUBAYA station there are a total of 24 escalators, 6 are downstream and 18 are upstream. Subway station service starts at 5 a.m. and finishes at 12 p.m. from Monday to Friday. On Saturday operates from 6 a.m. to 12 p.m., and on Sundays from 7 a.m. to 12 p.m. In last years, escalators have presented failures that have critically affected the passenger's mobility due to tens meters of deep. The aim of this paper is to develop an agent-based simulation model to minimize the impact of escalator failures on user's mobility making some escalators bidirectional allowing the flow of passengers to be continuous. First, a conceptual model of TACUBAYA subway station is developed. Second, it is implemented in computer using AnyLogic™ software. Then, some simulation scenarios are designed to evaluate the impact of escalator's failures on the mobility of passengers.

Keywords: escalator failures, urban mobility, subway, subway stairs, Mexico City.

1. INTRODUCTION

TACUBAYA subway station interconnects subway Lines 1,7, and 9 of Mexico City subway system (see Fig. 1). Line 1 PANTITLAN-OBSERVATORIO comprises 20 stations, and the annual ridership was more than 267 million in 2015. This subway line has a length of more than 16 km. underground and just one km. at ground level. While, Line 7 EL ROSARIO-BARRANCA DEL MUERTO operates with 14 stations. The annual ridership of Line 7 in 2015 was almost 101 million. Line 7 has a length of 18 km. underground and 0.6 km. at ground level. Meantime, Line 9 TACUBAYA-PANTITLAN is currently operated with 12 stations, and recorded an annual ridership of almost 120 million in 2015. Line 9 has a length near to 10 km. underground and 5 km. elevated. TACUBAYA subway station is located in the Miguel Hidalgo Borough of Mexico City. It is considered as one of the busiest subway stations. From January to March 2016, this stations recorded fifty-two thousand ridership.



Figure 1: Geographical location of TACUBAYA subway station

TACUBAYA station was built on many levels to accommodate the three subway lines. There are a total of 24 escalators, 6 are downstream and 18 are upstream, that communicate all levels. The comfortable escalator's speed is 0.5meters /second. From Monday to Friday Subway station service starts at 5 a.m. and finishes at 12 p.m. On Saturday operates from 6 a.m. to 12 p.m. and on Sundays from 7 a.m. to 12 p.m. TACUBAYA station has two busy periods, one at the morning and the other one at the evening. In last years, escalators have presented failures that have critically affected passenger's mobility due to tens meters of deep (see Fig. 2). The aim of this paper is to develop an agent-based simulation model to minimize the impact of escalator failures on user's mobility making some escalators bidirectional allowing the flow of passengers to be continuous.

This paper is prepared as follows: a conceptual model of TACUBAYA subway station is developed considering the flow of passengers in Section 2. An agent-based simulation model is implemented using AnyLogic™ software and validated via a sensitivity analysis in Section 3. The evaluation of the impact of escalator's failures on the mobility of passengers and its optimization is presented in Section 4. Conclusions are outlined in Section 5.



Figure 2: Passengers on escalators in TACUBAYA subway station when one escalator failures

2. THE CONCEPTUAL MODEL

We develop the conceptual model of the processes involved when a user stay in TACUBAYA subway station and decide to transboard from Line 7 and 9 to Line 1. For instance, if passengers get off Line 1-OBSERVATORIO direction, they decide to take between escalators and stairs. Then, passengers decide to take Lines 7 or 9. Contrary, they exit the system. In the case that passengers go to Line 7 or 9, they take again escalators or stairs (see Fig. 3). The flow chart of passenger's activities in TACUBAYA subway station when they get off Line 7 and Line 9 are showed in Fig. 4 and Fig. 5, respectively.

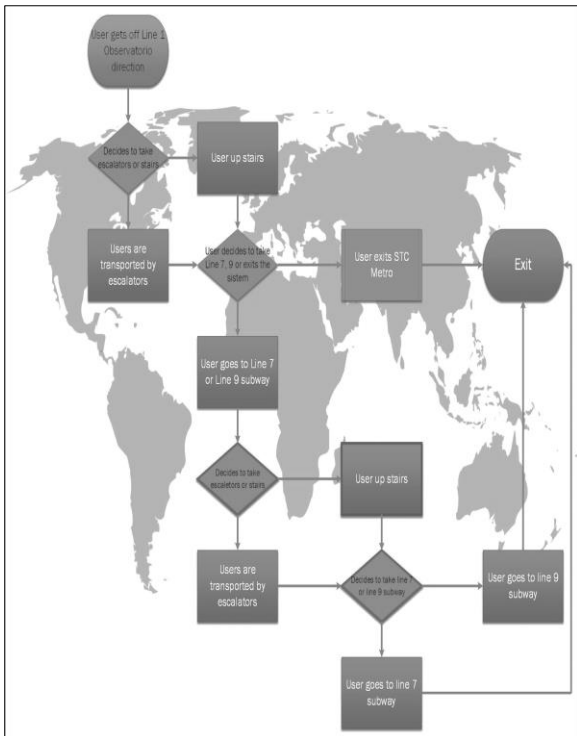


Figure 3: Flow chart of passenger's activities in TACUBAYA subway station, Line 1 direction



Figure 4: Flow chart of passenger's activities in TACUBAYA subway station, Line 7 direction



Figure 5: Flow chart of passenger's activities in TACUBAYA subway station, Line 9 direction

3. THE SIMULATION MODEL OF TACUBAYA SUBWAY STATION

We develop the simulation model of pedestrian activities in the TACUBAYA subway station using discrete-event simulation (DES) approach and agent-based modeling and simulation (ABMS) approach. The simulation model was implemented using AnyLogic™ Personal Learning Edition. In this study, we use two Anylogic libraries: Pedestrian (see Fig. 6) and Rail (see Fig. 7). It is

important to note that in models created with Pedestrian Library, pedestrians move in continuous space, reacting on different kinds of obstacles (walls, different kinds of areas) and other pedestrians (Anylogic 2016).

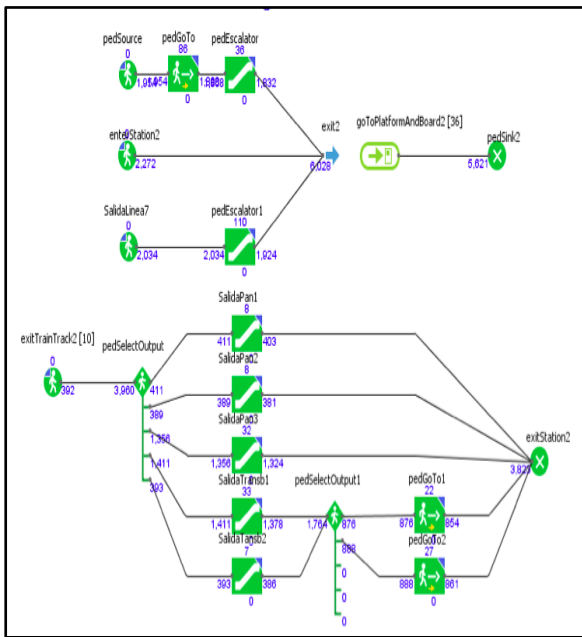


Figure 6: Pedestrian library blocks-TACUBAYA subway station

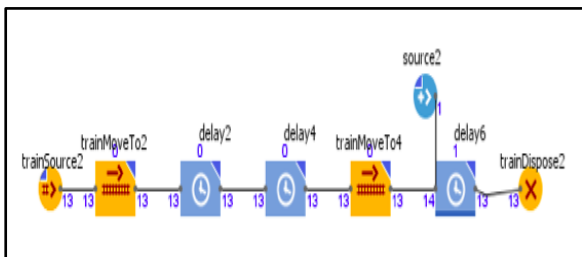


Figure 7: Rail library blocks- TACUBAYA subway station

3.1. Input data analysis

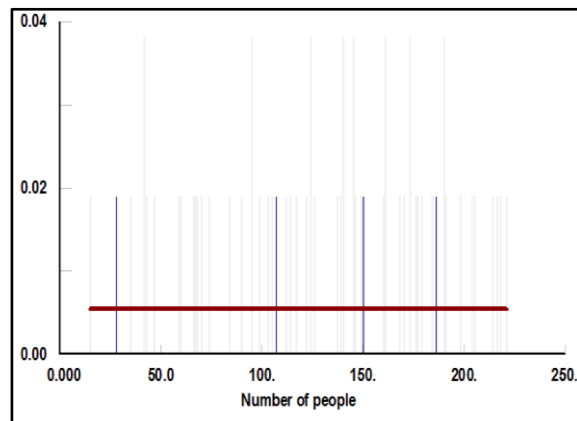
During two weeks, we did some technical visits in TACUBAYA station to collect data about the number of passengers in escalators during one busy hour on Thursday. Then, we used Stat::Fit® version 3.0.1.0 Standard, developed by Geer Mountain Software (Benneyan 1998) to auto fit the appropriate probability distributions. As is observed in Fig. 7, passengers in escalators from Line 7 and Line 9 fit very well to a Discrete Uniform and a Negative Binomial distributions. In this direction, the simulation model was calibrated using the Negative Binomial distribution.

data points	60
minimum	15
maximum	221
mean	128.983
median	138
mode	42
standard deviation	55.5858
variance	3089.78
coefficient of variation	43.0953
skewness	-0.219258
kurtosis	-1.01848

a)

distribution	rank	acceptance
Discrete Uniform(15, 221)	100	do not reject
Negative Binomial(5, 0.0373)	86.1	do not reject
Geometric(0.00769)	0.211	reject
Poisson(129)	0	reject
Logarithmic(0.999)	0	reject

b)



c)

Figure 7. Passengers in escalators from Line 7 and Line 9, a) descriptive statistics, b) Auto fit of distributions and c) graphic of discrete uniform distribution

3.2. Simulation model verification

The verification of simulation model is the process of determination of whether the computer implementation of the conceptual model is correct. (Banks 1998). For the purposes of this study, the Anylogic™ code was checked by other persons considered expertise using Anylogic™ software. Additionally, both the input data and the output data were verified to be reasonable and the trace tool included in Anylogic™ was free of warnings.

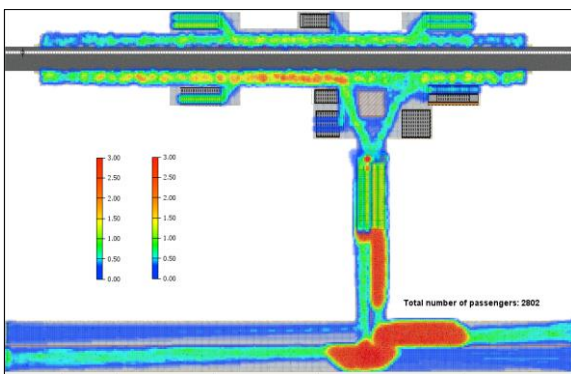
3.3. Simulation model validation

The validation of simulation model is the process of determination of whether the conceptual model can be substituted for the real system for the purposes of experimentation (Banks 1998). To validate the simulation model we design a sensitivity analysis considering changes in the escalator speed. We propose two extreme values, one low and the other one high, and observe with “what happened” scenarios with the flow of passengers (see Table 1).

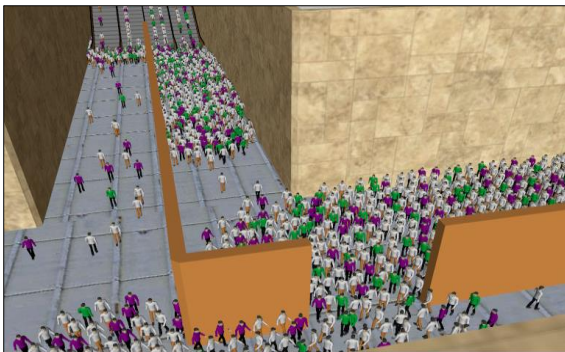
Table 1: Input variables for designing the sensitivity analysis

Input variables		
	Low	High
Escalator speed	0.1 m/s	1 m/s

In the first case, when the escalator speed is very low, 0.1 m/s, we observe two conflicting areas on escalators, where the passenger's density is very high reaching 3 passengers per square meter (see Fig. 8). Contrary, when the escalator speed is high, 1 m/s, we observe the escalator with a low passenger's density, almost 1 passenger per square meter (see Fig. 9), but we observe a conflicting area on the subway platform. Under extreme conditions of escalators speed, the dynamic of the simulation model is similar to the dynamic of the real system.



a)

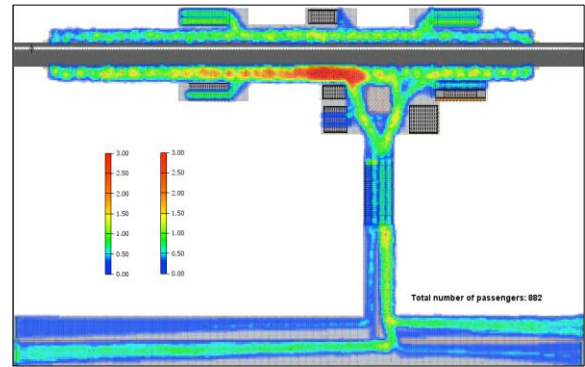


b)

Figure 8: Escalators speed at low value, a) pedestrian density (pedestrian/m²), b) pedestrian dynamics

3.4. Design of simulation experiments

We conduct a 2³ factorial design (Montgomery, 2005) of simulation experiments. We take into account three factors (see Table 2 and Table 3): escalator's speed, passenger's speed, and passenger's flow, with two different values, one minimum and one maximum. The simulation run time is during one hour. In Fig. 10, the main effect of factors on the number of passengers that up escalators are showed. From Fig. 11 to Fig. 18 the pedestrian density (pedestrian/m²) is shown related to each simulation experiment.



a)



b)

Figure 9: Escalators speed at high value a) pedestrian density (pedestrian/m²), b) pedestrian dynamics

Table 2: Factors for design of simulation experiments

Input variables and parameters			
Parameter	Input variable	Low value	High value
A	Escalator's speed	0.1 m/s	1 m/s
B	Passenger's speed	0.1 m/s	1 m/s
C	Passenger's flow	400	4000

Table 3: Design of simulation experiments

Input parameters			
A	B	C	Number of passengers that up escalators
-1	-1	-1	765
-1	+1	-1	775
-1	-1	+1	4661
-1	+1	+1	4800
+1	-1	-1	777
+1	+1	-1	792
+1	-1	+1	4739
+1	+1	+1	7827

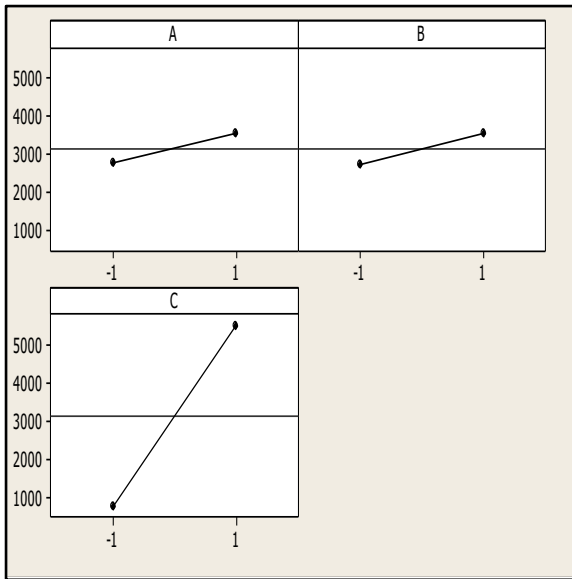


Figure 10: Main effect of factors on passengers that up escalators

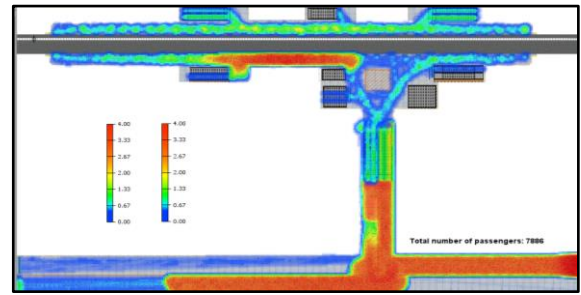


Figure 14: Pedestrian density (pedestrian/m²), simulation experiment 4

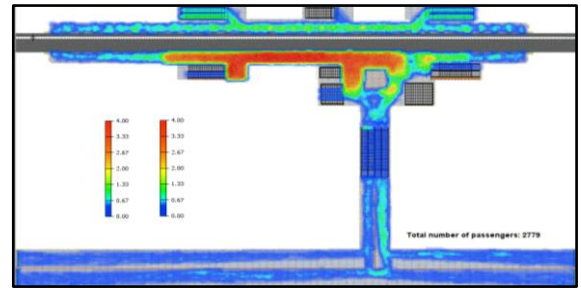


Figure 15: Pedestrian density (pedestrian/m²), simulation experiment 5

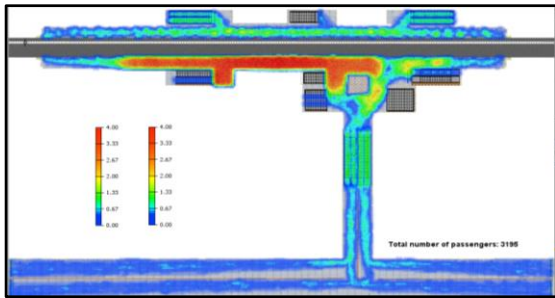


Figure 11: Pedestrian density (pedestrian/m²), simulation experiment 1

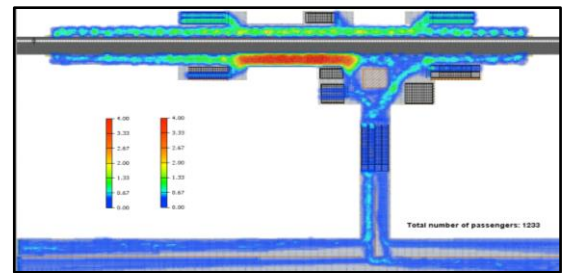


Figure 16: Pedestrian density (pedestrian/m²), simulation experiment 6

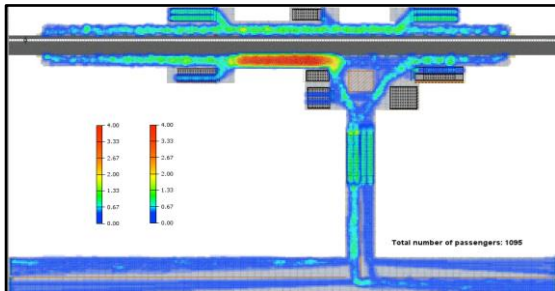


Figure 12: Pedestrian density (pedestrian/m²), simulation experiment 2

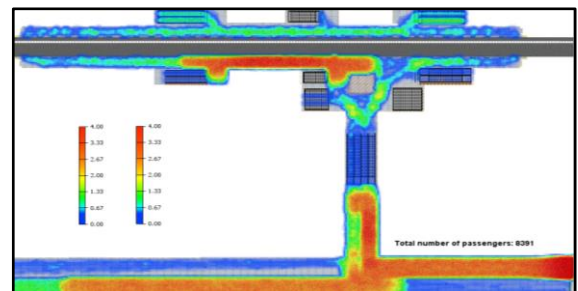


Figure 17: Pedestrian density (pedestrian/m²), simulation experiment 7

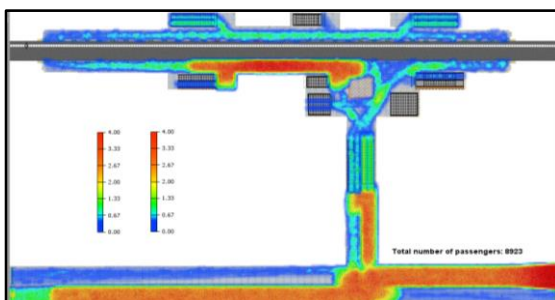


Figure 13: Pedestrian density (pedestrian/m²), simulation experiment 3

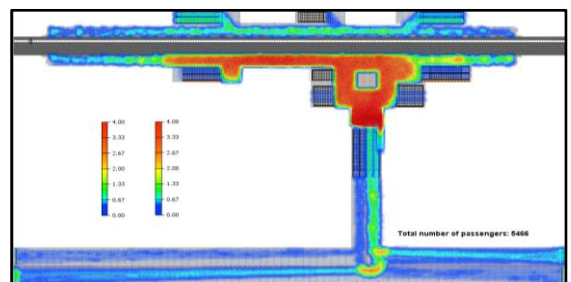
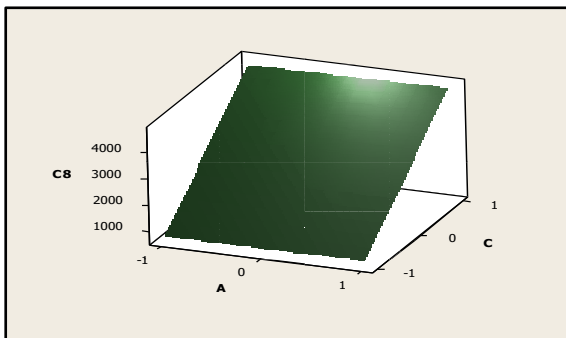
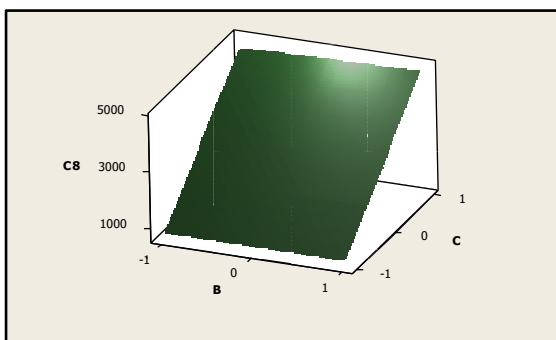


Figure 18: Pedestrian density (pedestrian/m²), simulation experiment 8

As is observed from Fig. 19, the passenger's flow effect on number of passengers that up escalators is larger than either the escalator's speed or the passenger's speed.



a)



b)

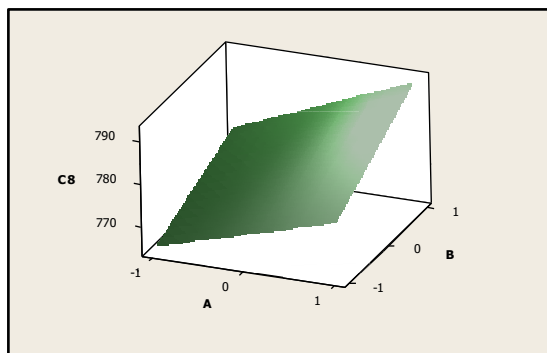


Figure 19: Main effect of factors on passengers that up escalators

4. MINIMIZING THE IMPACT OF ESCALATOR FAILURES ON USER'S MOBILITY

We design some simulation scenarios to evaluate the impact of escalator's failures on the mobility of passengers. We take into account the escalator's speed and passenger's speed equal to 0.5 m/s and the passengers from Line 7 and Line 9 equal to 2000. When escalator failures (see Fig. 20), there are three zones of conflict. One just in front of the escalator that does not work. One solution for this situation is to change from unidirectional to bidirectional an additional escalator (see Fig. 21). As result, the pedestrian density on the

escalator zone remains between 2 and 2.5 pedestrian/m², without traffic of pedestrians. The optimum situation is when all escalators work well and additionally one escalator works in a bidirectional way (see Fig. 22). In this case, pedestrian density remains less than 2 pedestrian/m².

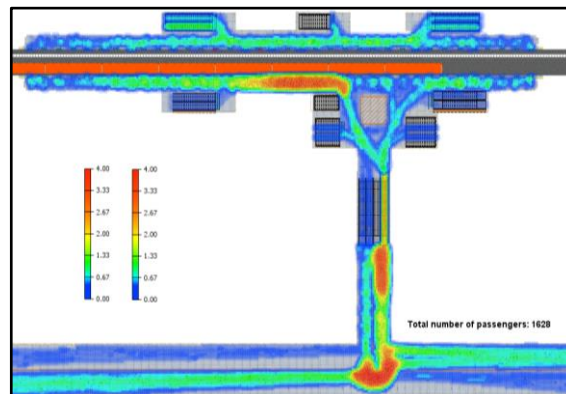


Figure 20: Pedestrian density (pedestrian/m²) when a escalator failures

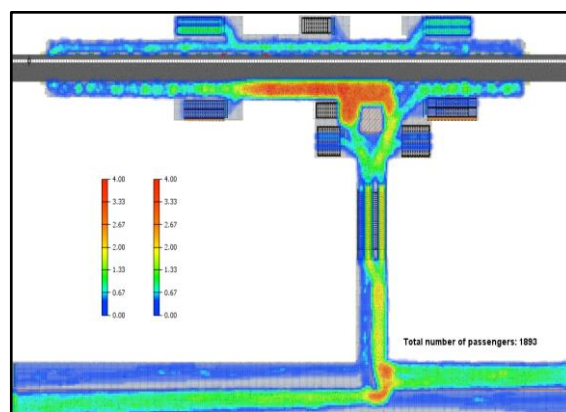


Figure 21: Pedestrian density (pedestrian/m²) with an additional bidirectional escalator

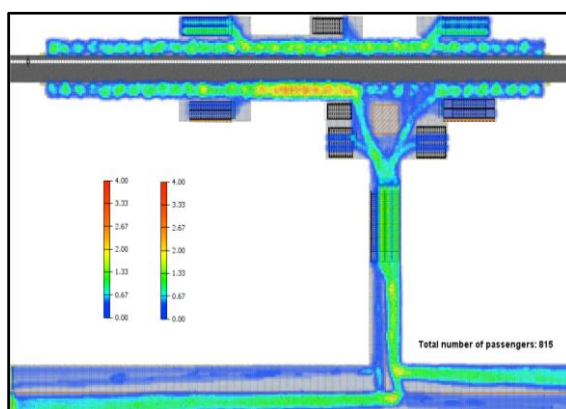


Figure 22: Pedestrian density (pedestrian/m²) with two escalators working well and an additional bidirectional escalator

5. CONCLUSIONS

We developed the conceptual model of the processes involved when a user stays in TACUBAYA subway station and decides to transboard from Line 7 and 9 to Line 1 using discrete-event simulation (DES) approach and agent-based modeling and simulation (ABMS) approach. The simulation model was implemented using AnyLogic™ Personal Learning Edition. During two weeks, we collected data about the number of passengers in escalators during one busy hour on Thursday, then we fitted it to probability distributions to calibrate the simulation model. To validate the simulation model we designed a sensitivity analysis considering changes in the escalator speed. We conducted a 2^3 factorial design taking into account three factors escalator's speed, passenger's speed, and passenger's flow. The passenger's flow effect on number of passengers that up escalators was larger than either the escalator's speed or the passenger's speed. Finally, we evaluated the impact of escalator's failures on the mobility of passengers. When escalator failures there are three zones of conflict. One solution for this situation is to change from unidirectional to bidirectional one escalator. However, the optimum situation is when all escalators works well and additionally one escalator works in a bidirectional way.

ACKNOWLEDGMENTS

The authors appreciate to Sistema de Transporte Colectivo Metro for the facilities in the data collection.

REFERENCES

- Anylogic 2016. Available from: <http://www.anylogic.com> [accessed 5 July 2016]
- Banks, J., 1998. Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Georgia: John Wiley & Sons, Inc.
- Benneyan, J.C., 1998. Software Review: Stat::Fit. OR/MS Today. Available from: <http://lionhrtpub.com/orms/orms-2-98/swr.html> [accessed 11 July 2016]
- Borshchev, A., 2013. The big book of simulation modeling. Multimethod modeling with Anylogic 6.
- Montgomery, D.C., 2005. Design and analysis of experiments. 6th ed. USA: John Wiley & Sons, Inc.

AUTHORS BIOGRAPHY

HECTOR OMAR ISRAEL BAUTISTA MARTÍNEZ currently is Master student of the Department of Systems Engineering at the School of Engineering, UNAM. His research interests are simulation of transport systems, logistics, supply chain and quality.

AIDA HUERTA BARRIENTOS received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the School of Engineering, UNAM and she is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

GERARDO DE JESÚS PORTILLO VILLASANA currently is Master student of the Department of Systems Engineering at the School of Engineering, UNAM. His research interests are security in urban transport, information and communication technologies, simulation and optimization.

AUTOMATIC SIMULATION MODELS GENERATION OF WAREHOUSES WITH MILKRUNS AND PICKERS

António Vieira¹, Luís S. Dias¹, Guilherme B. Pereira¹, José A. Oliveira¹, M. Sameiro Carvalho¹ and Paulo Martins¹

⁽¹⁾⁽²⁾ University of Minho, Campus Gualtar, 4710-057, Braga, Portugal.

^(1,2) (antonio.vieira, lsd, gui, zan, Sameiro, pmartins)@dps.uminho.pt

ABSTRACT

To help a company of the Bosch Group to reduce its costs (both in time and space) with its supermarket, a micro simulation model was developed in Simio. Particularly, the tool is able to model pickers riding milk runs to collect containers of products, from a supermarket, to satisfy the needs of the production lines. Practitioner may benefit from this tool, since it is able to model different supermarket scenarios, for instance different storage strategies. Additionally, the supermarket itself is automatically created, through an Add-in of Simio that was developed in C#, which implements the API of Simio. Conclusions and future work are discussed.

Keywords: Warehouse, milk run, picking, Micro simulation, Simio.

1. INTRODUCTION

In recent years, the Bosch Group has been applying concepts of the Toyota Production System (TPS) (Monden, 1998) and of the Lean Manufacturing (Womack et al., 1990, Womack and Jones, 1996), designated as Bosch Production System (BPS) (Yildiz et al., 2010, Costa et al., 2011). Its purpose is to “eliminate waste in production and all related business processes. BPS provides the basis for continuous improvements in quality, costs, and supply performance” (Bosch, 2014).

A significant part of the costs of a company are its supermarkets (Baker and Canessa, 2009). Since one of the objectives of the BPS is to reduce costs, the need, to study alternatives to the current design and picking system of the supermarket on the company Bosch Car Multimedia Portugal in Ferreiros, Braga, arose.

In this context, a micro simulation model, using Simio, was developed. The tool is able to model pickers riding milk runs to collect containers of products, from the channels of a supermarket, to satisfy the needs of the production lines. A Channel is the basic unit for storage in this supermarket. Each has the capacity to hold several containers. On the other hand, a container holds many units of one type of product.

The storage strategy used in this supermarket is the dedicated. This is the most simple that can be used, since it consists on having a channel dedicated to a single type of product (Bartholdi and Hackman, 2008). One of its great advantages, resides on the fact that, since the locations of the product don't change, the pickers can memorize them, making the picking process more efficient (Bartholdi and Hackman, 2008). Nevertheless,

the problem with this strategy is that “it does not use space efficiently. In fact, it is expected that, on average, the storage capacity is about 50%” (Bartholdi and Hackman, 2008), which represents a high amount of costs associated. To overcome this problem, other strategies can be considered (e.g. random storage). Thus, the simulation model must be able to model several storage strategies. In addition to that, the quantity of requests a picker gets per shift, the time between shifts, the number of types of products, the arrival rate of requests, and the number of milk runs and pickers need to be configurable.

Additionally, the supermarket is composed by circulation corridors for milk runs that gives them access to corridors of racks. In its turn, each rack is composed by a variable number of channels, in height and in width, whereby it is necessary to create several layouts of the supermarket. To do so, the API of Simio is being used to create an add-in, in C#. The latter reads data from an excel file, where the user is able to specify several inputs, e.g. the number of corridors, their positions, their rotation angles, the number of channels on each rack (in height and in width), among others. Nevertheless, the creation of the add-in will not be covered in this paper. Regardless of that, the simulation model was built so that several layouts of the supermarket could be modelled. Thus, this paper intends to document the first part of the simulation model developed, which consists on the pickers receiving requests and riding their milk runs to collect the respective containers from the supermarket. The return of the leftover containers and the restock processes are not yet modelled.

Section 2 presents a review over the literature. In section 3 and 4, the development of the simulation model; the data collected and validation of the model will be covered. Section 5 addresses the development of the add-in to automatically create the supermarket and, in the final section the main conclusions are discussed.

2. LITERATURE REVIEW

According to Coyle et al. “Warehousing provides time and place utility for raw materials, industrial goods, and finished products, allowing firms to use customer service as a dynamic value-adding competitive tool” (1988). Thus, supermarkets represent a very important role on modern supply chains (Baker and Canessa, 2009).

In fact, “whilst supermarkets are critical to a wide range of customer service activities, they are also significant from a cost perspective. Figures for the USA indicate that the capital and operating costs of

supermarkets represent about 22% of logistics costs (Establish, 2005), whilst figures for Europe give a similar figure of 25%” (Baker and Canessa, 2009). These costs impel us to understand the problematic and to use the storage space as efficiently as possible (Bartholdi and Hackman, 2008).

Thus, the need to provide companies with methods capable of improving the performance of supermarkets arises. According to Gu et al., some of these methods include simulation, analytical methods and benchmarking. The former is the most used whether in literature or in practice (2010). One example is the simulation model developed by Costa et al. using Arena. The authors conducted experiments to identify changes that could be made on a material delivery system to improve the efficiency and precision of the logistic train functioning they were modelling (2008).

Since the number of simulation tool options can be very high, tool comparison becomes a very important task. However, most of scientific works related to this subject “analyse only a small set of tools and usually evaluating several parameters separately avoiding to make a final judgement due to the subjective nature of such task” (Dias et al., 2007).

Hlupic and Paul (1999) compared a set of simulation tools, distinguishing between users of software for educational purpose and users in industry. In his turn, Hlupic (2000) developed “a survey of academic and industrial users on the use of simulation software, which was carried out in order to discover how the users are satisfied with the simulation software they use and how this software could be further improved”. Dias and Pereira et al. (2007, 2011) compared a set of tools based on popularity on the internet, scientific publications, WSC (Winter Simulation Conference), social networks and other sources. “Popularity should never be used alone otherwise new tools, better than existing ones would never get market place, and this is a generic risk, not a simulation particularity” (Dias et al., 2007). However, a positive correlation may exist between popularity and quality, since the best tools have a greater chance of being more popular. According to the authors, the most popular tool is Arena and the good classification of the Simio is noteworthy. Based on these results, Vieira et al. compared both tools taking into consideration several factors (2014a).

Simio was the chosen tool for this project. It is based on intelligent objects (Sturrock and Pegden, 2010, Pegden, 2007, Pegden and Sturrock, 2011). These “are built by modellers and then may be used in multiple modelling projects. Objects can be stored in libraries and easily shared” (Pegden, 2013). Unlike other object-oriented systems, in Simio there is no need to write any programming code, since the process of creating a new object is completely graphic (Pegden and Sturrock, 2011, Pegden, 2007, Sturrock and Pegden, 2010). The activity of building an object in Simio is identical to the activity of building a model. In fact there is no difference between an object and a model (Pegden, 2007, Pegden and Sturrock, 2011). A vehicle, a customer or any other

agent of a system are examples of possible objects and, combining several of these, one can represent the components of the system in analysis. Thus, a Simio model looks like the real system (Pegden and Sturrock, 2011, Pegden, 2007). This fact can be very useful, particularly while presenting the results to someone non-familiar to the concepts of simulation.

In Simio the model logic and animation are built in a single step (Pegden and Sturrock, 2011, Pegden, 2007). This feature is very important, because it makes the modulation process very intuitive (Pegden and Sturrock, 2011). Moreover, the animation can also be useful to reflect the changing state of the object (Pegden, 2007). In addition to the usual 2D animation, Simio also supports 3D animation as a natural part of the modelling process (Sturrock and Pegden, 2010). To switch between 2D and 3D views the user only needs to press the 2 and 3 keys of the keyboard (Sturrock and Pegden, 2010). Moreover, Simio provides a direct link to Google Warehouse, a library of graphic symbols for animating 3D objects (Sturrock and Pegden, 2010, Pegden and Sturrock, 2011).

3. MODEL DEVELOPMENT

Throughout this section, some terms will be used that may be unknown for a user not familiar with Simio. For those, a reading of the paper written by Vieira et al. (2014a) would be advisable.

For this simulation project, 4 types of entities and 5 models (4 sub-models and a main one) were created. In the first section of this chapter, the former will be presented, while the models will be analysed on the following sections. Particularly, the main goals, the properties and the external view of the sub-models will be presented, so that it becomes easier to understand their use on the main model, which will be addressed in the last section. The 4 created types of entities were:

3.1. Types of Entities

- **Picker:** Represents the pickers of the system. Their functions are to collect Requests at the beginning of a shift and take Containers from Channels of the Supermarket to place them on the milk run.



Figure 1: Symbol of the Picker entity

- **milk run:** Represents the milk runs of the system. Its only purpose is to transport the Picker and the selected Containers between the Supermarket.

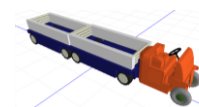


Figure 2: Symbol of the milk run entity

- **Request:** Represents the request of the system



Figure 3: Symbol of the Request entity

- **Container:** Represents the containers of the system.

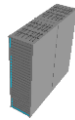


Figure 4: Symbol of the Container entity

3.2. Models

In this subsection, the developed models will be explained. The first developed model is the GoTomilk run. Its only purpose is to transfer a Picker to the riding station of the respective milk run. This way, the Picker will seem to be riding the milk run, while the Containers will stay at the wagon of the milk run. Figure 5 presents the external view of the model



Figure 5: External view of the GoTomilk run model

To model the milk run stopping; the Picker leaving the milk run to collect containers, as well as the return of the Picker and to evaluate if the Picker needs to return to the Channels or not - the picker collects 1 container at a time – the StopPlace model was developed. The properties defined for this model were:

- **Place:** Numeric property that works as an identifier number of the instances of this model placed on the Supermarket.
- **Rack:** String property that identifies the Rack that this model belongs to.
- **LastOfCorridor:** Boolean property to indicate whether this model is the last of a corridor or not.
- **ConnectTo:** Object property to specify instances of this model. Used when a corridor has sets of channels on both sides.

The Facility of this model is presented on Figure 6 and its external view is presented on Figure 7.

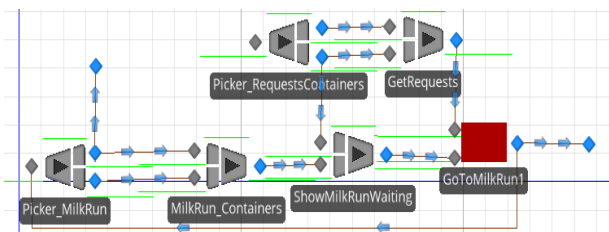


Figure 6: Facility of the StopPlace object

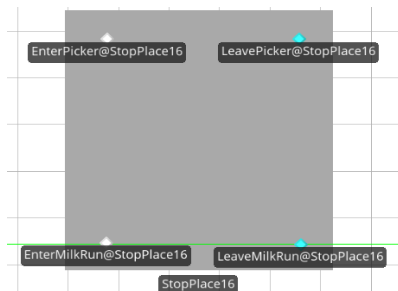


Figure 7: External view of the StopPlace model

For animation purposes, a new model – whose only purpose is to show a copy of the picker in front of the rack of channels – while the original travels inside the channel to collect the intended containers, was developed. Moreover, had this model not been developed and, after entering the Channel, the Picker would disappear for some time, before returning with the selected Container. Figure 8 displays the external view of this model.



Figure 8: External view of the StopPlace_Channel model

To store containers and model the behaviour of the Pickers, when they analyse a channel to select the container they want, the channel model was developed. The external view of this model is presented on Figure 9. The properties defined for this model were:

- **Position:** Numeric property that works as an identifier number of the instances of this model placed on the Supermarket.
- **TotalProducts:** Expression property that indicates the number of types of products to be modelled.
- **StockPolicy:** Expression property. This property indicates the stock strategy to be modelled. Since the restock process is not yet modelled, the containers are being created inside each Channel. Thus, this property indicates if the type of each container being created should be in accordance to the channel (dedicated storage) or not.
- **StopPlace:** Numeric property. The value of this property must be equal to the Place property of the StopPlace that allowed the Picker to reach this model.

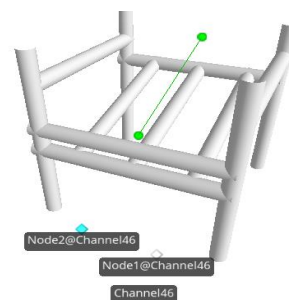


Figure 9: External view of the Channel model

All the presented models compose the supermarket itself. The properties of the supermarket model are:

- **Numbermilk runs:** Expression property that indicates the number of milk runs and Pickers to be modelled.
- **StockPolicy:** Expression property. It indicates, to all instances of the Channel model, the storage strategy being used.
- **NumberRequests:** Expression property that defines the way the Pickers add Requests to their batches.

- **TotalProducts:** Expression property that indicates, to all instances of the Channel model, the number of types of products to be modelled.
- **RequestsIntensity:** Expression property that defines the average Interarrival time of Requests to the system.

To access all the Channel and StopPlace objects placed on the Supermarket – among other reasons, to be able to consult the containers stored in each channel - two data tables were developed: one to gather all the channels and another to gather all the StopPlace models. Each object occupies an index of the data table correspondent to its Place or Position property, respectively. Additionally, a Corridors data table was created to gather all the information of all the corridors. Figure 10 shows an example of a Corridors data table.

Corridors			
	Corridor ID	Number Of Ways	Image Index_Rotation Angle
▶ 1	1	2	1
2	2	2	0
3	3	1	0
4	4	1	0

Figure 10: Corridors data table

As can be seen, the data table holds information relative to the rotation angle of the corridors, the number of ways and the identifier numbers. The former will not be analysed on this paper.

After being created, Pickers and milk runs will travel through the supermarket in a series of picking shifts, by executing a set of processes. Figure 11 shows one of such processes.

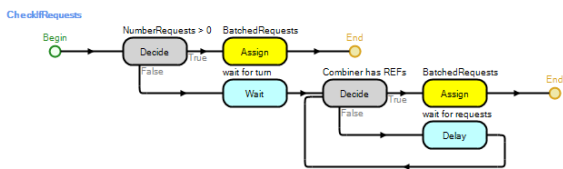


Figure 11: Process CheckIfRequests

The goals of this process are to make the picker wait for its turn and to specify the quantity of Requests to be allocated. If the property has a negative value, the Picker waits an amount of time (in minutes) equal to the module of that value. In this case, the number of Requests that are on the Combiner object, is saved to the BatchedRequests state. This way, when the Picker enters the Combiner object itself, that number of Requests are added to its batch. On the other hand, if the value is positive, the associated token will save that number to the BatchedRequests state of the Picker. Once on the Combiner, it will wait the time needed for that amount of Requests to be added to its batch. After the batch is formed, an associated will execute the process displayed on Figure 12.

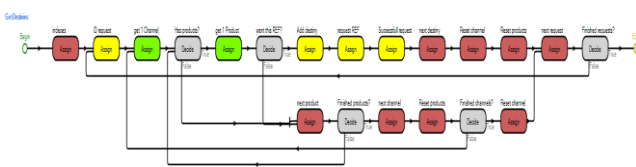


Figure 12: Process GetDestinies

The goal of this process is to save all the Channels that have the Containers correspondent to the Requests added, on an object array of the Picker. This way, each Picker has its own array of destinies. When analysing the Requests, the token saves the number identifier of the Picker on the state Requested of each Request. By doing so, it is ensured that there will be no exchanges of Requests during a picking shift. Lastly, when analysing each Container, the token also saves the ID of the Picker on their Requested state. This way, since the Containers are requested, it is ensured that the destinies of the Picker are the right ones and that no other Picker will take the Container requested.

Once the process ends, the Picker enters the GoTomilk run object, where the corresponding milk run is. In this object the Picker will be transferred to the riding station of the milk run. Additionally, on the Process property of this object the value GetStopPlaces is inserted, i.e., the milk run will have an associated token execute that process. Figure 13 shows the process GetStopPlaces. Similarly, to the process GetDestinies, this intends to save the StopPlaces where the milk run needs to enter, to an array of objects of each milk run. To that end, the StopPlace, with a value on the Place property equal to the value of the StopPlace property of the Channel on the array of destinies of the Picker, will be added. It should be noted that no repeated objects are added. Once the process ends, the milk run leaves the object and initiates its picking shift.

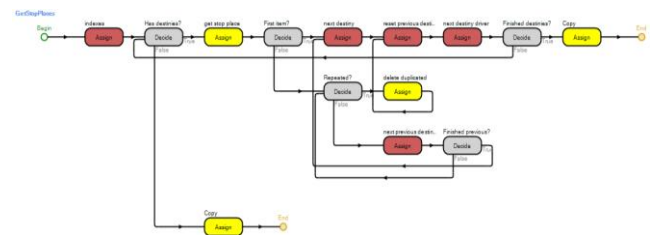


Figure 13: Process GetStopPlaces

The milk runs can travel through two types of corridors. In the first, they only have access to corridors of racks on one side of the corridor. In the second type, they have access to corridors of racks on both sides. Figure 14 and Figure 15 display examples of corridors of type 1 and 2, respectively. It should be noted that the placement of the objects this way only intends to make it simpler to understand the way the corridors work.

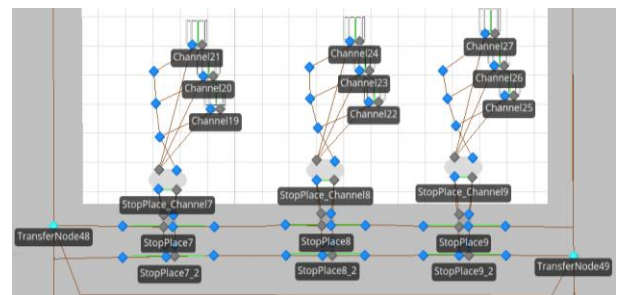


Figure 14: Example of a corridor of type 1

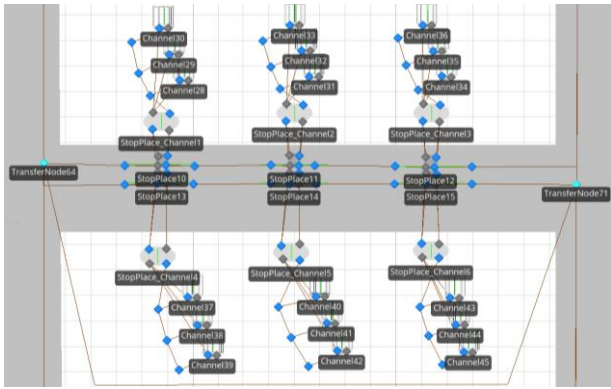


Figure 15: Example of a corridor of type 2

As can be seen, regardless of the type of corridor, both have two entry nodes. When the milk run enters one of the entry nodes of a corridor, it executes the process represented on Figure 16.

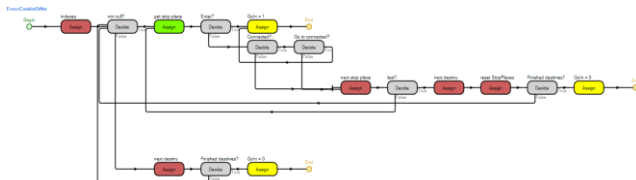


Figure 16: Process EnterCorridorOrNot

This process intends to evaluate if a milk run enter a corridor, or not, by evaluating if a StopPlace of that corridor belongs to the array of destinies of the milk run, or not. Since the token needs to know if the StopPlace being evaluated is the last of the corridor, the property LastOfCorridor needs to be checked. Lastly, in the possibility of being a type 2 corridor the ConnectTo property of all StopPlaces needs to be analysed. If there is an object specified on that property, the token also needs to evaluate if it exists on the array of destinies of the associated milk run.

Once inside a corridor, the milk runs enter a succession of StopPlaces, where each one accesses a different set of Channels. To better understand the objects that need to be used on a corridor of type 1, Figure 18 was created. Once again, the placement of the objects the way the figure displays only intends to make it simpler to understand the way they work and thus, it is not the final result of the animation of the model.

As can be seen, for any circulation direction, there is a TransferNode before and another after a StopPlace. Thus, on the first node, the process illustrated on Figure 17 is executed.

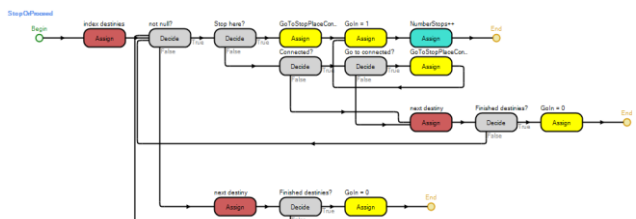


Figure 17: Process StopOrProceed

In this process, the milk run checks if the StopPlace belongs to the array of destinies of the milk run. Once the process ends, the milk run will select the Path based on the value on a state of the picker. For instance, considering that a milk run enters the TransferNode42 (from Figure 18) and executes the StopOrProceed process, if its GoIn state has the value 1, the milk run will select the Path that takes it to StopPlace8. Otherwise, it will choose the Path that takes it directly to the TransferNode43.

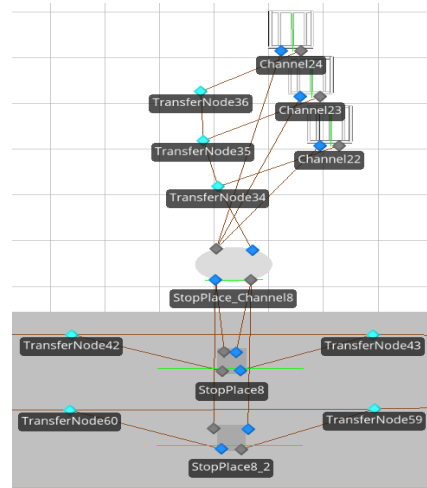


Figure 18: Objects used alongside a StopPlace and the corresponding set of Channels on a corridor of type 1

When a milk run enters a StopPlace, it will wait for the respective Picker to return from the set of Channels. In this context, if the corridor is of type 1 (e.g. Figure 18), the Picker will choose the Path that takes it to StopPlace_Channel8. However, if the corridor is of type 2, the Picker will choose its destiny based on the value on its GoToStopPlaceConnected. To help clarify this situation, Figure 19 was created.

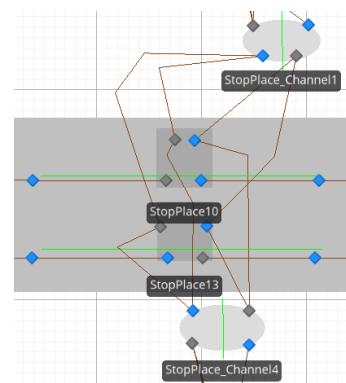


Figure 19: Pair of StopPlaces of a corridor of type 2

The Picker always returns to the StopPlace where its milk run is waiting. As soon as the Picker enters a StopPlace_Channel object, the remaining logic until it returns to it, is the same for both types of corridors.

Considering Figure 18 again, it is possible to see that there is only one TransferNode that gives access to a Channel (e.g. TransferNode 34 to Channel22 and TransferNode35 to Channel23 and) and, after leaving a Channel, the Picker will necessarily return to the

StopPlace_Channel object, i.e., it can only take one Container at a time. When a Picker enters a node that gives access to a channel, it executes the process presented by Figure 20.

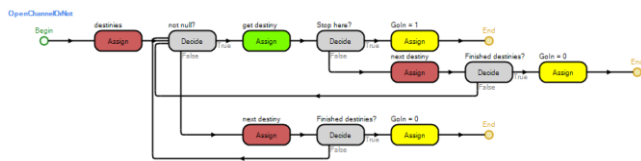


Figure 20: Process OpenChannelOrNot

The purpose of this process is to evaluate if the Channel in question belongs to the array of destinies of the Picker. To that end, the associated token consults the Channels data table on the index returned by the previously updated target state of the Picker and evaluates whether the returned object is one of the destinies of the Picker or not. If it is a destiny, the token assigns the value 1 to the GoIn state of the Picker, otherwise, the value 0. Afterwards, the Picker selects the next Path, based on the value of its GoIn state. Thus, if the value is 1, it selects the Path that takes it to the Channel. Conversely, if the value is 0, it selects the Path that takes it to the next TransferNode, updating its target state again, once on this Path. It should be noted that, since the milk run only enters a StopPlace if any of the Channels (that the StopPlace gives access to) is a destiny of the Picker, it is guaranteed that the Picker will at least enter one Channel.

Once inside a Channel object, the Picker selects the required Container and adds it to its batch. After leaving the Channel, the object is removed from its array of destinies. Then, the Picker returns to the StopPlace_Channel and, after that, to the StopPlace.

Naturally, before leaving the StopPlace object, the Picker needs to be transferred to the riding station of its milk run. Therefore, on the Facility of the StopPlace, there was the need to use a GoTomilk run object (Figure 6). On its Process property, the name of the process displayed by Figure 21 is inserted.

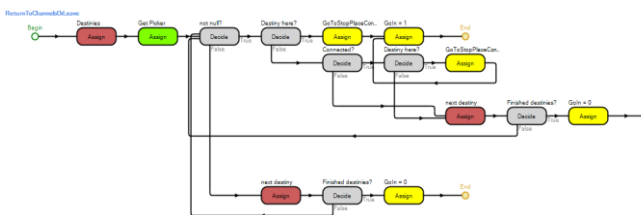


Figure 21: Process ReturnToChannelsOrLeave

The purpose of this process is to verify if the Picker needs to return to the set of Channels or not. To that end, the associated token verifies the StopPlace property of every Channel on its array of destinies. If any of those properties has a value equal to the Place property of the StopPlace where the Picker is at, the token saves the value 1 to the GoIn state of the Picker. Otherwise, it saves the value 0. Additionally, if the StopPlace has an object on its ConnectTo property, the token needs to repeat the verification to that object. This way, the GoToStopPlaceConnected state of the Picker will be

updated, to ensure that the Picker chooses the Path that takes it to the correct StopPlace_Channel.

Once the Picker has placed all the required Containers on the batch of the milk run, the latter removes the StopPlace from its array of destinies and resumes its route. When all the Containers from all the corridors have been collected, the milk run returns to the start point to restart a new shift.

3.3. Automatic Creation of Simulation Models

To make it simpler for the user to introduce the data related to the warehouse he wants to create, it was established that he would only have to introduce the data on an Excel spreadsheet. Table 1 shows an example of the content of the mentioned file and in this section the cells that the user needs to fill will be covered.

Table 1: Input Excel table

New corridor?	Size			Coordinates		Symbol index	Directions	Rack description	Channels per column								
	Length	Width	Height	x	y (z in Simio)				1	2	3	4	5	6	7	8	
1	0,23	0,42	0,58	-50	-50	0	2	AP	3	3	3	3	3	3	3	3	3
0								AO	3	3	3	3	3	3	3	3	3
0								AN	3	3	3	3	3	3	3	3	3
0								AM	3	3	3	3	3	3	3	3	3
0								AL	3	3	3	3	3	3	3	3	3
0								AJ	3	3	3	3	3	3	3	3	3
0								AK	3	3	3	3	3	3	3	3	3
0								AI	3	3	3	3	3	3	3	3	3
0								AH	3	3	3	3	3	3	3	3	3
0								AG	3	3	3	3	3	3	3	3	3
0								AF	3	3	3	3	3	3	3	3	3
0								AE	3	3	3	3	3	3	3	3	3
0								AD	3	3	3	3	3	3	3	3	3
0								AC	3	3	3	3	3	3	3	3	3
0								AB	3	3	3	3	3	3	3	3	3
0								AA	3	3	3	3	3	3	3	3	3
2								BE	3	3	3	3	3	3	3	3	3
0								BD	3	3	3	3	3	3	3	3	3
0								BC	3	3	3	3	3	3	3	3	3
0								BB	3	3	3	3	3	3	3	3	3
0								BA	3	3	3	3	3	3	3	3	3
0								AZ	3	3	3	3	3	3	3	3	3
0								AY	3	3	3	3	3	3	3	3	3
0								AX	3	3	3	3	3	3	3	3	3
0								AW	3	3	3	3	3	3	3	3	3
0								AV	3	3	3	3	3	3	3	3	3
0								AU	3	3	3	3	3	3	3	3	3
0								AT	3	3	3	3	3	3	3	3	3
0								AS	3	3	3	3	3	3	3	3	3
0								AR	3	3	3	3	3	3	3	3	3
0								AQ	3	3	3	3	3	3	3	3	3

In order to allow the user to specify any number of racks per corridor, it was established that on each line of the excel file, the user inserts data related to a single rack. Therefore, to start a new corridor, the user has to enter the value “1” on the column “New corridor?”. Conversely, if the user wants to keep adding racks to a corridor, he just has to keep entering the value “0” on the corresponding rows, on the same column. Additionally, for each corridor, the user can choose one of two types: a simple corridor, which is comprised by one or more racks; and a set of two corridors that are disposed inwards, so that a milk run traveling it may collect containers from both corridors of its left and right. To make it simpler to refer to these corridors, on the remaining sections of this document, these will be referred as simple and double, respectively. In this sense, to specify a double corridor, the user needs to assign the value “2” to the row corresponding to its first rack.

In the columns “Size” and “Coordinates”, the user can specify the size of the channels (length, width and

height) and the position on which the corridors starts to be built. These values are only read if the user entered the value “1” on the “New corridor?” column of that row, since it was assumed that this information does not vary in the same corridor. The same approach applies for the “Symbol index” and “Directions” columns. On the first, the user can specify a symbol, from an array of symbols, to be assigned to the channel. The only difference between the symbols on this array is its rotation angles. This approach had to be considered, since the API of Simio does not provide methods for rotating a fixed object and, for animation purposes, it was very important to rotate the corridors and its channels. However, this approach has a couple of flaws. Firstly, since the waiting queue of the object is not considered part of the symbol, it is not “rotated”, i.e., despite the fact that a different symbol is assigned to an object, its queue remains with the rotation as the original. Lastly, the possible rotation angles have to be previously assigned. For this case, rotations of 45 degrees were considered (e.g. 1 means a rotation of 45 degrees, 2 means a rotation of 90 degrees and so on). On the “Directions” column the user can define the number of ways through which the milk runs can travel on the corridor. On the last column, “Channels per column”, the user can define any number of columns per rack and any number of channels per column, depending on the number of cells that have values and the values on each of those cells, respectively. On the “Rack description” column, the user can specify a string that, as the name implies, indicates the rack description of the rack in question.

To create an object using the Simio API the user needs to call the `CreateObject` method. This method takes a string and a `FacilityLocation` as arguments. The later defines the coordinates x, y and z in Simio and the first is the name of the object that is supposed to be created on the specified location. This object can be any one of the Standard library of Simio, any other created by a user (e.g. a sub model) or even the object that represents an entity or a worker. Thus, to create the developed Simio sub-models, which have already been discussed [6], this method is used. Notwithstanding, to create a path, a conveyor, a time path or a connector between objects a different method is used, even though these are also objects in Simio. In these cases, the method `CreateLink` has to be used. Examples of both methods are given below:

```
createObject("Channel",new FacilityLocation (xx,yy,zz)) as IFixedObject;
CreateLink(String, INodeObject, INodeObject, IEnumerable<FacilityLocation>)
```

As can be seen, this method takes a string, two `INodeObjects` and a collection of `FacilityLocations` as arguments. The first corresponds to the object being created, while the following two arguments correspond to the two nodes the method is supposed to connect. Lastly, the collection of `FacilityLocations` is a list of coordinates used to create the vertexes of the object. If the user does not want to specify any vertexes, the value `null` can be passed through this argument.

Apart from creating objects, the Simio API may also be useful for other reasons, such as editing object properties. In many cases, to accomplish this, it is necessary to know the name of the property and use the following code line:

```
Object.Properties["PropertyName"].Value="NewValue";
```

However, there are some properties that require other means to edit them, like the name of the object, its size, symbol index, location, among others. Nonetheless, knowing the name of the property in question is not always a simple task, due to the lack of information concerning the Simio API available. In fact, when a user interacts with the tool and edits an object property, the name presented by Simio for that property is actually the display name. To confirm this situation Figure 2 shows the properties inherited by an object of the standard library of Simio.

As can be seen, the name of the selected property is “EnteredAddOnProcess”, while its display name is “Entered”. Thus, to learn the name of this property, the user would have to access the list of properties of the object and check its name, which is very troublesome.

To create different orientations for the corridors, or simply to create two corridors faced inwards, composing a single corridor, it would be necessary to use a Simio method that could rotate an object, just like it is possible to do when interacting with the tool itself. However, the API does not provide any method for this task, so other workarounds were considered. The solution adopted for this task was to assign different symbols to the objects, each one representing a different rotation angle. Nonetheless, this does not affect the queue of the objects. This fact can be seen on Figure 6 and on Figure 7 (chapter 4), where all the queues, of all the channels, of the two faced inwards corridors, are facing the same direction. Thus, the queues of the channels on the second set of channels are facing an opposite direction to where the pickers and the milk runs travel.

When the add-in starts its execution, all the data is read from the excel spreadsheet to avoid having to make multiple communications with the application. The method created to that end is given below.

```
public string[,] getData(int firstRow){
    Excel.Application app=new Excel.Application();
    Excel.Workbook workbook = app.Workbooks.Open("C:\\Path\\file.xlsx");
    Excel.Worksheet sheet = (Excel.Worksheet) workbook.Worksheets.get_Item(1);
    Excel.Range range = sheet.UsedRange;
    string[,] data = new string [range.Rows.CurrentRegion.EntireRow.Count -
    firstRow+1,range.Columns.CurrentRegion.EntireColumn.Count];
    for(int i=firstRow;i<= range.Rows.CurrentRegion.EntireRow.Count;i++){
        for(int j=1;j<=(range.Rows[i] as Excel.Range).CurrentRegion.EntireColumn.Count; j++){
            if((range.Cells[i,j] as Excel.Range).Value2!=null)
                data[i-firstRow,j-1]=(range.Cells[i,j] as Excel.Range).Value2.ToString();
            else data[i- firstRow,j-1]="NoData";
        }
    }
    workbook.Close();
    excelApp.Quit();
    return data;
}
```

As can be seen, the variable `app` is used to start Excel. Afterwards, the `workbook` variable opens the

intended excel file, by providing it with the correct path. Lastly, the sheet variable accesses the pretended worksheet (the first of the opened workbook) and the range variable gets the range currently being used. At this point, to read data from a cell of the opened sheet, it was necessary to use the following expression:

```
(range.Cells[i, j] as Excel.Range).Value2.ToString()
```

As the purpose of this method is to save the data contained on the excel sheet to a multidimensional array, the remaining code lines search through the cells with content and saves its string value to the respective position on the array to be returned. Once all the data is read, the communication with Excel can be terminated.

After retrieving the data, the add-in can start building the supermarket. In this section, the code for this task will be explained as pseudo-code, given below.

```
Executes getData() method
For each row read from the excel file{
  If data[countn, 0] = 1{
    Read size of the Channels
    Read coordinates, symbol index (rotation based on this) and number of ways of corridor
    CorridorData + Execute GetCorridorData() method
    do for i = 0 to 1 times{//One for each Position of the CorridorData array
      do for each list RACK in in CorridorData[i]{
        do for each node in RACK{
          if first node of RACK then save the rack Description
          else {
            creates StopPlaces and edits their properties
            create other needed objects and edits their properties
            if last StopPlace then draw Paths between corridors and edits their properties
            ChannelsToDraw + read the number of Channels accessible by the drawn StopPlaces
            do for j = 0 to ChannelsToDraw{
              if first loop then{
                Create StopPlace_Channel
                create other needed objects and edits their properties
              }
              else{
                create Channel and edit the properties
                create other needed objects and edits their properties
              }
            }
          }
        }
      }
    }
  }
}
Add objects to their respective object tables
}
```

As can be seen, the algorithm runs through the retrieved multidimensional array of strings, with the contents retrieved from the excel spreadsheet, and searches for the value “1” on the first column of every row it searches. Once it finds it, executes the GetCorridorData method, which is displayed below.

```
public List<List<string>>[] GetCorridorData(int n, string[,] data){
  List<List<string>>[] corridor = new List<List<string>>[2];
  int j = 0, flag = 1, foundDouble = 0;
  List<string> channels = new List<string>();
  List<List<string>> racks0 = new List<List<string>>(), racks1 = new List<List<string>>();
  while ((n + 1) <= data.GetLength(0) && ((data[n, 0] != "1" || (flag == 1))) {
    flag = 0;
    j = 5;
    while ((j < data.GetLength(1)) && (data[n, j] != "NoData")){
      channels.Add(data[n, j]);
      j++;
    }
    if(data[n,0]=="2")
      foundDouble = 1;
    if (foundDouble == 0)
      racks0.Add(channels);
    else
      racks1.Add(channels);
    channels = new List<string>();
    n++;
  }
  corridor[0] = racks0;
  corridor[1] = racks1;
  return corridor;
}
```

The purpose of this method is to get all the information related to a corridor and store it on a single data structure. This method had to be used, since the way defined to build a simple corridor is different from the way defined to build a double one. Moreover, to make it simple for the user to introduce data on the excel spreadsheet, he only needs to assign the value “2” on the first rack of the second corridor of the double corridor. Thus, to know if the corridor in question is a simple one or a double one, it is necessary to read all the rows belonging to the same corridor.

To store the data related to a corridor, the authors defined an array with only two positions of lists of lists of strings. The strings are the data retrieved from the excel spreadsheet, while the list of strings (channels in the code given above) stores the data related to the number of channels to create, per rack (values of the column “Channels per column” of Table 1). All the information related to racks belonging to the same corridor is stored on the remaining list (racks0 on the code above). Nonetheless, if the value “2” is found, the values are saved on a different list (racks1 in the code above.). After running through all the rows of a corridor, the two lists are saved on the respective array positions, and the final data structure is returned. Once again, considering the data on Table 1, the data structures resulted from executing the GetCorridorData for the first corridor is illustrated on Figure 3.

4. DATA INPUT AND VALIDATION

The system being modelled consists on an advanced supermarket, **supermarket**, which is located near the production lines and stores a number of containers in each **channel** (shelf-like structure that stores containers in depth, usually in FIFO order). In its turn, each container can store several product units of a single type. Containers are sent to the supermarket, for later being collected by pickers that travel through the supermarket, driving milk runs. After collecting the intended containers, the pickers deliver them to the respective production lines. These consume the required material and, when it is necessary to start consuming a different type of product, a reference change occurs. In some cases, this phenomenon can result on a container being returned to the supermarket with the leftover product units inside of it.

Through many meetings, the authors were able to obtain the data required for the simulation model to efficiently model the system in analysis. Among others, values for the speed of the milk run, the picker, devolution rates, production times, number of shifts per day, number of production lines, time to remove containers from their channels and others were collected. This process is important, since it increases the confidence level in the developed model.

To build a supermarket with a layout corresponding to the real one, the authors used the developed Simio add-in that automatically creates simulation models ready to perform simulation experiments (Vieira et al., 2015b).

This add-in receives data input from excel spreadsheets, in which the user can specify any physical layout.

The created supermarket corresponds to a single corridor of two sets of channels that can be accessed by a picker who travels in between them. This supermarket corresponds to a size of 930 channels. Each channel has the capacity to hold 6 containers. More examples of automatically created simulation models, using the Simio add-in developed for this project, can be found online (Vieira et al., 2015b).

After analysing the raw data provided by the company, the authors were able to produce, using VBA, the required excel files that would “feed” the developed simulation model with real data. To do so, this data was imported into Simio. Figure 22 shows a diagram that represents the mentioned data and its purpose in the system in analysis.

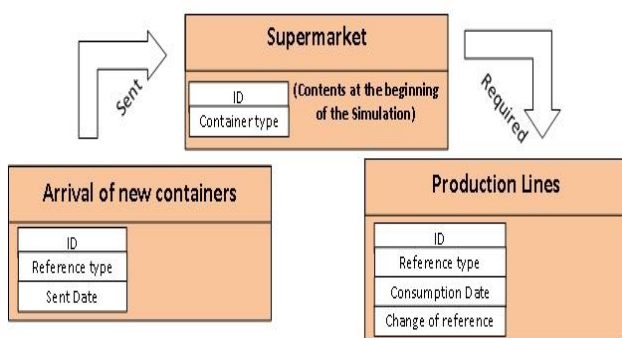


Figure 22: Representation of the system being analysed

Each of the 3 big areas, displayed in the diagram, represent a different excel spreadsheet imported into Simio. The data contained in those excel files is described inside each area. To specify containers that are located in the supermarket, at the beginning of the simulation (Supermarket area at the centre of the diagram), the excel files contain a unique identifier of each container and the correspondent type of that container. Similarly, to specify containers that are sent into the supermarket, the excel file must contain the unique identifier of each container, the type of the referred container and the date on which it is supposed to be sent into the supermarket. Lastly, the excel files that describe the containers required by the production lines must contain the unique identifier of each container, its type, the date on which it is required and a Boolean that indicates whether each container will originate, or not, a reference change. This data is important to indicate if a specific container has a chance of being returned to the supermarket, or not.

5. SIMULATION RESULTS AND ANALYSIS

In this section, the modelled storage strategies will be introduced. Many references to these strategies will be made throughout the paper. Therefore, the authors assigned short names to them that can be consulted in Table 2.

Table 2: Storage strategies definition

Short name	Storage Strategy
A	<ul style="list-style-type: none"> • Single-product channels;
C	<ul style="list-style-type: none"> • Multi-product channels; • Driven by consumption;

Strategy A corresponds to the one that is currently being used on the supermarket of the case study, on which the channels are dedicated (**single-product**). This is the most simple case, since it consists on having channels dedicated to a single type of container (Bartholdi and Hackman, 2008). One of its great advantages, resides on the fact that, since the locations of the containers do not change, pickers can memorize them, making the picking process more efficient (Bartholdi and Hackman, 2008). Moreover, it should be expected that single-product strategies require a higher quantity of channels to work, since it does not store different types of product on the same channel. In other words, the problem with this strategy is that “**it does not use space efficiently**”. In fact, it is expected that, on average, the storage capacity is about 50%” (Bartholdi and Hackman, 2008), which represents a high amount of costs associated. To overcome this problem, other strategies can be considered. Figure 23 displays the simulation running, while modelling a single-product storage strategy, where different colours were assigned for each type of container. As can be seen, all containers stored within the same channel have the same colour.

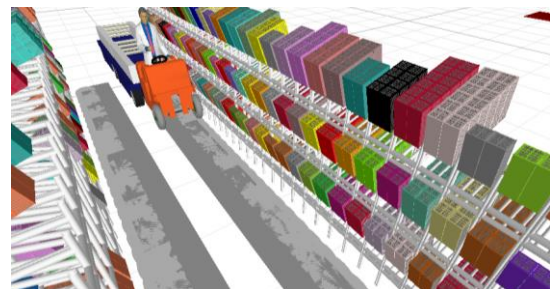


Figure 23: Single product storage strategy

Alternatives to strategy A would have to allow containers of different types to be mixed within the same channel (**multi-product**), whereby some companies oppose to its implementation. The main reason for this is that the Information System (**IS**) would have to be much more complex, to avoid picking from the non-first position of a channel and to guide pickers to the proper channel (and possibly also to the right position), once they would no longer have the advantage of having memorized the location of the containers. In a situation where the IS cannot handle this issue, the pickers would have to search for the container through all the positions of all the channels of the supermarket, which would negatively affect the picking system. Considering the afore mentioned, the proposed alternatives for the company have to analyse if, by **allowing mixes** of containers of different types on each channel, the picking of the containers will always be made in the **first position** on each channel. In this sense, **strategy C**

consists on storing containers, based on their consumption date, by giving priority to the channels that already have containers of the same type. Thus, a container can be stored in a channel with containers of other types, as long as the container to be stored has a posterior consumption date. Additionally, and similarly to the previous storage policy, an established limit of the number of containers of different types allowed per channel has to be respected. Since consumption dates are exposed to prediction errors, the impact of eventual errors has to be made. Figure 24 shows the simulation model in execution, while modelling a multi-product storage strategy. As can be seen, in this case, containers of different colour can be seen within a same channel.

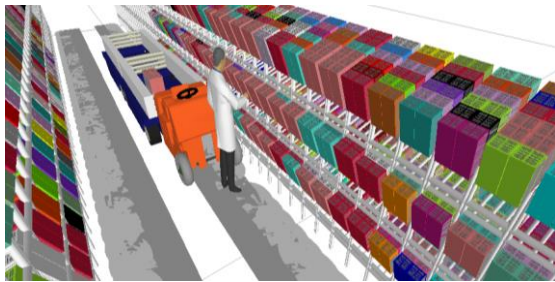


Figure 24: Multi-product storage strategy

Some simulation scenarios were defined for this problem, however only some of them will be presented in this chapter. The conducted experiments were run with a simulation time of one week. It should be noted, however, that the results presented within this chapter are directly dependent to the input data referred at chapter **Erro! A origem da referência não foi encontrada.** As such, some of the conclusions withdrawn from this comparison should not be generalized, since it corresponds to a specific studied case.

The simulation experiments conducted in Simio considered several performance indicators: Nonetheless, the most important **KPI** (Key Performance Indicators) considered were: the average total time spent in a picking shift in seconds, the average position from which containers are removed from the channels (depth), the total amount of channels that were never used throughout the simulation and the average number of stops per milk run per picking shift. In its turn, the error assigned to the prediction errors of the consumption dates and the time interval of different types of containers in each channel were defined as the **properties** of the simulation experiments.

In an attempt to quantify the different simulation scenarios, rather than using an explicit **multi-criteria** approach, weights were assigned to the four KPI, to define a score that considers all KPI values of all scenarios. Thus, and taking into account the main objective of the company, the weights 3 and 2 were respectively assigned to the number of channels not used and depth KPI. The remaining KPI had a weight of 1.

Currently, the storage strategy being used at the company of the case study is the single-product one. This strategy is the one that has the lower number of properties that can be changed in our simulation model, since it

allows a single type of container per channel. Table 3 shows the obtained results for this strategy.

As the results indicate, the pickers always collected the containers from the first position (depth), which is one of the perks of using this strategy. However, this affects the number of channels that were not used, which is lower than the same KPI on the remaining strategies, as will be shown in the next sections. Table 4 shows the obtained results for this strategy.

Scenario	Strategy	Time gap	Error	Milkruns	Different types of containers	Total time	Number of stops	Unused channels	Depth	Global Classification
4	A	0	Random.Uniform(0,0)	4	1	243,7	3,90	148	1,000	45%

Table 3: Simulation results for the modelled strategy A

Scenario	Strategy	Time gap	Error	Milkruns	Different types of containers	Total time	Number of stops	Unused channels	Depth	Global Classification
73	C	0	Random.Uniform(0,0)	4	6	214,9	1,72	473	1,000	91%
78	C	6	Random.Uniform(0,0)	4	6	215,6	1,75	428	1,000	87%
79	C	12	Random.Uniform(0,0)	4	6	215,0	1,73	403	1,000	85%
80	C	24	Random.Uniform(0,0)	4	6	215,2	1,73	401	1,000	84%
81	C	48	Random.Uniform(0,0)	4	6	215,9	1,78	357	1,000	80%
82	C	0	Random.Uniform(-01,01)	4	6	219,4	1,73	472	1,044	87%
83	C	6	Random.Uniform(-01,01)	4	6	215,3	1,74	425	1,000	87%
84	C	0	Random.Uniform(-12,12)	4	6	239,2	1,71	467	1,242	68%
85	C	6	Random.Uniform(-12,12)	4	6	222,7	1,73	422	1,075	79%
86	C	12	Random.Uniform(-12,12)	4	6	218,1	1,77	396	1,024	81%
87	C	24	Random.Uniform(-12,12)	4	6	216,3	1,81	369	1,000	81%
88	C	0	Random.Uniform(-24,24)	4	6	246,5	1,75	461	1,307	60%
89	C	12	Random.Uniform(-24,24)	4	6	221,8	1,78	395	1,058	78%
90	C	24	Random.Uniform(-24,24)	4	6	218,0	1,82	369	1,015	79%
91	C	48	Random.Uniform(-24,24)	4	6	216,1	1,80	360	1,000	80%
92	C	0	Random.Uniform(-48,48)	4	6	251,0	1,80	441	1,345	55%
93	C	24	Random.Uniform(-48,48)	4	6	225,4	1,84	370	1,084	73%
94	C	48	Random.Uniform(-48,48)	4	6	219,4	1,84	336	1,025	75%
95	C	72	Random.Uniform(-48,48)	4	6	217,0	1,81	349	1,006	78%
96	C	96	Random.Uniform(-48,48)	4	6	216,1	1,79	351	1,000	79%
97	C	0	Random.Uniform(-72,72)	4	6	253,5	1,82	424	1,365	51%
98	C	24	Random.Uniform(-72,72)	4	6	231,8	1,83	364	1,150	66%
99	C	48	Random.Uniform(-72,72)	4	6	223,1	1,83	337	1,065	71%
100	C	72	Random.Uniform(-72,72)	4	6	219,7	1,83	337	1,031	75%
101	C	96	Random.Uniform(-72,72)	4	6	218,1	1,84	346	1,013	77%
102	C	120	Random.Uniform(-72,72)	4	6	216,2	1,79	357	1,002	80%
103	C	144	Random.Uniform(-72,72)	4	6	215,7	1,77	356	1,000	80%

Table 4: Simulation results for the modelled strategy C

When analysing these results, the first thing to consider is that, similarly to the previous strategy, to significantly affect the system, the gap should be in the order of the days, rather than hours. Another aspect that should be noted is that the gap between containers of different types stored in the same channel – only for the scenarios without prediction errors - mainly affects the number of unused channels. The consequence of this fact was already addressed in the previous storage policy. When analysing the impact of the property that defines the prediction errors, the data showed that, when the errors were lower than the interval gaps, the depth values were always equal to 1 and the average picking time decreased

Assuming that a company can accurately predict the consumption date of their containers, scenario 73 (global score of 91%) can be considered the best solution. In

comparison to scenario 4 (the policy current being used at the case study), scenario 73 corresponds to a reduction of roughly **30 seconds** per trip (12% of reduction) on the average time per picking shift, **2 stops** per picking shift and per milk run (reduction of roughly 55%) and a reduction of around **69%** in the supermarket size (average difference of about 325 channels). All these gains were achieved by maintaining the rule stating that containers should be collected from the first position of any channel.

6. CONCLUSIONS

Warehouses are critical to a wide range of customer service activities and yet, they are also quite significant from a cost perspective. One of the goals of the Bosch Production System (BPS), implemented at Bosch, is to provide “the basis for continuous improvements in quality, costs, and supply performance” (Bosch, 2014). Thus, the opportunity to develop a **micro simulation** model in **Simio** that could help the Bosch Car Multimedia Portugal in Ferreiros, Braga arose. Particularly, this tool needs to be able to design several layouts of the supermarket and use them to test different scenarios of their **picking system**. In this on-going work, the present paper documents what was done to model the picking system observed at the Bosch Car Multimedia Portugal.

With the developed model, practitioners may benefit by using it to model different types of **warehouses**, not only supermarkets. Since the simulation model can be automatically created, the user only needs to insert the data correspondent to the layout and generate the intended simulation model. Afterwards, the model can be used to test different scenarios for the warehouse. Researchers may also benefit from the tool by using it to simulate different types of warehouses. The quality of the **animation** is quite perceptive, as the several figures illustrated throughout the document suggest.

Nonetheless, while interacting with Simio, some downsides were noted. Vieira et al. had already stated some of them (Vieira et al., 2014b). Moreover, the very useful expression editor feature that Simio offers, is not always enabled. For instance, on an Assign step, to define the StateVariableName property, the user can only select the state from a limited list of options. While it is true that it keeps it simpler for new users, it is also troublesome to have to use the expression editor where it is enabled to write a complex expression and then copy it to the actual place we want to use it. This is also true for other properties such as the StationName property of a Transfer step.

ACKNOWLEDGMENTS

This work has been co-supported by SI I&DT project in joint-promotion nº 36265/2013 (HMIEXCEL - 2013-2015 Project) and by FCT – *Fundação para a Ciência e Tecnologia* in the scope of the project: PEst-OE/EEI/UI0319/2014.

REFERENCES

- BAKER, P. & CANESSA, M. 2009. Warehouse design: A structured approach. *European Journal of Operational Research*, 193, 425-436.
- BARTHOLDI, J. J. & HACKMAN, S. T. 2008. *Warehouse & Distribution Science: Release 0.89*, The Supply Chain and Logistics Institute.
- BOSCH. 2014. *consulted online at: <http://www.bosch.com/en/com/home/homepage.html>* [Online]. [Accessed].
- COSTA, B., DIAS, L. S., OLIVEIRA, J. A. & PEREIRA, G. Simulation as a tool for planning a material delivery system to manufacturing lines. Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International, 28-30 June 2008 2008. 1-5.
- COSTA, P., ALVES, A. C. & SOUSA, R. M. 2011. Implementação da metodologia Quick ChangeOver numa linha de montagem final de auto-rádios: para além da técnica SMED.
- COYLE, J. J., BARDI, E. J. & LANGLEY, C. J. 1988. *The management of business logistics*, West Pub. Co.
- DIAS, L., PEREIRA, G. & RODRIGUES, G. 2007. A Shortlist of the Most Popular Discrete Simulation Tools. *Simulation News Europe*, 17, 33-36.
- GU, J., GOETSCHALCKX, M. & MCGINNIS, L. F. 2010. Research on supermarket design and performance evaluation: A comprehensive review. *European Journal of Operational Research*, 203, 539-549.
- HLUPIC, V. Simulation software: an Operational Research Society survey of academic and industrial users. Simulation Conference, 2000. Proceedings. Winter, 2000 2000. 1676-1683 vol.2.
- HLUPIC, V. & PAUL, R. 1999. Guidelines for selection of manufacturing simulation software. *IIE Transactions*, 31, 21-29.
- MONDEN, Y. 1998. Toyota Production System – an integrated approach to Just-In-Time. Institute of Industrial Engineers, Norcross, Georgia.
- PEGDEN, C. D. Simio: A new simulation system based on intelligent objects. Simulation Conference, 2007 Winter, 9-12 Dec. 2007 2007. 2293-2300.
- PEGDEN, C. D. 2013. Intelligent objects: the future of simulation.
- PEGDEN, C. D. & STURROCK, D. T. Introduction to Simio. Proceedings - Winter Simulation Conference, 2011 Phoenix, AZ. 29-38.
- PEREIRA, G., DIAS, L., VIK, P. & OLIVEIRA, J. A. 2011. Discrete simulation tools ranking: a commercial software packages comparison based on popularity.
- STURROCK, D. T. & PEGDEN, C. D. Recent innovations in Simio. Proceedings - Winter Simulation Conference, 2010 Baltimore, MD. 21-31.
- VIEIRA, A., DIAS, L., PEREIRA, G. & OLIVEIRA, J. 2014a. COMPARISON OF SIMIO AND ARENA SIMULATION TOOLS. *ISC*. University of Skovde, Skovde, Sweden.
- VIEIRA, A., DIAS, L., PEREIRA, G. & OLIVEIRA, J. 2014b. Micro Simulation to Evaluate the Impact of Introducing Pre-Signals in Traffic Intersections. *ICCSA*. University of Minho at Guimarães - Portugal.
- VIEIRA, A., DIAS, L. S., PEREIRA, G. A. B., OLIVEIRA, J. A., CARVALHO, M. S. & MARTINS, P. 2015a. Automatic Generation of 3D Simulation Models: Warehouses Performance Boosting. *business sustainability. Póvoa de Varzim, Portugal*.
- VIEIRA, A., DIAS, L. S., PEREIRA, G. A. B., OLIVEIRA, J. A., CARVALHO, M. S. & P., M. 2015b. Using Simio to Automatically Create 3D Warehouses and Compare Different Storage Strategies. *Faculty of Mechanical Engineering Transactions*, 43, 335-343.
- WOMACK, J. P. & JONES, D. T. 1996. *Lean Thinking*. Siman & Schuster, New York, USA.
- WOMACK, J. P., JONES, D. T. & ROOS, D. 1990. *The machine that changes the world*. Rawson Associates, NY
- YILDIZ, H., RAVI, R. & FAIREY, W. 2010. Integrated optimization of customer and supplier logistics at Robert Bosch LLC. *European Journal of Operational Research*, 207, 456-464.

A SIMULATION MODEL FOR ASSIGNING SECURE WAITING AREAS ON SUBWAY PLATFORMS TO MINIMIZE ACCIDENTS

Gerardo de Jesús Portillo-Villasana^(a), Aida Huerta-Barrientos^(b)

^{(a)(b)}Department of Systems Engineering
National Autonomous Mexico University

^(a)gerardojpv@hotmail.com, ^(b)aida.huerta@comunidad.unam.mx

ABSTRACT

Due to the high levels of passenger's traffic inside of the subway stations and the uncertainty of the exact position of the convoy arrival on the platform, events such as accidents occurs every day due to the lack of secure areas assigned for passengers to wait the convoy arrival. Then, the passenger's agglomeration in the doors causes delays along the entire subway platform creating dislikes among passengers. The aim of this paper is to develop an agent-based simulation model of ZOCALO subway platform to assign secure waiting areas for passengers in order to minimize accidents. First, a conceptual model of subway platform that includes the process of ascent and descent of passengers is presented. Second, the conceptual model is implemented in computer using AnyLogicTM software. Then, a simulation scenario is designed to test accidents minimization due to the implementation of secure waiting areas on the subway platform.

Keywords: subway platforms, Mexico City, secure waiting areas, accidents.

1. INTRODUCTION

Mexico City subway is a system of public transport that runs 12 lines that connect Mexico City from North to South and from East to West. A total of 195 subway stations operate every day. In the last months, the Mexico City subway recorded in its statistical of affluence a considerable increment in the demand of passengers who prefer this public transport for arriving at their final destination in an economical way. In this direction, during 2016 the demand in the most busy subway stations has reach about 4.5 million users daily (see Table 1). Unfortunately, due to the high levels of passenger's traffic inside of the subway stations and the uncertainty of the exact position of the convoy arrival on the platform, events such as accidents occurs every day due to the lack of secure areas assigned for passengers to wait the convoy arrival. Then, the passenger's agglomeration in the doors causes delays along the entire subway platform creating dislikes among passengers (see Fig. 1). Additionally, the human suicide rate has been increasing in the subway platforms causing operational problems to the subway. One suicide for instance causes a delay that reaches almost 40 minutes and an accident 10-20 minutes.



Figure 1: Passengers waiting the convoy at subway platform in ZOCALO station

Table 1: The busiest subway stations in Mexico City during January-March 2016

Subway line	Subway station	Daily total passengers
A	Pantitlan	126,029
3	Indios Verdes	122,723
2	Cuatro Caminos	113,470
8	Constitución de 1917	98,328
3	Universidad	82,671
2	Taxqueña	81,959
1	Observatorio	77,205
1	Insurgentes	67,730
B	Ciudad Azteca	67,073
2	Zocalo	65,942
B	Buenavista	57,544
1	Chapultepec	57,171
1	Zaragoza	56,482
9	Tacubaya	52,063
9	Chilpancingo	51,452
1	Pantitlan	49,260
1	Merced	44,855
7	Barranca del Muerto	44,441

The aim of this paper is to develop an agent-based simulation model to assign secure waiting areas for passengers on the subway platforms in order to minimize accidents and reduce the doors obstructions to let a free ascent and descent of passengers. We study the case of ZOCALO station.

This paper is prepared as follows: the literature review about measures of prevention of accidents in public transport is presented in Section 2. A conceptual model of Mexico City subway platform including the process of ascent and descent of passengers is developed in Section 3. An agent-based simulation model of subway platform is implemented using AnyLogic™ software and presented in Section 4. A simulation scenario is designed to test the accidents reduction rate due to the implementation of secure waiting areas in the simulation model and the results are showed in Section 5. Concluding remarks are draw in Section 6.

2. LITERATURE REVIEW

In the case of railway, an increased proportion has been observed of suicide attempts in the last years after several days without sunlight (Kadotani, Nagai and Sozu 2014). Several authors have reported security measures such as physical barriers in public transport. One of them is the called platform screen door (PSD). Installing OSDs on train platforms is considered to be one of the most effective measures to prevent railway suicide (Ueda et al. 2015). Law et al. (2009), analyzed the effectiveness of installation of PSD in the railway in Hong Kong. In this case, over the 11-year study period, a total of 76 railway suicide cases were reported. A significant reduction of 59.9% in railway suicides was found after the PSD installation. Soeta and Shimokura (2012) argued that the PSD might affect train noises in stations, but the effects are not well understood. They carried out a study to clarify the effects of PSDs on acoustic characteristics. Their results suggest that PSDs made train noises more diffused because PSDs blocked the lower frequency components of train noises.

Other studies focused in prevention. For example, Ichikawa et al. (2014), reconsidered the effects of blue-light installation for prevention of railway suicides. They pointed out that the installation of blue lights on platforms, even were they to have some effect in preventing railway suicides at night, would have a much smaller impact than previously estimated. Of 5841 total reported suicide attempts, 43% occurred both within the station premises, 43% occurred at night, and 14% occurred both within the station premises and at night. Havarneanu et al. (2015) presented a review based on 139 relevant publications, to identify the past and current trend in the prevention of suicides and trespassing accident in railway. The need for a unified approach is outlined, and on the importance to consider the effect mechanism of the measures in order to design better interventions. Mishara and Bardon (2016) critically reviewed research on railway suicides to inform suicide prevention initiatives and future studies. They argued that most research focuses on the incidence and characteristics of events and attempters and suggested that better quality research is needed, particularly studies that investigate why people use railways to kill themselves and how railway suicides can be effectively prevented.

Chung et al. (2016) analyzed the PSDs in the subway context. Their findings were based on the data from a single operator for a limited period of time. They found that installing physical barriers at subway stations can be an effective strategy to reduce the number of subway suicides; however, the barriers should be made high enough so that nobody can climb over them.

3. THE CONCEPTUAL MODEL OF ZOCALO SUBWAY PLATFORM

The conceptual model of ZOCALO subway platform includes the process of ascent and descent of passengers and two main agents are considered: passengers and the convoys (see Fig. 2). All interactions between these two agents are on the platform subway. The ascent process of a passenger starts when he enter the subway station. Then he walks to the subway platform and select a location for waiting the arrival of a convoy. When the convoy arrives to the subway platform, it fully stop and after that, the convoy's doors are open. At this point, there are no a coordination between the ascent and descent of passengers. That means, all passengers move at the same time, ascending and descending. If a passenger do not ascent the convoy, he continues waiting on the platform. The process finish when the passenger exit the subway platform.

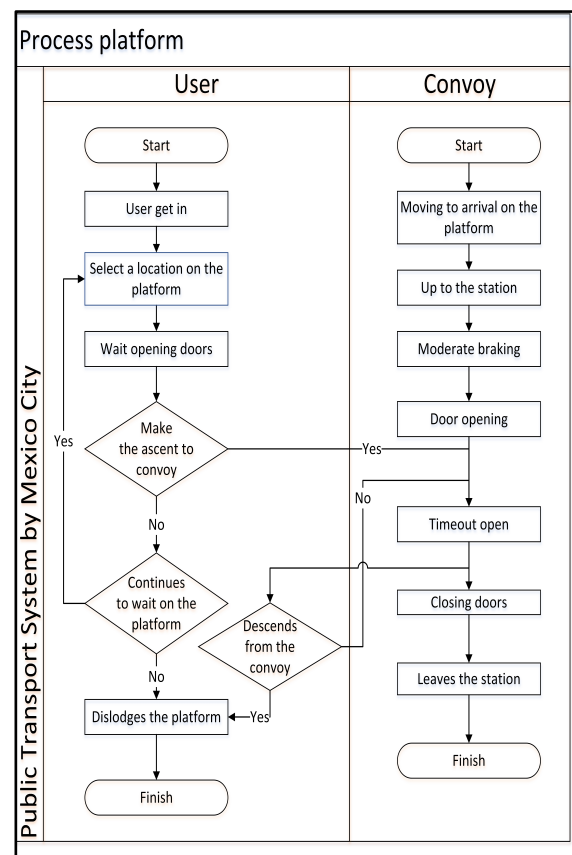


Figure 2: Flow chart of user and convoy interactions inside the subway platform

In the case of an incident or accident on the subway platform, it happens normally when a passenger exceeds the yellow safety line on the platform and he falls into the subway tracks (see Fig. 3). In this situation, the accident can be fatal if the convoy arrives at the same time of the fall. If it occurs, the body is hit by the convoy. So, the normal operation is affected until the body is removed from the tracks. Contrary, if a passenger falls into the subway tracks when the convoy does not arrive yet, it can be helped by the subway authorities.

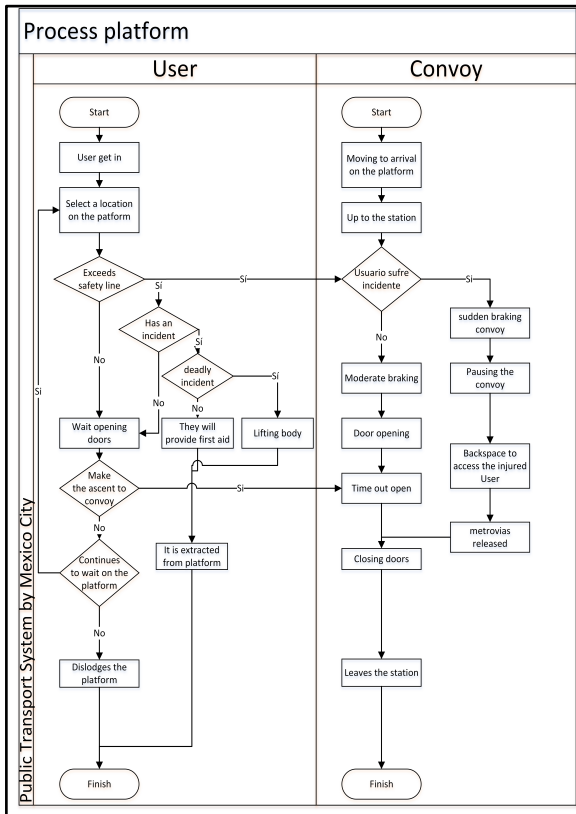


Figure 3: Flow chart for fatal incidents on the subway platform

4. THE SUBWAY PLATFORM AGENT-BASED SIMULATION MODEL

In order to analyze the dynamic of passengers in ZOCALO subway platform, we develop an agent-based simulation model (ABMS). The ABMS simulation model is implemented using AnyLogic™ Personal Learning Edition software. For the purpose of this study, we use the Pedestrian and Rail Anylogic libraries. On the one hand, the AnyLogic Pedestrian Library is dedicated to simulate pedestrian flows in “physical” environment creating models of pedestrian buildings (like subway stations, security checks etc.) or streets (big number of pedestrians) (Anylogic 2016). It is possible to collect statistics on pedestrian density in different areas, to assure acceptable performance of service points with hypothetical load. Additionally, it can

be estimated lengths of stay in specific areas, detected potential problems with interior geometry – for example the effect of adding obstacles and many other applications. The pedestrian library also supports 3D animation (Borshchev 2013).

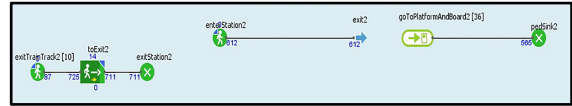


Figure 4: The Anylogic Pedestrian library blocks used in the simulation model – TASQUEÑA Direction

On the other hand, the Anylogic Rail Library supports detailed and accurate modeling (dimensions of individual cars, exact topology of tracks and switches, accelerations and decelerations of trains), as consequence the simulations it produces are very high performance (Borshchev 2013). Additionally, this library supports also 2D and 3D animation of tracks, switches, and rail cars. Since the Pedestrian Library also supports 3D animation, the model of the subway platform was created full dynamic 3D.

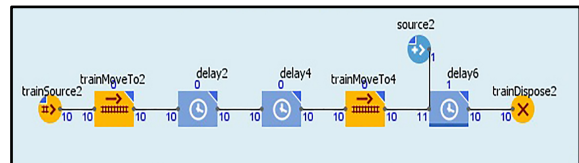


Figure 5: The Anylogic Rail library blocks used in the simulation model- TASQUEÑA direction

We design a simulation model user-interface to support the decision making (see Fig. 6). The passenger density, in passengers per square meter unit, is allocated to visualize its evolution under certain operational conditions. The layout of ZOCALO subway platform is designed considering real physical dimensions. ZOCALO station has two directions: TASQUEÑA and CUATRO CAMINOS. The subway tracks are traced between the subway platforms of the two directions.

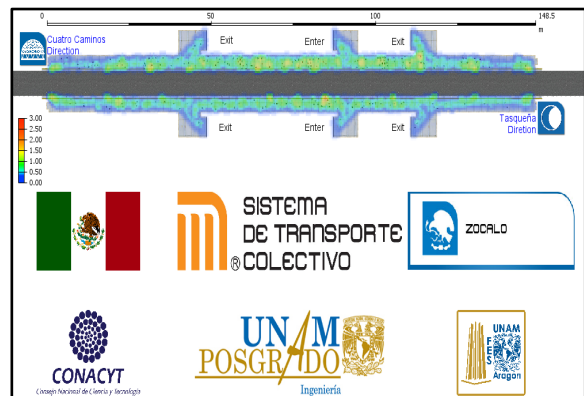


Figure 6: The interface to support the decision making

Additionally, the user-interface includes six cameras to visualize 3D animation of dynamic of ascent and descent of passengers in CUATRO CAMINOS (see Fig. 7.a) and TASQUEÑA direction (see Fig. 7.b)

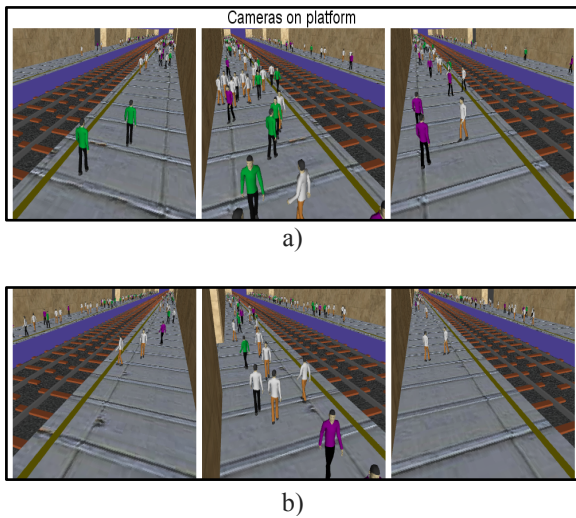


Figure 7: The simulation model of passengers ascending and descending from the convoy- subway platform, a) CUATRO CAMINOS direction, b) TASQUEÑA direction

4.1. Input data analysis

To calibrate the simulation model, we collected data about the number of passengers entering to the subway platform, measured in passenger per minute unit, as well as the inter arrival time of convoys in the two directions, during one hour. Then, we used Stat:Fit® version 3.0.1.0 Standard, developed by Geer Mountain Software (Benneyan 1998) to fit the appropriate probability distributions. The number of passengers on the subway platform CUATRO CAMINOS direction was fitted to a Negative Binomial distribution (see Fig. 8).

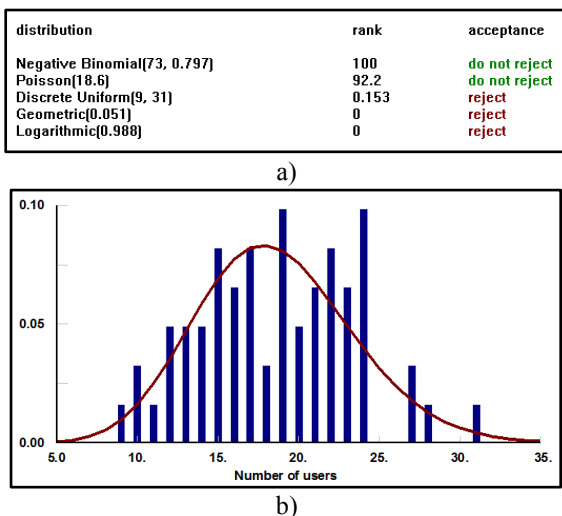


Figure 8. Passengers per minute entering to subway platform-CUATRO CAMINOS direction, a) autofit of distributions, b) graphic of the Negative Binomial distribution

The number of passengers on the subway platform TASQUEÑA direction was fitted also to a Negative Binomial distribution (see Fig. 9).

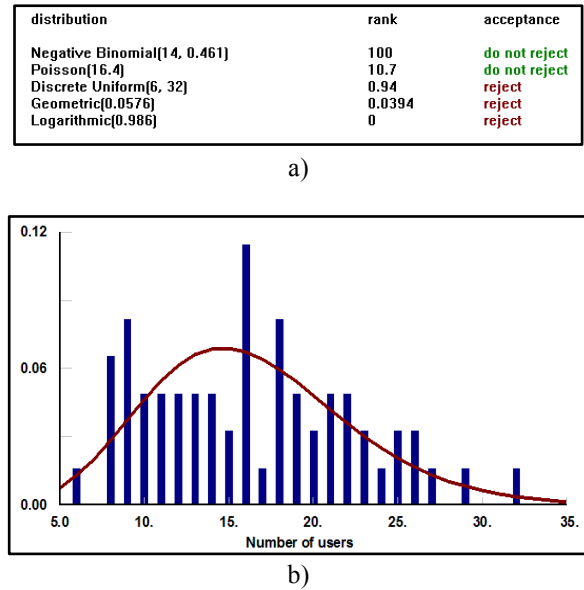


Figure 9. Passengers per minute entering to subway platform -TASQUEÑA direction, a) autofit of distributions, b) graphic of the Negative Binomial distribution

The inter arrival time of convoys CUATRO CAMINOS(see Fig. 10) and TASQUEÑA (see Fig. 11) direction were fitted to a Negative Binomial distribution.

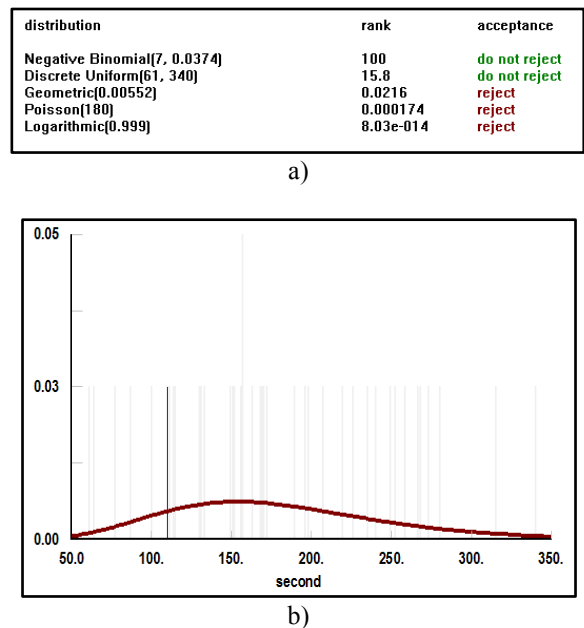
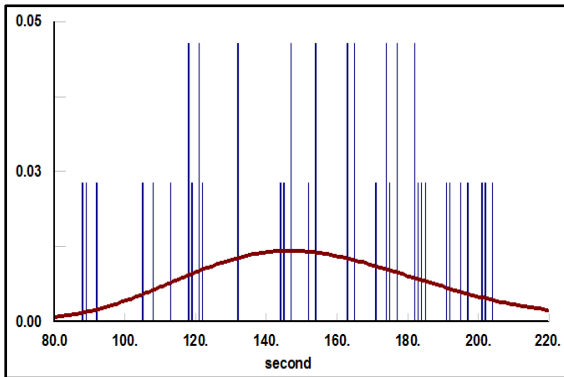


Figure 10. Interarrival time of convoys at ZOCALO station –CUATRO CAMINOS direction, a) autofit of distributions, b) graphic of the Negative Binomial distribution

distribution	rank	acceptance
Negative Binomial(23, 0.13)	100	do not reject
Discrete Uniform(88, 204)	61.8	do not reject
Poisson(154)	0.0215	reject
Geometric(0.00645)	4.18e-005	reject
Logarithmic(0.999)	2.85e-014	reject

a)



b)

Figure 11. Interarrival time of comboys at ZOCALO station-TASQUEÑA direction, a) autofit of distributions, b) graphic of the Negative Binomial distribution

4.2. Simulation model verification

Model verification is defined as ensuring that the computer program of the computerized model and its implementation are correct (Sargent 1999). As Banks (1998) points out, the animation is a very useful verification tool. Using the 3D animation, we detected some actions that firstly were illogical, mainly the passengers movements. Also, we observed conflict areas due to the wrong scale of physical dimensions of passengers. Additionally, all logical errors were adjusted using the internal debugger included in the simulator.

4.3. Simulation model validation

A model is considered valid for a set of experimental conditions if its accuracy is within its acceptable range, which is the amount of accuracy required for the model's intended purpose (Sargent 1999). In this direction, we conduct a sensitivity analysis to validate that the simulation model reproduces the dynamic of passenger's ascent and descent in the ZOCALO subway platform. For this, we propose two extreme values for passenger's comfortable speed (see Table 2) and execute during one hour the simulation model. In the case when the passenger's comfortable speed is very low, it is observed that the passenger's density on the subway platform is very high (see Fig. 12). In fact, passenger's congestion is observed in all enter and exit.

Contrary, in the situation when the passenger's comfortable speed is large, it is observed that the passenger's density on the subway platform is very low (see Fig. 13), as it happens in the real system, according to the technical visits done on the subway platform in the ZOCALO station during this study.

Table 2: Input variables for designing the sensitivity analysis

Input variables and parameters			
Parameter	Input variable	Low	High
A	Passenger's comfortable speed	Uniform (0.0, 0.5) m/s	Uniform (0.5, 1.5) m/s

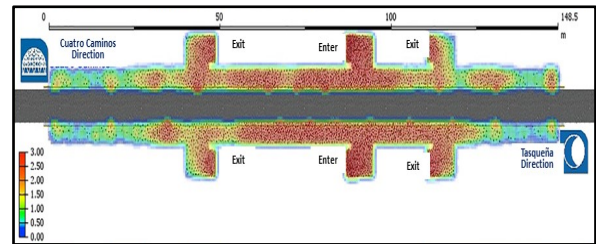


Figure 12: Passenger's density (Passengers per square meter) when their comfortable speed is at low value

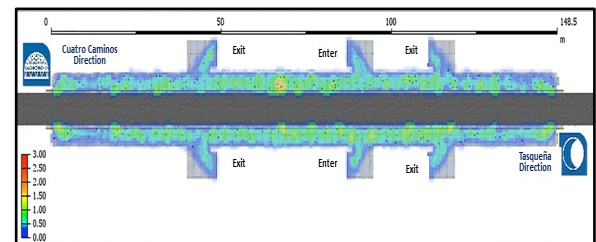
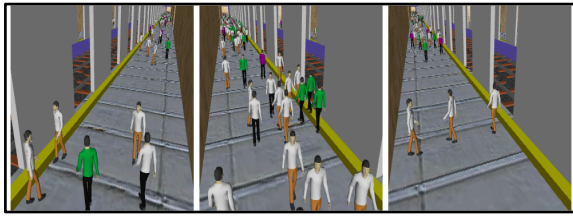


Figure 13: Passenger's density (Passenger per square meter) when their comfortable speed is at high value

5. IMPLEMENTING SECURE WAITING AREAS ON THE SUBWAY PLATFORM

Once the simulation model about the passenger's dynamic has been developed, we are interested on investigate if secure areas can prevent some accidents on the subway platform in the ZOCALO station and their impact on the dynamic of passengers. The secure areas proposed are designed installing physical barriers that let the ascent and descent of passengers only at specific points on the platform. We include the physical barriers in the simulation model (see Fig. 14) and execute it during one hour to analyze the dynamic of passengers. First, we observe that passengers remained in the secure areas until the convoy arrived, so the accidents are reduced. Second, the physical barriers contributed to the free flow of passengers through the platform avoiding the congestion points because passengers modified their direction according to the situation they observed, preferring the empty security areas (see Fig. 15).



a)



b)

Figure 14: Physical barriers installed in the secure waiting areas to minimize subway platform accidents, a) CUATRO CAMINOS direction, b) TASQUEÑA direction

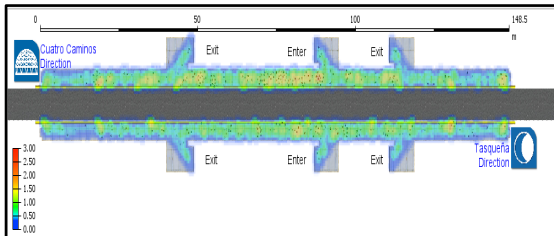


Figure 15: Passenger's density (Passenger per square meter) while physical barriers are installed

6. CONCLUSIONS

We developed an agent-based simulation model of ZOCALO subway platform to assign secure waiting areas for passengers in order to minimize accidents. The simulation model was implemented using AnyLogic™ using the Pedestrian and Rail libraries. To calibrate the simulation model, we collected data about the number of passengers entering to the subway platform, as well as the inter arrival time of convoys and fitted it using Stat::Fit software. Then, we conducted a sensitivity analysis to validate the simulation model and we observed the effect of the passenger's comfortable speed on the passenger's density on the subway platform. After that, the secure areas were proposed installing physical barriers on the subway platform that let the ascent and descent of passengers only at specific points. In the simulation model including such secure areas, we observed that passengers remained on it until the convoy arrived, so the accidents were reduced. Additionally, the congestion points on the platform were avoided by passengers because they preferred the empty security areas. We consider that the present study provide insights about that the installation of physical barriers at subway stations can be an effective strategy to reduce the number of accidents at subway stations in Mexico City.

ACKNOWLEDGMENTS

The author appreciates the partial support by the National Autonomous Mexico University. The author is also grateful to Sistema de Transporte Colectivo Metro (Mexico City subway) for the facilities in the data collection.

REFERENCES

- Anylogic 2016. Available from: <http://www.anylogic.com> [accessed 5 July 2016]
- Banks, J., 1998. Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Georgia: John Wiley & Sons, Inc.
- Benneyan, J.C., 1998. Software Review: Stat::Fit. OR/MS Today. Available from: <http://lionhrtpub.com/orms/orms-2-98/swr.html> [accessed 5 July 2016]
- Borshchev, A., 2013. The big book of simulation modeling. Multimethod modeling with Anylogic 6.
- Chung, Y., Kang, S., Matsubayashi, T., and Sawada, Y., 2016. The effectiveness of platform screen doors for the prevention of subway suicides in South Korea, 194, 80-83.
- Havarneanu, G., Burkhardt, J., and Paran F., 2015. A systematic review of the literature on safety measures to prevent railway suicides and trespassing accidents. Accident Analysis and Prevention, 81, 30-50.
- Ichikawa, M., Inada, H., and Kumeji, M., 2014. Reconsidering the effects of blue-light installation for prevention of railway suicides. Journal of Affective Disorders, 152-154, 183-185.
- Kadotani, H., Nagai, Y., and Sozu T., 2014. Railway suicide attempts are associated with amount of sunlight in recent days. Journal of Affective Disorders, 152-154, 162-168.
- Law, C. K., Yip, S. F., Chan, S. C., Fu, K., Wong, P., and Law, Y. W., 2009. Evaluating the effectiveness of barrier installation for preventing railway suicides in Hong Kong. Journal of Affective Disorders, 114, 254-262.
- Mishara, B. and Bardon C., 2016. Systematic review of research on railway and urban transit system suicides. Journal of Affective Disorders, 193, 215-226.
- Sargent, R., 1999. Validation and verification of simulation models. Proceedings of Winter Simulation Conference, pp. 39-48. December 5-8, Phoenix.

Soeta, Y. and Shimokura, R., 2012. Change of acoustic characteristics caused by platform screen doors in train stations. *Applied acoustics*, 73, 535-542.

Ueda, M., Sawada, Y., and Matsubayashi, T., 2015. The effectiveness of installing physical barriers for preventing railway suicides and accidents: evidence from Japan. *Journal of Affective Disorders*, 178, 1-4.

AUTHORS BIOGRAPHY

GERARDO DE JESÚS PORTILLO-VILLASANA currently is Master student of the Department of Systems Engineering at the School of Engineering, UNAM. His research interests are security in urban transport, information and communication technologies, simulation and optimization.

AIDA HUERTA-BARRIENTOS received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the School of Engineering, UNAM and she is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

PAYSIM: A FINANCIAL MOBILE MONEY SIMULATOR FOR FRAUD DETECTION

Edgar Alonso Lopez-Rojas^(a), Ahmad Elmir^(b), and Stefan Axelsson^(c)

^{(a),(b)} Blekinge Institute of Technology, ^(c)The Norwegian University of Science and Technology

^(a)edgar.lopez@bth.se ^(b)ahel11@bth.se ^(c)stefan.axelsson@hig.no

ABSTRACT

The lack of legitimate datasets on mobile money transactions to perform research on in the domain of fraud detection is a big problem today in the scientific community. Part of the problem is the intrinsic private nature of financial transactions, that leads to no public available data sets.

This will leave the researchers with the burden of first harnessing the dataset before performing the actual research on it. This paper propose an approach to such a problem that we named the PaySim simulator.

PaySim is a financial simulator that simulates mobile money transactions based on an original dataset. In this paper, we present a solution to ultimately yield the possibility to simulate mobile money transactions in such a way that they become similar to the original dataset. With technology frameworks such as Agent-Based simulation techniques, and the application of mathematical statistics, we show in this paper that the simulated data can be as prudent as the original dataset for research.

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

1. INTRODUCTION

Obtaining access to data sets of mobile transactions for research is a very hard task due to the intrinsic private nature of such transactions. Scientists and researchers must today spend time and effort in obtaining permits and access to relevant data sets before they can research on such data set. This is time consuming and distracts researchers from focusing on the main problem which is performing experiments on the data and finding novel ways to solve problems such as the problem that inspired this paper which is the fraud detection on financial data.

The work presented in this paper provides a tool and a method to generate synthetic data with the help of a simulator that we named *PaySim*. *PaySim* generates synthetic datasets similar to real datasets from mobile money transactions. This will be done by the means of computer simulation, in particular, agent based simulation. Agent based simulation is of great benefit in this context, this partially because the models created represent with accuracy the human behaviour during transactions and are flexible enough to easily be adapted to new constraints.

PaySim simulates mobile money transactions based on a sample of real transactions extracted from the logs of a mobile money service implemented in an African coun-

try. With the help of statistic analysis and social network analysis *PaySim* is able to generate congruous results with the original data set.

The scope of this paper covers the design and construction of the simulator as well as the evaluation of the quality of the data generated. The injection of malicious fraud behaviour and the application of different fraud detection methods are outside the scope of this paper and are the topics for further work with the *PaySim* simulator.

Outline This paper is structured as follows: Section 2. presents the background and previous work in simulating financial data. Section 3. states the problem and during sections 4. and 5. we present the implementation of *PaySim* and the results of the simulations. Finally section 6. present the conclusions and future work.

2. BACKGROUND AND PREVIOUS WORK

In many parts of Africa the adoption of mobile money as a means of sending & receiving funds have improved the life of merchants and customers alike. In Tanzania for instance, which according to the world bank is one of the fastest growing economies in the world, the adoption of mobile money as a solution for creating payments has induced a positive effect on the overall economy. During December 2013 alone, 100 million transactions were made in total netting a volume of \$1.8 billion dollars (Seetharam and Johnson, 2015).

The domain of Mobile Money Transfer has grown substantially in the last few years and have attracted greater attention from users, specifically in areas in which banking solutions may not be as procurable as in developed countries. Many solutions have been employed in many places for this purpose. There are existing mobile money services in more than 10 African countries which coverage of 14% of all mobile subscribers (Rieke et al., 2013).

The ever growing usage of mobile money has increased the chances and likelihood of criminals to perform fraudulent activities in an attempt to circumvent the security measures of mobile money transfers services for personal financial gain. There is therefore a great amount of pressure on researching the potential security pitfalls that can be exploited with the ultimate goal to develop counter-solutions for the attacks.

Due to the large amount of transactions and the ever changing characteristics on fraud, the current measures against fraud lack effectiveness. Many current system still base their detection mechanism on simple thresholds

assigned arbitrarily. Therefore there is a need to push forward and investigate the effect of fraud and stop the wrongdoers from fraudulent profit.

With *PaySim*, we aim to address this problem by providing a simulation tool and a method to generate synthetic datasets of mobile transactions. The benefits of using a simulator to address fraud detection was first presented by (Lopez-Rojas and Axelsson, 2012b). This research states the problem of obtaining access to financial datasets and propose using synthetic datasets based on simulations. The method proposed is based on the concept of MABS (Multi Agent Based Simulation). MABS has the benefits that allows the agents to incorporate similar financial behaviour to the one present in domains such as bank transactions and mobile payments.

The first implementation of a simulator for financial transaction was introduced by (Lopez-Rojas and Axelsson, 2012a) with a mobile money transactions simulator. This simulator was implemented due to the difficulties to implement a proper fraud detection control on a mobile money system that was under development. This paper was the first to present an alternative to the lack of real data problem. The synthetic dataset generated by the simulator was used to test the performance of different machine learning algorithms in finding patterns of money laundering.

The work by (Gaber et al., 2013) introduced another similar technique to generate synthetic logs for fraud detection. The main difference here was that this time there was available real data to calibrate the results and compare the quality of the result of the simulator. The purpose of this study was to generate testing data that researchers can use to evaluate different approaches. This works differs significantly from our work because we present a different method for analysing the data place special attention on evaluating the quality of the resultant synthetic data set.

There has been some work done in the domain of financial transactions for retail stores. The most prominent of which is the work done by (Lopez-Rojas et al., 2013). The work done in that paper is very similar to the work done in this paper. A large collection of data was gathered from Sweden's biggest shoe-retailer, and techniques involved complex machine-learning algorithms in an attempt to find fraudulent behaviour in clients. The paper showed among other things results from Social Network which described the relationship between the clients and the sellers for each store. A definition of what was perceived as "fraudulent" was made and based on that the machine-learning algorithms were trained to detect that type of behaviour.

Public databases of financial transactions are almost non-existent. However the work of (Lopez-Rojas and Axelsson, 2014) during the implementation of a simulator called BankSim presents a MABS of financial payments. BankSim is implemented in a similar way as the RetSim simulator and our simulator using in addition to statistical analysis a social network analysis. BankSim is based on

the aggregated financial information of payments during 6 months of the two main cities of Spain that was provided by a bank in Spain with the purpose of developing applications of different kinds that benefit from this sort of data. Our work differs from this work because the source of the data and the characteristics of bank payments and mobile transactions are different as presented later in the following sections.

The key common aspect on previous work is the use of the paradigm of "Multi Agent Based Simulation" approach which incorporates into the behaviour of the agents the main customer logic to reach similar results as the real world. It is important to recognize that a simulation is not an actual "replication" of the original data set. Rather, a simulation will with the aid of statistical methods generate a very similar data set of the original data set. The degree in variance will largely be dependent on how the data on the original data set is structured, hence, different simulations based on different seeds will generate different output data sets but consistent with the real world.

3. PROBLEM

The problem formulation for this research paper tackles the issue of whether the generation of synthetic financial data is sufficient to supersede real financial data whilst simultaneously yield commensurable results if the synthetic data is used as the source data set for any research. This is of primary concern for any researcher that wish to perform scientific tests but does not or have limited access to a real financial data set.

The main focus and goal for the simulation is to yield another completely self-sufficient data set with the goal of having similar statistical properties as the original data set. To yield such results, the simulator must go through several steps to be able to complete.

In order to simulate the mobile money service, we need to properly simulate the different kind of transactions that the system supports. We decided to cover 5 of the most important transaction types: CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.

CASH-IN is the process of increasing the balance of account by paying in cash to a merchant.

CASH-OUT is the opposite process of CASH-IN, it means to withdraw cash from a merchant which decreases the balance of the account.

DEBIT is similar process than CASH-OUT and involves sending the money from the mobile money service to a bank account.

PAYMENT is the process of paying for goods or services to merchants which decreases the balance of the account and increases the balance of the receiver.

TRANSFER is the process of sending money to another user of the service through the mobile money platform.

There are other types of transactions that we decided to exclude from the simulation due to the low percentage of data found in the sample.

4. MODEL AND IMPLEMENTATION

PaySim uses the MABS toolkit called MASON version 19 which is implemented in Java (Luke, 2005). We selected MASON because it is: multi-platform, supports parallelisation, and fast execution speed in comparison with other agent frameworks. This is especially important for multiple running and computationally expensive simulations such as PaySim (Railsback et al., 2006).

4.1. Overview, Design and Details (ODD)

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

4.1.1. ODD Overview

The purpose of this simulator is to simulate payments done in the realms of mobile transactions. The simulator should ultimately perform simulations in such a way that synthetic data in regards to mobile transactions can be generated. The simulator should generate synthetic data that is very similar to a batch of real transactional Data provided by Ericsson. The goal is to have a generator that can generate data on the fly that can later be used by the scientific community in an attempt to research more about fraud detection.

The model has one primary type of Entity which is *Client*. Each client has a profile that describes the allowed behaviour for the client such as the limit on transactions daily/yearly, the transaction limit and the maximum balance for the client. Furthermore the number of transactions, withdrawals, transfers and deposits is stored for each client. The client can further be classified by age to be young, adult or senior. Each client has a base currency in which the transactions are based upon. The client can perform transactions in the form of deposits, withdrawals and transfers. For every transaction that is made, it is stored and saved within the client.

4.1.2. Process Overview and Scheduling

The client has several processes that alter their internal states. For each step that is made by the simulator, based on a random variable that is contingent on calculated probabilities, a type of transaction that is to be performed by the client is chosen. A deposit transaction will increase the balance of the client, a withdrawal will decrease the balance of the client and a transfer transaction will withdraw money from the original client and then deposit them to the destination client in question.

The concepts that are behind the model are based on statistical analysis of a large batch of real data. From this batch of data, probabilities of each action were calculated and incorporated into the model to generate synthetic information as close as possible to the real data. The client agent has some adaptive behaviours that will alter their way of acting; for instance if the client has reached its daily limit it cannot withdraw money any more for that day. This adaptive behaviour is a direct result of the *transfer* process mentioned above. There is interaction

between agents since there is a probability that at a particular step of the simulation, an agent might transfer money to another agent and thus alter its and the other agents state.

4.2. Inputs

There are multiple inputs required in order for the simulator to function smoothly. As initial input, the number of clients neighbours for each agent is assigned. The profile for each agent is then further attached based on a probability. Their location on the spatial space along with their neighbours is also initialized.

- **Parameter File** This is the file that contains all of the needed parameters that the simulator needs to initiate. Among these parameters we find the seed and perhaps the most relevant of which is the paths for where the input files and the output files are placed on the current machine.
- **Aggregated Transaction File** This file contains the distribution of the transactions from the original data set. More precisely, it contains how many transactions were made at any given day/hour combination (step). what is the average price for that, what type of transaction it was etc. This is of paramount importance for the simulator since statistical data is generated from the information gathered from this file.
- **Repetitions File** This file contains the frequency of transactions that the original clients had per type of transaction. This means that some of the agents are schedule more than others based on a social network analysis of the indegree and outdegree of the customers.

Since the simulator is using MASON as the framework for performing the simulation, it is of paramount importance to define how each step is to be regarded. For this simulation we defined that each day/hour combination represents one step. At each step, a Client that represents the agent for the simulator is generated. The client will be placed in an environment in which it is to make decisions based on the information it perceives. The Client is created with the statistical distribution of the possibilities to perform each transaction type for a specific day/hour combination. The client then randomly perform (based on the distribution initiated) different transaction types in relation to the other clients on the simulator. Also, for each client generated, there is a probability P for the client to make future transactions at later steps. This probability is gathered from the database of the original data set.

4.3. Initiation Stage

In this stage, the PaySim simulator must load the necessary data needed from the original dataset:

- **Load The Parameters** The first and most important step is to load the values for each parameter in the

parameter file. These will among other things contain the file paths for the source data inputs that the simulator needs to load.

- **Load Aggregate File** This is the original Aggregate File that will be used as a base point for the simulator to generate statistically similar results in terms of "What to simulate, at what day, at which amount" etc. One such extraction could be for instance: At Day 1 and hour 15, simulate 8703 transactions of the type PAYMENT with the average transaction size of 180000 and the standard deviation of 15000.
- **Load Initial Balance Container** Apart from the statistical distribution for each transaction type input to the client, there is another important input, namely, the "Initial Balance" of the client. Upon the generation of each client in the simulation, there must be an initial "Balance" attached to that client. This balance is generated with the "Balance Container" file as base point. The Balance Container consists of the different probabilities that will generate different initial balance ranges.
- **Load Maximum Repetitions File** As mentioned previously, each client has a probability P of making future transactions in future steps in the simulation. What this file does is to make sure that each client does not make more repetitions that is allowed. Like the Balance Container, the probabilities in this file is also yielded from the database of the original data set.

4.4. Execution Stage

Upon completion of the Initiation Stage and all of the parameters are successfully loaded, the simulator can now proceed to the execution stage. It is at this stage that the simulator will perform the actual simulation, and yield the simulated transaction results:

4.4.1. Generating The Clients

The agents are the founding blocks of the "Agent Based Simulator". The agent in this context, resembles the "Client". Upon each step of the simulation, the PaySim simulator will convert each step to a "Day/hour" combination. This will then be used as an input to extract the statistical distributions from the original data set. Based on the *Aggregated Transaction File*, PaySim harness the probability P of performing each each transaction in the simulator and save it into the model of the client. With this information, the client now has gained more knowledge and will know the following important things:

- **Number Of Transactions** This is the total number of transactions that this generated client will do.
- **Make Future Steps** This is the information of whether the client is to participate in future steps. Which means scheduling the tasks of performing more transactions during further steps.

- **Statistical Distribution** This is the different probabilities that the client will have loaded into it which entails the probability P of performing each action.
- **Initial Balance** This will be the initial balance that the client will have once generated.

4.4.2. Performing the transactions

After each client is generated, the client will make the decision of what type of transaction it will ultimately make, again this is completely derived from the distribution loaded. The client is in an environment which allows it to freely interact with other clients in the simulation. There are some types of transaction types that is based on that, like "TRANSFER" for instance. The "TRANSFER" type is exchange of money from one client to another; hence, the client will have to interact with other clients to simulate the actual exchange of funds.

4.5. Finalization Stage

After each of the clients have completed their role in the simulation and performed all of the transactions allotted the results must be saved. There are 4 outputs generated for each simulation made. All of which serve a specific purpose which will allow for the exact repetition of the simulation with the exact initial properties.

- **Logfile** Each transaction that is made will contain a record with the meta-data for that transaction. Data such as what client performed which action, to which other client, the sum of the transaction, and the delta in balance for all clients involved. Each such record will be saved in a logfile unique for the specific simulation.
- **MySQL Database** Apart from the logfile, the record for each transaction will also be saved into a MySQL database. The purpose of which is to allow for easier queries when the analysis of the results is to be made.
- **Aggregate Dump** An aggregatedump that is similar to the original aggregatedump from the original data set will also be generated. It is these two files that will be used to generate the plots and graphs resembling the results of the transactions.
- **Parameter File History** This file will contain the exact properties needed for the simulation to be able to reproduce the exact same results again. This is important because each simulator must be able to be reproduced again, and without the original "seed" used, it will not be possible.

5. RESULTS

We ran PaySim several times using random seeds for 744 steps, representing one month of real time data. Each run took around 30 minutes in a i7 intel processor. We selected a datasets that contained the lowest difference in

Figure 1: Visualization of transaction type CASH-IN

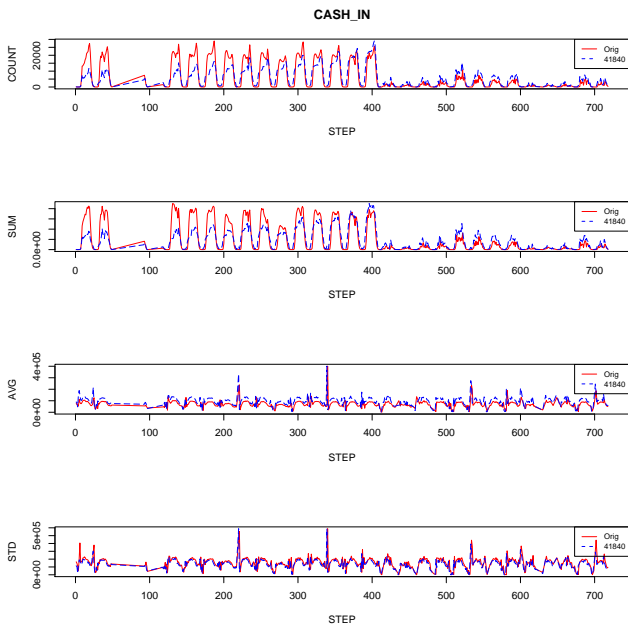


Figure 3: Visualization of transaction type TRANSFER

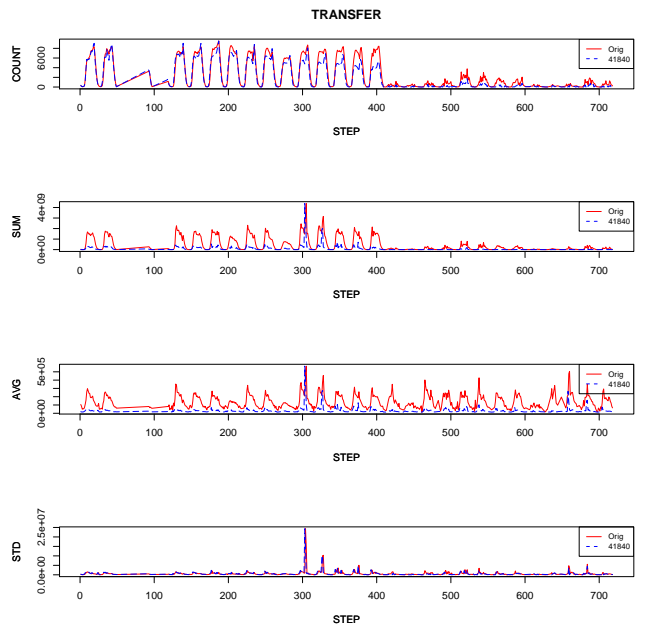


Figure 2: Visualization of transaction type CASH-OUT

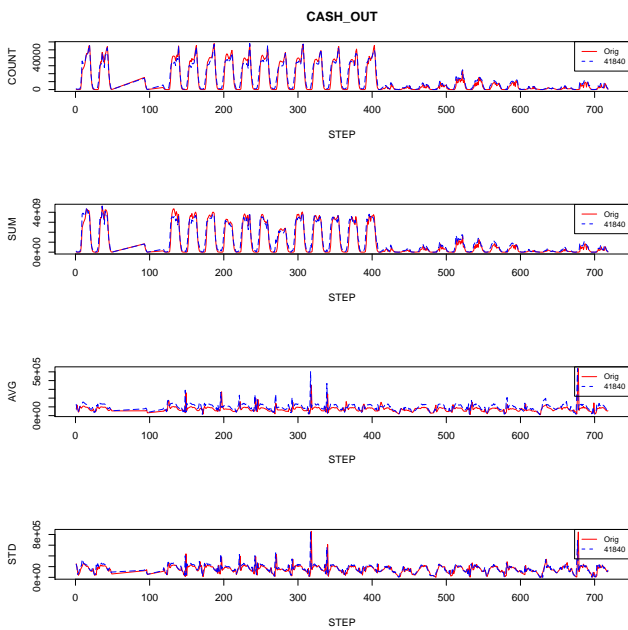


Figure 4: Visualization of transaction type PAYMENT

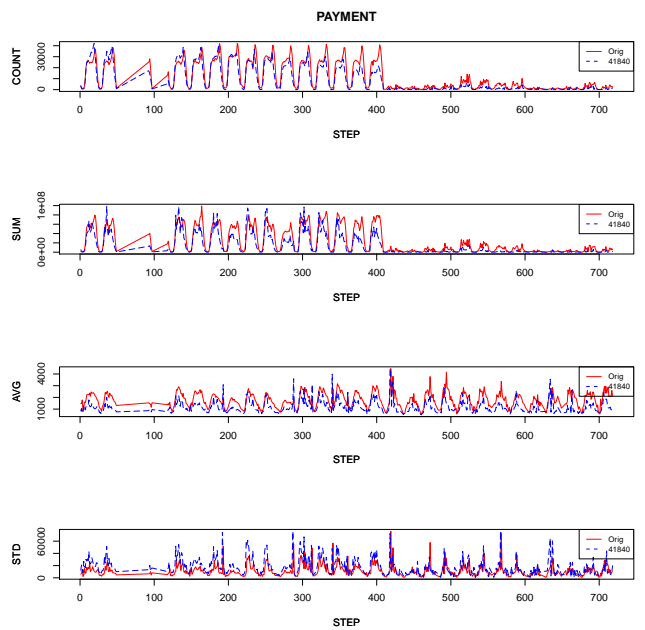
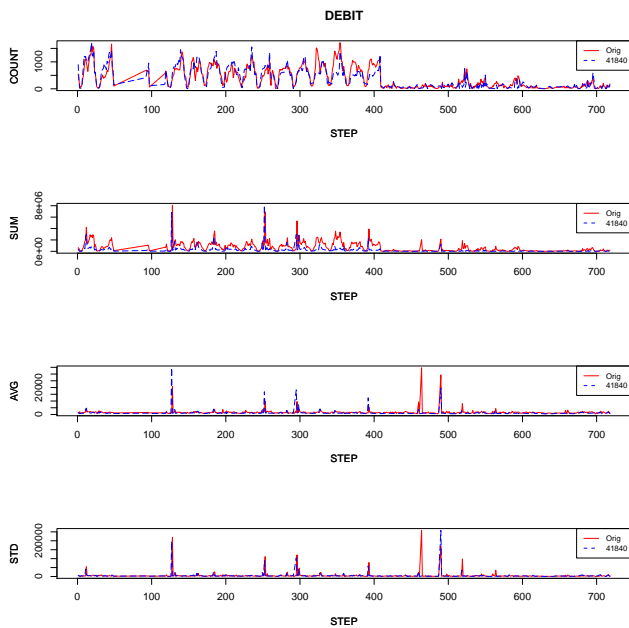


Figure 5: Visualization of transaction type DEBIT



values according to the original data set provided. The selected synthetic dataset was arbitrary named *PS41840*. *PS41840* contains around 23 million records divided into the 5 types of categories presented before. Table 1 shows the types of transactions, count and average amount generated with the simulator. The amount values are given in a currency that we can not disclose.

Table 1: Simulated PS41840

TYPE	Count	avgAmount
CASH-IN	4 496 947	153 019
CASH-OUT	9 014 407	155 989
TRANSFER	2 030 969	630 810
PAYMENT	8 955 794	10 793
DEBIT	139 935	5 016

The evaluation of the quality of the database was first calculated using the sum of square error (SSE) method on the quantities of the different datasets. The one with the lowest error was *PS41840*.

In order to verify that the simulation was working properly we plotted the distributions to visually identify significant differences between the original and the synthetic dataset. Figures 1, 2, 3, 4 and 5 show the visualization per type of transaction. Each figure contains the output for each step regarding the count of transactions, the total sum of transaction, the average and the standard deviation. The red continuous line represent the original data distribution and the blue dashed line represent the synthetic dataset *PS41840*.

Something we noted is that the first 14 days of the simulation the activity in the system is higher compared to the remaining days. This is perhaps a phenomenon present

due to the introduction of income during the first days of the month.

6. CONCLUSIONS

PaySim is a simulation of mobile money transactions with the objective to generate a synthetic transactional data set that can be used for research into fraud detection. The data sets generated with PaySim can aid academia, financial organisations and governmental agencies to test their fraud detection methods or to compare the performance of different methods under similar conditions using a common public available and standard synthetic data set for the test.

We argue that PaySim is ready to be use as a tool to generate synthetic transactions that resemble the original and private data set supplied. By using PaySim we protect the privacy of the customers of the service at the same time that interesting results are possible to share with other researchers without the constrains and legal boundaries of the original data.

The results presented in the section 5. help to visually appreciate that the generated dataset captures the process and the frequencies of the different transaction types of the mobile money service.

Future work on the simulator is to add to the model fraudulent agents and run different scenarios to test the efficacy and accuracy of diverse fraud detection methods. We also want to make a synthetic data set available to other researchers and be able to compare and share diverse results.

ACKNOWLEDGMENTS

This work is part of the research project "Scalable resource-efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.

References

- Chrystel Gaber, Baptiste Hemery, Mohammed Achemlal, Marc Pasquet, and Pascal Urien. Synthetic logs generator for fraud detection in mobile transfer services. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 174–179. IEEE, may 2013. ISBN 978-1-4673-6404-1. doi: 10.1109/CTS.2013.6567225.
- Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K. Heinz, Geir Huse, Andreas Huth, Jane U. Jepsen, Christian Jørgensen, Wolf M. Mooij, Birgit Müller, Guy Pe'er, Cyril Piou, Steven F. Railsback, Andrew M. Robbins, Martha M. Robbins, Eva Rossmannith, Nadja Rüger, Espen Strand, Sami Souissi, Richard a. Stillman, Rune Vabø, Ute Visser, and Donald L. DeAngelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115–126, September 2006. ISSN 03043800. doi: 10.1016/j.ecolmodel.2006.04.

023. URL <http://linkinghub.elsevier.com/retrieve/pii/S0304380006002043>.

Edgar Lopez-Rojas and Stefan Axelsson. Multi agent based simulation (mabs) of financial transactions for anti money laundering (aml). In Audun Josang and Bengt Carlsson, editors, *Nordic Conference on Secure IT Systems*, pages 25–32, Karlskrona, 2012a.

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

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Social Simulation of Commercial and Financial Behaviour for Fraud Detection Research. In *Advances in Computational Social Science and Social Simulation*, Barcelona, 2014. ISBN 9789172952782.

Edgar Alonso Lopez-Rojas, Stefan Axelsson, and Dan Gorton. RetSim: A Shoe Store Agent-Based Simulation for Fraud Detection. In *The 25th European Modeling and Simulation Symposium*, number c, page 10, Athens, Greece, 2013.

S. Luke. MASON: A Multiagent Simulation Environment. *Simulation*, 81(7):517–527, July 2005. ISSN 0037-5497. doi: 10.1177/0037549705058073. URL <http://sim.sagepub.com/cgi/doi/10.1177/0037549705058073>.

S. F. Railsback, S. L. Lytinen, and S. K. Jackson. Agent-based Simulation Platforms: Review and Development Recommendations. *Simulation*, 82(9):609–623, September 2006. ISSN 0037-5497. doi: 10.1177/0037549706073695. URL <http://sim.sagepub.com/cgi/doi/10.1177/0037549706073695>.

Roland Rieke, Maria Zhdanova, Jurgen Repp, Romain Giot, and Chrystel Gaber. Fraud Detection in Mobile Payments Utilizing Process Behavior Analysis. In *2013 International Conference on Availability, Reliability and Security*, pages 662–669. IEEE, sep 2013. ISBN 978-0-7695-5008-4. doi: 10.1109/ARES.2013.87.

Balachandran Seetharam and Drew Johnson. *Mobile Money's Impact on Tanzanian Agriculture*. 2015.

AUTHORS BIOGRAPHY

MSc. Edgar A. Lopez-Rojas

Edgar Lopez is a PhD student in Computer Science at Blekinge Institute of Technology in Sweden and his research areas are Multi-Agent Based Simulation, Machine Learning techniques with applied Visualization for fraud detection and Anti Money Laundering (AML) in the domains of retail stores, payment systems and financial transactions. He obtained a Bachelors degree in

Computer Science from EAFIT University in Colombia (2004). After that he worked for 5 more years at EAFIT University as a System Analysis and Developer and partially as a lecturer. He obtained a Masters degree in Computer Science from Linköping University in Sweden in 2011 and a licentiate degree in computer science (a degree halfway between a Master's degree and a PhD) in 2014.

MSc. Ahmad Elmir

Ahmad obtained a master in computer science with speciality in security from the Blekinge Institute of Technology. His master's thesis was about the design and construction of PaySim under the supervision of the main author of this paper. Previously he have studied natural sciences in the gymnasium for three years. His speciality was at computer science and programming. He has a keen interest for scientific inquiry in the domain of security as it is in his opinion an ever developing field. He have also worked for 9 months with software development in a corporation.

Dr. Stefan Axelsson

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

ANALYSIS AND EVALUATION OF PERFORMANCE ISSUES OF PARALLEL SOFTWARE ON MULTI-CORE PROCESSORS

Franz Wiesinger^(a), Mustafa Tunca^(b), Michael Bogner^(c)

^{(a), (b), (c)} University of Applied Sciences Upper Austria – Embedded Systems Design,
Softwarepark 11, A-4232 Hagenberg, AUSTRIA

(a) franz.wiesinger@fh-hagenberg.at, (b) mustafa.tunca@fh-hagenberg.at, (c) michael.bogner@fh-hagenberg.at

ABSTRACT

For decades, the processor manufacturers have attempted to achieve performance gains by increasing the clock frequency on single-core processors. But physical problems – such as the high power dissipation – lead to the release of the first multi-core processors on the market in 2005.

To benefit from the multi-core architecture, parallel programming is required. However, this programming model requires a different approach and is associated with certain risks and pitfalls.

This paper focuses on modelling of certain test scenarios for two common multi-core specific problems, namely oversubscription and false sharing. Various simulations and tests offered solutions and design patterns to avoid such problems. Results have shown that the problems have a fatal impact on the execution time, so that the performance gain on the multi-core system is nearly nonexistent. Thence, any software developer must have in-depth knowledge of the used hardware and software to benefit as much as possible from multi-core architectures.

Keywords: multi-core, parallel software, performance analysis, oversubscription, false sharing

1. INTRODUCTION

For decades, processor manufacturers have tried to achieve more performance of their processors mainly by increasing the clock frequency. But despite technical countermeasures, this methodology of improving the performance of single-core processors has led to an increase of the power consumption and to an unmanageable heat generation when using conventional cooling techniques.

This fact and the growing demands of customers induces the large processor manufacturers such as Intel, IBM and AMD to further improve the development of single-core processors and to focus on the development of multi-core processors.

This was achieved by increasing the number of transistors – the main component of a processor – by improving the manufacturing techniques. As a result, these additional transistors serve in the processor development for many architectural advantages (Rauber and Runger 2012). Instead of further increasing the chip packing density on a single-core processor and making it even

more complicated, the available surface area on the silicon is now used for multiple cores which allows the distribution of the workload across these cores. Thereby, it is possible to increase the throughput and in further consequence, the performance referring to the power consumption is improved (Gove 2010).

The first multi-core processors reached the end-user in 2005 – thus, a new era of information processing was born. However, this resulted also in a new challenge for software developers. Because so far, the processor manufacturers were primarily responsible for performance improvements on their processors. Existing software was automatically faster with each generation of processors due to technological advances. But now having multi-core chips, it is no longer possible to get such improvements by only replacing the processor on a system. Instead, the software engineers have to use parallel software development to benefit from the new hardware architecture and to gain from performance improvements.

The parallel software development existed long before the era of the multi-core processors and therefore had been used by a minority of software developers in multi-processor environments. But now the multi-core processors being the new standard, all software developers have to deal with the issue of parallelism. Generally spoken, parallelism requires, inter alia, the decomposition of a given problem to a set of threads, which process the problem on different cores of the processor. However, this approach of development differs from the former sequential execution on a single-core processor, so in turn the software development requires a new way of thinking in modelling parallel software, which further yield to various difficulties and pitfalls.

As mentioned above, these problems and pitfalls are not really new and are known from former multi-processor systems. For this reason, the main focus of this paper is to check whether the performance-reducing problems also have an impact on current multi-core systems. Therefore, we have modelled some typical test scenarios which are most common. In this paper, we present our results for the issues oversubscription and false sharing. We have simulated and analyzed the run-time behavior of our software implementations and have derived some solutions, which will be explained in detail in this paper.

2. RELATED WORK

As the multi-core systems became the new standard, numerous well-documented papers about these systems are available, especially from the processor manufacturers. Since the described issues already existed before the multi-core processors got popular, there are many well-documented works which deal with the general issues on parallel systems (Duffy 2008, Tanenbaum 2009).

Since the developers tend to use too many threads in their applications, the common problem of oversubscription arises. Blumofe and Leiserson (1999) analyze and discuss a way of scheduling multithreaded applications on parallel systems and create the so-called work-stealing technique. Yanyong Zhang et al. (2003) provide a deeper analysis of the oversubscription problem, by comparing existing techniques and migrating the solutions into large scale parallel systems.

In contrast to the active community for scheduling problems on parallel systems, there is sparsely research regarding the problems on hardware-level on such systems. The work of Torrellas et al. (1994) are one of the first in this field analyzing the issue of false sharing, one of the problems occurring in shared-memory parallel systems.

However, regarding problems on hardware and software in parallel systems, most of them do not deal with the new hardware platform and present no in-depth discussion according to the special issues of concurrent multi-core systems. On the other hand, there are works which discuss this issues also for multi-core systems, but they are mainly theoretically, do not focus on special hardware and derive no suggestions or design patterns (Akhter and Roberts 2006, Gove 2010, Gleim and Schüle 2012, Rauber and Rüniger 2012, Williams 2012).

3. TEST ENVIRONMENT

To analyze the problems precisely, the measurements have to be made according to specific criteria, which are defined in the following sections.

3.1. Test System

The test system uses a multi-core processor with Sandy Bridge microarchitecture with the model name Intel Core i7 2670QM. The processor has four physical cores and supports eight logical processors. Furthermore, the test system has 6 GB of memory and three cache levels, wherein a cache line in each cache-level has the size of 64 bytes. The L3-cache is shared among all cores and allows inter-core communication on the chip. Windows 7 Professional (64-bit) has been used as the operating system.

3.2. IDE and Profiler

For the implementation of the test scenarios the integrated development environment Microsoft Visual Studio 2013 has been used.

The analysis of the run-time behavior has been done with the profiler integrated in the IDE, the Visual Studio Profiling Tools.

In addition to the integrated profiler, an official plug-in from Microsoft for Visual Studio 2013 called Concurrency Visualizer has been used for visualization.

3.3. Libraries and Programming Language

The elaboration of the test scenarios takes place in the programming language C++. For managing the threads the `std::thread` class has been used, which is part of the C++11 language standard.

The renunciation of the operating system interface e.g. to create threads or synchronize them, means that the source code is independent of the operating system. Thereby, it is possible to run the test scenarios on any system with a compiler that supports the C++11 language standard. In our research, we used the compiler Visual C++ 2013.

3.4. Class Design

For our simulations, we decided to use an extensible software design to allow further implementations of other critical problems. The software design encapsulates the concrete implementations by using the strategy design pattern. So the developer can easily exchange the implementation of the concrete problem at run-time.

Figure 1 illustrates this structure represented in a UML class diagram. There are two important interfaces. The interface *CriticalProblem* offers methods which can be overridden by the implementation of a test scenario for a specific problem. And the class *ScenarioStarter* has methods, which have a concrete strategy such as for oversubscription to run its operations.

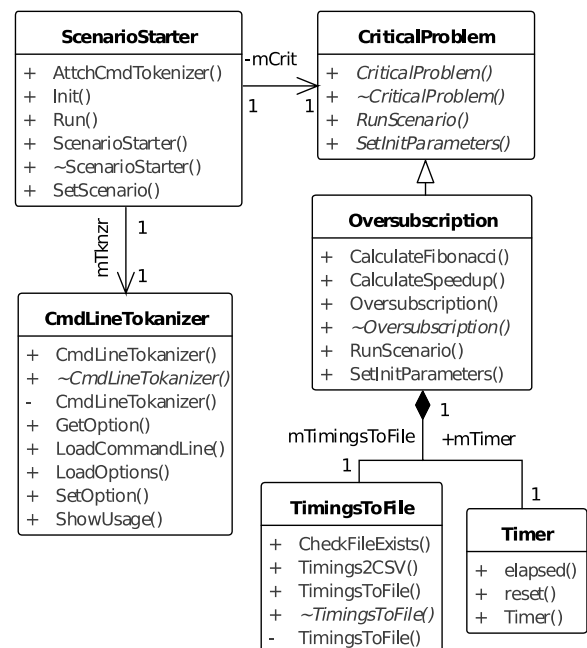


Figure 1: Class diagram modelling a general structure for implementing the test scenarios.

Listing 1 shows an example where the design pattern illustrated in Figure 1 is applied. The object *tokenizerObj* gets the user-given command line arguments respectively initial values for the test scenario and ana-

lyzes it for correctness. Then a test scenario *oversubscriptionObj* is attached to *starter*. The test scenario is finally initialized with the user-given options from the *tokenizerObj* and can be started with the *Run()* method. After the test scenario has finished another type of test scenario can be started directly afterwards.

Listing 2: Program code showing a typical usage of the class design.

```

1 int main(int argc, char** argv)
2 {
3     CmdLineTokenizer tokenizerObj;
4     ScenarioStarter starter;
5     Oversubscription oversubscriptionObj;
6     FalseSharing falsesharingObj;
7
8     tokenizerObj.LoadCommandLine(argc, argv);
9     starter.AttachCmdTokenizer(&tokenizerObj);
10    starter.SetScenario(&oversubscriptionObj);
11    starter.Init();
12    starter.Run(); // run and measure run-time
13
14    starter.SetScenario(&falsesharingObj);
15    starter.Init();
16    starter.Run(); // run and measure run-time
17    return 0;
18 }
```

3.5. Time Measurement

In addition to the integrated profiler, time measurements have been done on specific sequences in the source code. To gain detailed and accurate measurement results, we have taken the class *std::chrono* with nano-second precision from the C++11 language standard (simple high resolution timer in C++ 2015). The class *Timer* encapsulates the time measurement functionality. It has been used by the concrete strategy respectively by the test scenario for every specific problem.

4. CRITICAL PROBLEMS

This section discusses the performance degradations caused by oversubscription and false sharing in the parallel software development. These will, of course, lead to correct program results and also to a certain progress in the program execution. However, by false assumptions and a wrong way of thinking belonging the development of parallel software for these multi-core systems, the performance potential given by the hardware cannot be fully exploited.

4.1. Oversubscription

Since the operating system is responsible for creating and managing the threads, this circumstance is very important for the performance of a parallel application (Akhter and Roberts 2006).

Depending on the concrete implementation of the scheduler, the user simply misses the desired performance in a program.

Oversubscription is a problem that occurs when there are more active threads – which are managed by the operating system and referred as software threads – than

hardware threads, which define the number of virtual processors (Akhter and Roberts 2006).

Most of the operating systems use the time slicing technique, also known as round-robin scheduling, which negatively influences the situation of oversubscription. Although blocked threads are not part of the active round-robin scheduling, since these threads are dequeued from the waiting queue, the problem still exists. This scheduling technique leads to context switches of all active threads in which the footprint of the interrupted process, such as the used registers are saved and the register data from the next thread that has received a time slice, is loaded (Tanenbaum 2009).

Since each thread gets a time slice, none of them suffers from starvation. However, the context switch causes a certain overhead by recovering and storing the thread conditions, which results – when having a high number of software threads – in poor performance (Akhter and Roberts 2006).

Not only by the context switches, but also at memory-level, oversubscription causes problems. Processors try to use the caches for storing frequently used data to avoid accessing the slower main memory. But as the cache memory is – compared to the main memory – very small, oversubscription forces to outsource the data from previous time slots to a slower memory. Accordingly, this critical problem forces the threads to compete for the cache memory, which in turn affects the performance of the program badly (Akhter and Roberts 2006).

4.1.1. Simulation

To reproduce the oversubscription problem, the following test scenario has been modelled:

The goal is to measure the time needed to count up a variable from zero to a user-specified limit in the main thread. During this counting up, a user-defined amount of worker threads calculate the Fibonacci numbers. By this simulation, we can measure how the worker threads impact the task of the main thread.

4.1.2. Measurement Analysis

A summary of the measurement results testing up to 70 threads is shown in Table 1, supplemented by the execution time and the speedup – a value for describing the ratio between the serial and multithreaded execution time. The results indicate that although the workload of the main thread does not change, the run-time behavior starts to get worse at 8 active threads.

Table 1: Summary of results of the oversubscription test

Run-time behavior of the test scenario		
Threads	Run-time [sec]	Speedup
1	0.0600001	1
8	0.155009	0.387074944
17	0.35402	0.169482233
25	1.00306	0.05981706
33	2.3088	0.025987569
49	5.60041	0.010713519
57	7.69081	0.007801532
65	10.3428	0.005801147
70	12.2304	0.004905817

Since 8 logical processors exist on the test system and hence 7 worker threads and the main thread can run simultaneously, oversubscription starts to occur only at 8 worker threads. This in turn describes the increase of the run-time exactly at this number of active threads. The graphical representation of the simulation results is shown in Figure 2. The horizontal axis shows the number of active threads and the vertical axis represents the execution time in seconds. The diagram shows that the execution time progressively rises up in a curve for an increasing number of threads. A detailed view is illustrated for better readability in Figure 3, which shows the execution time up to 10 threads, where the run-time is almost nearly constant up to 7 threads.

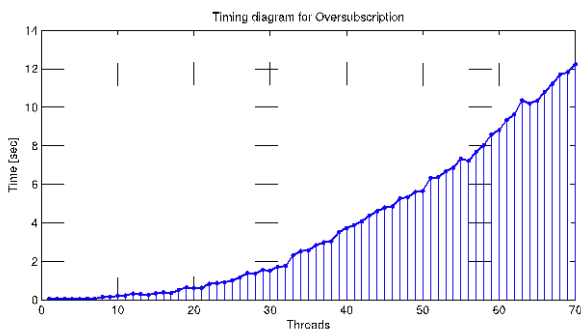


Figure 2: Run-time behavior of the oversubscription test scenario illustrated in a diagram.

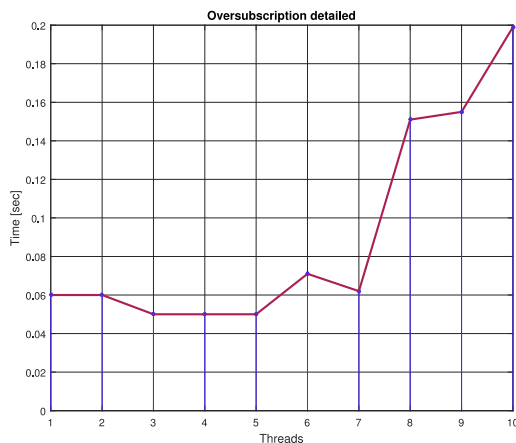


Figure 3: Detailed view of the execution time up to 10 threads.

To analyze what exactly happens when too many threads are active, a simulation with 40 threads has been started in the Concurrency Visualizer. The result illustrated in Figure 4 shows the various thread states at program execution. It shows that 77 % of the execution time is spent for the preemptions caused by the context switches and only 21 % of the run-time is actually used for the execution of the worker threads.

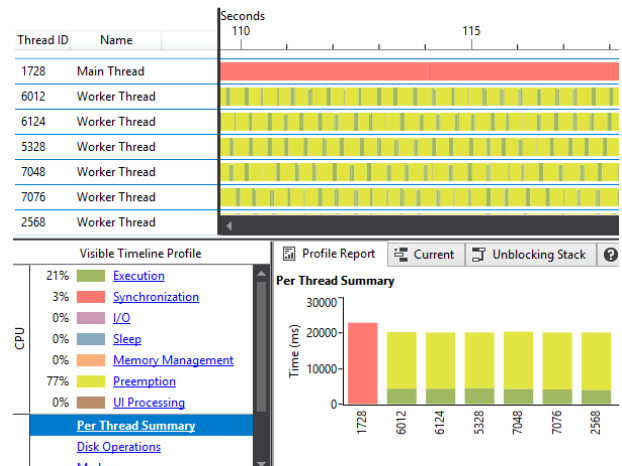


Figure 4: Concurrency Visualizer analyzing the test scenario for oversubscription.

4.1.3. Problem and Solution

The simulations show, how drastically the round-robin scheduling technique affects the run-time behavior. Generally spoken, oversubscription can be eliminated by setting the maximum amount of active threads to the actual number of logical processors.

Although, this does not affect the implemented test scenario, but a reason for an intentional or unintentional high number of active threads used in a program can also be due to poor load distribution. To avoid the side effects of the time slicing technique one can make use of the so-called work-stealing technique. The aim of this technique is to avoid unloaded threads as much as possible. The specific procedure based on (Hwu 2011) is shown in Figure 5 as a flow chart.

In the illustration, all the threads have their own work-pool. If a thread has already processed a task, then it tries to take another one of its own tasks. If during run-time new subtasks have been created, then these are also stored into the thread's collection. Since the subtasks do often share data with the main task a better cache utilization is achieved (Hwu 2011).

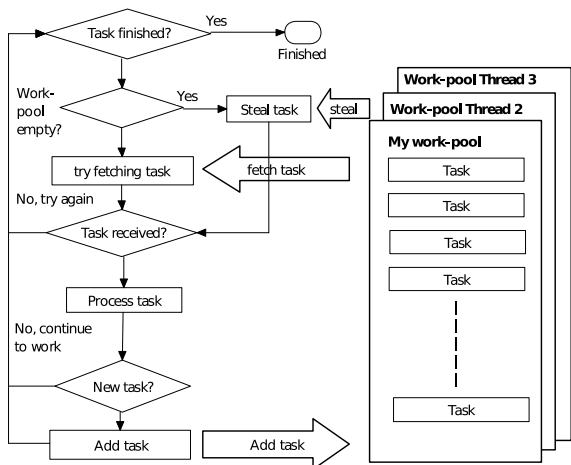


Figure 5: Flow chart illustration of the work-stealing technique.

If a thread has already executed all tasks in its own collection, then it tries to “steal” tasks from another thread by randomly looking into the collections of other threads. Since this search of tasks causes a certain amount of effort, it is the goal of this technique to provide the “thieves” big tasks to keep them busy as long as possible.

4.2. False Sharing

Due to the special hardware architecture of multi-core systems and the limited amount of memory, performance bottlenecks can also lead to problems on software-level or operating system level, and also on the hardware-level.

The cache offers a very fast memory for outsourcing frequently used data of the processors. Since the modern cache-memories are well suited for sequential programming, these fast memories are one of the main causes for substantial side effects in the parallel software development. One of these effects respectively problems is cache line ping-pong (Akhter and Roberts 2006).

This problem occurs when multiple hardware threads try to perform operations on data in the same cache line. Therefore, it can happen that during an ongoing operation on two or more cores, the cache line gets copied like in a ping-pong game between the different cores of the processor through the memory bus (Akhter and Roberts 2006).

One of the causes of cache line ping-pong is false sharing. Since the cache line is the smallest storage unit in data transfer through the memory bus, multiple hardware threads have often a local copy of exactly the same cache line. This does not represent an issue when only a reading access is performed on the cache-line, but is very critical on write operations (Gleim and Schüle 2012).

Although, in false sharing each hardware thread works in disjoint memory positions on their own copy of the cache line, all changes get invalidated in all copies to maintain the cache coherency (Gleim and Schüle 2012).

Thus, the cache coherence protocol ensures that the cache line gets exchanged between the cores on every change of the cache line.

For this purpose, the modified line has to be written back to the main memory, so it can be re-copied to all cache memories of the other hardware threads (Tian and Chiu-Pi Shih 2012).

This sequence is shown with reference to (Tian and Chiu-Pi Shih 2012) in Figure 6:

Two hardware threads, T1 and T2, load a cache line from the shared L2-cache in their own L1-cache. Thread T1 modifies a part of the cache line. This line gets invalidated by the cache coherency protocol and set to dirty (Tian and Chiu-Pi Shih 2012). Thread T2 wants to change an element in a disjoint memory location in the same copy of the cache line, but to maintain the cache coherency, the cache line from the core 0 is copied through the memory bus into the shared L2-cache. T2 loads the updated line from the L2-cache in its own L1-cache and writes its data into the memory location (Tian and Chiu-Pi Shih 2012).

False sharing, therefore, strains the memory bus, although looking at the source code no dependencies of the data can be seen at the first glance.

Furthermore, the advantage of using the cache for fast data storage gets nullified, whereupon subsequently there is a substantial performance loss (Gove 2010).

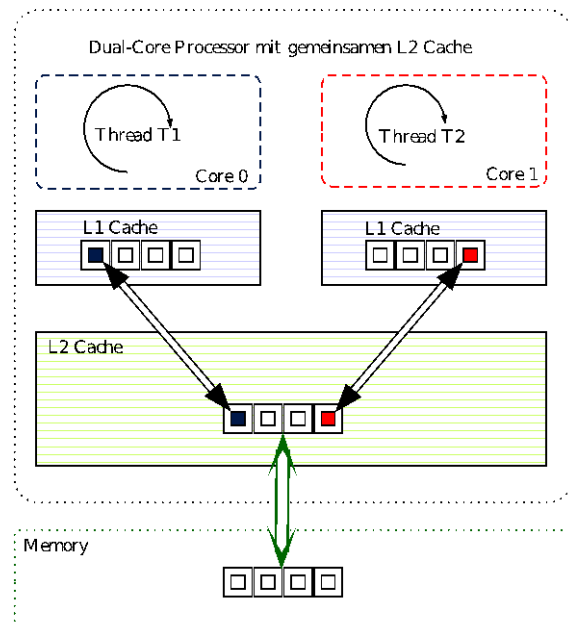


Figure 6: Visualization of the false sharing problem on a dual-core processor with shared L2-cache.

4.2.1. Simulation

In order to reproduce the false sharing problem, the created threads have to share the same cache line but work in different memory locations. For this reason, the following scenario has been modelled:

Several robots in a factory which are simulated by threads can only move in a rectilinear direction, which are controlled by an external program. This controlling

program fetches at certain intervals the current distance compared to the last measurement and stores for each robot the received data into disjoint memory locations into a random access container. In case of failure, these received values are used to reset all the robots back to their starting position, so the robots can proceed with their task again.

The data type for the distance is *int64_t* which has the size of 8 bytes on the test system. The cache line size is 64 bytes. Thus, up to 8 threads can work in the same cache line.

4.2.2. Measurement Analysis

In Table 2 a summary of the run-time behavior and the speedup can be found. The measured values show that the run-time increases linearly with the increasing number of threads, although the threads – of a semantic point of view – work on independent variables having write access to their memory locations.

This worsening of the run-time is also shown in Figure 7, in which on the horizontal axis the number of threads and the run-time in seconds on the vertical axis is plotted.

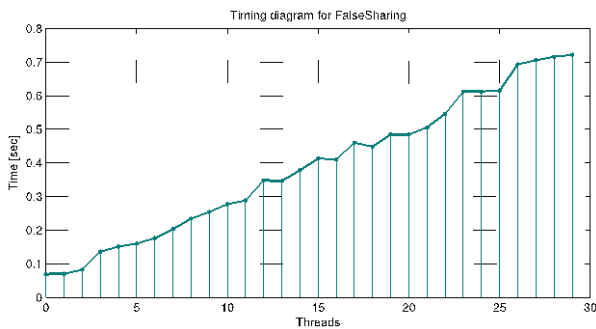


Figure 7: Run-time behavior for the false sharing test scenario illustrated in a diagram.

Table 2: Summary of results of the false sharing test

Run-time behavior of the test scenario		
Threads	Run-time [sec]	Speedup
0	0.0690473	1
1	0.070049	0.98570001
2	0.0820556	0.841469686
4	0.151098	0.456970311
7	0.203136	0.339906762
8	0.234155	0.294878606
10	0.277185	0.249101863
20	0.484326	0.142563686

4.2.3. Problem and Solution

The biggest disadvantage with false sharing is that it is not easily recognizable at the first glance. Figure 8 shows the measurement results of the Concurrency Visualizer. The profiler indicates that the program is up to 72 % busy executing the program and the remaining 20 % of the run-time is lost due to the console output through the main thread. Even further analysis with the

Concurrency Visualizer still does not give any hints about false sharing.

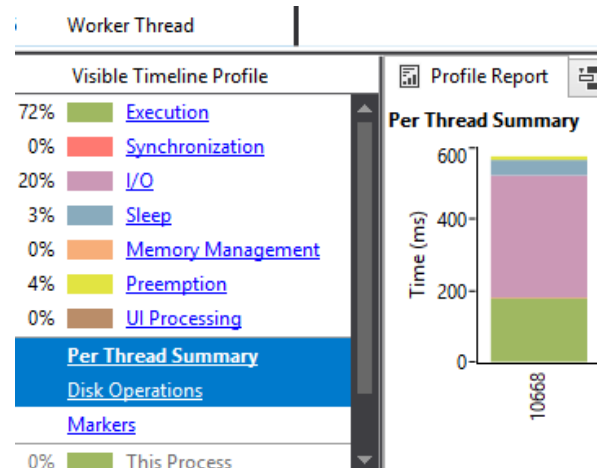


Figure 8: Results of the Concurrency Visualizer does not give any hints for false sharing.

Therefore, there is still the possibility to make use of the so-called hardware performance counters which are supported by many modern processors. Using suitable tools like the Visual Studio profiling tools, this counter can be used to determine the amount of cache misses. A high value of cache misses indicates false sharing.

The general solution for false sharing referring to the simulated test scenario is the following:

The implemented example uses a container of type *std::vector*, in which each thread respectively robot has write access to an element in the container. But each element is not large enough to contain a whole cache line, which in turn results in the cache line being constantly copied between the hardware threads on every change to preserve the cache coherency. For this reason, the software developer must take care of that each thread gets its own cache line for its changes (Gleim and Schüle 2012).

The solution to fix this problem is illustrated in Figure 9. In the implemented example, the basic data type of the container has been *int64_t*, which comprises 64 bits respectively 8 bytes on the test system. Since the test system has a cache line size of 64 bytes, to solve the problem we have to use the technique of spacing, in which we have to enlarge the container at least by a factor of 8. Thereafter, each thread can change an element with the multiplicative increment of 8 and with the ascending index of the threads. So the threads can access their own cache line without getting in conflict with the other threads.

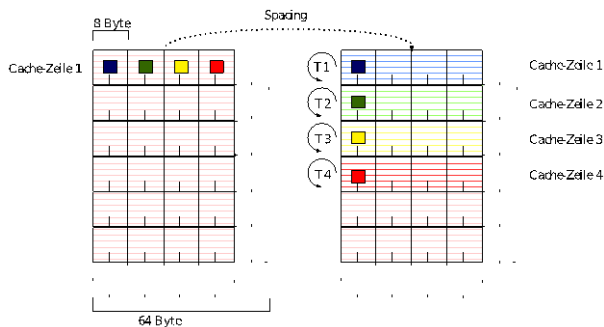


Figure 9: One possible way to solve the false sharing problem by using the spacing technique.

5. CONCLUSION

The detailed problem analysis of the selected multi-core specific problems has on the one hand shown that the well-documented problems of the former multi-processors are still applicable also for multi-core processors. On the other hand, it has to be remarked that it is not sufficient to only know and avoid general problems such as deadlocks and race conditions. The software developers need also to know about the processor, software and operating system used to avoid unconscious performance bottlenecks. They have to be familiar with modern parallel software development and must also have in-depth knowledge distributing tasks to multiple threads. Furthermore, the developers must be able to identify data dependencies and have to skillfully coordinate the involved threads by using appropriate synchronization techniques. The development trend of multi-core processors also shows that the number of such processors will increase further in the coming years. Thus, the developer must also learn to cope with this trend, so that the application does not only scale well with the problem size, but also with the increasing number of cores on such processors.

REFERENCES

- Akhter S. and Roberts J. 2006. Multi-Core Programming: Increasing Performance through Software Multi-Threading. USA: Intel Press.
- Blumofe R. and Leiserson C., 1999. Scheduling multithreaded computations by work stealing. *Journal of the ACM (JACM)* 46:720-748.
- Duffy J., 2008. *Concurrent Programming on Windows*. Boston: Addison-Wesley.
- Gleim U. and Schüle T. 2012. *Multicore-Software: Grundlagen, Architektur und Implementierung in C/C++, Java und C*. dpunkt. Heidelberg: dpunkt.
- Gove D., 2010. *Multicore Application Programming: For Windows, Linux, and Oracle Solaris*. Crawfordsville: Addison-Wesley Professional.
- Hwu W.W., 2011. *GPU Computing Gems, Volume 2*. Boston: Elsevier.
- Rauber T. and Rüniger G. 2012. *Parallele Programmierung*. Heidelberg: Springer-Verlag.
- Simple high resolution timer in C++ Available from <https://gist.github.com/gongzhitaa/7062087> [March 2015]
- Tanenbaum A.S., 2009. *Moderne Betriebssysteme*.

München: Pearson Deutschland GmbH.

- Tian T. and Chiu-Pi Shih., 2012. *Software Techniques for Shared-Cache Multi-Core Systems*. Available from <https://software.intel.com/en-us/articles/software-techniques-for-shared-cache-multi-core-systems> [July 2015]
- Torrellas J., Lam H.S., Hennessy J.L., 1994. False sharing and spatial locality in multiprocessor caches. *IEEE Trans. Comput.* 43:651–663.
- Williams A., 2012. *C++ Concurrency in Action: Practical Multithreading*. Shelter Island: Manning.
- Yanyong Zhang Y., Franke H., Moreira J., Sivasubramanian A., 2003. An integrated approach to parallel scheduling using gang-scheduling, backfilling, and migration. *IEEE Trans. Parallel Distrib. Syst.* 14:236–247.

BALANCING OF A MANUFACTURING WORKSTATION, USING A SIMULATION AND OPTIMIZATION MODEL

Chaves Almanza F.D.^(a), Wellens A.^(b)

^{(a),(b)}Facultad de Ingeniería, UNAM-MÉXICO

^(a)fabiodaniel@comunidad.unam.mx, ^(b)wann@unam.mx

ABSTRACT

This paper describes the balancing of a manufacturing workstation in a metalworking industry, using simulation to predict the productivity of a workstation consisting of ten pressure welding machines, and afterwards optimizing the operation time, by assigning welding tasks to the different machines depending on the demand requirements. The operation time depends on the number of welding points that are achieved at the time, thus machine productivity is mainly related to the configuration required for each new product. Production and engineering teams of the facility are interested in knowing what would be the best workstation performance for future projects, considering possible machine settings. A simulation model was developed to get estimates of future machine productivity, obtaining the required parameters to optimize the workstation's overall operation time by means of an optimized task schedule.

Keywords: machine productivity, optimization, assignment problem, simulation

1. INTRODUCTION

Welding is the joining of metals applying heat and/or pressure, with or without the addition of an appropriate metal. Specifically, pressure welding implies the application of both heat and pressure, resulting in a macrodeformation of the base material to produce coalescence. For a detailed review of pressure welding see Jeffus (2011). This process is applied by the metalworking facility where this study took place, for manufacturing metal parts for the automotive industry, specifically for chair components. The facility deals with major automotive clients who often have high levels of customer demands. To fulfill these requirements the factory has a workstation consisting of ten pressure welding machines, similar between them but coming from different manufacturers and with varying size and/or possible machine settings.

As car design changes constantly, car components have to be adjusted accordingly by design and engineering specialists of the automotive companies. These new parts are to be manufactured by the metalworking factory, and for each part the pressure welding machines need to be configured differently. This includes the design and creation of a specific guiding tool or production mold for each new part; its

specification and cost has to be included in the project description.

The ten welding machines have slightly different specifications, and not all new automotive parts can be manufactured indistinctly in each of the machines, basically because of machine size constraints. Depending on client requirements and part size, generally two or three different machine configurations can be considered. For example, to join two screws to a foil part it is possible to operate the machine once (fixing both screws at the same time) or two times (fixing the screws one by one in two independent operations), depending on the size and position of the foil part and the location of the machine's adjusting tools to fix the foil. Of course, doing the operation in two independent steps will include each of them part handling, and overall operation time will increase and productivity will decrease consequently.

In order to provide decision criteria that can be used by the metalworking facility, the present study proposes the best workload assignment for the ten pressure welding machines, obtaining thus the optimal operation time. To achieve this, machine productivity was simulated in PROMODEL® for different setup scenarios, using as a basis historical production data (3 months) for similar parts. Afterwards, a linear programming model was applied to optimize production task assignment in order to obtain the smallest overall operation time and thus generate savings for the metalworking facility.

2. CONCEPTUAL APPROACH

The present research was based on the approach of Law (2015) to develop a simulation study. This approach involves a complete methodology that includes not only the construction of a simulation model, but also other important stages to develop a complete study. The set of stages defined by Law is: i) formulate the problem, ii) collect information and data and build a conceptual model, iii) validate the conceptual model, iv) program the model, v) validate the programmed model, vi) decide, make and analyse experiments, and vii) document and summarize the simulation results. Law is one of the most recognized authors in the field of simulation and his methodology is known to be consistent and effective in simulation studies all over the academic ambit.

In the present study, simulation results are used to construct a linear programming model that optimizes overall operation time. Because of the conditions of the problem studied, the linear programming model is based on the classical assignment problem. This problem is one of the fundamental examples of combinatorial problems in operations research, and it consists of finding a minimum weight perfect matching in a weighted bipartite graph (Brualdi, 2006).

In this study the instance of the problem considers ten metallic parts to be spot welded in ten pressure welding machines, working in parallel and without sequential operations. One of the parts involves a demand of 316,000 units, whereas other four parts each have demands of 32,000 units. The remaining five parts require manufacturing 79,000 units each, completing a total production of 839,000 units for the ten parts; it takes about 14 months to be completed in non-optimal conditions. The complete production of each part must be assigned to only one machine, depending basically on sizing and shape constrains and considering a specific required machine configuration for each part. Many variations of the original problem have been studied. The most recognized should be the cost associated problem, quadratic assignment and the inverse assignment. Furthermore, about the solution methods, the Hungarian algorithm is one of many algorithms that have been devised that solve the linear assignment problem within time bounded by a polynomial expression. Other algorithms include adaptations of the primal simplex algorithm, and the auction algorithm. See Burkard et al. (2012) for a complete review of the assignment problems. The next section presents the state of the art of current applications of the assignment problem, simulation studies and the use of simulation and optimization together.

Therefore, the methodological approach in the present research is based on the preceding concepts, used for solving a problem in the context of the metalworking industry. This approach is summarized in Figure 1.

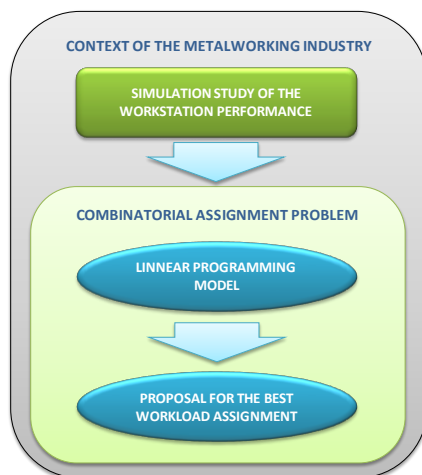


Figure 1. Methodological approach

3. LITERATURE REVIEW

As simulation has been one of the most important available methodologies to investigate systems behaviour, nowadays it is extensively used as a cutting edge technology in critical sectors as defense, airspace, industry and supply chain (Longo, 2011).

Smith (2003) resumes 30 years of simulation applications. In his survey, he reviews and classifies literature on the use of discrete event simulation for manufacturing systems design and operation problems. Similarly, the recent survey published by Jeon and Kim (2016) presents the state of the art of simulation modelling techniques in production planning and control (PPC), from 2002 to 2014. The authors classify the applications reviewed in three types of simulation techniques (system dynamic, discrete event simulation and agent-based simulation) and eight PPC issues (facility resource planning, capacity planning, job planning, process planning, scheduling, inventory management, production and process design, purchase and supply management).

Some of the most recent applications described in literature on assignment problems in supply chain and industry include the work of Ruiz et al. (2012), who proposes a decision-making model that optimizes allocation of demand across a set of suppliers considering expected losses associated with delivery defaults, purchasing costs, unforeseen costs of non-delivered supplies, and supplier management costs. Pérez Becerra (2006) develops a generic model for assigning orders to a set of machines, also applicable to the assignment of workers or materials. Salazar et al. (2013) propose a two-phase approach based on quadratic-quartic assignment and AHP (Analytic Hierarchy Process) as an integrated methodology in flexible manufacturing environments to solve the problem of cell formation and facility layout in the metallurgical sector. Medina et al. (2009) use heuristics for solving the problem of forming lots in a flexible manufacturing system, for a specific assignment of operations and tools to machines.

Finally, it is interesting to review some of the most recent researches that use simulation and optimization together as complementary methodologies. Many applications of this type found in literature are related to the chemical industry, for instance the work of Scenna et al. (1999), Puigjaner et al. (2006), and Rodríguez et al. (2005). Also in finance there are many applications like portfolios selection (Torres, 2004) and development of financial models (Gutiérrez Carmona, 2008). Other applications include the simulation and optimization of a gas separation and stabilization factory (Pan-Echeverría et al., 2009), a coloured Petri net based methodology for the optimization of logistics systems through simulation (Narcizo et al., 2005), the optimization of clinical trials of medication by means of simulation of discrete events (Monleón, 2005), and the development of an air-cooled high efficiency absorption machine for its use in cooling houses or transportation vehicles, such as trucks, busses or boats (Marcos,

2008). Fu, Glover and April (2005) describe the most important approaches in the use of simulation and optimization together, proposing a multi-category classification framework and a survey on related software available.

In the local context of the Mexican industry, where the present study takes place, simulation and optimization have been used in industry and logistics applications like vaccines distribution (Lara, 2016), supply chain optimization (Villareal, 2011), optimal ambulance allocation (Vindel, 2015) and production planning (García, 2013).

4. THE SIMULATION AND OPTIMIZATION MODEL

Following the simulation methodology, the performance of the pressure welding workstation was conceived in the conceptual model shown in Figure 2. This preliminary model was validated in the facility with the production and engineering teams, concluding that the model is an appropriate representation of the real workstation performance.

Foils and other parts that are to be assembled are stored in a materials warehouse. The workstation is nearby the warehouse, so the supply of material can be considered immediate. The complete demand of one part is provided in small quantities from the warehouse to the workstation. Each kind of part supplied to the workstation must specify the machine to be used, as machine settings are specific for each part. The ten machines can work simultaneously. As an important assumption for the simulation, a generic batch of product is considered to be a unique entity, which has the associated attributes “kind of product” and “batch size”. The latter corresponds to the demand of the part. Therefore, considering a project that consists of processing 10 parts, the production cycle ends when all ten batches are processed (one batch per part). One production cycle corresponds to one run in the simulation.

In order to determine the probability distributions of the machine productivities needed in the model programming, 3 months of historical data for the manufacturing of similar parts was used. The main measure of interest is the productivity of each machine, which is to be analyzed in different feasible scenarios for the machine configuration. Another important performance measure is the average batch time of the system, which permits to validate the model through a comparison test for population means. It is also necessary to review the overall operation time in each scenario, as earnings increase when this time is minimized.

Once all interested parties validated and accepted the conceptual model, simulation model programming took place with the PROMODEL® software, installed in a Samsung Serie 7 Chronos laptop with an Intel Core i5-3210M CPU 2.50 GHz processor. Eleven locations were used (10 welding machines and 1 materials

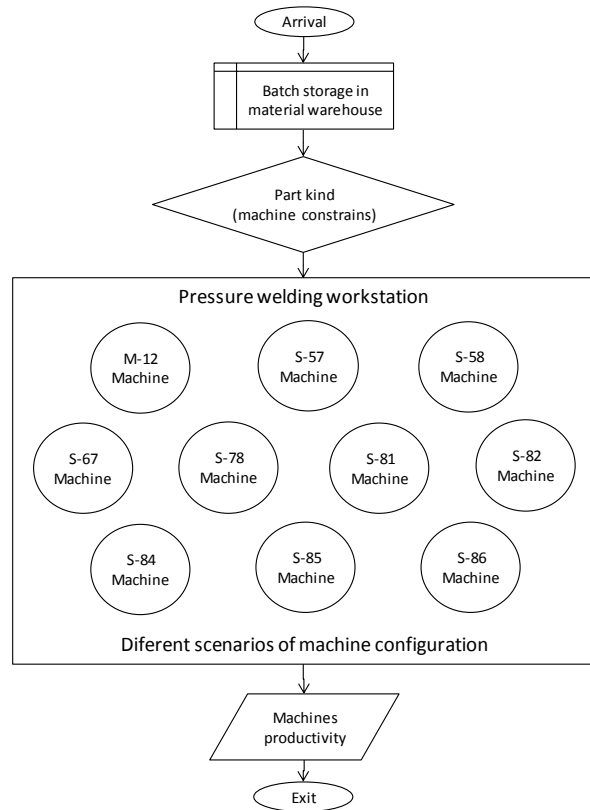


Figure 2. Conceptual model

warehouse), one entity (product batch), five attributes (3 for the entity and 2 for the operation locations), eighteen variables, 12 resources (10 machine operators and 2 setup operators), one external file (work shift), and transversely the programming logic according to the author’s approach and experience.

Each probability distribution included in the model associates an independent chain of random numbers tested on independence, homogeneity and goodness of fit, using software StatFit®. Facility’s work shifts were designed using PROMODEL’s Shift Editor, a complementary tool that generates an external .sft file linkable to the model through the shift assignments function.

Verification of the programmed model permitted to review its correspondence with the proposed conceptual model and its assumptions, and if the study objectives in the facility were being achieved. Additionally, the simulated average batch time in the system was tested against historical data for the statistical validation of the model. For historical data, sampling standard ANSI.ASQC Z1.4 was used to establish the sample size required for the test, bearing in mind a general level II inspection (ANSI, 1993). For model data, the number of replications was used as the sample size. The number of replications was determined by means of the expression (1):

$$n = \left(\frac{t_{(n-1, \alpha/2)} S}{e} \right)^2 \quad (1)$$

where $t_{(n-1, \alpha/2)}$ is the critical t-student value for $n = 10$ and $\alpha = 0.05$, the size of a preliminary sample and the significance level, respectively. S is the standard deviation of the preliminary sample and e is the estimation error margin defined by the engineering and production teams of the facility. The number of replications was determined to be 15 while the historical data sample size was 20. A t -test for comparison of population means was applied assuming corresponding variances as unknown and different. Table 1 presents the test results; sample averages are given in working days.

	FACILITY MODEL	
Average	44.900	43.533
Variance	215.989	28.838
Observations	20	15
Degrees of freedom	25	
t -statistic	0.383	
P(T<=t), one queue	0.352	
Critical value of t (one queue)	1.708	
P(T<=t), two queues	0.704	
Critical value of t (two queues)	2.059	

Table 1. Test t of means for two samples assuming unknown and different variances

As the absolute value of the t -statistic is lower than the critical t -value, there is no statistical evidence that the population means of historical and simulated data are different, so they are assumed to be equal. In conclusion, the simulation model can be considered a valid representation of the welding workstation.

In the second stage of the study, the linear programming model proposed for optimizing the overall operation time, is defined in terms of a binary variable x_{ij} that equals 1 if the i -th part is assigned to the j -th machine, and equals 0 in any other case, as follows:

$$\text{Minimize } Z = \sum_{i=1}^m \sum_{j=1}^n \frac{d_i}{c_{ij}} x_{ij} \quad (2)$$

Subject to:

$$\sum_{i=1}^m x_{ij} = 1 \quad (3)$$

$$\sum_{j=1}^n x_{ij} \leq 2 \quad (4)$$

where c_{ij} is the productivity of the j -th machine to process the i -th part, simulated in the first stage of the study, d_i is the demand required for the i -th part, n is the number of welding machine in the workstation and m is the total amount of parts to be manufactured. Expression (3) is the constraint that ensures that the

complete production of one part is assigned to only one machine, and the constraint given in expression (4) ensures that the production of maximum two parts is assigned to each machine, as required in the facility. In cases where the assignment is restricted because not every machine j can process every part i , the corresponding machine productivity c_{ij} is a huge value assigned randomly.

5. EXPERIMENTS AND RESULTS

Reproducing in the simulation model the production scheme according to presently observed assignation rules based on experience, the machine productivities obtained are in a range between 1.67 and 5 units per minute, with an average of 2.64 units per minute; this scenario is called the *present situation*. The minimum overall operation time obtained for this scenario is 3643 man-hours, which corresponds to 405 working days, bearing in mind one daily shift, as required by the facility.

The facility was interested in analyzing different scenarios to configure the welding workstation for the production of the ten parts considered. On one hand, a new supplier of welding molds offered a different configuration, increasing in 20% the productivity of machines S-58 and S-67 but also decreasing in 50% the productivity of machine S-78. On the other hand, the engineering team was considering taking a guiding tool from machine S-84 and installing it on machine S-82, so more screws could be fixed per unit in machine S-82, which would increase its productivity in 33% but would also decrease the productivity of machine S-84 in 50%. The facility considered a third alternative, combining both supplier and engineering staff proposals. These three feasible scenarios were analyzed in the simulation model by adjusting the parameters in the corresponding probability distributions. The summary of the scenarios is presented in table 2.

Scenario	Increase (%)	Machines	Decrease (%)	Machines
Supplier proposal	20%	S-58, S-67	50%	S-78
Engineering proposal	33%	S-82	50%	S-84
Combined proposal	20%	S-58, S-67	50%	S-78, S-84
	33%	S-82		

Table 2. Feasible scenarios for workstation configuration

Table 3 summarizes the estimations of the machine productivities obtained by simulating the present and the proposed scenarios. The range and average values observed for all machines and all replications are presented.

Scenario	Productivity range (units/min)		Average machine productivity (units/min)
	Min	Max	
Present situation	5	1.67	2.64
Supplier proposal	6.25	1.67	2.67
Engineering proposal	5	1.25	2.50
Combined proposal	6.25	1.25	2.52

Table 3. Machine productivities: range and average

These results are illustrated in figure 3. The scenario proposed by the new supplier implies the largest range of machine productivity, as well as the maximum average.

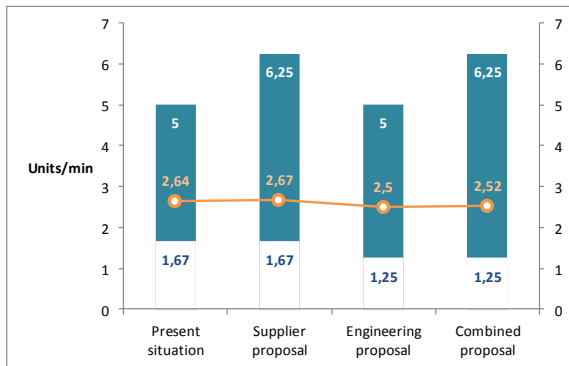


Figure 3. Machine productivities: range and average

Although optimal productivity is achieved for the supplier proposal, average batch time and overall operation time were not considered in the corresponding analysis. An additional analysis based on these performance measures was carried out to be able to decide which is the best scenario. The overall operation time and the average batch time obtained for the present situation and the feasible scenarios are presented in table 4, both in man-hours and working days. The same information is presented graphically in figure 4.

Scenario	Overall operation time		Average batch time	
	man-hours	working days	man-hours	working days
Present situation	3643	405	364	40
Supplier proposal	3581	398	358	40
Engineering proposal	3537	393	354	39
Combined proposal	3581	398	358	40

Table 4. Overall operation time and batch averaged time in system

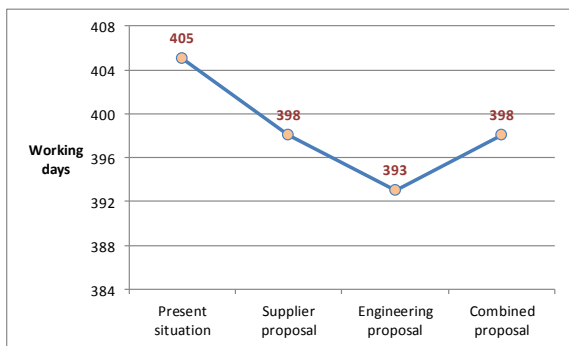


Figure 4. Overall operation time

As figure 4 shows, the optimal overall operation time is obtained in the scenario proposed by the engineering team, which achieves a minimum of 393 working days for the complete production of the 10 parts. Therefore, this alternative represents the best savings for the facility.

Although the facility considered the combination of the supplier proposal with the engineering proposal as an

enhanced scenario, the assignment model demonstrates that the overall operation time obtained in the mixed scenario is the same as the time obtained in the single supplier proposal. The engineering proposal implies a minimum overall time, and it is cheaper to carry out.

Additionally, even though the supplier proposal involves the highest machine productivity of all proposed scenarios, the best alternative for the welding workstation is the machine configuration suggested by the engineering team due to the best performance of the overall operation time, even while machine productivity was the lowest in this scenario. These results are complemented by the performance of the average batch time in the system (figure 5), which has a minimum value of 39 days for the engineering proposal.

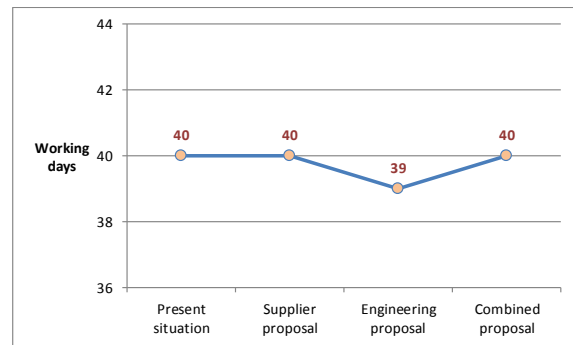


Figure 5. Average batch time in the system

As the average batch time in system corresponds to a central tendency measure, its value is found to be very similar for the different scenarios; still, it presents an optimum value in the engineering proposal, corroborating this scenario as the best choice for the facility.

CONCLUSIONS

The present simulation and optimization study in the metalworking facility conducted successfully to a simulation model that represents the pressure welding workstation's behavior, and an assignment model that optimizes the most relevant performance measures for the facility. The combination of simulation and optimization offers important quantitative and qualitative technical criteria for the facility's decision making.

The present situation in the welding workstation revealed a non-optimal assignment of the production requirements, because the assignment planning was based only on empirical knowledge. The proposed methodology permitted identifying the assignment scenario that offers the best benefits to the facility.

With the successful culmination of this project, the advantages of using simulation were confirmed, as the cost associated with this simulation and optimization study is lower than the cost for implementing the scenarios in the real system.

REFERENCES

- ANSI 1993. Sampling procedures and tables for inspection by attributes ANSI/ASQC Z1.4-1993. Wisconsin: American national standard institute - American society for quality control.
- Brualdi R., 2006. Combinatorial matrix classes. Encyclopedia of Mathematics and Its Applications 108. Cambridge: Cambridge University Press.
- Burkard R., Dell'Amico M., Martello S., 2012. Assignment Problems. Revised reprint. Society for Industrial and Applied Mathematics.
- Fu M.C., Glover F.W., April J., 2005. Simulation optimization: a review, new developments, and applications. Proceedings of the 37th Winter Simulation Conference, pp. 83-95. Dic 4-7, Orlando (FL, USA).
- García J.A., 2013. Metodología para la planeación de la producción utilizando simulación y programación entera. Magister dissertation. Universidad Nacional Autónoma de México.
- Gutierrez Carmona, J., 2008. Modelos financieros con Excel. Bogotá. 2a Ed. Colombia: Ecoe Ediciones.
- Jeffus L., 2011. Welding: principles and applications. 7th Ed. Delmar, USA: Cengage Learning.
- Jeon S.M., Kim G., 2016. A survey of simulation modeling techniques in production planning and control (PPC). Production Planning and Control, 27 (5), 360-377.
- Lara J., 2016. Simulación y optimización para una estrategia de distribución de vacunas contra la A(H1N1). Magister dissertation. Universidad Nacional Autónoma de México.
- Law A., 2015. Simulation Modeling and Analysis. 5th ed. Boston: McGraw-Hill.
- Longo F., 2011. Advances of modeling and simulation in supply chain and industry. Simulation, 87 (8), 651-656.
- Marcos J.D., 2008. Prototipo de máquina de absorción de LiBr/H₂O de doble efecto condensada por aire. Simulación, optimización y resultados experimentales. Doctoral dissertation. Universidad Carlos III de Madrid.
- Medina P.D., Cruz E.A., Restrepo, J.H., 2009. Problema de formación de lotes de fabricación en un sistema de manufactura flexible: Heurística de selección de partes. El Hombre y la Máquina, 32, 68-75. Universidad de Barcelona.
- Monleón T., 2005. Optimization of the clinical trials of medicaments by means of simulation of discrete events, modeling, validation, monitoring and the improvement of the quality of its information. Doctoral dissertation. Universidad de Barcelona.
- Narciso M.E., Piera M.A., Figueras Jové J., 2005. Optimización de sistemas logísticos mediante simulación: Una metodología basada en redes de Petri coloreadas. Revista iberoamericana de automática e informática industrial, 2(4), 54-65.
- Pan-Echeverría G., Gaumer-Araujo T., Pacho-Carrillo D., 2009. Simulación y optimización de una planta de separación y estabilización de gas y condensados. Tecnol. Ciencia Ed. (IMIQ), 24 (1), 66-75.
- Pérez Becerra J.A., 2006. Modelo de asignación aplicado a la manufactura en una cadena de abastecimientos, Bachelor dissertation. Universidad Mayor de San Marcos, Perú.
- Puigjaner L., Ollero P., De Prada C., Jiménez L., 2006. Estrategias de modelado, simulación y optimización de procesos químicos. Madrid: Editorial Síntesis.
- Rodríguez S.L., Vega Granda, A.B., 2005. Simulación y optimización avanzadas en la industria química y de procesos: HYSYS. Oviedo: Universidad de Oviedo.
- Ruiz Torres A.J., Ablanedo Rosas, J.H., Ayala Cruz J., 2012. Modelo de asignación de compras a proveedores considerando su flexibilidad y probabilidad de incumplimiento en la entrega. Estudios gerenciales, 28 (122), 29-48.
- Salazar A.F., Vargas L.C., Añasco C.E., Orejuela J.P., 2013. Propuesta de distribución en planta bietapa en ambientes de manufactura flexible mediante el proceso analítico jerárquico. Revista EIA, 7(14), 161-175.
- Scenna N.J., Aguirre P.A., Benz S.J., Chiotti O.J., Espinosa H.J., Ferrero M.B., Montagna, J.M., Mussati M.C., Pérez G.A., Rodríguez J., Salomone H.E., Santa Cruz A.S.M., Tarifa E.E., Vega J., 1999. Modeling, simulation and optimization of chemical processes. Argentina: Universidad Tecnológica Nacional.
- Smith J.S., 2003. Survey on the use of simulation for manufacturing system design and operation. Journal of Manufacturing Systems. 22 (2), 157-171.
- Torres C.L.M., Munera J.A.R., Henao, J.D.V., 2004. Selección de portafolios usando simulación y optimización bajo incertidumbre. Dyna, 71 (141), 35-57.
- Villareal J.E., 2011. Simulación y optimización de la cadena de suministro con programación lineal entera mixta. Magister dissertation. Universidad Nacional Autónoma de México.
- Vindel J.A., 2015. Localización óptima de ambulancias en Ciudad Universitaria empleando simulación. Magister dissertation. Universidad Nacional Autónoma de México.

AUTHORS BIOGRAPHY

Fabio Daniel Chaves Almanza is an industrial engineer from the National University of Colombia, Bogotá, in 2006, and he received the degree as Production and Logistics Engineering Specialist from the Districtal University "Francisco José de Caldas", Bogotá, Colombia, in 2013. He has achieved 10 years of career in the field of management systems, in private and public sectors. Actually he is a graduate student in Industrial Engineering at the National University of Mexico (UNAM) in Mexico City.

Ann Wellens is a chemical engineer with postgraduate studies in Industrial Administration (KUL, Belgium) and a master degree in Environmental Engineering (UNAM, Mexico). At the moment she is a full-time lecturer in the Systems Department of the Industrial and Mechanical Engineering Division of the National University of Mexico (UNAM). She has been working in air pollution issues for the last 15 years, dictating courses, collaborating in research projects and participating in conferences related with mathematical modelling of air pollution dispersion and statistics.

SERVICE LEVEL DEFINITION FROM FORECAST ERROR COST EVALUATION

Sosa A.^(a), Wellens A.^(b)

^{(a),(b)}Facultad de Ingeniería, UNAM-MÉXICO

^(a)armandososas@yahoo.com.mx, ^(b)wann@unam.mx

ABSTRACT

Although forecast errors can imply important losses, frequently its financial impact in business is neglected; some researchers have studied this issue but it is not being used as a common tool for evaluation of the forecasting process. The present paper describes the assessment of the forecast error cost in order to define the most appropriate service level for an inventory control system, considering a product classification that is able to focus on those articles that need more management attention. A key point is the cost function, which allows to quantify the over forecasting cost represented by excess inventory and the under forecasting cost represented by lost sales.

Keywords: forecasting, service level costs, inventory optimization

1. INTRODUCTION

Forecasting for business is a very important process in any industry; finding a formal method to assess the result is particularly of interest. In order to improve the forecasting process, many metrics have been developed, as for example MAPE, Bias, MAD, tracking signal among others; for a complete overview please see Makridakis (1997).

In the last decade, different authors have implemented approaches where the cost associated to forecast errors was introduced as a metric. Some of these methods are very general, as for example the determination of an associated index in balanced scorecards. Others are more specific; for example, Kahn (2003) takes into account monetary costs for lost sales, operational inefficiency and transport costs for urgent stock replenishments, quantifying the effect of a 1% over or under deviation of the prediction during one year. Götz and Köhler (2006) define a metric to quantify the cost of error based on lost sales and the cost of residual inventory; however, they do not consider safety stock so they find many lost sales as a consequence. Catt (2007) proposes a cost function where holding cost and lost sales are the main components of the cost of forecast error estimation (CFE). Singh (2013) shows a forecast accuracy impact in both financial and operational KPI's, and establishes the relationship between forecast accuracy, customer service and shipping cost and finally simulates the cost of forecast error.

In this paper, the CFE metric proposed by Catt (2007) is preferred as it considers both relevant costs and poor

service, and even when some assumptions are subjective (for example, the percentage of lost sales), it is a very practical approach that has been well received by experts in the field.

A correct forecast is difficult to achieve, as many variables impact the final result and randomness is present at different levels. In an intend to find a realistic approach to assess the financial impact of forecast error that can be used to define the optimal service level, this paper presents a practical guide to get more accurate results in the forecasting process and measure its effect in the company profits.

1.1. Product Classification

A product classification is crucial for making appropriate planning decisions, as the products in a catalogue not necessarily pertain to the same category, and the inventory should be constructed accordingly (Broeckelmann, 1998). The ABC classification is very extended and well known; it is frequently used as service level tag, due to a strong desire to have better service levels for the items at the top of the scale, and leaving lower service levels for C parts. However, it can lack information, as some items with high demands but low costs can be ranked low, while a high cost article with low volume can be ranked at the top. A fixed service level specification by product category does not consider the accuracy of the forecast and therefore can cause unnecessary costs, for example, considering a low service level for a highly predictable B or C part or a high service level for an A or B part that is very difficult to forecast. In this way, the information obtained by the forecast technique can be inconsistent with the service level specification suggested by the ABC classification. To overcome this inconvenient, this paper proposes a wider classification in order to strengthen the link within the information for every item, by adding a third factor, volatility, and combining all three factors. In this way, articles become more relevant if one of these elements is important, whichever the cause is: cost, volume or volatility.

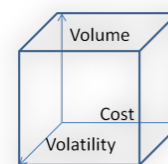


Figure 1: Tridimensional VCV Classification

1.2. Product Life Cycle

In a context of shorter and shorter product life cycles, it is relevant to point out that the proposed methodology is applied to articles in a maturity stage, or at the end of the growth phase, which is when its growth rate is decreasing significantly.

1.3. Volume, Cost and Volatility

Having three factors for the classification of any product in high, medium or low-level items defines 27 different classification options. The *volume* corresponds to the number of parts sold during a given period of time, usually one year. In general, the range in *cost* for different items in a catalogue can be huge; consider for example the difference in cost for a machine and a spare part. The concept of *volatility* is given by a statistical measure, and corresponds to the variation coefficient, which is the standard deviation of the past sales divided by the corresponding mean:

$$CV = \frac{\sigma}{\mu} \quad (1)$$

This measure shows how scattered the data are with respect to the mean, so a high volatility or a high scattering means a potentially less accurate forecast; on the opposite, a low volatility allows a higher accuracy and therefore less forecast error and a lower forecast error cost. Previous studies indicate that a high correlation exists between the volatility and the forecast error, so considering this measure in the final classification can give a clue about the articles that need special attention in order to increase the forecast accuracy.

1.4. VCV Matrix

To be able to use the volume-cost-volatility matrix (VCV matrix – Figure 1) as a standard for comparison, it is necessary to define for each product its level for these three components; normally *low*, *medium* and *high* are used, but depending on the type of facility a different range can be proposed. In this paper, the mean and standard deviation of each set of data for volume and cost will be used to define the ranges corresponding to low (first 2σ), medium ($\mu \pm \sigma$) or high (last 2σ) levels. Limits for volume and cost are defined in terms of business data, while the category range for volatility is dimensionless and is valid in any case. Different authors consider different ranges where volatility can be considered high; in this study a range of 0.2 – 0.6 is considered for medium volatility. Table 1 shows the proposed definition of the ranges for volume, cost and volatility.

Table 1: Ranges for the Definition of the VCV Matrix

	Volume / Cost	Volatility
Low	$x < (\mu - \sigma)$	$x < 0.2$
Medium	$(\mu - \sigma) < x < (\mu + \sigma)$	$0.2 < x < 0.6$
High	$x > (\mu + \sigma)$	$x > 0.6$

2. FORECAST MODELS

Choosing the right forecasting technique is vital for the prediction process. The next step is to compare different models in order to select the best fit for every item. Common models with reliable results are single exponential smoothing or SES, double exponential smoothing or DES and the Winters method which considers all components in demand patterns, being the level component, trend and seasonality. Generally, a naïve method is recommended as a starting point as it can be used without extra costs in readily available spreadsheet software as for example Excel. Its results can be compared with the results of the other methods; any additional forecast effort has to add value to the process (Gilliland, 2010).

An additional step is to optimize the smoothing parameters for every component in the model; for SES the parameter corresponds to alpha, for DES they are alpha and beta, and for the Winters method parameters are alpha, beta and gamma. All of them are optimized by a nonlinear method in order to get the minimum error and therefore the best fit. The objective function is represented by a forecast performance metric like the mean absolute deviation (MAD), which is minimized. In addition, restrictions for non negativity and values below 1 have to be set.

3. METRICS

Different metrics in forecast evaluation are available; calculating more than one allows a better understanding of the forecasting process. Some of the most useful and less biased metrics are presented here, although from a practical viewpoint only one is needed when assessing the forecast; in this case, either MAD or the standard deviation are proposed.

The most used metrics are the forecast accuracy, the bias, the mean absolute deviation and the tracking signal.

3.1. Forecast Accuracy

A measure of forecasting performance that has the useful characteristic of always being scaled between 0% and 100% is the *forecast accuracy*, *FA*. It uses the maximum within the actual or forecast as the denominator (Hawitt 2010). Kolassa and Martin (2011) show that there exists less bias when using the maximum of the actual and forecast instead of using the forecast or actual in the denominator, as the mean absolute percentage error (MAPE) does.

$$FA = \left(1 - \frac{\sum |F - A|}{\sum \max(F, A)} \right) * 100 \quad (2)$$

3.2. Bias

The *bias* is a measure of forecasting performance that indicates whether the forecast is chronically too high (positive bias) or too low (negative bias).

$$Bias = 100 * \left[\left(\frac{\sum F}{\sum A} \right) - 1 \right] \quad (3)$$

If the loss function of overestimating or underestimating is asymmetric, a bias in a given direction would be preferable.

3.3. Mean Absolute Deviation

The *mean absolute deviation*, MAD , is the average of the absolute forecasting errors through time, given as follows:

$$MAD = \frac{1}{n} \sum_{t=1}^n |F_t - A_t| \quad (4)$$

It is the most common metric to evaluate forecasts and its relationship with the corresponding standard deviation has been proved to be a factor of approximately 1.25.

3.4. Tracking Signal

The *tracking signal* for a given period t , TS_t , is the ratio of bias and MAD in this time period and shows when a forecast is out of control. In that case a change in forecast model is recommended.

$$TS_t = \frac{Bias_t}{MAD_t} \quad (5)$$

4. INVENTORY MODELS

Depending on the inventory model definition, different costs are considered relevant; any model can be addressed, but generally EPQ models are used for production environments, while EOQ models with backorder and lost sales are the most appropriate ones for commercial services. It is important to evaluate in which case a customer would be willing to wait for a product which is out of stock and in which cases the customer just cannot wait. For retail companies, costs of inadequate stock can be huge. Corsten and Gruen (2004) report results of a survey of 71,000 retail consumers worldwide. They found that if the desired item is out of stock, 31% of the customers will leave the store to buy it elsewhere, while another 9% will choose not to make the purchase at all; this consideration depends on the brand and type of product, as a commodity can be replaced very easily, but a specialized item cannot. In general, three different situations can occur for any out of stock product demanded by a customer: the customer can wait which means a backorder, the customer cannot wait which means a lost sale and finally the brand or store can substitute the necessity of the customer with a similar product. In this case, the company does not lose a customer but sales records will be biased and thus will not reflect the real needs of the customer.

It is necessary to express the occurrence probability for each of these situations and to distinguish between different classes of items; in this paper, item categories include commodities, specialized articles and intermediate products for which more than one

competitor offers an available solution. A possible probability distribution is shown in Table 2.

Table 2: Probability Distributions of Stock Out Customer Reactions

Item category	Backorder	Lost sales	Substitute
Commodity	0	0.8	0.2
Medium	0.2	0.5	0.3
Speciality	0.8	0.1	0.1

Depending on the type of item (for example, rice vs. iPhone) and the strength of the brand in the market, these probabilities can vary, and an appropriately conducted research at the point of sales will be needed to show how willing clients are to substitute their preferred product.

5. INVENTORY COSTS

The inventory cost can be divided in three different types, being the holding cost, the setup cost and the shortage cost. On the other hand, also a cost due to forecast error can be defined.

5.1. Holding Cost

The holding cost is the cost due to having the item available in stock; it is an opportunity cost for capital tied up in inventory and in this sense it is important not to include in the inventory expensive slow movers whose holding cost could provide a higher return when using an alternative investment. The dominating part of the holding cost usually corresponds to the capital cost. Other costs included are material handling, storage, damage and obsolescence, insurance, and taxes. All costs that vary with inventory level should be included (Axsäter 2006).

5.2. Ordering or Setup Cost

There are usually fixed costs associated with replenishment (independent of the batch size). In production, common reasons are setup and learning costs. Also administrative costs and fixed costs associated with transportation and material handling have to be included and can decrease if order size increases. The setup cost may be especially high if an expensive capacity constrained machine has to stop during the setup. In some situations, though, most of the setup can take place outside the machine, which reduces the costs. When ordering from an outside supplier there are also various fixed costs for an order, like costs for order forms, authorization, receiving, inspection, and handling of invoices from the supplier (Axsäter 2006).

5.3. Shortage Cost

If an item is demanded and cannot be delivered due to a shortage, various costs can occur. There are situations where a customer agrees to wait while his order is backlogged, but also situations when the customer chooses some other supplier. If the customer order is backlogged, there are often extra costs for administration, price discounts for late deliveries,

material handling, and transportation. If the sale is lost, the contribution of the sale is also lost. In any case, this usually means a loss of goodwill that may affect sales in the long run. Most of these costs are difficult to estimate. Shortage costs in production are, in general, even more difficult to estimate. For example, a missing component can cause a chain of negative consequences like delays, rescheduling, etc.

There are also situations when shortage costs are easy to evaluate. Assume, for example, that a missing component can be bought at a higher price in a store next door. The additional cost can then be considered as the shortage cost. Because shortage costs are so difficult to estimate, it is very common to replace them by a suitable service constraint. Of course it is also difficult to determine a suitable service level, but yet it is regarded to be somewhat simpler in most practical situations. Still, the motivation for a service constraint is nearly always some underlying shortage cost (Axsäter 2006).

5.4. Cost of Forecast Error

The cost of forecast error (CFE) was introduced by Catt (2007), in addition to the common metrics that measure the forecast error. This metric measures the impact in cost and indicates if it is more convenient to under- or overestimate the inventory needed to satisfy the customers' requirements, so it can be used to plan the inventory in a more efficient way. The cost of forecast error (CFE) is represented by:

$$CFE = \left(SS * v * r + \frac{B_s * m_p * \sigma \sqrt{L+R} * G_u(k)}{R} \right) * P \quad (6)$$

CFE = annual cost of forecast error, \$

SS = safety stock, units

v = unit cost, \$/unit

r = inventory carrying charge, \$/month

B_s = lost sales margin per unit short

m_p = product margin, \$

$\sigma \sqrt{L+R}$ = standard deviation of forecast error, units over the replenishment cycle

$G_u(k)$ = loss function used to calculate shortages per replenishment cycle

R = review period

P = period multiplier to convert from months to year

It is common to refer to the total inventory as total stock (TS). The total stock is partitioned into two components: cycle stock (CS) and safety stock (SS). The cycle stock is needed to fill the forecast of demands from the customers, and the safety stock is needed in case the demand exceeds the forecast or supply fluctuations.

Assuming a normal distribution for the demand, the security factor can be found as the z-value corresponding to the service level or probability not to stock out. The loss function is than the probability to be out of stock (Table 3).

Table 3: Security Factor and Loss Function as a Function of the Service Level

Service Level	Security Factor (k)	Normal Loss Function
90%	1,28	0,0475
91%	1,34	0,0417
92%	1,41	0,0358
93%	1,48	0,0306
94%	1,55	0,0261
95%	1,64	0,0211
96%	1,75	0,0161
97%	1,88	0,0116
98%	2,05	0,0074
99%	2,33	0,0033

The safety stock is the product of the service factor given a desired service level and the standard deviation of forecast error and lead time:

$$SS = k\sigma\sqrt{L} \quad (7)$$

$$\sigma \approx 1.25MAD \quad (8)$$

6. SIMULATION

Sometimes it can be difficult to choose suitable inventory control methods; for example, if the service level is suddenly changed, it can take weeks before the consequences are observed. If the change turns out to increment unsustainably the inventory levels or provokes lost sales due to missing items, the change in service level can highly impact financial results.

One way to evaluate different control methods without taking them to practice is to use simulation. In simulation, the real system is replaced with a mathematical model. Simulation experiments can be carried out very quickly and can be used for analyzing the system. Using simulation, it is possible to evaluate the system's long time behaviour in a few seconds; in this case, it is easy to make various test runs with different inventory control techniques. A simulation model has, however, the same type of limits as other mathematical models and can never give a complete illustration of the real system.

7. STUDY CASE

The proposed VCV classification and posterior simulation was applied in a specific department of a retail store in Mexico City, where data for the 300 items of the product portfolio of a specific department were available, including two years of historical sales for most of the items. Currently, and according to top management guidelines, the service level is set to 95%. At present, the moving average forecast model is used, which is a very simple model. The retail store is interested in comparing the forecast error cost for their actual policy and wants to establish if a new service level policy is necessary after balancing the cost of overestimating against the cost of underestimating.

At present, the service level policy is the same for each item, although an optimal service level for different types of products may lower overall costs. As a part of the methodology and to focus on some specific items, a classification is proposed based on cost, volume and volatility where H stands for *high*, M for *medium* and L for *low*. The first letter corresponds to *volume*, the second one to *cost* and the last one to *volatility*, so HML stands for example for a product with high volume, medium cost and low volatility. The VCV classification for the items in the study case is presented in Table 4.

Table 4: VCV Classification for the Study Case

VCV Classification	items	MAD/mean ratio	CFE Share	Management attention
MHM	12	28%	24%	special
MHH	8	40%	20%	special
MMM	29	21%	12%	special
HHM	2	20%	7%	special
LHM	16	27%	6%	special
MML	25	16%	5%	regular
HML	8	16%	5%	regular
LMM	37	25%	4%	regular
LMH	16	35%	3%	regular
MLM	33	20%	2%	regular
HMM	2	19%	2%	regular
HLL	14	15%	2%	regular
MMH	3	35%	2%	regular
HHL	1	16%	2%	regular
MLH	14	31%	1%	regular
MLL	35	15%	1%	regular
LHH	6	45%	1%	regular
HLM	6	18%	1%	regular
MHL	1	23%	0%	regular
LML	10	17%	0%	regular
LLH	6	36%	0%	regular
LLM	11	20%	0%	regular
LHL	1	19%	0%	regular
LLL	4	19%	0%	regular
HHH	0	NA	NA	NA
HMH	0	NA	NA	NA
HLH	0	NA	NA	NA
All items	300	20%	100%	

Table 4 includes the number of items in each VCV classification and the corresponding level of management attention required; it also includes the MAD/mean ratio in order to recognize those categories with highest forecast error, and finally the cost share per classification. Medium and low volatilities are observed for more than 80% of the products, making its forecasting less complicated. It can be observed from table 4 that items with high volatility present high MAD/mean ratios (37% on average), being low volume, high cost and high volatility the worst combination for accurate forecasting. However, forecast accuracy is not always the best criterion to determine priorities, as items with low forecast accuracy not necessarily have

the biggest influence on the cost of forecasting. To define the items that require special attention, the CFE share is proposed in addition to the MAD/mean ratio, finding 5 categories (MHM, MHH, MMM, HHM and LHM) that include 22% of the total number of items and represents 68% of the total cost of forecast error; items with these characteristics will require more attention and effort to improve the forecast process.

7.1. Forecast Model Results

The presently used moving average model was compared to an alternative method, being *single exponential smoothing* (SES) the best fit. SES is an appropriate forecast model for the study case data, as a regular pattern for demand without trend and with low volatility can be observed and SES fits well. Moreover, the purpose of this investigation, in addition to the assessment of the results of the forecast technique, is to measure the financial impact of a forecast deviation, so a more complex model like ARIMA can be evaluated in future research.

As a next step, the smoothing factor α for the level component of the historical sales was optimized through a non linear process using solver, minimizing the MAD subject to a non negativity restriction and ensuring it to take values between 0 and 1. In general, the common method of moving averages works well due to the fact that the data shows a constant pattern and a low or medium volatility for most items. The average reduction of MAD using the SES method with regard to the moving average method is of about 12%, which is a considerable improvement. Based on informal comparisons with real demands in the past, the forecast model presently used by the store's administration seems to forecast with acceptable results; however, as no formal comparison or evaluation was ever carried out, the improvement obtained with the proposed method cannot be quantified.

Figure 2 shows the last two years of sales at an aggregated level, showing a steady behaviour, calculating a coefficient of variation of 0.12, being considered low.

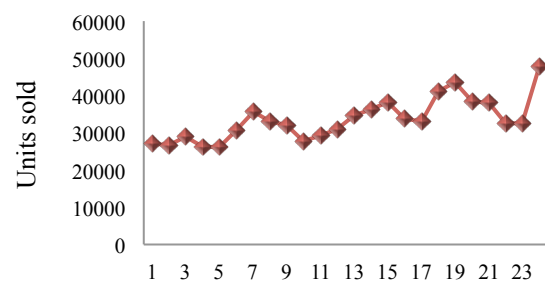


Figure 2: Aggregated Sales, 2013 – 2015.

The final step of the forecasting process is its evaluation. As mentioned before, although different metrics can be used, MAD is preferred for two reasons. In the first place because the current planning system determines MAD as a default output and according to equation (8) there is a proportional relationship between

the amount of safety stock required and MAD, so no additional calculation step is needed.

7.2. Inventory Model Results

As the study case corresponds to a retail store where the products come from several suppliers, the model that best fits is the classical economic order quantity (EOQ), assuming a constant demand and an immediate supplier response. Furthermore, the portfolio does not include any special product, neither a commodity; instead, for most of the products moderate competence exist, so all products can be considered in the second category defined in Table 2 which corresponds to a 50% probability of lost sales if a stock out takes place.

7.3. Costs

Once the forecast has been optimized, the next step is to determine the costs inherent to forecast error. As discussed before, many factors can be involved but two of them are the most important, being the inventory holding cost and the cost of lost sales.

7.3.1. Inventory Holding Cost

The most important factor is the opportunity cost, as the average gross utility margin in the retail store in study is about 50%. A higher rate could result if an investment in a different product category is considered, although alternative investments always involve a risk. Another component of interest is the storage cost, as the store is located in an area of expensive rental rates. Although it is difficult to apportion the holding cost among individual products, generally a good estimation is 35% of the capital cost per year per unit.

7.3.2. Cost of lost sales

As reported by Corsten and Gruen (2004), the loss of utility margins in retail can be up to 40%; additionally, a penalty of 10% for loss of future margins as a result of customer dissatisfaction will be considered, so 50% of loss margin can be reasonable for a product provided by more than one competitor (Table 2). Of course, this can differ depending on the brand, the product, accessibility to other stores and customer loyalty.

7.3.3. Cost of forecast error

The purpose of this study is to measure the financial impact of forecast deviation in order to provide an optimal service level, and many variables are involved. The safety stock results from the desired service factor and the standard deviation or MAD of the forecast errors through time. The higher the MAD and the service factor, the higher will be the amount of stock needed; on the other hand, the loss function of normal distribution will be smaller, as well as the cost of lost sales.

The unit cost, the stock out penalty and the margin can vary with every product in a given portfolio, and therefore also the holding cost, which makes the computation more complex. The cost function is the key step in order to balance the overestimating and

underestimating costs and thus optimize the total cost using the proper service level.

The cost function involves two components: one of them represents the forecast overestimation which increases the inventory holding cost, while the second part represents the underestimation of real sales and therefore lost sales. A key point is to understand to which kind of deviation the cost is more sensible in order to assure that if a forecast bias exist, it is always in the less expensive way.

7.3.4. Simulation

Given the difficulty to optimize the service level, the MAD and loss function for 300 items each month, a practical and cheaper method to find a solution is to simulate the impact of service level variations and possible MAD reduction, estimating the cost reduction through an improvement in the forecasting process. Changes in the following variables will lead to different scenarios:

- Service levels from 90% to 99% with 1% gradual increments.
- Hypothetical MAD reductions from 1% to 50% with variable increments (considering that more accurate forecast methods may lower the MAD still more than the 12% obtained for SES forecasting).

7.4. Some Scenarios and Results

Once the components of the costs function are defined, the estimated cost can be simulated varying the service level and therefore the service and loss function factors. Two different methods were defined.

1. One service level common to all items, which facilitates the master data administration.
2. Choose the service level with minimum cost per item.

After running the first method, the service level with minimum cost of forecast error was found to be 98%, which means 6.9% in savings compared to the current service level of 95% (Figure 3). Additionally, the improvement of 12.6 % in MAD leads to a 397,802 MNX or 18.4% in savings, which cannot be disregarded (Figure 4). As said before, different MAD reductions may be achieved with different forecast techniques.

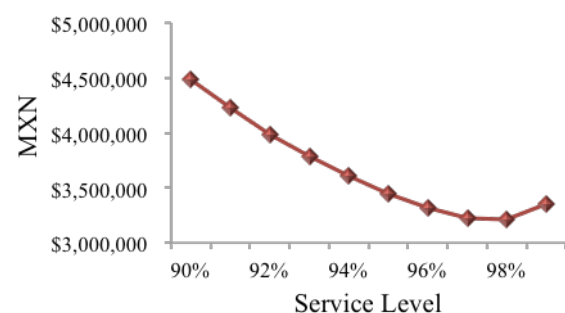


Figure 3: Optimal Unique Service Level using CFE

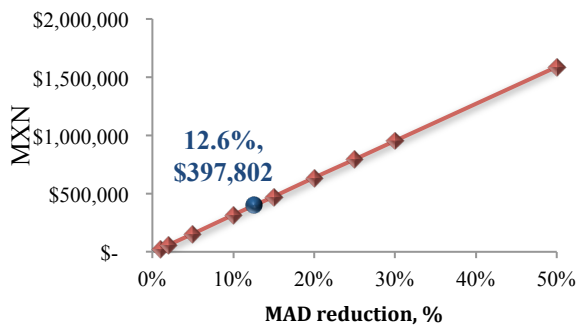


Figure 4: Savings as Function of Hypothetical MAD Reduction

For the second method, choosing an individual service level that leads to the minimum cost for each item, yielded results are presented in Table 5. For 189 items or 63% of the items, the optimal service level of 98% corresponds to the unique service level if chosen the same one for all products.

Table 5: Optimal Service Level, Individual Products

Service level	Items
96%	14
97%	93
98%	189
99%	4

The difference of the total cost determined when using individual service levels as compared to a generalized service level is less than 0.1%, so due to easy handling the first method is preferred. However, the present service level of 95% does not result to be the most efficient; a unique service level of 98% is proposed to reduce overall costs.

According to Figure 3, the cost function is more sensible to underestimating the sales, as the cost increases rapidly towards the left of the service level scale, while the cost increases more slowly if the inventory on hand is enough for a higher service level, so overestimating the real sales is preferred to underestimating them.

8. CONCLUSIONS

A VCV product classification is proposed as an alternative to the classical ABC classification, to have more information in a single tag and focus in those articles with high volume, cost or volatility, in order to link the forecasting process and service level definition. It is possible to determine the service level policy from the assessment of forecast error cost, even though to get an optimal result it is necessary to estimate the cost function properly, particularly the holding cost and cost of lost sales, otherwise the result can be biased. This proposal shows an alternative way to design the service level policy through analysis of the financial impact of forecast error beyond the strategic definition that most of the time is subjective.

The reduction in forecast error can be quantified in real savings; it can be used for managers as an argument for investments to improve the process, such as training, information systems and planning processes like S&OP. Using the VCV classification along with the CFE metric leads to focus on relevant categories.

According to the presented study case a low service level can be more expensive than having a high inventory, in which case overestimating is better than underestimate the demand.

9. REFERENCES

- Axsäter S., 2006. Inventory Control, 2nd edition, Springer.
- Broeckelmann R., 1998. Inventory Classification Innovation, CRC Press.
- Catt P., 2007. Assessing the cost of forecast error, Foresight, 7, 5-10.
- Corsten, D. and Gruen, T., 2004. Stock-outs cause walkouts, Harvard Business Review, 82(5), 26-28.
- Gilliland, M., 2010. The Business Forecasting Deal, John Wiley & Sons, NJ.
- Götz, N. and Köhler, C., 2006. The new SAP forecasting and replenishment solution: Is it an improvement over mySAP ERP? Foresight, 5, 45-50.
- Hawitt, D., 2010. Should you report forecast error or forecast accuracy? Foresight, 18, 46.
- Kahn, K.B., 2003. How to Measure the Impact of a Forecast Error on an Enterprise, Journal of Business Forecasting, 22, 21-25.
- Kolassa, S. and Martin R., 2011. Percentage Errors Can Ruin Your Day, Foresight, 23, 21-27.
- Makridakis S. G., Wheelwright S. C. and Hyndman R. J., 1998. Forecasting: Methods and Applications, 3rd edition, Ed. Wiley, NY.
- Singh S., 2013, What is the cost of your forecast error? Journal of Business Forecasting, 32(1), 26-29.
- Thomopoulos N.T., 2015. Demand Forecasting for inventory control, Springer.

AUTHORS BIOGRAPHY

Armando Sosa obtained his bachelor degree in Industrial Engineering at the Instituto Politécnico Nacional in México and is studying at present a Master course on Systems Engineering.

Ann Wellens is a chemical engineer with postgraduate studies in Industrial Administration (KUL, Belgium) and a master degree in Environmental Engineering (UNAM, Mexico). At the moment she is a full-time lecturer in the Systems Department of the Industrial and Mechanical Engineering Division of the National University of Mexico (UNAM). She has been working in air pollution issues for the last 15 years, dictating courses, collaborating in research projects and participating in conferences related with mathematical modelling of air pollution dispersion and statistics.

SUPPLIER SELECTION: SINGLE SOURCING VS. DUAL SOURCING WITH SUSTAINABILITY CONSIDERATIONS

Mistral Accorsi^(a), Roberto Montanari^(b), Marta Rinaldi^(c), Piera Centobelli^(d), Eleonora Bottani^(e)

^(a) FRANCHI S.r.l. (Società Unipersonale) - v. Ca' De' Bruni 61, 26041 Casalmaggiore (CR), Italy

^{(b),(c),(e)} Department of Industrial Engineering, University of Parma – viale G.P.Usberti 181/A, 43124 Parma, Italy

^(d) Dipartimento di Ingegneria Chimica, dei Materiali e della Produzione Industriale, University of Naples Federico II – Piazzale Tecchio 80, 80125 Napoli (Italy)

^(a) mistral.accorsi@studenti.unipr.it, ^(b) Roberto.montanari@unipr.it, ^(c) marta.rinaldi@unipr.it, ^(d) piera.centobelli@unina.it,
^(e) eleonora.bottani@unipr.it

ABSTRACT

This paper presents an integrated approach to supplier selection, which takes into account sustainability considerations and allows for the analysis of single sourcing vs. dual sourcing. A set of criteria and sub-criteria are identified from the literature and used for the evaluation of the suppliers. In line with the purpose of taking into account sustainability considerations, social sustainability and environmental sustainability criteria are included in the analysis, together with service criteria which reflect the traditional criteria used to evaluate and rank suppliers.

The model starts with the analysis of the single-sourcing strategy, where a single vendor (*supplier #1*) is available for a given supply. On the basis of the selection criteria, as well as of a fuzzy logic controller model, a score is computed for *supplier #1*. Such a score can be compared with the performance of *supplier #2*, who is taken into account when the dual sourcing strategy is pondered.

The proposed model is applied to the case of an Italian company, to show its usefulness in practical cases.

1. INTRODUCTION

In an increasingly competitive and interconnected market, companies need to strengthen their relationships with all the supply chain players, from the suppliers to the final consumers (Moqri, Moshref Javadi, & Yazdian, 2011). Market and demand are changing continuously and diversification of the needs and demands necessarily leads to an increase in the operating costs of a company and, consequently, to a decrease in profits. Purchasing decisions thus become strategically crucial to ensure the profitability and the long-term survival on the market.

Companies often forget that suppliers are actually a part of their supply chain and, as such, relationships with them have the potential to affect the costs and profitability of an enterprise. In many industries, the purchase volume, i.e. the total amount of money that a company spends in the purchase of parts, materials, components and services from external resources, is

estimated to account for approx. 60% of the company's income (Monczka, Trent and Handfield, 2005; Weele, 2000). Right purchasing decisions allow these costs to be reduced and the final consumers to be satisfied, responding to their needs and expectations (Dulmin & Mininno, 2003; Xia & Wu, 2007; Figuera, Greco, & Ehrogott, 2005).

Supplier selection is currently recognized as one of the key decision-making processes of a company for gaining competitive advantage. Suppliers can be considered as a part of the company's asset. Indeed, choosing the "right" supplier enhances the quality of the incoming material, resulting in better quality of the finished product. At the same time, it also reduces the internal management costs, because of the lower need to run checks on the purchased items and lower purchasing cost that can result from increase in the purchasing volume.

Typically, companies can choose from a large number of potential suppliers. Accordingly, supplier selection strategies are distinguishable between single sourcing, dual sourcing and multiple sourcing. As the name suggests, in the single sourcing strategy a company selects only one source, from a set of available suppliers that can provide the purchased item. Conversely, in the multiple sourcing strategy, more suppliers will be selected for the provision of the same product or service. Obviously, in this context, several providers are available for the provision of the same product or service. At the same time, these providers can have different characteristics in terms of market coverage, price, reliability and sustainability (Tomlin and Wang, 2005). Finally, in the case of dual sourcing, the company will select exactly two providers for a given supply.

Several techniques are available in literature to help decision-making processes in supplier selection. These techniques can be either deterministic or stochastic, linear or non-linear, mono-criteria or multi-criteria. Nonetheless, typically supplier selection has been approached as a multi-criteria decision-making (MCDM) problem (Sahraeian, Tafti, & Nobar, 2010).

In this case, the decision-maker will identify a set of criteria and sub-criteria to evaluate and rank different candidate suppliers to reach a defined target. MCDM methodologies provide solutions that can help decision-makers in manipulating complex data to identify the best supplier. In addition, MCDM techniques are effective in those situations where the number of alternatives involved and decision criteria is relatively high.

Recently, supplier selection issues are gradually embracing the concepts of sustainability. Indeed because of the growing knowledge about sustainability in enterprise, sustainable supplier selection would be the central component in the management of a sustainable supply chain (Amindoust et al. 2012). For this purpose, appropriate criteria and sub-criteria should be included in the supplier selection process, to ensure that the ranking of the suppliers will take into account sustainability considerations.

On the basis of the premises above, this paper presents an integrated approach to supplier selection, which takes into account simultaneously both sustainability considerations and the evaluation of single sourcing vs. dual sourcing strategies. Because sustainability considerations are often difficult to be translated into quantitative performance, the model is supported by a fuzzy logic controller (FLC), which is used to gather all the performance into a set of linguistic values, to derive a synthetic judgment. The model is tested on the case of a real company, which typically adopts a dual sourcing strategy, to evaluate its practical usefulness.

The remainder of the paper is organized as follows. Section 2 details the steps of the decision model developed. Section 3 shows the application of the model to the real case company. Section 4 concludes by describing the main pros/cons of the proposed model, as well as the key implications of the study, and by outlining future research steps.

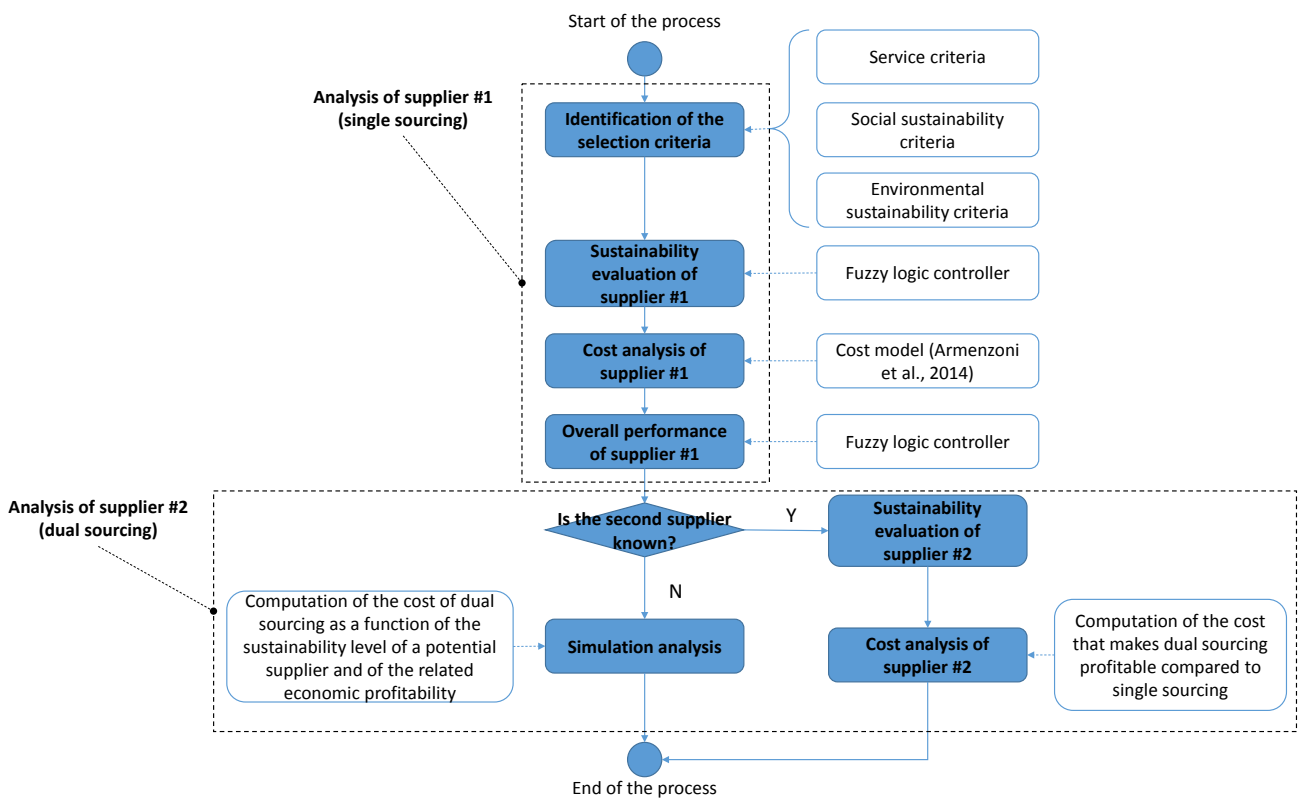


Figure 1: scheme of the approach proposed.

2. THE DECISION MODEL

The detailed scheme of the approach is proposed in Figure 1.

The starting point of the decision model is the analysis of a single supplier (*supplier #1*), i.e. of the single sourcing strategy. For this analysis, the first step is to identify the criteria and sub-criteria useful to the process of supplier evaluation. The selection criteria can be shared into three sets, i.e.: (1) service criteria; (2) social sustainability criteria; (3) environmental sustainability criteria.

More precisely, social sustainability and environmental sustainability criteria are recommended to be included in the analysis in line with the purpose of taking into account sustainability considerations when evaluating potential suppliers. Conversely, service criteria reflects the traditional criteria used to evaluate and rank suppliers. Obviously, these criteria cannot be defined in general terms, but rather the relevant factors should be identified on the basis of the specific context in exam. Scientific literature can provide useful guidelines for the identification of the selection criteria.

On the basis of the supplier's performance against the whole set of criteria, a preliminary evaluation of supplier #1 can be made. The supplier's performance are likely to be expressed with different measurement units, e.g. in terms of cost for the economic sustainability criteria or in qualitative terms for the social sustainability criteria. To gather the performance values into a synthetic index, a FLC is used to support the evaluation, and in particular to transform the different evaluations into the same linguistic scale (Amindoust et al. 2012). This leads to a synthetic judgment of the social and environmental sustainability of supplier #1. To complete the evaluation, the economic perspective of the triple bottom line approach to sustainability (Elkington, 1997) should be evaluated. This means that the cost of single sourcing at supplier #1 should be assessed. To this purpose, we adapt the model by Armenzoni et al. (2014), who have proposed a tool for the analysis of the cost of single sourcing vs. dual sourcing as a function of the reliability level of supplier #1. This completes the evaluation of supplier #1.

The second part of the model considers the analysis of *supplier #2*, which reflects the case of dual sourcing. The aim is to compare the cost of this latter strategy with that of the single sourcing, to assess the corresponding economic profitability. To carry out the analysis, it is paramount to know whether a second supplier is actually available for the given supply and therefore its performance can be evaluated and compared to that of supplier #1. In the case the company knows supplier #2, the sustainability analysis is repeated on that supplier, applying the same criteria used for the evaluation of supplier #1 (for consistency). Similarly, the cost analysis of supplier #2 can be easily carried out. Under that circumstance, the company will own the complete set of information useful to assess the profitability of the sourcing strategies (i.e., single sourcing vs. dual sourcing). In the case supplier #2 exists, but its performance is not known to the company, some simulation analyses can be carried out exploiting both the fuzzy logic model and the cost model by Armenzoni et al. (2014). More precisely, the total cost of the dual sourcing strategy can be estimated as a function of the sustainability level, to assess the economic profitability of that strategy compared to single sourcing.

3. APPLICATION EXAMPLE

The proposed model is applied to the case of a real Italian company, *Franchi S.r.l. Società unipersonale*, which was experiencing a problem of supplier selection on some critical provisions. Franchi S.r.l. operates as a manufacturer of home and office furniture. We consider the provision of chipboards, which are one of the critical raw materials used by Franchi S.r.l. for manufacturing. For this provision, Franchi S.r.l. can exploit two suppliers. Supplier #1 is a local supplier, quite sensitive to sustainability practices; conversely, supplier #2 is a big company, headquartered in Austria

and operating worldwide. Franchi S.r.l. has already worked with both vendors, therefore it knows the second supplier and the related performance.

To apply the model to the problem under examination, we start by analyzing supplier #1. The relevant criteria identified for the evaluation of the suppliers are:

- (1) Service criteria: delivery reliability; quality; delivery lead time; delivery flexibility;
- (2) Environmental sustainability criteria: emissions from the production of chipboards; use of eco-design practices; availability of an environmental management system; use of raw materials and resources;
- (3) Social sustainability criteria: employment practices; health and safety at the workplace; stakeholders influence; employee's rights.

For all the criteria listed above, a performance judgement is assigned to supplier #1, using a 1-10 linguistic scale (1 = very low; 10 = very high). Performance judgements are listed in Table

Table 1: Performance of supplier#1 against the service, environmental sustainability and social sustainability criteria

Criterion	Sub-criterion	Performance (1-10)
Service	delivery reliability	6
	quality	8
	delivery lead time	2
	delivery flexibility	8
Environmental sustainability	emissions from the production of chipboards	8
	use of eco-design practices	6
	availability of an environmental management system	9
	use of raw materials and resources	7
Social sustainability	employment practices	10
	health and safety at the workplace	10
	stakeholders influence	5
	employee's rights	10

By processing these performance by means of the FLC, developed *ad hoc* under Microsoft Excel™, we got scores of 8.71, 8.50 and 7.50 respectively for the service, economic sustainability and social sustainability of supplier #1 (aggregated score: 8.64).

With respect to the cost analysis, the model by Armenzoni et al. (2014) requires the following input data: order cost [€order]; inventory holding cost [€pallet/day]; stock-out cost [€pallet]; reliability level of supplier #1 [%]. These data were computed for

supplier #1 (see Table 2) and led, overall, to a total cost of the single sourcing strategy of 2,210.60 €day.

Table 2: cost components for supplier #1

Parameter	Numerical value	Measurement unit
Order cost	152.21	€order
number of orders	1	order/year
inventory holding cost	314.80	€pallet/day
stock-out cost	3365.04	€pallet
reliability	0.70	-

By aggregating the economic and sustainability performance of supplier #1, once again exploiting the FLC model, we got an aggregated performance of 6.00 (on a scale from 1 to 10).

The same analysis, carried out on supplier #2, led to scores of 8.50, 8.71 and 8.50 respectively for service, economic sustainability and social sustainability (aggregated score: 8.50). The total cost of the dual sourcing strategy, computed again exploiting the model by Armenzoni et al. (2014), is estimated to account for 3.313,18 €day. As both the performance and the total cost of supplier #1 outperform that of supplier #2, it is suggested that Franchi S.r.l. exploits a single sourcing strategy (supplier #1) for the supply under examination.

4. DISCUSSION AND CONCLUSIONS

This paper has presented an integrated approach to the supplier selection problem, which takes into account sustainability considerations and allows for the analysis of single sourcing vs. dual sourcing strategies.

The proposed approach has some strong points. First, it considers simultaneously the three main sustainability pillars during the supplier selection process, i.e. the environmental, social and economic points of view. To be more precise, the first two pillars are evaluated by means of a specific set of criteria and sub-criteria, which can be defined by the decision maker depending on the specific problem in exam (with possible suggestions from the literature). The economic pillar, instead, is evaluated by means of a detailed model, adapted from Armenzoni et al. (2014). Moreover, the proposed approach takes into account the traditional evaluation criteria for suppliers (service criteria). Overall, the resulting supplier selection process complements the traditional ones, which typically consider only the service criteria and the economic performance of the supplier, by including several sustainability considerations.

The methodological approach is new as well, since it grounds on the use of a fuzzy logic controller (FLC), which is useful to gather a set of performance, with different measurement units, into a synthetic performance index of the supplier.

Finally, the model allows for the comparison of different sourcing strategies, and in particular of the

single sourcing strategy and the dual sourcing one. From the use of the model, therefore, a company can also identify the most appropriate sourcing strategy.

The application of the model to the case study company considered allows deriving some practical considerations. The first one is that the successful application of the model requires a wide set of input data, including the definition of the relevant criteria and sub-criteria for the evaluation of the supplier, the supplier performance against these criteria, as well as some economic parameters (e.g., order cost, inventory holding cost or stock-out cost). All these input data cannot be defined in general terms, but rather should be identified on the basis of the specific company and vendor selection problem. This means that the use of the model should be supported by some company's representatives. Apart from this shortcoming, the use of the model allows a company to derive several interesting outcomes about its suppliers, namely three performance parameters (service, environmental sustainability and social sustainability) and the total cost of the sourcing strategy, thus enabling a very detailed evaluation and ranking of its suppliers.

Future research activities could address the comparison of the model with more traditional supplier selection procedures (where sustainability considerations are not made) or the extension of the model to the case of multiple sourcing.

ACKNOWLEDGEMENTS

This research was carried out thanks to the financial support of the Italian Ministry of University and Research (MIUR), under the ESCALATE « Economic and environmental Sustainability of Supply Chain and Logistics with Advanced Technologies » project (Scientific Independence of Young Researchers – SIR 2014 funding program - Decreto Direttoriale 23/01/2014).

REFERENCES

- Amindousta, A., Ahmeda, S., Saghafiniab, A., Bahreininejad, A., 2012. Sustainable supplier selection: a ranking model based on fuzzy inference system. *Applied Soft Computing*, 12, 1668–1677
- Armenzoni, M., Rinaldi, M., Montanari, R., Bottani, E., & Solari, F., 2014. Single sourcing vs. double sourcing: a simulation approach for supplier selection. In: Affenzeller, M., Bruzzone, A.G., Jiménez, E., Longo, F., Merkurjev, Y., & Zhang, L. (eds.), *Proceedings of the 26th European Modeling & Simulation Symposium (EMSS 2014)*, ISBN 978-88-97999-32-4 (paper), ISBN 978-88-97999-38-6 (pdf), pp.543-549. September 10-12, 2014, Bordeaux (France).
- Dulmin, R., Mininno, V., 2003. Supplier selection using a multi-criteria decision aid method. *Journal of Purchasing & Supply Management*, 9, 177-187

- Elkington, J., 1997. *Cannibals with forks: the triple bottom line of twenty-first century business*. Capstone, Oxford.
- Figuera, J.R., Greco, S., Ehrogott, M., 2005. *Multiple Criteria Decision Analysis. State of the Art Survey*. New York: Springer.
- Monczka, R.M., Trent R.J., Handfield, R.B., 2005. *Purchasing and supply chain management* (2nd ed.). Mason: South-Western.
- Moqri, M., Moshref Javadi, M., Yazdian, S., 2011. Supplier selection and order lot sizing using dynamic programming. *International Journal of Industrial Engineering Computations*, 2, 319–328
- Sahraeian, R., Tafti, A.F., Nobar, M.N., 2010. Selecting the best supplier by considering supply chain management strategy and utilizing uncertainty parameters. *Proceedings of the 4th International Conference on Industrial Engineering and Industrial Management XIV Congreso de Ingeniería de Organización Donostia - San Sebastián*, (p.1819-1829). San Sebastian.
- Tomlin, B., Wang, Y., 2005. On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing & Service Operations Management*, 7(1), 37–57.
- Xia, W., Wu, Z., 2007. Supplier selection with multiple criteria in volume discount environments. *Omega*, 35(5), 494-504.
- Weele, A., 2000. *Purchasing and supply chain management* (4th ed.). London: Business Press.
- Yu, H., Zeng, A., Zhao, L., 2008. Single or dual sourcing: decision-making in the presence of supply chain disruption risks. *Southeast University and Worcester Polytechnic Institute*.

AUTHORS BIOGRAPHY

Mistral ACCORSI graduated (with distinction) in Industrial Engineering and Management in 2016. She currently works at Franchi S.r.l. (S.U.) as accountant and assistant of procurement and production manager. Her interests concern especially supply chain management and logistics.

Roberto MONTANARI is Full professor of Mechanical Plants at the University of Parma. He graduated (with distinction) in 1999 in Mechanical Engineering at the University of Parma. His research activities mainly concern equipment maintenance, power plants, food plants, logistics, supply chain management, supply chain modelling and simulation, inventory management. He has published his research in approx. 70 papers, which appear in qualified international journals and conferences. He acts as a referee for several scientific journals, is editorial board member of 2 international scientific journals and editor of a scientific journal.

Marta RINALDI is research fellow of the University of Parma. She graduated (with distinction) in Industrial Engineering and Management in 2011, and got her Ph.D. in Industrial Engineering in 2015, both at the

University of Parma. She currently works on discrete event simulation and its application to industrial plants, logistics, supply chain management, supply chain modelling and simulation, inventory management, manufacturing systems and business processes. She is author (or co-author) of more than 10 papers published in international journals.

Eleonora BOTTANI is Associate professor of Industrial Logistics at the Department of Industrial Engineering of the University of Parma. She graduated (with distinction) in Industrial Engineering and Management in 2002, and got her PhD in Industrial Engineering in 2006, both at the University of Parma. Her research interests are in the field of logistics and supply chain management. She is the coordinator of a national project, called ESCALATE (Economic and environmental Sustainability of Supply Chain and Logistics with Advanced Technologies) related to supply chain sustainability. She is author (or co-author) of approx. 120 scientific papers, referee for more than 60 international scientific journals, editorial board member of five scientific journals, an Associate Editor for one of those journals, and editor-in-chief of a scientific journal.

Piera CENTOBELLI received with honours the M.Sc. degree in Management Engineering at the University of Naples Federico II. In 2016 she was awarded a PhD in Production Technologies and Systems at the Department of Chemical, Materials and Production Engineering of the same university. Currently she is Research Fellow at the University of Naples Federico II. Her research interests focus on operations management, logistics and supply chain management, knowledge and technology management, and development of decision support systems.

A PROPOSAL FOR OPTIMIZING URBAN MOBILITY IN MEXICO CITY BASED ON THE PUBLIC TRANSPORT NETWORK

Idalia Flores De La Mota^(a), Aída Huerta Barrientos^(b)

^{(a),(b)}Department of Systems Engineering
National Autonomous Mexico University

^(a)idalia@unam.mx, ^(b)aida.huerta@comunidad.unam.mx

ABSTRACT

At global level, last year Mexico City ranked the highest congestion level on the road network, causing more than 90% extra travel time for citizens during busy hours. The traffic congestion impacts directly on the quality of life, however due to government policies that have encouraged the use of private transport, citizens prefer to use it instead of the public transport network. Nevertheless, public transport network offers a poor service. The aim of this study is to carry out a diagnosis about the public transport network in Mexico City by proposing a pertinent theoretical tool to optimize its operations on a daily basis. First, we present the official statistics about urbanization of Mexico City. Then, we present the official statistics about mobility. After that, we describe the components of the Mexico City public transport network. Finally, we propose a methodology to optimize operational aspects.

Keywords: urban mobility; optimization – simulation; synthetic microanalysis; Mexico City.

1. INTRODUCTION

“Adding highway lanes to deal with traffic congestion is like loosening your belt to cure obesity. – Lewis Mumford, The Roaring Traffic’s Boom.

The activities of managing and planning of services in a context of cities entails a lot of work and participation of experts in different areas. Such is the case of transport that currently represents a challenge for researchers from different areas. There are three measures used for transportation analysis: traffic, mobility and accessibility (Litman 2011). As is observed in Figure 1, the aspects taken into account to compare the three measures are: definition of transportation, unit of measure, modes considered, assumptions concerning what benefits consumers, consideration of land use and favored transport improvement strategies.

On the other hand, politically Mexico City is divided by sectors (see Fig. 2), each sector has its own local responsibilities of management. We consider that a research about certain aspects of Mexico City should be conducted at different levels. For example, the first one can be done considering the complete city, at macro level.

	Traffic	Mobility	Access
<i>Definition of Transportation</i>	Vehicle travel.	Person and goods movement.	Ability to obtain goods, services and activities.
<i>Unit of measure</i>	Vehicle-miles and vehicle-trips	Person-miles, person-trips and ton-miles.	Trips.
<i>Modes considered</i>	Automobile and truck.	Automobile, truck and public transit.	All modes, including mobility substitutes such as telecommuting.
<i>Common performance indicators</i>	Vehicle traffic volumes and speeds, roadway Level of Service, costs per vehicle-mile, parking convenience.	Person-trip volumes and speeds, road and transit Level of Service, cost per person-trip, travel convenience.	Multi-modal Level of Service, land use accessibility, generalized cost to reach activities.
<i>Assumptions concerning what benefits consumers.</i>	Maximum vehicle mileage and speed, convenient parking, low vehicle costs.	Maximum personal travel and goods movement.	Maximum transport options, convenience, land use accessibility, cost efficiency.
<i>Consideration of land use.</i>	Favors low-density, urban fringe development patterns.	Favors some land use clustering, to accommodate transit.	Favors land use clustering, mix and connectivity.
<i>Favored transport improvement strategies</i>	Increased road and parking capacity, speed and safety.	Increased transport system capacity, speeds and safety.	Improved mobility, mobility substitutes and land use accessibility.

Figure 1: Comparing transportation measurements, reproduced from Litman(2011)



Figure 2: Mexico City sectors, reproduced from <http://mapamexicodf360.com.mx/carte/image/es/mapa-delegaciones-mexico.jpg>

While the second one should take into account the relationship between the different political sectors. Additionally, one should analyze each sector individually, at micro level. Last year Mexico City ranked the highest congestion level on the road network, causing more than 90% extra travel time for citizens during busy hours. The traffic congestion impacts directly on the quality of life, however due to government policies that have encouraged the use of private transport, citizens prefer to use it instead of the public transport network.

The aim of this study is to carry out a diagnosis about the public transport network in Mexico City for proposing a pertinent theoretical tool to optimize its operations on a daily basis.

This paper is prepared as follows: the official statistics about urbanization of Mexico City are presented in Section 2. The official statistics about mobility in Mexico City are shown in Section 3. The public transport network is described and a methodology to optimize its operational aspects is proposed in Section 4. Conclusions are drawn in Section 5.

2. URBANIZATION OF MEXICO CITY

As in the United States and Brazil, the majority of the Mexican population is urban (78% of population lives in cities), though this population grows at a slower rate (1.2% annually) compared to cities in China and India. Like in many countries around the globe, urban population in Mexico is growing at higher rates compared to the total population, making Mexican cities local engines for national growth, Varela (2015).

Table 1: Urbanization and economic growth, adapted from Tsay and Herrmann (2013)

	Countries				
	Br	C	I	Mx	USA
Population (billions)	0.5	1.3	1.2	0.1	0.3
Annual growth rate of population	0.83 %	0.46 %	1.28 %	1.07 %	0.9% %
Urban population	87 %	47 %	30 %	78 %	82% %
Change on annual level of urbanization (2010 – 2015)	1.10 %	2.30 %	2.40 %	1.20 %	1.20 %
GDP per capita (in U.S Dollars)	12,000	9,100	3,900	15,300	49,800
GDP growth rate per capita (annual percent in 2011)	1.80 %	8.00 %	4.90 %	2.70 %	1.00 %

Br-Brasil, C-China, I-India, Mx-Mexico, USA- United States of America

Moreover Varela (2015) points that “As competitiveness and growth in Mexican cities are increasingly compromised by congestion, air quality problems, and increased travel times; city officials not only face the challenge of accommodating a growing urban population but also sustaining a constant provision of basic urban services (e.g. clean water, health, job opportunities, transportation, and education). Unfortunately, periods of high growth without effective planning and increasing motorization, have pushed Mexican cities towards a “3D” urban growth model: distant, disperse, and

disconnected. The 3D model is a direct result of national policies subsidizing housing projects in the outskirts of urban agglomerations, managing urban and rural land poorly, and prioritizing car-oriented solutions for transportation. As a result of this policies and trends, over the past 30 years Mexico City’s population has doubled and its size has increased seven-fold.” Now the alternative is a 3C urban growth model: Compact, connected and coordinated as is suggested by Floater et al. (2014).According to Varela (2015) Mexico City is the most populated metropolitan area in the western hemisphere. It concentrates 17% of the national population and account for 17% of the national GDP.

Table 2: Socio-economics KPI’s of Mexico City, adapted from CAF (2011a)

Administrative organization	The metropolitan area of Mexico is composed of 16 Delegations in the Federal District, 58 municipalities in State of Mexico, and 1 Municipality in State of Hidalgo.
Population (2008)	Federal District: 8.8 million Metropolitan area (Federal District and State of Mexico): 19.2 million
Area (2010)	Federal District: 1,487 km ² Metropolitan area: 7,180 km ² (40.1% of which is urbanized)
Population density (2010)	Federal District: 5,958 people/km ² Metropolitan area: 6,671 people/km ²
Annual population growth rate (2005 - 2010)	Federal District: 1.49% Metropolitan area: 3.96%
GDP and growth (2011)	163.6 billion USD (17% of the national GDP, Federal District only) Annual GDP growth (2008-2011): 4%
Unemployment rate (2011)	6.5%

3. MOBILITY IN MEXICO CITY

Mobility in Mexico City is a huge problem since its size makes it insoluble. For instance, Table 3 shows the trips in order the reader has a better picture. This table is very important for the analysis since it can be used at the different levels, micro, mezzo and macro. Just as it was established in Table 1, the emphasis has variations in accessibility, mobility and traffic. From the macro level perspective it is necessary to consider the whole city as a complex network system, in such a case the connectivity has a major role, in order to show

this in next section public transportation networks maps are presented and analyzed.

4. THE PUBLIC TRANSPORT NETWORK IN MEXICO CITY CONTEXT

As can be observed from Fig. 3 to Fig. 4, the public transportation networks are constituted by other networks.

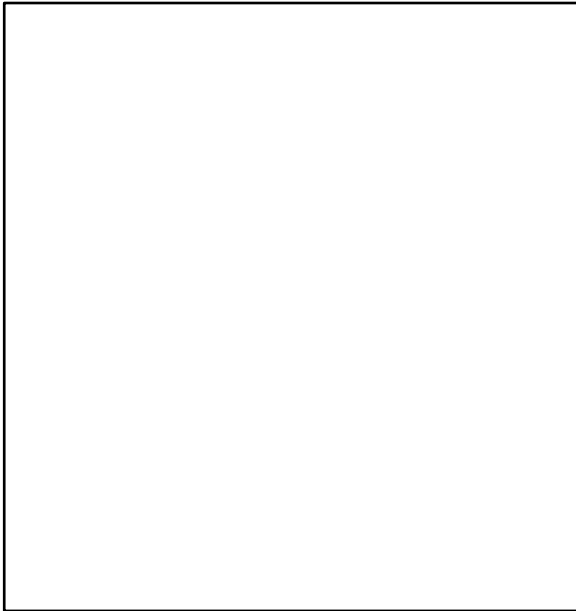


Figure 3: Metro and Metrobus networks, reproduced from <http://www.juliotoledo.com>

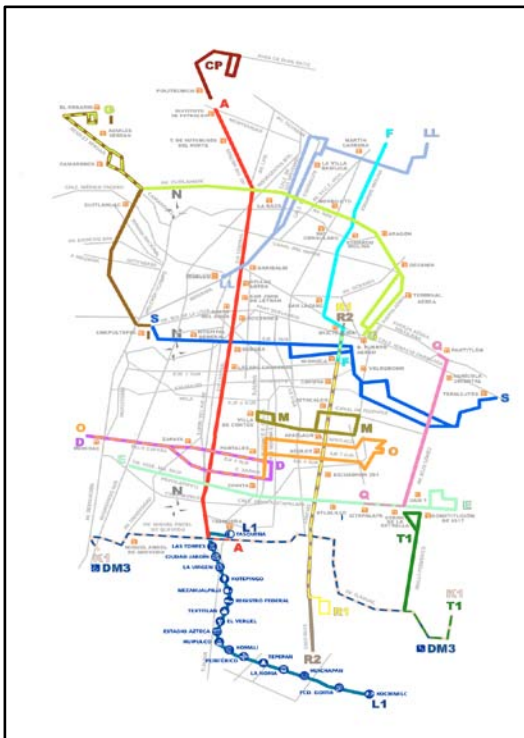


Figure 4: Electric bus network, reproduced from <http://www.juliotoledo.com>

According to Varela (2015), the management of the public transport in Mexico City is fragmented and this makes it difficult to establish some planning policies for its improvement (see Fig. 5).

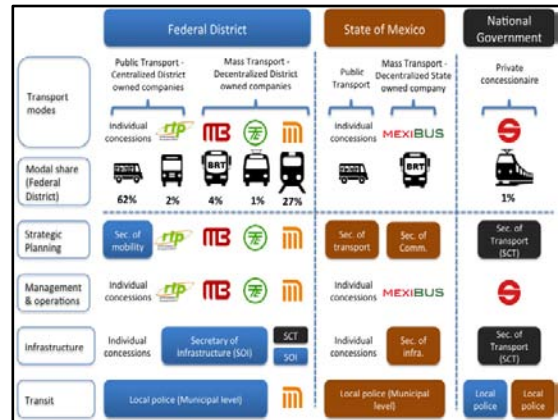


Figure 5: System of governance for public transport in Mexico City, adapted from Varela (2015)

“Such institutional and operational fragmentation has significant implications especially for users. In Buenavista - an area of Mexico City where three modes of transport converge – travellers must walk up to 1.5 km to transfer from one mode to another. Close to 150,000 users everyday use this disconnected transport hub with significant costs for users as well as for operators” Varela (2015).



Figure 6: Modal connection in Buenavista, adapted from Varela (2015)

The public transport networks in Mexico City are considered complex and both vulnerability and resilience are important factors for taking into account. As Reggiani (2015) states, the following questions that need to be answered:

1. Is a complex network a necessary condition for the emergence or presence of transport resilience and vulnerability?
2. Several indicators of resilience and vulnerability co-exist; are these differences related to specific fields of transportation research?
3. (c) Can connectivity or accessibility be considered as a unifying framework for understanding and interpreting – in the transport literature – the concepts of resilience and vulnerability?

In relation to the first question, we should bear in mind that “the term ‘complexity’ embeds both the assemblage of different units in a system and their intertwined dynamics. In other words, the term ‘complexity’ is strictly related to the concept of networks” (Reggiani, 2014, p. 814). Furthermore, connectivity, i.e. the ability to create and maintain a connection between two or more points in a spatial system, is one of the essential elements that characterize complex networks. Given the relevance of the connectivity pattern in complex networks, it may seem plausible that complex networks – and connectivity – are a sine qua non for the development of resilience and vulnerability in transport systems. Recent studies show how the topological properties of a network can offer useful insights into the way a transport network is structured and into the question of which are the most critical nodes (hubs). Resilience and vulnerability conditions associated with such hubs can then impact on the resilience/ vulnerability of the whole network. For optimizing urban mobility in Mexico City, we propose to carry out a network analysis based on traffic, mobility and accessibility aspects to develop a simulation model. Additionally, we consider that the lack of connections in the multimodal network could be tested based on some algorithms. According to Sochi (2011), three main methods can be used for this purpose: Direct inspection, Node Mapping and Segment Mapping. The segment mapping method is highly efficient method, the search for connectivity starts from seeding a list of connected nodes by the two nodes of a randomly selected segment. By going through the remaining segments and adding the node of any segment whose other node is found on the connected nodes list, a connected partition, which possibly comprises the whole network, will gradually build up. All segments whose nodes are added to the list are removed from the segments list either directly or by the use of a labeling mechanism such as a Boolean array to mark the status of thesegments as being removed or not.

The inspection of the segments list are repeated until happens one of the two conditions:

1. The segments list is empty (in which case the network is totally connected).
2. The inspection of all the remaining segments in the list in one of the iteration cycles returns no new nodes to be added to the connected nodes list (in which case the network is dismembered and partially connected).

In the latter case, this inspection process is repeated iteratively to identify all the partitions of the network until the exhaustion of the entire segments list. The input data required for this method is a list of the network segments where each segment is identified by the indices of its two end nodes. Sochi (2011).

5. CONCLUSIONS

Recent studies have been made on mobility in Mexico City have a transportation management approach and proposals for improvements in mobility, mobility indicators and policies to follow. While they are important, they have not used a systemic approach and made use of quantitative tools such as simulation and optimization.

As we can see the analysis proposed is innovative because it considers the transport network as a complex network that has as an important issue the disconnection.

At a macro level connectivity analysis is verified through an optimization algorithm, then the simulation is used to design simulation scenarios and search for feasible solutions with respect to environmental, land use, travels data and budget constraints.

For future research some results in the macro level will be available, a mezzo and micro level analysis will be developed, always considering aspects of traffic, mobility, and accessibility. This study will be done for a sector of the city of Mexico with data provided by the respective authorities.

REFERENCES

- Auyang S., 1999. Foundations of complex-system. Theories in economics, evolutionary biology, and statistical physics. Cambridge: Cambridge University Press.
- CAF, 2011a. Desarrollo urbano y movilidad en América Latina and INEGI 2013.
- CAF, 2011b. Desarrollo urbano y movilidad en América Latina. CTSEMPARQ México, Based on 2007 mobility survey. Estadísticas SETRAVI. Fideicomiso para el mejoramiento de las vías de DF. 2001. Secretaría del Medio Ambiente. 2008. Inventario de emisiones de la ZMVM, 2006.
- Helbing D., 2012. Social Self-Organization, Understanding Complex Systems. Berlin: Springer-Verlang. DOI: 10.1007/978-3-642-24004-1_10

- Litman, T., 2011. Measuring Transportation Traffic, Mobility and Accessibility. Victoria Transport Policy Institute. www.vtpi.org Info@vtpi.org
- Reggiani, A., Nijkamp, P., and Lanzi, D., 2015. Transport resilience and vulnerability: The role of connectivity. *Transportation Research Part A* 81, 4–15.
- Sochi, T., 2011. Testing the Connectivity of Networks. <https://arxiv.org/pdf/1111.2527.pdf>
- Sterman, J. D., 2001. System Dynamics Modeling: Tools for Learning in a Complex World,” *California Management Review* 43(4), 8-28.
- Toledo, 2016. <http://www.juliotoledo.com/mapas%20juliotoledo/mapas%20transporte%20df.html>. Accessed on May 2016.
- Tsay, S. and Herrmann, V., 2013. Rethinking Urban Mobility: Sustainable Policies for the Century of the City.
- Varela, S. 2015. Urban and suburban transport in Mexico City: Lessons learned implementing BRTs lines and suburban railways for the first time. *Integrated Transport Development Experiences Of Global City Clusters*, International Transport Forum, 2-3 July 2015, Beijing China.

AUTHORS BIOGRAPHY

IDALIA FLORES DE LA MOTA received a Master with honors, being awarded the Gabino Barreda Medal for the best average of her generation, in the Faculty of Engineering of the UNAM, where she also obtained her Ph.D. in Operations Research. Dr. Flores is a referee and a member of various Academic Committees at CONACYT as well as being a referee for journals such as *Journal of Applied Research and Technology*, the *Center of Applied Sciences and Technological Development*, UNAM and the *Transactions of the Society for Modeling and Simulation International*. She is a full time professor at the Posgraduate Program at UNAM and her research interests lie in simulation and optimization of production and service systems.

AIDA HUERTA BARRIENTOS received her Ph.D. in Operations Research from National Autonomous Mexico University (UNAM), and currently is Associate Professor of the Graduate Department of Systems Engineering at the School of Engineering, UNAM and she is an invited young researcher at the Center for Complexity Sciences, UNAM, in the Program for Social Complexity.

Table 3: Mobility in Mexico City, adapted from CAF (2011b)

Mobility KPI's for Mexico City during 2001, 2007 and 2010							
Total trips per day (2007)		48.8 million (Metropolitan area) and 32.0 million (Federal District)					
Daily trips per person (2007)		2.5 (Metropolitan area) and 3.6 (Federal District)					
Trips and modal share in the Federal District (2007)	Mode	Trips	% Total	%Public transport			
	Non motorized	8,600,000	26.9%				
	Private vehicles	4,800,000	15.0%				
	Microbuses	9,448,800	29.5%	50.8%			
	Metro	4,984,800	15.6%	26.8%			
	Autobuses	1,878,600	5.9%	10.1%			
	Taxis	1,041,600	3.3%	5.6%			
	Metrobus	762,600	2.4%	4.1%			
	Trolley (RTP)	204,600	0.6%	1.1%			
	Suburban train	167,400	0.5%	0.9%			
	Light train	111,600	0.3%	0.6%			
Total	32,000,000	100.0%	58.1%				
Road network (2007)		10,200 km (91% local roads)					
Total vehicles (Federal District, 2001)	Cars					4,460,386	
	Taxis					225,302	
	Motorcycle					11,920	
	Microbuses					20,459	
	Buses					8,240	
	Combis					3,519	
	Metrobus – articulated buses					322	
	Metrobus – regular buses					54	
	Metrobus – biarticulated buses					27	
	Totals					4,730,228	
	Road safety (2010)	Total number of accidents		14,729			
Number of deaths		1,026					
Involved vehicle in deaths		3.5% Microbus 81.0% Car 5.6% Truck					
Involved victim in deaths		14.0% Motorcycle driver 52.0% Pedestrian 20.0% Car driver					
Emissions (contribution by vehicle type) (2006)	<i>Pollutant</i>	<i>Cars</i>	<i>Taxis</i>	<i>Microbuses</i>	<i>Buses</i>	<i>Motorcycles</i>	<i>Trucks</i>
	CO2	58.0%	6.2%	13.2%	1.9%	6.0%	14.7%
	NOx	46.2%	6.9%	11.4%	10.2%	1.0%	24.3%
	PM2.5	14.2%	2.4%	1.6%	25.2%	1.5%	55.1%

SIMULATION OPTIMIZATION FOR AN AGE-BASED VACCINES DISTRIBUTION STRATEGY AGAINST THE SPREAD OF INFLUENZA EPIDEMIC

Javier Lara de Paz^(a), Idalia Flores^(b)

^{(a),(b)}Engineering School, Universidad Nacional Autónoma de México

^(a)javier_lara@unam.mx, ^(b)idalia@unam.mx

ABSTRACT

Through a small world complex network simulation epidemic lifespan of the influenza virus outbreak is analyzed considering an age-based vaccines distribution preference. Mexico's City population was divided in three different groups:

- all children aged 4-12 years;
- persons aged 2-64 with underlying chronic medical conditions;
- adults aged 65 years and older;

The simulation model is based in the vaccines distribution to this three groups in a differential way, exploring which proportion would show the shortest epidemic duration, therefore being that one the more efficient distribution strategy. The results obtained through the simulation, would help to minimize epidemic damage as in 2009 with the A(H1N1) strain. Also any other influenza strain should be considered in order to develop a containment strategy. To set up the simulation model a homogeneous social contact network was considered, where every population group presents its own social behavior forming small clusters and also presenting a probability of being in contact with an agent of a different cluster, as happens in social groups. Based on Mexico's city population, estimations show that the 28% presents underlying chronic medical conditions, health care institutions consider only the most acute disease such as diabetes mellitus, heart problems and some kinds of cancer. The population group aged between 4-12 years represents approximately also the 28% of the entire population, while adults aged 65 years and older represents hardly a 6.6%. These three considered groups represents the more vulnerable population to be infected by a respiratory transmission disease during an epidemic threshold situation. The agent based simulation was set up in the NetLogo software and was run under a full randomized factorial experiments design, results were analyzed by an analysis of variance. Final results show that epidemic duration lasts less when the schoolchildren population received the biggest proportion of vaccination rather than giving priority to persons with underlying chronic medical conditions as done now by the actual containment strategies. These results are consistent with the knowledge that an epidemic outbreak starts between the household population, and due to the higher social interactions that occur at schools, it is more probable that this population

affects more in the epidemic propagation.

The 2009 new AH1N1 influenza virus strain showed the weakness of the public health authorities in for planning a pandemic containment strategy. The outbreak of the new strain had expanded rapidly around the world by the current connectivity conditions. An optimal strategy was needed to design an adequate containment policy at the proper outbreaks moment. The most common ways to control an outbreak are vaccination and isolation, the first one represents public investments in acquiring enough amounts of doses, the decision of social isolation involves less expenses as much as promoting, but imply stopping many productive activities. One of the more affected sectors, as an example, would be the touristic sector, which in Mexico has represented the third economical income in the past years (Mexico's National Institute for Statistics Geography and Information, INEGI 2014 report). This means that before such a critical arrangement could be implemented, other containment strategies should be carry out. While influenza vaccine is the most effective tool for preventing

Keywords: complex network, epidemiology, spreading models, small world networks, vaccination strategies.

INTRODUCTION

health complications, the distribution and use of pandemic influenza vaccine will differ from that of the annual vaccine in several ways. Pandemic influenza vaccine would not be available due to production lags for at least 4-6 months after the pandemic virus had been identified and the supply of vaccine from manufacturers would be insufficient for some time after that. Another critical issue in planning pandemic influenza mitigation strategies is the delay between the arrival of the pandemic in a community and the availability of an effective vaccine. The likely scenario, born out in the 2009 pandemic, is that a newly emerged influenza pandemic will have spread to most parts of the world before a vaccine matched to the pandemic strain is produced. It is likely that even after a vaccine is invented against a new pandemic, its near-term availability would be highly limited until its mass production. For a severe pandemic, additional rapidly activated intervention measures will be required if high mortality rates are to be avoided. It is also of considerable concern that government strategies of recommending that vaccine be

concentrated primarily in high-risk groups and distributed to those people who keep the health system and social infrastructure operating. Because of A(H1N1) influenza vaccine shortage, a plan was enacted to distribute the limited vaccine stock to these groups first. This vaccination strategy, based on direct protection of those most at risk, has not been very effective in reducing influenza morbidity and mortality (Donghyun, et al. 2015). The present research plan, supported by agent based simulation models, would be to concentrate vaccine in schoolchildren, the population group most responsible for transmission, while also covering the reachable high-risk groups, who would also receive considerable indirect protection. In conjunction with a plan to ensure an adequate vaccine supply, this alternative influenza vaccination strategy would help control inter-pandemic influenza and be instrumental in preparing for pandemic influenza as the case of the A(H1N1) virus presented in 2009. The main objective of the present research is to find an optimal age-based vaccine distribution strategy considering three priority groups by means of an agent model simulation. To achieve this objective, a propagation model should be designed. Therefore an epidemiological SIR model in order to evaluate their parameters of contagious conditions has been explored. Also a model for A(H1N1) spread can be estimated by a small-world-like network. Interactions within a population are studied by means of different kinds of social networks, mathematically based on the structure of the population, which in a SIR model presents three different subpopulations (Susceptible, Infected and Recovered) that interact in proportion to their sizes. With these zero dimensional models, it has been possible to study the epidemic features, the asymptotic solution for the density of infected people, the effect on stochastic fluctuations on the modulation of an epidemic situation as well as the thresholds values. (Lara & De la Mota, 2014). Another classical approach to the epidemic spread describes spatially extended subpopulations, such as elements on a lattice. We consider a population rarely falls into either of these categories, being neither well mixed nor lattices. Recently Watts and Strogatz (Watts & Strogatz, 1998) introduced the small world network in order to study many social processes such as disease spreading, formation of public opinion, distribution of wealth, transmission of cultural traits, etc. In the case of epidemiological models, it has been shown that small world networks present a much faster epidemic propagation than other studied epidemic models, such as reaction-diffusion models, or discrete models based on regular lattices of a social network (Vázquez 2006). In this study the epidemiological SIR model is used to acquire the epidemic features and with these parameters the pandemic spread was simulated by means of a small-world network which, as mentioned before, presents a much faster epidemic propagation. Therefore the epidemic propagation model considers a SIR model set up on a small-world-network, and another subgroup is added to the interacting population which is the

vaccinated group. As the three vaccinated groups present a different social dynamic and different respiratory disease resistance as well, a simulation of different vaccination distribution strategies would show an optimal scenario in order to reduce the pandemic propagation time.

2. METHODOLOGY

The first part of the project consists in defining a complex system where interactions between agents represent the dynamical in a complex small world social network where three kinds of nodes are set up in the network. One kind of nodes are Susceptible, other Infected and the third kind represent the Recovered as epidemiological SIR model which considers homogeneous contacts between an specific deterministic population within a one-dimensional lattice. The random network was built upon a topological ring with N vertices and coordination number $2K$. Each link connecting a vertex to a neighbor, considering clockwise sense, is then rewired at random, with probability p , to any vertex of the system. It is assumed any self-connection and multiple connection as prohibited, obtaining a regular lattice at $p=0$ and progressively random graphs for $p>0$. The system model was programmed in NetLogo5.1.0. Once the system was established, the pandemic evolution takes part as any infected node contacts any other susceptible node and with a defined probability of getting sick the susceptible node gets infected, and after some steps with a defined resistance gain chance this node turns into a recovered node. Basic reproduction number R_0 is being read at every timestep, and once it becomes $R_0=1.4$, the containment vaccination strategy is triggered, and susceptible subgroup is then divided in three other different subgroups based on an age-based division to whom vaccines will be supplied. The shortest duration time of the epidemical propagation is being sought by means of simulation. Finding the optimal vaccine distribution strategy based on three target age groups is sought by means of the epidemic length considering three different levels of vaccines doses distribution to each age group. And through a design of experiments and its later results ANOVA for three factors, which are represented by the target groups, the optimal influenza vaccine distribution plan is obtained.

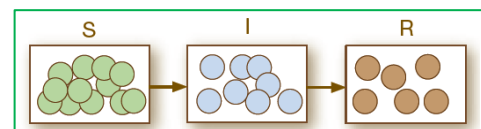


Figure 1. SIR model used to establish the system

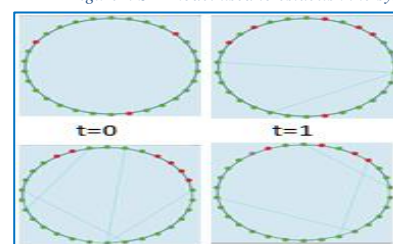


Figure 2. Small world network where social contacts were carried out

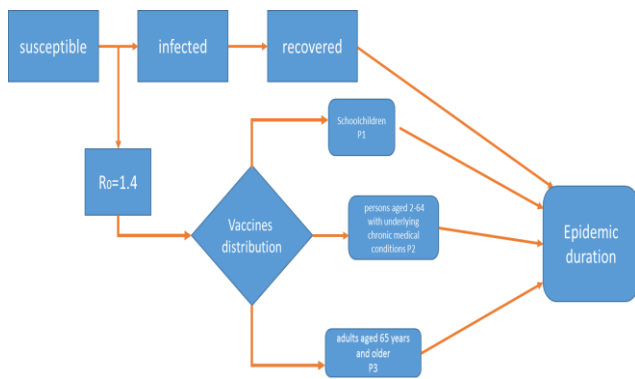


Figure 3. Methodology

2.1. The model

To establish the model it is considered that target aged subgroups, (school aged, elderly and population with underlying chronic medical conditions), have its own social affairs patterns. As an elder individual, having more than 64 years old, its more common to have a more isolated and sedentary way of life, interacting with persons of the same age range, having less participation in labour force and also present less outdoor activities. Hence in a social contact network topology this age group as nodes present less probability to interact with other social cluster nodes. This age group will be represented in the small-world network with a lower rewiring probability than established in the network. The considered range is between 0.6-0.4 of the general rewiring probability.

About school aged group, its social dynamic is considered opposite to the previous mentioned age group, as school aged population have a lot of social affairs within its educational centers and between their household related. These interactions causes a highly dynamic social contact network, which can trigger a more harmful disease propagation. Therefore, in the model, a higher rewiring probability level is assigned to this subgroup ranging between 1.2-1.3 times the assigned rewiring probability.

Population which presents chronic medical conditions, concerning this conditions as diabetes mellitus, heart illnesses and some other kind of cancer, interact the same way as the general population does, in most of the cases, so its rewiring probability is the one used for the entire small-world network. About the illness susceptibility, an assumption done for building the model is that these three targets population subgroups present most likely the same probability of being infected by an influenza contagion, young children don't have their immunity system well developed and elderlies have also a weak immune system due to natural causes, as well as chronic disease subjects which have an affected immune system due to medication. This characteristic is represented in the model by the gain resistance chance.

Table 1. Rewiring probabilities and corresponding percentages for each target age group used to set up the model. Source INEGI 2015 and Mexico's Public Health Care System.

Age population subgroup	Rewiring probability parameter	Population proportion
School aged (4-12 years old)	1.2 - 1.3	27 %
Elderlies (65 years and older)	0.4 – 0.6	7.2 %
Persons aged 2-64 with chronic medical conditions	For this group it is consider the same rewiring probability of the whole population	28 %

Once the population dynamic was included in the model through the rewiring probability, epidemiological parameters for an influenza epidemic were also considered. These parameters indicate the disease propagation not the individual risk of being infected, therefore weakness or strength of the different immunity systems of each objective group could not be represented by the epidemical parameters. The three age subgroups are also distinguished as Susceptible, Infected or Recovered individuals, as in a SIR model, and each one with a corresponding probability of becoming infected, or recovered. If vaccination strategy is assigned to an age group, some members of that group will become recovered with the corresponding probability established by the vaccination plan. Every individual is represented by a node in a complex small-world network and at every time step a node has a probability of getting infected by another adjacent infected node and after some time steps and based in a recovery chance this infected node would become recovered. Also through every tick some network links are rewired with a rewiring probability P which represent the social dynamic in the contact network resulting in different epidemic lengths. Considering that every age group has different social dynamic besides the general one, epidemic duration would vary if a vaccination strategy is established.

2.2. Model parameters

Simulation was set up with a design of experiments where vaccination distribution proportion among the three target groups took three levels for each group, $L1=30\%$, $L2=60\%$ and $L3=90\%$. After running simulation based on a complete randomized factorial design an ANOVA was studied in order to determine which level combination could show significantly less epidemic duration. Other considered parameters were about the target groups behavior as the school aged population presents more mobility, but it is expected that population portion which present chronic medical conditions behaves with normal social interactions, different as the older group ($65 <$) whose social affairs are lower. It is assumed that epidemical conditions are

the same for the three groups. The transmissibility, parameters are considered in the gain resistance chance, recovery chance, and mobility by the rewiring probability due to the kind of network it is used. In the following chart the mentioned parameters are shown:

Table 2. Epidemiological parameters and their corresponding values used in the model for an influenza disease. (Fraser, 2009),(Wu, JT et al, 2006)

Parameters	Definition	Estimated value
$1/\gamma$	Infectious period	1.5 days
$1/\kappa$	Latency period	1.9 days
Average recovery time	Total time since the contagious moment	3.4 days (84 hours)
Recovery chance	Inverse of mortality rate	98%
Attack Rate Transmission ART	Is the infection chance	7.5%
Transmission risk	Corresponds to the infection chance. This parameter affects directly the transmission rate β	is adjusted during the simulation process to acquire the corresponding R_0 threshold value

2.3. Simulation

Once we established the parameters, NetLogo 5.1.0 software was used to perform the simulation for different scenarios of propagations disease. Also the versatility of this software was proven given that it contains several libraries in which it is possible to develop different kinds of models with remarkable simplicity without losing generality. Besides, it has not been enough employed in the epidemic spread simulation field. Thereby we were able to perform our simulation, to obtain the needed scenarios for a later searching of the optimal epidemical mitigation strategy.

The model was programed in NetLogo 5.1.0. Epidemic and vaccine distribution controls were set up. Population and age subgroups were represented in a ring network shape, each group with different radius, and each with different rewiring probability. Although every subpopulation had the same node degree $K=4$.

First performed simulation were run with vaccine settings = 0 and other model controls set as shown below:

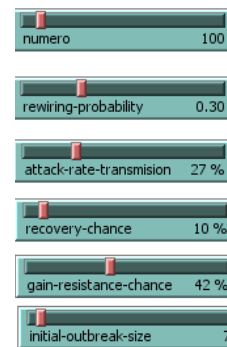


Figure 3. Epidemiological model settings.

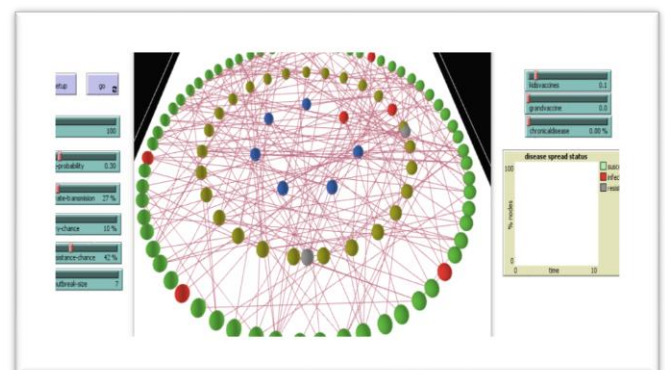


Figure 5. Model interface developed with NetLogo 5.1.0. At the right side epidemiological controls are established. In the center part population is represented as circle nodes. Green nodes represent Susceptible population including subjects with chronic medical conditions. Yellow circles represent school-aged population and blue circles are elderlies subjects. Infected nodes are red colored and gray nodes represent Recovered as well as vaccinated subjects.

Rewiring probability P was varied in order to find system stability. Four different values were used (0.20, 0.40, 0.60 and 0.80) the rewiring probability with the less variation would be considered as a steady state. For each P value 60 runs were performed.

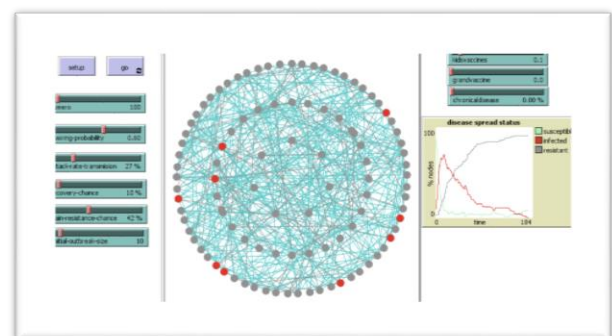


Figure 6. Simulations run for a rewiring probability $P = 0.60$. Four different adjustments were established P to find steady state of the system (0.20, 0.40., 0.60 and 0.80). 60 runs were performed in every case.

The simulations outgoing results were first analyzed through an ANOVA in order to find the lowest variation

in the epidemic duration. The results and its analysis are presented in section 3.

Once the stability was found model validation was considered by the literature (Tuite, et al. 2010). Moreover simulations implementing vaccine distribution strategy were performed. As the main objective of the present project is to establish an optimal vaccine distribution strategy among the target age groups simulations of the model were run according to a full factorial design of experiment with three factors corresponding to the mentioned age groups and each factor with three levels, which correspond to vaccine coverage in each group. The chosen levels were 30%, 60% and 90% of vaccine coverage. Next the corresponding statistical analysis of variance showed the optimal vaccine distribution levels for each population target age group.

The DOE implemented is shown below:

Table 3. Multilevel factorial design of experiment with 3 factor and 3 levels for each factor.

Multilevel Factorial Design			
Factors	3	Replicates	3
Base Runs	27	Total runs	81
Blocks	1	Total blocks	1
Number of levels	3, 3, 3		

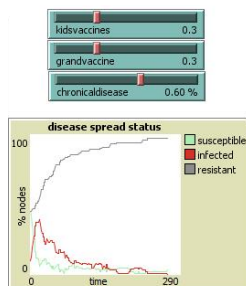


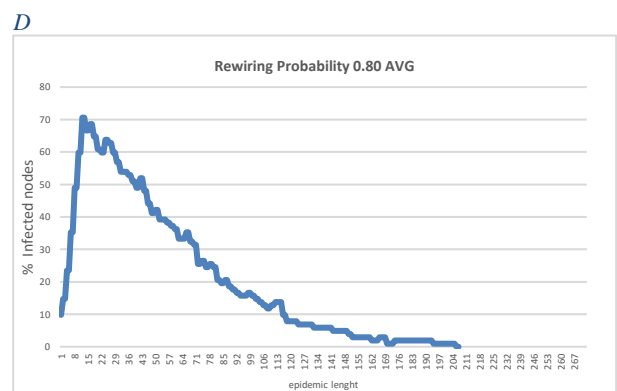
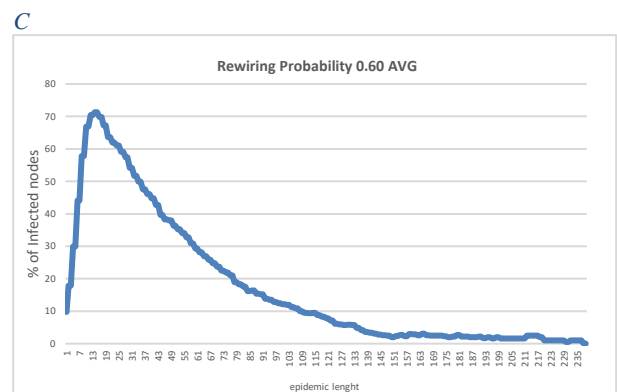
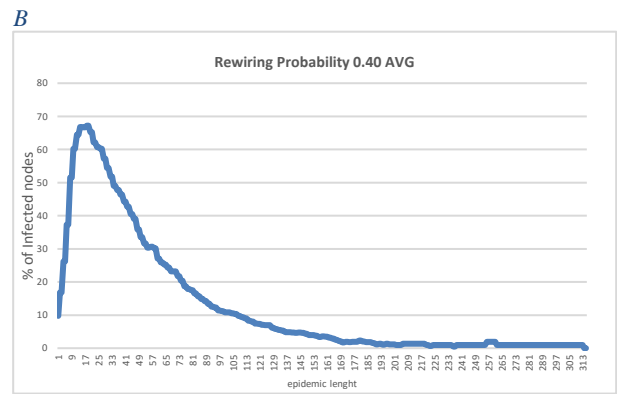
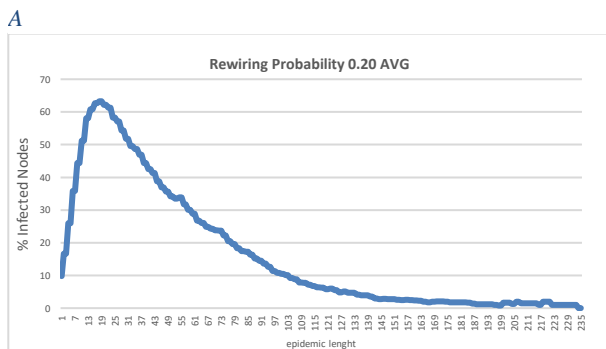
Figure 7. Vaccination strategy settings.

Corresponding results are discussed in next section as well.

3. RESULTS AND ANALYSIS

3.1 Simulation setup results

Results for the first simulation designed for system steady state are showed below for every chosen P level with 60 replicates for each run.



Plots A-D; Error! No hay texto con el estilo especificado en el documento. Epidemic duration corresponds to the horizontal axe and Infected population dynamic is plotted in vertical axe.

Plots A-D show epidemic status through the percentage of infected nodes including the targets groups. These results were obtained with no vaccine strategy in order to obtain performance characteristics of the model. The system behaves normally respect to an epidemic outbreak with an asymptotic infected population grow and a slow decay. As social contact dynamic grows it is expressed by rewiring probability, epidemic length last more, hence for a level $P = 0.80$ a longer epidemic duration should be observed it verified through the following ANOVA's.

About the media and variance analysis an one-way media ANOVA test was performed in order to find if there exist significant differences between the chosen levels, and as mentioned the scope of the lower variability. Results are discussed next.

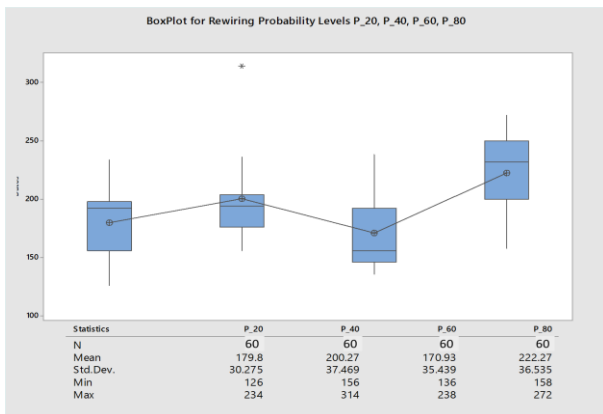


Figure 8. Boxplot Graph for each P level with corresponding statistical summary.

With figure 8 Boxplot graph is shown and statistical summary is added. It can be seen that level presenting lower standard deviation corresponds to $P = 0.20$, additional information should be analyzed to assert a significant media and variability difference.

With a significant level $\alpha = 0.05$ difference meaning test is performed

Table 4. Means differences using Tukey Method

#	Sample	Differs from
1	P_60	4
2	P_20	4
3	P_40	
4	P_80	1, 2

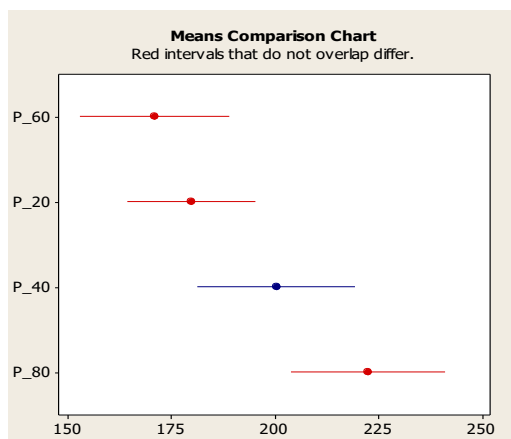


Figure 9. Mean comparison chart. Confidence intervals are drawn in different colors to show significant difference between means for P selected levels

Table 5. Descriptive summary for each P level.

Sample	N	Mean	Std.Dev.	Mean 95% CI
P_20	60	179.8	30.275	(22.975,45.983)
P_40	60	200.27	37.469	(17.404,92.792)
P_60	60	170.93	35.439	(23.698,60.963)

P_80	60	222.27	36.535	(28.232,54.384)
------	----	--------	--------	-----------------

In figure 9 one can observe that there are difference among the mean at the 0.05 level of significance. Also none of the red intervals overlap which indicates difference among each other. About information showed in table 5 about Confidence Intervals it can't be asserted that there is a significant difference in variability for each sample.

3.2 Optimization Results

The spread model is then used to determine an optimal vaccine strategy. The corresponding DOE was performed using a rewiring probability $P = 0.20$, due to a better performance in the simulation and also by literature considerations (Lara & De la Mota 2014, Jiahao 2005), although there they consider a random wire probability of 0.30 for a social contact small world network.

For the ANOVA design the corresponding results as shown as follows:

Table 6. ANOVA results from the DOE.

Source	DF	Seq SS	Adj MS	F	P
Chronicalmedic al affected	2	24132	12066	11.65	0.00
elderlies	2	548	274	0.26	0.768
Scholar age	2	25114	12557	12.13	0.000
Error	74	76632	1036		
total	80	126425			

$S = 32.1803$ $R-Sq = 39.39\%$ $R-Sq(adj) = 34.47\%$

For a significant level 0.05, vaccination school aged population is significant as well as population underlying chronic medical conditions, different to vaccination to elderlies which is not significant. It would be confirmed through the main effects Plot for the epidemic duration.

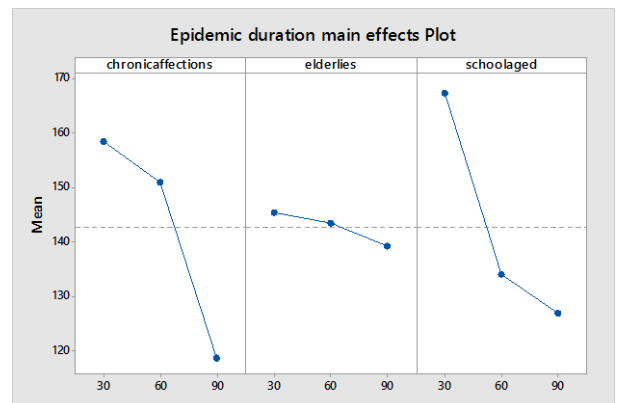


Figure 10. Main effect Plot for epidemic duration based on three level vaccination strategies.

For chronic medical affected population significant difference in the vaccine coverage is observed, same as a school aged population vaccine policy it is shown that there would be considerable difference in vaccine

coverage. For the third age group corresponding to elderlies, there's no significant effect on epidemic duration.

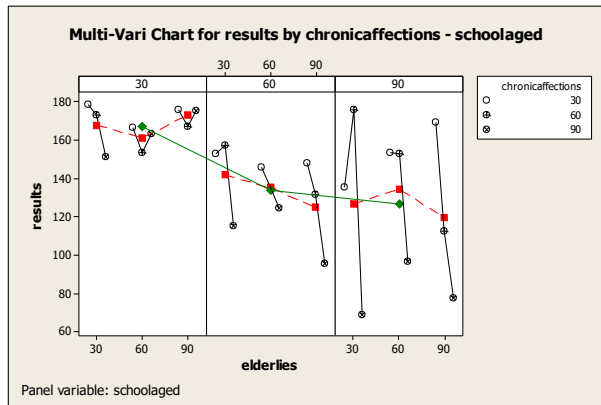


Figure 11. Multi-Vari Chart for epidemic length by chronic medical condition population group – school aged and elderlies. Green line represent school aged dynamic for three different vaccination levels. Red line represents elderlies infection behave under the same vaccination policy. Third group is represented by three different marks for each level. (reference is placed in the plot).

The above graph shows effects each factor has for the corresponding vaccination strategy. Red line indicates effect that elderlies vaccination strategy has in epidemic duration when the others groups are in the corresponding level. For instance if school ages group vaccine coverage is about 30% (green line in the first block). With this plot an optimal vaccine distribution strategy could be implemented. Also through a contour graph analysis it would be able to assert corresponding policy.

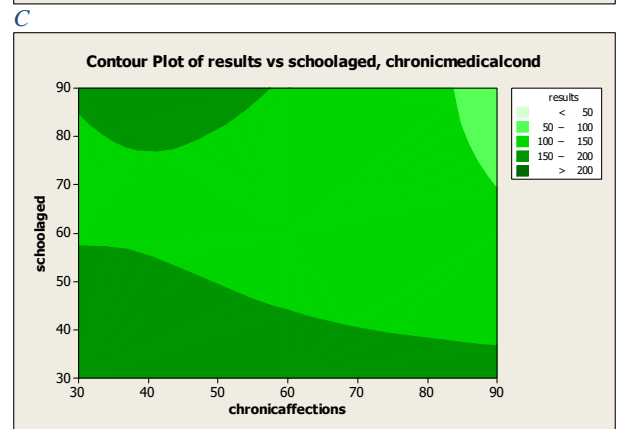
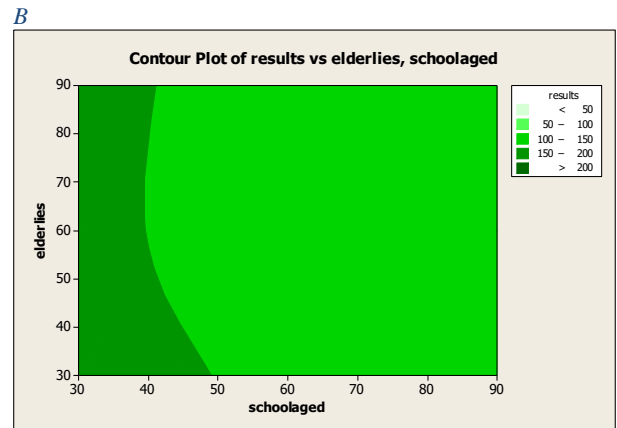
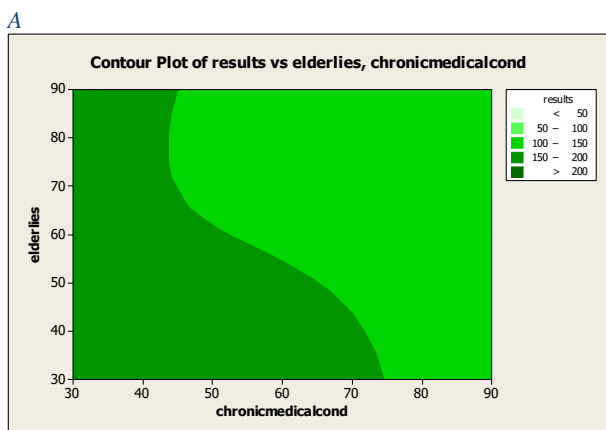


Figure 12. Three-dimensional Contour Plots for effects vs epidemic duration. Plot A represents interaction between elderlies and population underlying chronic medical conditions vaccine distribution policies. Plot B represents interaction between elderlies and school aged vaccine distribution policies. Plot C represents interaction between school aged and population underlying chronic medical conditions vaccine distribution policies.

By the scope of contour plots it is possible to determine for which vaccine distribution coverage an optimal policy should be implemented. In plot A epidemic duration is more affected would last less with a 40% coverage of population with chronic medical conditions but with a coverage of 70 to 80 % of elderlies. And comparing with plot B in the same coverage range for elderlies (70-80 %), school aged population could have an enough coverage of the 40-50 %. Comparison with other scenarios would show good policies.

CONCLUSIONS

A influenza strain propagation model was developed based on complex interactions between social groups, structuring the system in a complex small world network, considering three different social groups (children aged 4-12 years; persons aged 2-64 with underlying chronic medical conditions; adults aged 65 years and older) with the assumption that these three population groups present same contagion risk but different number of social interactions, represented in the small world network as the probability of having contact with other cluster far from their educational centers. The model present good approximation to previous works

done by authors (Lara & De la Mota 2014) and presented an according behavior to an epidemic spread (Fraser 2010), although an numerical goodness of fit could not be done because the lack of information but nevertheless the model is adequate for simulating different epidemic scenarios, and moreover to design an optimal vaccine distribution policy among these population subgroups using statistical technics being that the simulation showed what was expected that due to school aged population dynamics prioritizing this population group in a vaccine coverage would diminish epidemic risk more than if an elderlies population vaccine distribution policy is implemented. The present research would support stakeholders during an influenza epidemic outbreak, in order to help decision making. As further works it is suggested to explore a linear program model using the outgoing information of the simulation of the presented system, considering costs of vaccination.

REFERENCES

Ashleigh R. Tuite, Amy L. Greer, Michael Whelan, Anne-Luise Winter, Brenda Lee, Ping Yan, Jianhong Wu, Seyed Moghadas, David Buckridge, Babak Pourbohloul and David N. Fisman, Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza, *CMAJ*. 2010 Feb 9; 182(2): 131–136.

Chen Jiahao, Epidemics on small-world networks. Unpublished term paper for PHYS, 2005.

D. Watts and S. Strogatz. "Collective dynamics of 'small world' networks", *Nature* 393, 440-442 (1998).

Donghyun Kim, Hao Guo, Wei Wang, Joong-Lyul Lee, Sung-Sik Kwon, Alade O. Tokuta, "On efficient vaccine distribution strategy to suppress pandemic using social relation". *Discrete Math. Algorithm. Appl.* 08, 1650010 (2016).

Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*. 2009;324:1557–61.

Public Mexico's Health Care System (Secretaria de Salud), 2014 influenza vaccination report.

INEGI, 2014 annual report.

Joel K Kelso, Nilimesh Halder and George J Milne, Vaccination strategies for future influenza pandemics: a severity-based cost effectiveness analysis, *BMC Infectious Diseases*, 2013, 13:81.

Lara Javier, De la Mota Idalia, Simulation optimization for a vaccine distribution strategy against the spread of A(H1N1) epidemic, *Proceedings EMSS* 2014.

Longini Ira, Jr. and M. Elizabeth Halloran, Strategy for Distribution of Influenza Vaccine to High-Risk Groups and Children, *American Journal of Epidemiology*, Volume 161, Number 4, p 303-306, 2005.

Moore C. and Newman M. E. J., *Phys. Rev. E* 61, 5678(2000).

Vazquez A., Spreading dynamics on small-world networks with connectivity fluctuations and correlations, *Physical Review E* 74, 056101 2006.

Terry White; Suzanne Lavoie, and Mary D. Nettleman, Potential Cost Savings Attributable to Influenza Vaccination of School-aged Children. *Pediatrics Journal*, V 103,6, e73, 1999.

Wilensky, U. (1999). *NetLogo*. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL.

Wu JT, Riley S, Fraser C, et al. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med*. 2006;3:e361.

Xiao Fan Wang, Guanrong Chen, *Complex Networks: Small-World, Scale-Free and Beyond*, IEEE circuits and systems magazine, First Quarter 2003.

Yang, C. and Wilensky, U. (2011). *NetLogo epiDEM Basic model*. <http://ccl.northwestern.edu/netlogo/models/epiDEMBasic>. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL.

AUTHORS BIOGRAPHY

Javier Lara de Paz studied Physics at the Faculty of Science of National Autonomous University of Mexico (UNAM), a Specialization in Applied Statistics at the UNAM. He has been professor of Science at the Technological Institute of Gustavo. A. Madero in Mexico City and Operation Research professor at the Engineering School at the UNAM. Currently he is studying a Ph.D. in Operational Research at UNAM.

Idalia Flores de la Mota She received her Ph.D. in Operations Research at the Faculty of Engineering of the UNAM. She graduated Master with honors and received the Gabino Barreda Medal for the best average of her generation. She has been a referee and a member of various Academic Committees at CONACYT. She has been a referee for journals such as *Journal of Applied Research and Technology*, the *Center of Applied Sciences and Technological Development*, UNAM and the *Transactions of the Society for Modeling and Simulation International*. Her research interests are in simulation and optimization of production and service systems. She is a full time professor at the Postgraduate Program at UNAM.

MODELLING A BIOMASS BOILER USING AN ARTIFICIAL NEURAL NETWORK

Paula Ruiz^(a), Gorka Sorrosal^(b), Cruz E. Borges^(c) and Ana M. Macarulla^(d)

^(a-d) DeustoTech, Deusto Foundation, Unibertsitate etorbidea 24, 48007, Bilbao, Spain
^(a-d) Faculty of Engineering, University of Deusto, Unibertsitate etorbidea 24, 48007, Bilbao, Spain

^(a) pruiz@opendeusto.es, ^(b) gsorrosal@deusto.es, ^(c) cruz.borges@deusto.es, ^(d) ana.macarulla@deusto.es

ABSTRACT

This paper is focused on the application of Artificial Neural Networks to model a recovery biomass boiler from the Joutseno paper mill (Finland). The cross-validation technique has been used to train the neural model. The validation phase has been carried out with new data, which have never been known or seen during the training procedure. As a result of the validation stage, the model has achieved only 1.77% of MAPE error for the main output variable (net steam), and reaching good performance metrics for the estimation of the main gas emissions. This work will be the basis of a future development using Artificial Neural Networks, in order to control and minimise the impact of the air emissions produced by the industrial plant.

Keywords: Artificial Neural Networks, recovery boiler, real plant data.

1. INTRODUCTION

Finland is a country of wood and trees and its main business is the industry of paper and pulp (Suhr et al. 2015). The Kraft process (see Section 2 for details) is a basic process of making paper or pulp from wood. The Kraft pulp process consists of two lines; while in the main has as principal activity the extraction of wood pulp for producing paper the secondary line has the recovering of chemical elements to be reused. This work is focused on the recovery boiler, a key infrastructure as it is the principal responsible of both recover the white liquor and to generate electricity for the plant. Moreover, as any boiler, various chemicals are emitted to the environment so it is fundamental to control and limit those emissions (Vakkilainen 2005).

The principal objective of this work is to develop intelligent control strategies able to improve their performance while keeping under control the emissions. Safety is an important aspect to take into account. While improving the recovery of chemicals, the air emissions shall never go over the safety limits. Therefore, it is very important to have models of the processes that allow us to carry out simulations or to be used within the controller itself. However, the complexity as well as the high computational cost of the mathematical models of these types of processes makes necessary the use of surrogated black-box models.

Artificial Neural Networks (ANN) strategies have been used for years, but there are not commonly implemented in the industry. In process control this black-box modelling technique has been used successfully, but it is still possible to achieve better results. It can be more widely applied to solve some of the most nagging process control problems (Yen-Di Tsen et al. 1996). For example, ANN have been used to model the extraction of lignosulfonates from barley straw (Serna-Diaz et al. 2014). Furthermore, neural modelling has been used to reduce the maintenance cost and to improve the maintenance efficiency of the electricity generation by linking operational data and plant conditions (Fast and Palmé 2009).

The purpose of this work is presented in Section 4 where a model based on ANN of a biomass boiler installed in the secondary line of a Kraft process. The objective is obtained a fast response model to control the air emissions of the boiler

Kraft process and biomass boiler introduction is performed and the main design parameters are presented in Section 2. Next Section describes the datasets that have been used. Then, Section 4 the modelling methodology is presented. Finally, the preliminary results and the obtained conclusions are illustrated in Section 5 and 6 respectively.

2. KRAFT PULPING PROCESS

The Kraft process was introduced in 1879 and currently it is the most distinguished pulp process worldwide. Gaining cellulose is the main purpose of this process (Gustafson et al. 1983). This type of plants use the maximum energy generated by raw materials, and minimises the amount of waste. We can consider these kinds of factories of minimal environmental impact due to the appliance of the energy recovery line.

Figure 1 shows the three main sections in which the manufacturing process is divided. Extraction of cellulose from the wood is considered as a main line. Recovery of chemical elements as a secondary sequence and finally, the last phase is the effluent treatment. The process begins with the logs. Their surface bark is removed and remaining wood is cut it in small pieces called chips. From the stockpile, the chips are extracted, classified and taken to the cooking process-continuous digester with white liquor. White liquor is an alkaline

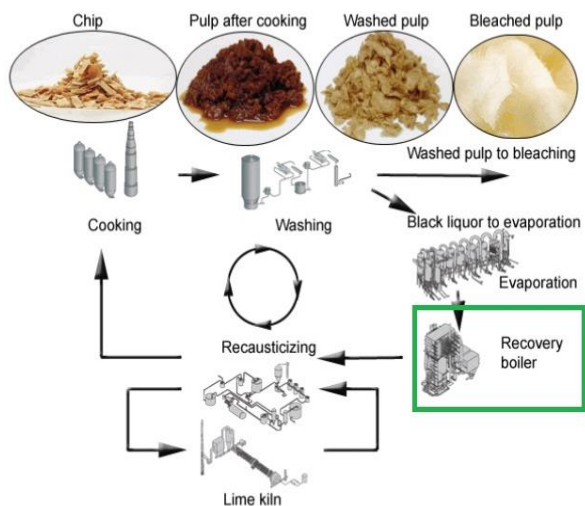


Figure 1: Scheme of the Kraft Process and Situation of the Modelled Recovery Boiler (courtesy of Katriina Mielonen)

solution of caustic soda and sodium sulfide, which reacts with the wood lignin. As a result of the cooking process two products are obtained: black liquor, which is the consequence of the reaction between the white liquor and the lignin, and cellulose pulp, which is classified, washed and bleached.

Once the black liquor is obtained, the recovery phase begins. The process is gradually becoming better known and spread throughout the world. The increment of the production, raised up the second production line, which is based on the recovery of chemicals. When production was expanding, they realized that it is cheaper to recycle the black liquid than to buy more chemicals. This operation is carried out in all Kraft factories; here it is why it is called a closed system, due to the recovery of the chemicals used in the beginning of the process.

The purpose of the recovery system is to return the inactivate sodium compounds at the end of the cooking process into activate compounds that can be reused at the beginning. The main operations for recovery the chemicals are based on the evaporation of black liquid, the black liquor combustion in the boiler, the causticizing of the sodium carbonate and the regeneration of lime mud. This recovery is also vital for a profitable process, without this step the waste would be very harmful for the environment and its treatment very expensive.

The acquired black liquor is classified in weak or strong liquor. The weak black liquor is evaporated to remove the water excesses, and the strong black liquor is taken to feed the recovery boiler. The black liquor is burned in the recovery boiler to produce heat and a smelt consisting of the inorganic sodium-containing cooking chemicals. The smelt is dissolved in water to produce green liquor that is recausticized to form the white liquor which constitutes the main fed in the Kraft cooking process (Vakkilainen 2005). Finally, the black liquor is burned in the boiler as combustible to produce heat water and steam. With the steam produced in the boiler, a steam turbine is fed and the generated energy

accomplishes to power the entire factory. The main parameters of the burner can be found in Table 1.

Table 1: Main Design Parameters of Joutseno Recovery Boiler (Vakkilainen and Sandegard 1999)

Continuous burning capacity (MCR)	3150tds/24h
Peak Load burning capacity	3500tds/24h
Steam Production	130kg/s
Main steam pressure	9.3e+6 Pascal
Main steam temperature	763.15K

3. MATERIALS AND METHODS

For this study real data obtained in the boiler shown in Figure 2, from the Joutseno pulp mill located in eastern of Finland has been used (Hamaguchi and Vakkilainen 2011). The main design features of Joutseno plant are shown in Table 1 (Vakkilainen and Sandegard 1999). 32 variables were collected hourly and stored electronically for three non-consecutive months from January 2007 to April 2008 (2016 samples in total). The data has been divided in 1512 samples for training and 168 samples for validation. As in any other dataset, failures always occur due to poor maintenance, human errors or device errors. Less than 4% of the data have been detected with anomalous behaviour, or above the physical possible limits. These values have been replaced by the linear interpolation of the closest valid measures. Table 2 shows the 32 variables, which are divided into sections and grouped by families. Chosen data include winter and summer months due to expert knowledge indicates that during the cold month, the amount of nitrogen embedded in the trees is lower, the quality of the wood is low, and therefore fuel quality decreases. On one hand, the variables related to the black liquor, air, steam, auxiliary flue and on the other hand the variables related to the emissions.

The ANN recognizes the measurable interrelationship between input and output parameters. Hence, the correct determination of training data, from the accessible raw plant data is essential for increase the precision of the developed model (Kaiadi et al. 2007). Once the datasets have been

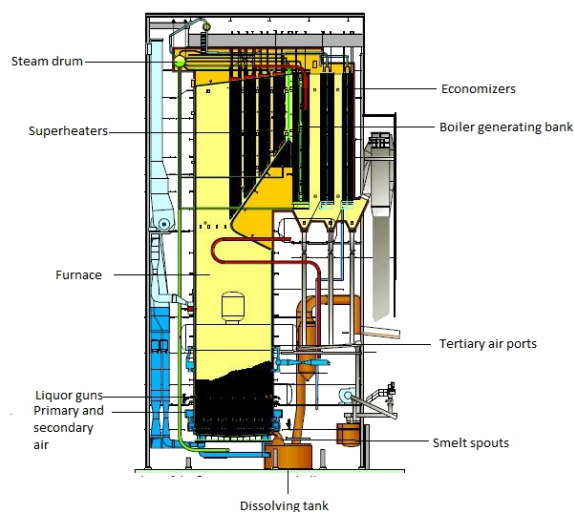


Figure 2: Scheme of the Modelled Recovery Boiler

Table 2: Main Variables of Joutseno Recovery Boiler

Variables to analyse	Units	Input/Output
Black Liquor		
Liquor flow	tds/d	Input
Ds at guns	%	Input
Density at guns	kg/l	Input
Temp in storage tank	°C	Input
Flow front	l/s	Input
Flow right	l/s	Input
Flow left	l/s	Input
Flow rear	l/s	Input
Water & Steam		
Feed Water	kg/s	Input
CBD	kg/s	Output
SB	kg/s	Output
Net steam	kg/s	Output
Auxiliar fuel in operation		
Load gas	m ³ n/s	Input
Start up gas	bar	Input
Load gas	bar	Input
Air		
Primary left	m ³ n/s	Input
Primary right	m ³ n/s	Input
Secondary left	m ³ n/s	Input
Secondary right	m ³ n/s	Input
Load burner air	m ³ n/s	Input
Tertiary front	m ³ n/s	Input
Tertiary rear	m ³ n/s	Input
Strong to combustion	m ³ n/s	Input
Strong to combustion	kg/s	Input
Flue gas		
CO left	ppm	Output
CO right	ppm	Output
SO ₂	mg/m ³ n	Output
TRS	mg/m ³ n	Output
NO _x	mg/m ³ n	Output
Dust	%	Output
O ₂ after BB left	%	Output
O ₂ after BB left	%	Output

pre-processed, a cross validation training procedure is carried out to obtain the neural model which has the best possible goodness fit. Figure 3 illustrates a summary of the steps followed. First a selection of the most relevant output variables have been made, then a reduction of the input variables have been carried out and, finally, the neural structure have been selected (Ruiz 2016).

Different tools have been used in this work: the pre-processing step has been carried out with Microsoft Excel® (version 2010), the selection of input and output variables have been made with the statistical software R (version 3.3.0) and, finally, the modelling procedure has been implemented using the Neural Network Toolbox package of the programming package MATLAB® (version 8.6, 2015b, Mathworks Company).

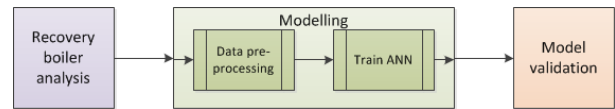


Figure 3: Experimental Methodology

4. NEURAL MODELLING

Correlation analyses between each of the variables have been performed to select the most influential variables in the behaviour of the boiler. These analyses have been used to choose the inputs of the neural model between all the available information and to avoid redundant information. High absolute value of the correlation coefficient (R^2) between two variables indicates the existence of a linear dependence; therefore, one of those variables would be removed because the election of both variables would be redundant. However, near zero R^2 values implies, that there are not linear relations between the studied variables.

First, the most important outputs of the biomass boiler have been chosen as the modelling objectives. The generated Net Steam has been chosen due to be the main product of the boiler in order to feed the steam turbine. The SO₂, NO_x and O₂ (after burning left) output gases are some of the variables that must be kept under control avoiding exceed certain limits. Although SO₂ and NO_x have not shown correlation with any input variable, they have been included to test if the neural model is able to relate them. Instead, the output O₂ (after burning right) has been discarded due to its high correlation with the included variable O₂ (after burning left).

Next, the input variables with higher correlation with the previously chosen outputs of the model have been selected. Those inputs variables not correlated with any output have been excluded. Moreover, some variables correlated with the outputs, such as the front, right, left and rear flows, have been finally discarded due to be highly correlated also with the Liquor Flow and Feed Water which are the main inputs variables of the boiler. Therefore, after the data pre-processing and variable selection step, only 10 variables were selected, 6 inputs and 4 outputs, due to its impact on the process behaviour. The Figure 4 represents the linear regression curve between the Liquor Flow and the Fed Water. The correlation coefficient is equal to 0.33887 between two analysed inputs, which means that both have varied behaviours and are needed on the final selection.

Another clear example can be seen in Figure 5 between an input and output variables of the boiler. As can be observed the correlation coefficient index is remarkably high, thus it can be deduced that the Feed Water is (obviously) highly related to the Net Steam output.

Table 3 resumes the finally selected variables to model the biomass boiler using the neural modelling approach. The main output variable is the Net Steam which will feed to the steam turbine. The other selected variables are divided in three groups: emissions to the air, air entries and liquid entries. Air emissions (SO₂, NO_x and

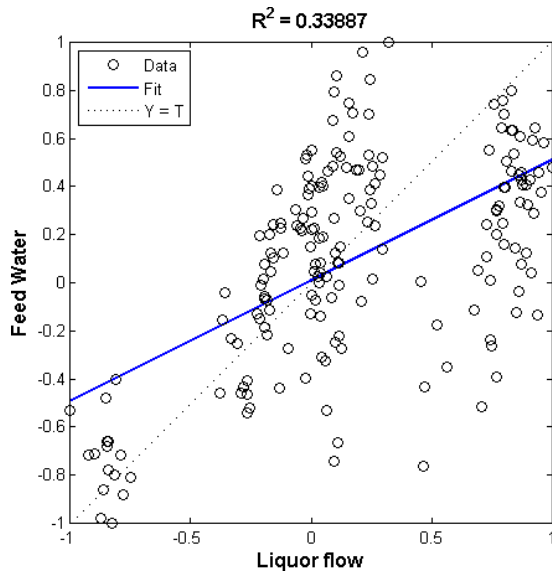


Figure 4: Regression Curve Between the Feed Water and Liquor Flow Inputs Variables

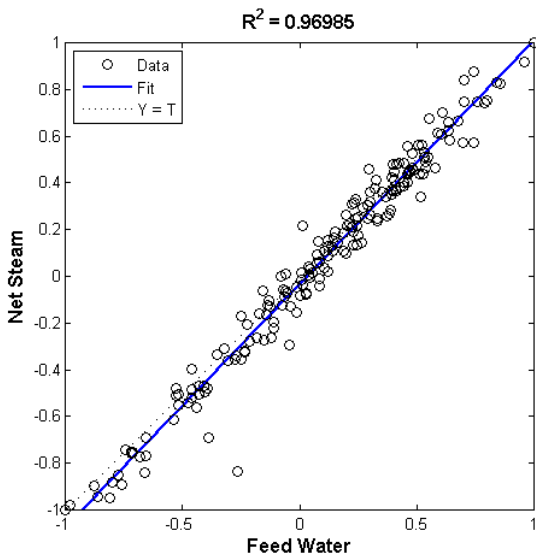


Figure 5: Regression Curve Between the Feed Water Input Variable and the Net Steam Output Variable

O₂) have been selected because one of the final objectives of the model is to maintain these emissions under control. Inputs related to liquids (Liquor Flow and Feed Water) have also a great significance because they are the main combustible of the boiler and, finally the family associated to the air inlets (secondary and tertiary air) are those who control the boiler operation for an optimal combustion.

Once finished the selection of the input and output variables, an iterative methodology modifying the number of layers and neurons in each layer has been carried out. Cross-Validation (CV) technique has been used in order to select the neural model structure that better fits the process. This technique consists of setting aside a set of data from the model training phase and only using them for the validation phase. The process is repeated until every single set of experiments is used in

the validation stage. 9 weeks have being used for the training and validation datasets and 1 week for the final testing purpose. Therefore, each training fold is a whole week of data consisting of 168 samples, using the “week” criteria as divider. The chosen weeks have been taken from winter and summer months because expert knowledge indicates that during the cold month, the amount of nitrogen embedded in the trees is lower, the quality of the wood is low, and therefore fuel quality decrease. This implies regimes changes in the behaviour in the variables of the boiler. Consequently, the available data are divided into training, validation and test datasets. The training and validation datasets are used for the selection of the model structure and the most suitable training algorithm. The first one is used to train several models with different structures and using different training algorithms.

The validation dataset consists of the excluded experiments from the training at each CV iteration and it is used to calculate the validation error that is used in the early stopping concept to avoid the overfitting. The training is stopped if the validation error increases in a sufficiently high number of consecutive iterations.

Table 3: Selected Variables from the Database to be Included in the Neural Model

Variables to analyse	Units	Input/Output
Black Liquor		
Liquor flow	tds/d	Input
Water & Steam		
Feed Water	kg/s	Input
Net steam	kg/s	Output
Air		
Secondary left	m ³ n/s	Input
Secondary right	m ³ n/s	Input
Tertiary front	m ³ n/s	Input
Tertiary rear	m ³ n/s	Input
Flue gas		
SO ₂	mg/m ³ n	Output
NOx	mg/m ³ n	Output
O ₂ after BB left	%	Output

Finally, the CV average error, calculated over the data that were excluded from the training at each iteration, is used to compare the model structures. This type of technique tests the generalization capability of a model structure (Kashani and Shahhosseini 2010). The error performance metric used in this procedure for the comparison of the model structures has been the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) (Equations 1, 2 and 3).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Act - Pred)^2}{n}} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^n |Act - Pred|}{n} \quad (2)$$

Table 4: Average Errors for Each Normalized Output Variable, Neural Structure and Training Function

	Training function	Best Structure	Output	RMSE	MAE	MAPE (%)
NARX	BR	6-100-4	Net steam	0.01096	0.00744	1.56
			SO ₂	0.02408	0.01469	2.00
			NO _x	0.09197	0.06130	83.17
			O ₂	0.02733	0.02120	2.81
	LM	6-4-4	Net steam	0.05637	0.04620	11.68
			SO ₂	0.12369	0.08945	41.28
			NO _x	0.28522	0.23312	166.52
			O ₂	0.08927	0.06856	9.77
	LM	6-11-4-4	Net steam	0.06593	0.05386	14.67
			SO ₂	0.12804	0.08995	40.58
			NO _x	0.26994	0.22028	155.11
			O ₂	0.09336	0.06855	9.57
Feed-Forward	BR	6-130-4	Net steam	0.00959	0.00636	1.23
			SO ₂	0.02272	0.01353	1.78
			NO _x	0.09016	0.05924	82.64
			O ₂	0.02550	0.01952	2.55
	LM	6-8-4	Net steam	0.06229	0.04947	12.05
			SO ₂	0.13700	0.10143	40.64
			NO _x	0.28147	0.22480	170.82
			O ₂	0.08958	0.06483	9.70
	LM	6-3-4-4	Net steam	0.08174	0.06570	16.59
			SO ₂	0.12523	0.08844	38.64
			NO _x	0.26801	0.22182	170.69
			O ₂	0.09411	0.07085	10.71

$$MAPE (\%) = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{Act - Pred}{Act} \right| \quad (3)$$

Once the neural model structure is selected and properly validated, the final model is trained with all the available data except for the test dataset which is used to estimate the generalization error of the fitted model. Nonlinear AutoRegressive with eXogeneous inputs (NARX) and Feed-Forward neural structures have been compared in this work in order to prove which is able to fit better the process. Furthermore, two different training functions have been test in both of the structures; Bayesian regularization backpropagation (BR) and Levenberg-Marquardt backpropagation (LM). In order to carry out properly the training procedure, giving equal weight to all the variables, all them have been normalized in the range [-1, 1]. Due to be a stochastic training procedure, all the training procedure has been repeated 10 times, in order to smooth the results Several neural structures have been test with neurons from 1 to 200 in the hidden layer and with 1 or 2 hidden layers. Note that due to its high computational cost, for the NARX structure trained with BR function, it has been tested only up to 100 neurons.

Based on the minimum average CV errors, the number of hidden neurons has been selected for each output. The RMSE has been selected as main indicator according to the literature (Chai and Draxler 2014).

Figure 6, represents the CV RMSE error of output variables versus the number of hidden neurons. As can

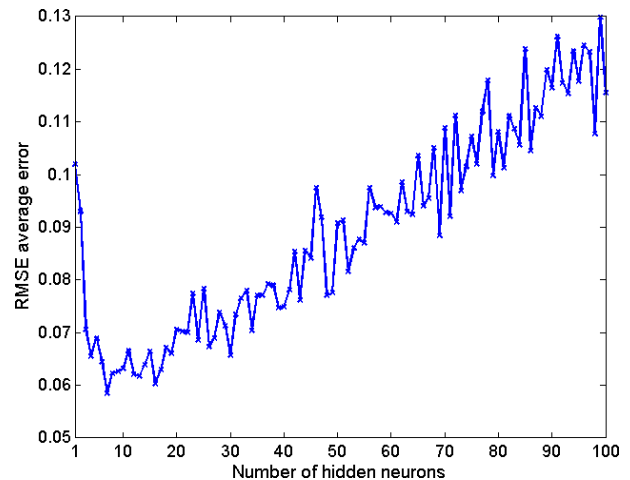


Figure 6: Evolution of the Cross Validation Average Error for the Net Steam Output with the Number of Neurons in the Hidden Layer

be seen, the CV error reaches a minimum value with the selected hidden neurons. However, each output reaches a minimum CV error with a different number of neurons. Thus, between the “best structure” for each output is selected one that has the lower average error of the CV errors for all the variables.

Table 4 illustrate the average CV results for each neural structure and for each output variable and training function. Hence, Figure 7 shows the selected ANN configuration which is a Feed-Forward neural network

with 1 hidden layer with 130 neurons, and trained with the BR function.

Finally, the above selected neural structure is trained with all the available data except from the test data and validated using that unseeing dataset in order to test its generalization ability and to measure the goodness of fit of the final model. Table 5 shows the validation errors of the final model for the denormalized test dataset. The obtained results presents reasonable error values that allow to stand that the neural network based model has been able to successfully assimilate the behaviour of the process with an acceptable accuracy. The behaviour of the Net steam output has been simulated with a MAPE error of 1.77%. Moreover, the NO_x and O₂ presents MAPE of 5.75 and 3.73% respectively. Only the SO₂ gas emissions present a MAPE error near the 15%. Figure 8 shows the real and the estimated data of all output variables for 1 week of simulation time (1 sample per hour). Note that the values are normalized in the range [-1, 1]. As expected, the selected model is able to follow the behaviour of each output variable, with more difficulties for the SO₂ and NO_x. Note that during the winter the nitrogen is stored in the root of the tree which radically changes the behaviour of the output variables depending the season of the year. Therefore, the operation regimes change with higher or lower values for the whole week, scrolling along the operation range.

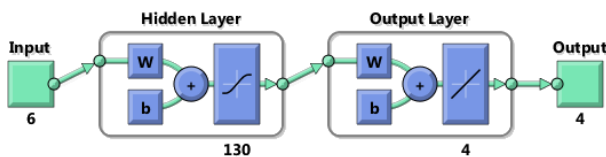


Figure 7: Neural Network Used to Model the Biomass Boiler

Table 5: Validation Errors of 1 Week Unseeing Data for the Output Variables Using the Selected Neural Network

Output	RMSE	MAE	MAPE (%)
Net steam (kg/l)	3.0543	2.6979	1.77
SO ₂ (mg/m ³ n)	0.4908	0.3607	14.72
NO _x (mg/m ³ n)	13.748	10.173	5.75
O ₂ (%)	0.1260	0.1006	3.73

Additionally, Table 6 shows the validation errors obtained with an alternative neural model for the denormalized test dataset. Note that this alternative model is a NARX neural network with 2 hidden layers, the first one with 11 neurons and the second one with 4 neurons, and trained with the LM function. The results obtained with this model are slightly worse but model size is noticeably smaller.

5. DISCUSSION AND FUTURE WORKS

In this work an Artificial Neural Networks based model of the recovery boiler of a Kraft process has been

presented. The results suggest that a Feed-Forward neural model is able to assimilate the dynamics of the

Table 6: Validation Errors of 1 week Unseeing Data for the Output Variables Using the Alternative Neural Network

Output	RMSE	MAE	MAPE (%)
Net steam (kg/l)	3.4975	3.3480	2.22
SO ₂ (mg/m ³ n)	0.9177	0.7807	31.00
NO _x (mg/m ³ n)	13.7195	9.9763	5.85
O ₂ (%)	0.1958	0.1688	6.10

process just with the information of the selected inputs: liquor flow, feed water and secondary and tertiary input air. The best results have been obtained with a neural model with 130 hidden neurons.

The implemented model has been able to estimate the main output variable (Net steam) over the denormalized test dataset with a MAPE error of only 1.77 percentage points. The model also presents good performance for the O₂ and NO_x. By contrast, it has not been able to properly model the behaviour of the SO₂ gas emissions. However, to obtain the best results a high number of neurons are needed. While more simple structures such as a NARX neural network with 2 hidden layers, are able to obtain similar errors for three of the output variables.

The final neural model will support future works in order to study the influence of the input variables and control set points into the electricity generation and contaminant emissions.

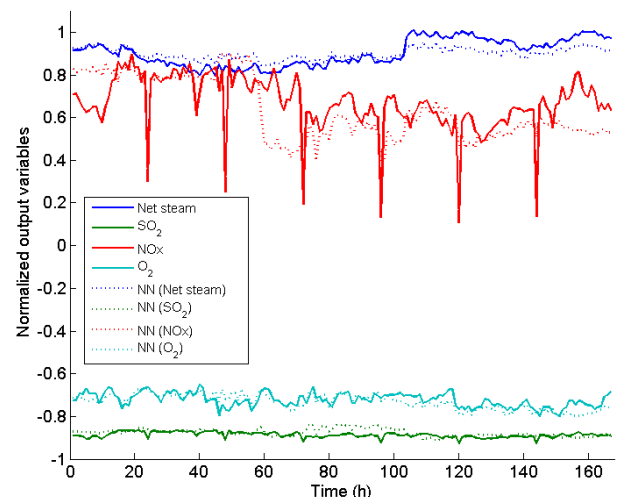


Figure 8: One Week Simulation of the Biomass Boiler for all the Normalized Output Variables with the Selected Neural Model

ACKNOWLEDGMENTS

The work has been carried out with the financial support of the Basque Government via the PI2013-40 project. The research team would also like to thank the contribution of the School of Energy Systems of the Lapperranta University of Technology. Special thanks

to Esa Vakkilainen for his collaboration and scientific support.

REFERENCES

- Chai, T. and Draxler, R.R. 2014, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247 -1250.
- De, S., Kaiadi, M., Fast, M. and Assadi, M., 2007. Development of an artificial neural network model for the steam process of a coal biomass cofired combined heat and power (CHP) plant in Sweden. *Energy*, 32(11), pp. 2099-2109.
- Fast, M. and Palme, T., 2010. Application of artificial neural networks to the condition monitoring and diagnosis of a combined heat and power plant. *Energy*, 35(2), pp. 1114-1120.
- Gustafson, R.R., Sleicher, C.A., McKean, W.T. and Finlayson, B.A., 1983. Theoretical model of the kraft pulping process. *Industrial & Engineering Chemistry Process Design and Development*, 22(1), pp. 87-96.
- Hamaguchi, M. and Vakkilainen, E., 2011. Effect of timescale on emission levels from pulp mills. 1(1).
- Kashani, M.N. and Shahhosseini, S., 2010. A methodology for modeling batch reactors using generalized dynamic neural networks. *Chemical Engineering Journal*, 159(1), pp. 195-202.
- Ruiz, P., 2016. Modelling a recovery boiler using artificial neural networks. Final Degree Project. University of Deusto.
- Serna-Diaz, M.G, Medina-Marin, J, Arana-Cuena, A, Seck-Tuoh-Mora, J.C, Tellez-Jurado, A, Mercado-Flores, Y, Jimenez-Gonzalez, A, Hernández-Romero, N, 2014, Modeling and optimization of the extraction of lignosulfonate from barley straw by using artificial neural networks. *Proceedings of the European Modeling and Simulation Symposium*, 978-88-97999-38-6; pp 469-473.
- Suhr, M., Klein, G., Kourti, L. and Gonzalo, M.R., 2015. Best Available Techniques (BAT) Reference Document for the Production of Pulp, Paper and Board.
- Tsen, A.Y., Jang, S.S., Wong, D.S.H. and Joseph, B., 1996. Predictive control of quality in batch polymerization using hybrid ANN models. *AIChE Journal*, 42(2), pp. 455-465.
- Vakkilainen, E., 2005. Kraft recovery boilers—Principles and practice.
- Vakkilainen, E., Sandegard, P., Ahlstrom Machinery Corporation, Veitola, J. and Oy Metsä Botnia A.B, JOUTSENO PULP, 2006. New Ahlstrom recovery boiler at Metsä-Botnia Joutsenomill.

AUTHORS BIOGRAPHY

Paula Ruiz, currently finishing her studies in Automatic and Industrial Electronic engineering at University of Deusto. She has been doing an internship in Lapeenranta in the Kraft pulping process area.

Gorka Sorrosal, graduated in Automatic and Industrial Electronic engineering at University of Deusto in 2011, and Master in Control, Automatic and Robotic Engineering at University of the Basque Country in 2013. Ever since 2009, he has been collaborating with DeustoTech Energy in projects related to computer vision and processes optimization. Currently, he is developing his Ph.D. Thesis in optimization and control of the BTO process.

Cruz E. Borges, received his diploma degree in Mathematics from University of La Laguna (Spain) and his Ph.D from University of Cantabria where he solve Symbolic Regression Problems using Genetic Programming and Numerical Methods. He has made a postdoc at DeustoTech where he develop forecasting techniques to predict the energy consumption of buildings (CENIT ENERGOS, ITEA2 IMPONET and ITEA2 NEMO&CODED) and have also developed Decision Support Systems for the evaluation of energy efficiency measures (MANUNET ENEPRO and HEFICAL). Currently, he leads the smart grid group of the Energy Unit of DeustoTech.

Ana M. Macarulla, industrial Engineer specialized in Electricity (University of Navarra, 1990) and Dr. in Automatic Control (University of Navarra, 1996). Researcher at CEIT-IK4 from 1990-1997, lecturer in the University of the Basque Country (1997-1999). Since 1999 she is lecturer in the Department of Industrial Technology at the Faculty of Engineering of the University of Deusto and researcher at Energy Unit of DeustoTech. His teaching activity has been focused especially in the field of automatic control, especially in the area of nonlinear, multivariable, adaptive and predictive control. His research work is related to the modelling and development of control algorithms. She is collaborating in activities related to climate change, energy policy and environmental sustainability.

COMPARISON OF PSO AND DE IN THE TASK OF OPTIMAL CONTROL OF CHAOTIC LOZI MAP

Roman Senkerik^(a), Michal Pluhacek^(b), Ivan Zelinka^(c), Adam Viktorin^(d)

^{(a),(b),(d)}Tomas Bata University in Zlin , Faculty of Applied Informatics, Nam T.G. Masaryka 5555, 760 01 Zlin, Czech Republic

^(c) Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic

^(a)senkerik@fai.utb.cz, ^(c)ivan.zelinka@vsb.cz

ABSTRACT

In this paper, evolutionary algorithm Differential Evolution (DE) is compared with swarm based technique PSO in the task of optimal evolutionary tuning of controller parameters for the stabilization of selected discrete chaotic system, which is the two-dimensional Lozi map. The novelty of the approach is that the most utilized examples of evolutionary/swarm based algorithms are compared directly on the highly nonlinear and complex multimodal optimization and simulation task. The simulations were performed for three different required final behavior of the chaotic system.

Keywords: Chaos control, Differential Evolution, PSO, optimization, evolutionary algorithms, swarm algorithms, soft computing.

1. INTRODUCTION

In many applications, one of the most challenging tasks is the controlling of highly nonlinear dynamical systems in order to either eliminate or synchronize the chaos. The first successful approach to control chaotic dynamics by means of a simple linearization technique was introduced in 1990s by Ott, Grebogi and Yorke (i.e. OGY method) (Ott, Grebogi, and York 1990). Later, the rapid development of methods for stabilizing of chaotic dynamics has arisen and more advanced modern techniques have been applied for chaos control and chaos synchronization including unconventional methods from the soft computing field.

The most current intelligent methods are mostly based on soft computing, representing a set of methods of special algorithms, belonging to the artificial intelligence paradigm. The most popular of these methods are neural networks, evolutionary algorithms (EA's) and fuzzy logic. Currently, EA's are known as a powerful set of tools for almost any difficult and complex optimization problem.

The interest about the connection between evolutionary techniques and (not only) control of chaotic systems is rapidly spreading. The initial research was conducted in (Zelinka 2009), whereas (Zelinka, Senkerik, and Navratil 2009) and (Senkerik et. al 2010) was more concerned with the tuning of parameters inside the existing chaos

control technique based on the Pyragas extended delay feedback control (ETDAS) (Pyragas 1995). Later works (Senkerik et. al 2013; Kominkova-Oplatkova et al. 2013; Senkerik et al. 2014) show a novel approach as to how to generate the entire control law (control method) for the purpose of stabilization of any chaotic system.

Other approaches utilizing the EA's for the stabilizing of chaotic dynamics have mostly applied the Particle Swarm Optimization algorithm (PSO) (Kennedy and Eberhart 1995) and multi-interval gradient-method (Abedini, Vatankhaha and Assadian 2012) or minimum entropy control technique (Sadeghpour et al. 2011). EA's have been also frequently used in the task of synchronization of chaos (Coleho and Grebogi 2012). In (Richter 2000) an EA for optimizing local control of chaos based on a Lyapunov approach is presented.

This work is an extension of previous aforementioned research focused on optimal stabilization of chaotic systems by means of evolutionary algorithms. The novelty of this work and motivation was to perform a direct comparison between swarm based algorithms and evolutionary algorithms in such a highly nonlinear and complex multimodal simulation and optimization task.

Firstly, a problem design is proposed. The next sections are focused on the description of used metaheuristics, experiment workflow, results and conclusion.

2. PROBLEM DESIGN

The brief description of used chaotic system and original feedback chaos control method, ETDAS is given.

2.1. Lozi Map Chaotic System

The chosen example of chaotic system was the two-dimensional Lozi map in form (1).

$$\begin{aligned} X_{n+1} &= 1 - a|X_n| + bY_n \\ Y_{n+1} &= X_n \end{aligned} \quad (1)$$

The Lozi map is a simple discrete two-dimensional chaotic map. The map equations are given in (1). The parameters used in this work are: $a = 1.7$ and $b = 0.5$ as suggested in (Sprott 2003). For these values, the system exhibits typical chaotic behavior and with this parameter

setting it is used in the most research papers and other literature sources (Aziz-Alaoui and Grebogi 2001). The x, y plot of the selected map is depicted in Figure 1. The chaotic behavior of the map, represented by the example of output iterations is depicted in Figure 2.

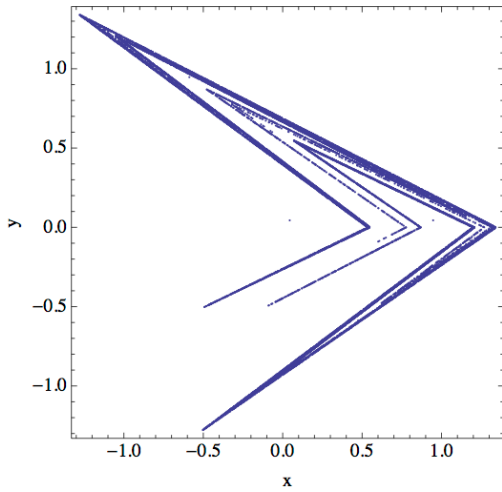


Figure 1: x, y plot of chaotic Lozi map

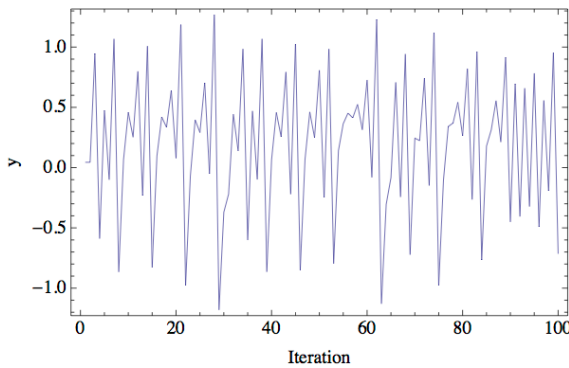


Figure 2: Iterations of the uncontrolled Lozi map.

2.2. ETDAS Control Method

This work is focused on the direct performance comparisons of DE and PSO algorithms in the task of tuning of parameters for ETDAS control method to stabilize desired Unstable Periodic Orbits (UPO). The original control method – ETDAS has form (2).

$$\begin{aligned} F(t) &= K[(1 - R)S(t - \tau_d) - x(t)] \\ S(t) &= x(t) + RS(t - \tau_d) \end{aligned} \quad (2)$$

Where: K and R are adjustable constants, F is the perturbation; S is given by a delay equation utilizing previous states of the system and τ_d is a time delay.

The original control method – ETDAS in the discrete form suitable for optimizations and connection with discrete maps has form (3).

$$\begin{aligned} F_n &= K[(1 - R)S_{n-m} - x_n] \\ S_n &= x_n + RS_{n-m} \end{aligned} \quad (3)$$

Where: m is the period of m -periodic orbit to be stabilized. The perturbation F_n in equations (3) may have arbitrarily large value, which can cause diverging of the system outside the interval $\{0, 1.0\}$. Therefore, F_n should have a value between F_{\min}, F_{\max} . To find the optimal value, it is a task of metaheuristics.

2.3. Cost Function

The idea of the basic cost function (CF_{Simple}), which could be used problem-free only for the stabilization of $p-1$ orbit (stable state of chaotic system – no oscillations), was to minimize the area created by the difference between the required state and the real system output on the whole simulation interval – τ_i . (4). This CF design is very convenient for the evolutionary searching process due to the relatively favorable CF surface and it is convenient for simple simulation based performance comparisons. Nevertheless, this simple approach has one big disadvantage, which is the including of initial chaotic transient behavior of not stabilized system into the CF value. As a result of this, the very tiny change in control method setting for extremely sensitive chaotic system (given by the very small change of CF value), can be suppressed by the above-mentioned including of initial chaotic transient.

$$CF_{Simple} = \int_{t=0}^i |TS_t - AS_t|, \quad (4)$$

where: TS - target state, AS - actual state

3. USED METAHEURISTICS

This work has utilized two metaheuristics as an examples of swarm based algorithm, which is PSO and population (evolution) based algorithm DE.

3.1. Particle Swarm Optimizer – PSO

Original PSO algorithms take their inspiration from behaviour of fish and birds (Kennedy and Eberhart 1995). The knowledge of the global best-found solution (typically denoted as $gBest$) is shared among the particles in the swarm. Furthermore, each particle has the knowledge of its own (personal) best-found solution (designated $pBest$). The last important part of the algorithm is the velocity of each particle, which is taken into account during the calculation of the particle's movement. The new position of each particle is then given by (5), where x_i^{t+1} is the new particle position; x_i^t refers to the current particle position and v_i^{t+1} is the new velocity of the particle.

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (5)$$

To calculate the new velocity, the distance from $pBest$ and $gBest$ is taken into account along with its current velocity (6).

$$v_{ij}^{t+1} = v_{ij}^t + c_1 \text{Rand} (pBest_{ij} - x_{ij}^t) + c_2 \text{Rand} (gBest_j - x_{ij}^t) \quad (6)$$

Where:

v_{ij}^{t+1} - New velocity of the i^{th} particle in iteration $t+1$.; (component j of the dimension D).

v_{ij}^t - Current velocity of the i^{th} particle in iteration t .; (component j of the dimension D).

$c_1, c_2 = 2$ - Acceleration constants.

$pBest_{ij}$ - Local (personal) best solution found by the i^{th} particle; (component j of the dimension D).

$gBest_j$ - Best solution found in a population; (component j of the dimension D).

x_{ij}^t - Current position of the i^{th} particle; (component j of the dimension D) in iteration t .

$Rand$ - Pseudo-random number, interval (0, 1).

3.2. Differential Evolution

DE is a simple and powerful population-based optimization method that works either on real-number-coded individuals or with small modifications on discrete type individuals (Price, Storn and Lampinen 2005), (Storn and Price 1997), (Price 1999). DE is quite robust, fast, and effective, with global optimization ability. This global optimization ability has been proven in many interdisciplinary researches. It works well even with noisy and time-dependent objective functions. The canonical basic principle is following.

For each individual $\vec{x}_{i,G}$ in the current generation G , DE generates a new trial individual $\vec{x}_{i,G}$ by adding the weighted difference between two randomly selected individuals $\vec{x}_{r1,G}$ and $\vec{x}_{r2,G}$ to a randomly selected third individual $\vec{x}_{r3,G}$. The resulting individual $\vec{x}_{i,G}$ is crossed-over with the original individual $\vec{x}_{i,G}$. The fitness of the resulting individual, referred to as a perturbed vector $\vec{u}_{i,G+1}$, is then compared with the fitness of $\vec{x}_{i,G}$. If the fitness of $\vec{u}_{i,G+1}$ is greater than the fitness of $\vec{x}_{i,G}$, then $\vec{x}_{i,G}$ is replaced with $\vec{u}_{i,G+1}$; otherwise, $\vec{x}_{i,G}$ remains in the population as $\vec{x}_{i,G+1}$.

Please refer to (7) for notation of crossover, and to (Price, Storn and Lampinen 2005) for the detailed description of used DERand1Bin strategy and all other DE strategies:

$$u_{i,G+1} = x_{r1,G} + F (x_{r2,G} - x_{r3,G}) \quad (7)$$

4. SIMULATION RESULTS

This research encompasses three case studies. Three different required behavior of the chaotic system were simulated in the following form:

- Case study 1: p-1 UPO, Lozi map as controlled system with CF_{Simple} .
- Case study 2: p-2 UPO, Lozi map as controlled system with CF_{Simple} .
- Case study 3: higher order p-4 UPO, Lozi map as controlled system with CF_{Simple} .

The ranges of all evolutionary estimated parameters are

given in Tab. 1.

Table 1: Estimated parameters and ranges

Parameter	Min	Max
K	-2	2
R	0	0.99
Fmax	0	0.9

Within the research a total number of 50 simulations for each experiments were performed in an environment of *Wolfram Mathematica*. All experiments used different initialization, i.e. different initial population was generated in each run of DE/PSO.

The parameter settings for both DE and PSO were given following way (see Table 2 and 3)

Table2: DE settings

Parameter	Value
PopSize	25
F	0.5
Cr	0.9
Generations	300
Max. CF Evaluations (CFE)	7500

Table 3: PSO settings

Parameter	Value
PopSize	25
c_1, c_2	2.0
Iterations	300
Max. CF Evaluations (CFE)	7500

All simulations were successful and have given new optimal settings for ETDAS control method securing the fast stabilization of the chaotic system at required behaviors, which were p-1 UPO (stable state), p-2 UPO (oscillation between two values) and finally p-4 UPO. The organization of the results is following:

Tables 5 -7 are focused on the performance comparisons between DE and swarm based PSO algorithm. These tables contain simple statistical overview of evolutionary optimization/simulation results i.e. average, median max, min (the best solution), std. dev. values for the particular cost function and for all 50 runs of both compared heuristics. Italic numbers represent the best result.

The chaos stabilization properties for the particular case studies are given in the description (caption) of the figures with the simulation of the best individual solutions. These chaos stabilization properties contain parameters set up for ETDAS control method, also the Istab. Value representing the number of iterations required for stabilization on desired UPO and further the average error between desired output value and real system output from the last 20 iterations.

Graphical simulation outputs of the best individual solutions for particular case studies are depicted in Figures 3, 5 and 7 whereas the Figures 4, 6 and 8 show the simulation output of all 50 runs of the best performing metaheuristic, thus confirm the robustness of this

approach. For the illustrative purposes, all graphical simulations outputs are depicted only for the variable x of the stabilized chaotic system.

The graphical comparisons for the performance analysis of DE and PSO within all 3 case studies are given in complex Figure 9. It shows the comparisons of time evolution of average CF values for all 50 runs of DE/PSO, thus it confirms the robustness of both used metaheuristic strategies within many repeated runs.

The values for desired UPOs of unperturbed chaotic Lozi map based on the mathematical analysis of the systems are given in Table 4.

Table 4: The values for desired UPOs.

UPO	Values of UPO of unperturbed system
p-1	$x_F = 0.454545$
p-2	$x_1 = -0.382166 ; x_2 = 0.700637$
p-4	$x_1 = -0.691899 ; x_2 = 0.334059$ $x_3 = 0.086151 ; x_4 = 1.020573$

4.1. Case study 1

In the simplest case, the performance of both heuristic is similar from the statistical point of view. Nevertheless, the DE has converged very fast towards optimal solution in 20 generations, as in Fig 9, (multiplied by population size, it means only 50 CF evaluations were required). All 50 runs of heuristic have given identical results (Fig 4), i.e. confirmed the robustness of the DE algorithm while searching in very nonlinear solution space. Convergence of PSO is much slower, DE is therefore suitable also for on-line control of nonlinear chaotic dynamics.

Table 5: Comparison for DE and PSO case study 1.

Statistical data	DE	PSO
	CF Value	CF Value
Min	0.520639	0.530679
Max	0.527132	0.573742
Average	0.520769	0.548688
Median	0.520639	0.549588
Std.Dev.	$9.18 \cdot 10^{-4}$	$1.07 \cdot 10^{-2}$

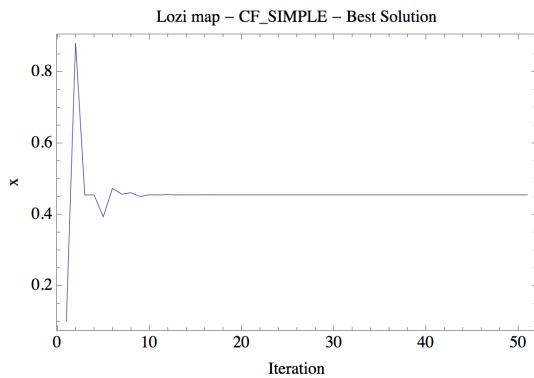


Figure 3: Simulation of the best individual solution – DE and Lozi map: Case study 1, $K = -1.11259$, $F_{max} = 0.9$, $R = 0.289232$, Istab. Value = 21, Avg. err. = $7.21 \cdot 10^{-15}$

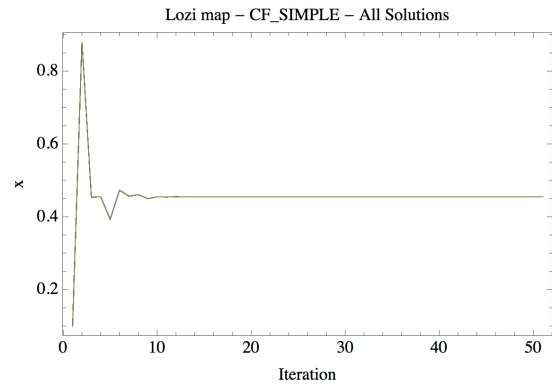


Figure 4: All 50 runs of EA – DE and Lozi map: Case study 1.

4.2. Case study 2

As in the previous case, results structure, as well as simulation outputs show similar features, even with a larger difference in favor of the algorithm DE.

Table 6: Comparison for DE and PSO case study 2.

Statistical data	DE	PSO
	CF Value	CF Value
Min	6.99829	7.29818
Max	7.33379	8.05143
Average	7.2827	7.67487
Median	7.33379	7.6944
Std.Dev.	$9.29 \cdot 10^{-2}$	0.216771

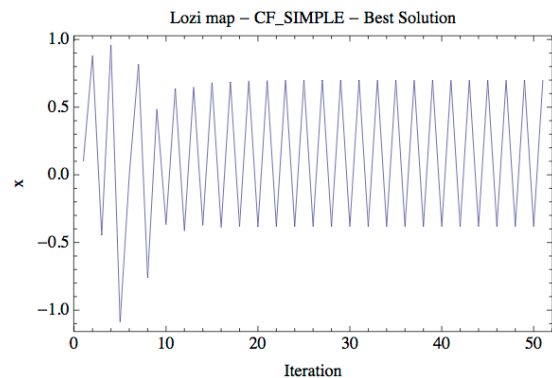


Figure 5: Simulation of the best individual solution – DE and Lozi map: Case study 2, $K = 0.5740$, $F_{max} = 0.4308$, $R = 0.4454$, Istab. Value = 22, Avg. err. = $2.98 \cdot 10^{-8}$

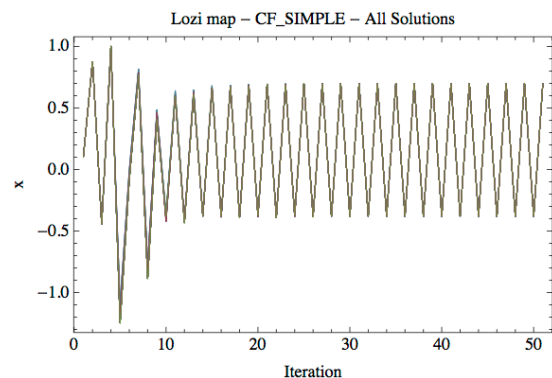


Figure 6: All 50 runs of EA – DE and Lozi map: Case study 2.

4.3. Case study 3

Table 7: Comparison for DE and PSO case study 3.

Statistical data	DE	PSO
	CF Value	CF Value
Min	13.7305	14.9151
Max	51.9391	29.6333
Average	15.4354	21.794
Median	14.2548	22.4428
Std.Dev.	5.4219	3.5719

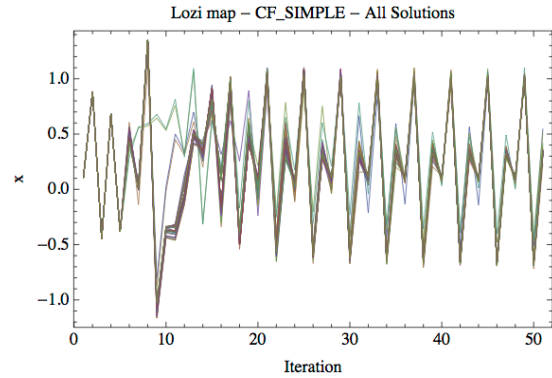


Figure 8: All 50 runs of EA – DE and Lozi map: Case study 3.

The last case study dealing with the most complex and highly nonlinear dynamics shows interesting features. Classical geometrical (vector crossover) based evolutionary algorithm has been stuck in many suboptimal solutions, with small chance to leave this area of solution space. While PSO (swarm based) algorithm was statistically better in searching process (lower standard deviation, range of CF). Even though the DE has found lower final CF value, the PSO seems to be a better and more robust choice for optimization in extremely nonlinear solution space due to its natural better exploration ability.

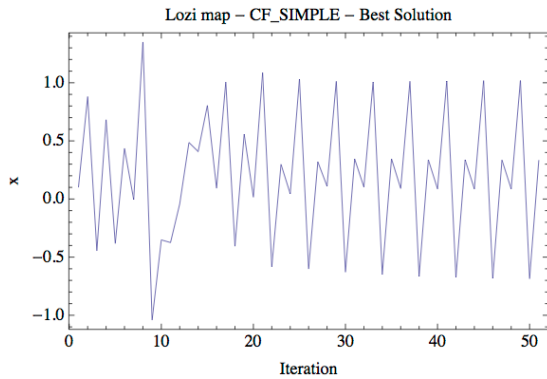


Figure 7: Simulation of the best individual solution – DE and Lozi map: Case study 1, $K = -0.8693$, $F_{max} = 0.25556$, $R = 0.4093$, Istab. Value = 46, Avg. err. = $2.96 \cdot 10^{-4}$

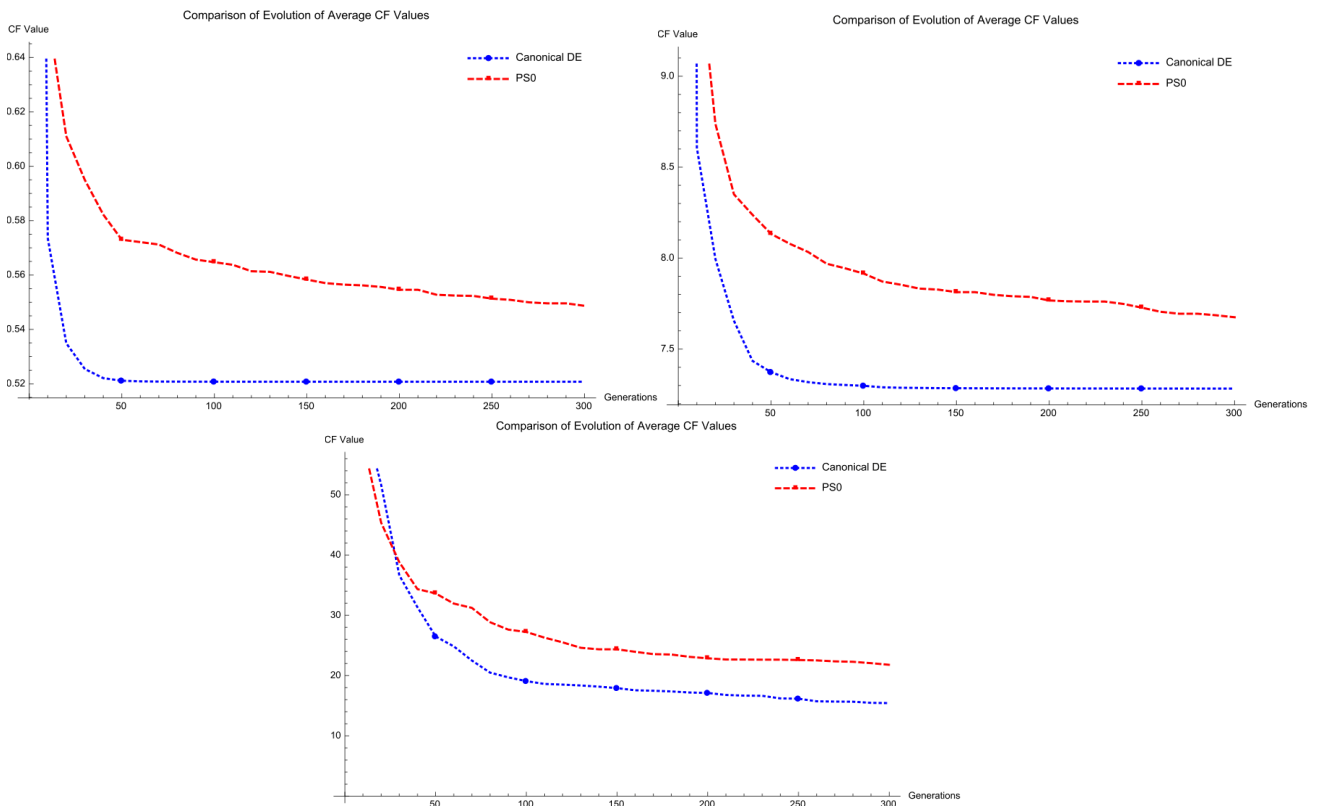


Figure 9: Comparisons of time evolution of average cost function values of the best solutions from all 50 runs of DE and PSO; from left to right and from above to bottom: Case study 1 – Case study 3.

5. CONCLUSION

In this paper, evolutionary algorithm Differential Evolution and swarm algorithm PSO were used for the evolutionary tuning of controller parameters for the stabilization of selected discrete chaotic system, which was the two-dimensional Lozi map.

The originality of the presented approach is that the examples of swarm based algorithms and evolutionary algorithms are compared in such a highly nonlinear and complex multimodal optimization task, which is optimal control of chaotic systems.

Comparisons between both DE versions and swarm based PSO algorithm show, that PSO is not good choice in the task of nonlinear/chaos control optimization.

Future research will be aimed at energy costs and more precise and faster stabilization and at the time-continuous systems, not only discrete chaotic maps.

ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund under the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089 and project IT4Innovations Centre of Excellence No. CZ.1.05/1.1.00/02.0070 and by Internal Grant Agency of Tomas Bata University under the project No.IGA/FAI/2012/037.

REFERENCES

- Abedini, M., R. Vatankhah, N. Assadian Stabilizing chaotic system on periodic orbits using multi-interval and modern optimal control strategies. *Communications in Nonlinear Science and Numerical Simulation*, 2012, 17(10), 3832-3842.
- Aziz-Alaoui, M. A., C. Robert, C. Grebogi Dynamics of a Hénon–Lozi-type map. *Chaos, Solitons & Fractals*, 2001, 12(12), 2323-2341.
- Coelho, L. D. S., R. B. Grebogi Chaotic synchronization using PID control combined with population based incremental learning algorithm. *Expert Systems with Applications*, 2010, 37(7), 5347-5352.
- Kennedy, J., R. Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks*. 1995, vol. 4, p. 1942-1948.
- Kominkova Oplatkova, Z., R. Senkerik, I. Zelinka, M. Pluhacek Analytic programming in the task of evolutionary synthesis of a controller for high order oscillations stabilization of discrete chaotic systems. *Computers & Mathematics with Applications*, 2013, 66(2), 177-189.
- Ott, E., C. Grebogi, J. A. Yorke Controlling chaos. *Physical Review Letters*, 1990, 64(11), 1196-1199.
- Price K., Storn R. M., Lampinen J. A., 2005, "Differential Evolution : A Practical Approach to Global Optimization", (Natural Computing Series), Springer; 1 edition.
- Price, K. V. An Introduction to Differential Evolution. In D. CORNE, M. DORIGO AND F. GLOVER eds. *New Ideas in Optimization*. McGraw-Hill Ltd., 1999, p. 79-108.
- Pyragas K., 1995. "Control of chaos via extended delay feedback", *Physics Letters A*, vol. 206, 1995, pp. 323-330.
- Richter, H., K. J. Reinschke Optimization of local control of chaos by an evolutionary algorithm. *Physica D: Nonlinear Phenomena*, 2000, 144(3–4), 309-334.
- Sadeghpour, M., H. Salarieh, G. Vossoughi, A. Alasty Multi-variable control of chaos using PSO-based minimum entropy control. *Communications in Nonlinear Science and Numerical Simulation*, 2011, 16(6), 2397-2404.
- Senkerik, R., I. Zelinka, D. Davendra, Z. Oplatkova Utilization of SOMA and differential evolution for robust stabilization of chaotic Logistic equation. *Computers & Mathematics with Applications*, 2010, 60(4), 1026-1037.
- Senkerik, R., Z. Kominkova Oplatkova, I. Zelinka, B. Chramcov, et al. Utilization of analytic programming for the evolutionary synthesis of the robust multi-chaotic controller for selected sets of discrete chaotic systems. *Soft Computing*, 2014, 18(4), 651-668.
- Senkerik, R., Z. Oplatkova, I. Zelinka, D. Davendra Synthesis of feedback controller for three selected chaotic systems by means of evolutionary techniques: Analytic programming. *Mathematical and Computer Modelling*, 2013, 57(1–2), 57-67.
- Sprott, J. C. *Chaos and Time-Series Analysis*. Edition ed.: Oxford University Press, 2003. 528 p.
- Storn, R., K. Price Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 1997, 11(4), 341-359.
- Zelinka, I. Real-time deterministic chaos control by means of selected evolutionary techniques. *Engineering Applications of Artificial Intelligence*, 2009, 22(2), 283-297.
- Zelinka, I., R. Senkerik, E. Navratil Investigation on evolutionary optimization of chaos control. *Chaos, Solitons & Fractals*, 2009, 40(1), 111-129.

MESOSCOPIC SIMULATION MODELS FOR THE AUTOMOTIVE INDUSTRY

Andreas Hennig^(a), Tobias Reggelin^(b), Daniel Wolff^(c)

^(a) Otto von Guericke University Magdeburg

^(b) Otto von Guericke University Magdeburg, Fraunhofer Institute for Factory Operation and Automation IFF Magdeburg

^(c) INPRO Innovationsgesellschaft für fortgeschrittene Produktionssysteme in der Fahrzeugindustrie mbH Berlin

^(a) andreas.hennig@st.ovgu.de, ^(b) Tobias.Reggelin@iff.fraunhofer.de, ^(c) daniel.wolff@inpro.de

ABSTRACT

Simulation models allow multiple options to analyse and improve production and logistic processes. The mesoscopic modelling approach is a relatively new concept which combines the advantages of the widespread continuous and discrete event simulation while minimizing the disadvantages. The paper verifies the feasibility of the mesoscopic concept in context of topics concerning the automotive industry. Therefore the approach is used to model a manufacturing process of components with the focus on resource efficiency. A discrete event model of the production line is used to validate and verify the mesoscopic model. Founded on a validated and verified mesoscopic model different scenarios are presented in order to optimize the aspect of resource efficiency. Based on the findings of the simulation study recommendations for action to improve the mesoscopic approach are provided. In addition, potential problems from the automotive sector are presented which can be examined in the future using the mesoscopic approach.

Keywords: mesoscopic simulation, resource efficiency, compressed air, production

1. INTRODUCTION

By using simulation models multiple tasks in production and logistics areas can be depicted and optimized. Due to an ongoing increase in the complexity of many production systems the relevance of simulation will most likely increase as well. There are two main modelling approaches to investigate these topics: System Dynamics (SD) and Discrete Event Simulation (DES). In most of the cases DES is used as modelling approach for production logistic aspects. But both mentioned approaches have some disadvantages which will be described in the following. To reduce these disadvantages the mesoscopic approach was developed.

So in this work, possibilities of the relatively new mesoscopic simulation approach for topics in the field of the automotive industry are presented and examined using a practical example concerning aspects of resource efficiency.

2. DESCRIPTION OF SIMULATION APPROACHES

2.1. System Dynamics

The System Dynamics approach was invented in the 1950s by Jay W. Forrester at the Massachusetts Institute of Technology (MIT) and is a continuous modelling approach (Borshchev and Filippov 2004). It is characterised by terms of stocks, flows between these stocks and information about the values of the flows. So Systems Dynamic models use an aggregated view and abstract from single events. This leads to the most significant downside of this approach: the results of the simulation cannot be very exact. But on the other hand it reduces the time for modelling and simulating. Some topical papers can be found which use System Dynamics in order to simulate production and logistic processes. (Herrera et al. 2014) use a System Dynamics model to describe the introduction of RFID technology and the associated impact on the supply chain processes of groceries. (Aschauer 2013) describes the application of a System Dynamics model for the development of a sustainable transport concept to reduce pollution.

2.2. Discrete Event Simulation

In contrast the widely spread Discrete Event Simulation approach is characterized by the concept of entities, resources and block charts which describe the entity flow and resource sharing (Borshchev and Filippov 2004). Every single unit of the system can be represented. These aspects enable the user to adjust the level of detail of the system at will. But especially for complex systems this results in high efforts for modelling the system and in long lasting simulation runs (Law and Kelton 2007). As stated above there are many examples of simulation studies using DES. For example (Witthaus et al. 2014) take a discrete event simulation as a planning assistance for a multiple stage distribution network. (Klaas and Klibi 2015) combine the discrete event simulation with mathematical optimization to strive for a holistic supply chain optimization.

2.3. Mesoscopic Simulation

The findings of this work are based on the mesoscopic approach which was developed for logistical purposes at the Otto-von-Guericke-University Magdeburg, Germany in cooperation with the Fraunhofer Institute for Factory Operation and Automation IFF in Magdeburg, Germany and is based on the Discrete Rate Simulation approach which is described in (Krahl 2009).

As mentioned before one of the aims of the mesoscopic simulation approach is to reduce the disadvantages of the Systems Dynamics and the Discrete Event approaches. So the approach should provide more exact results than System Dynamics models and decrease the time for modelling and simulation in comparison to discrete event models at the same time.

In order to reach the aims the mesoscopic approach uses aspects of both aforementioned approaches. This is achieved by using a medium level of detail which is sufficient for most tasks. Unlike in discrete event models not every single unit is depicted. Instead the approach uses flow rates like Systems Dynamics models do. But here they are only piecewise constant. So a new calculation is only necessary when the rate changes, e.g. after reaching a predefined limit and can be predicted. This aspect leads to an immense decrease in calculations during a simulation run. In addition to this the mesoscopic approach offers the possibility to trigger impulses with includes the aspects of discrete event models (Reggelin 2011).

3. DESCRIPTION OF THE SIMULATION STUDY

3.1. Classification and Motivation

As mentioned in the introduction the mesoscopic approach was investigated with focus on topics in the automotive industry. Therefore at first a representation of typical planning, analysis and design tasks in the automotive field was given. Here a subdivision into the following four categories can be made:

- business simulation,
- supply simulation,
- supply chain simulation and
- traffic flow simulation.

For detailed information on these categories see (Müller-Sommer and Strassburger 2009).

To illustrate the practical relevance of the investigated topic in the simulation study the reasons for the simulative consideration of resource use strategies were worked out. Here, both economic and ecological aspects are taken into account. Due to a decreasing amount of natural raw materials companies are confronted with increasing purchase prices. In order to improve their balance sheets companies are willing to optimize their usage of raw materials. On the other hand the user awareness of ecological aspects grew dramatically over the last years so companies can get a unique selling

proposition through an environmentally friendly production system.

Because the practical example specifically relates to the compressed air consumption of a component manufacturing area of an automobile manufacturer, various measures were introduced additionally which can be used to optimize a compressed air system. These measures can be divided into the following categories:

- generation,
- processing,
- allocation and
- application.

The most significant reductions can be achieved by adjusting the dimensioning of the installed generators in combination with a superordinate control (EnEffAH, 2012).

3.2. Initial Situation and Objective Target

The aim of the simulation study was to verify whether the mesoscopic approach is suitable to analyse and optimize the resource efficiency of a component production of an automobile manufacturer (see Figure 1). If the applicability of the mesoscopic approach can be confirmed the optimisation of the usage of compressed air was another goal. Usually all analyses in this branch are performed using discrete event simulation. Therefore the production was already modelled as a discrete event model in the simulation tool Plant Simulation. Due to the importance and the increasing awareness of aspects of resource efficiency it can be assumed that proof of use will lead to many further application possibilities in this area.

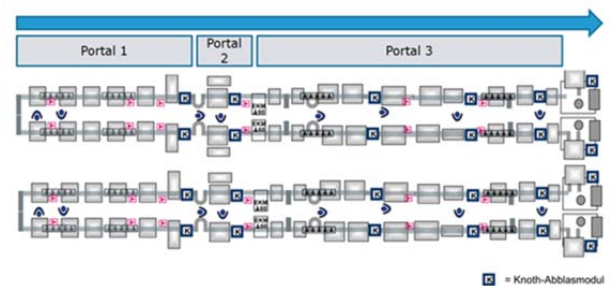


Figure 1: Schematic Overview of the Production

3.3. Procedure

To test the applicability of the mesoscopic simulation approach a simulation study on the basis of "VDI guideline 3633" in the simulation tool ExtendSim has been performed. So first of all the system was analysed and all relevant data for the mesoscopic approach from the DES model needed to be extracted and edited. The production process corresponds to a highly automated production line with 20 process steps. Six of these steps with so called dry cleaners are particularly significant since they use compressed air. Moreover they are only few variants produced which meets the requirements of the mesoscopic approach (Reggelin 2011).

As mentioned before the investigated system was already modelled in Plant Simulation. That is why the relevant data can be extracted from the existing model. Due the fact that the DES model is frequently used and under maintenance the timeliness of data is provided. Nevertheless the extracted data needed to be checked for plausibility to avoid transmission errors.

After collecting the data the actual systems needed to be transformed into a conceptual model. As shown in Figure 2 the original concept was to model the system in a low level of detail. The production line was divided into four black boxes which were defined by the used means of transportation. Based on this concept the system was modelled in ExtendSim (see Figure 3) and the first test runs were executed.

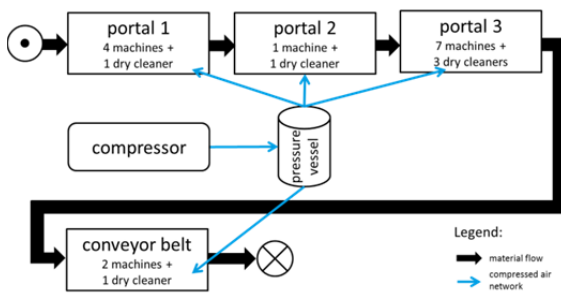


Figure 2: Abstract Conceptual Model

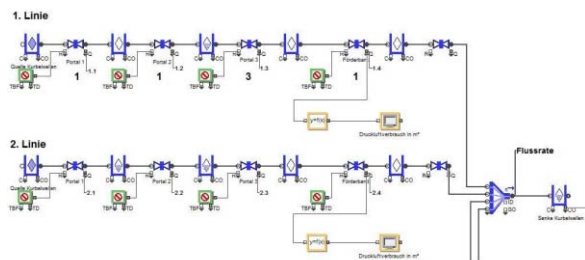


Figure 3: Detail of Abstract Modelling Approach

Within the test runs it became obvious that this approach would lead to a deviation in the throughput per production line of about 8.5 % compared to the DES model which was not satisfying. In order to reduce the deviation the level of detail of the mesoscopic model was increased (see Figure 4).

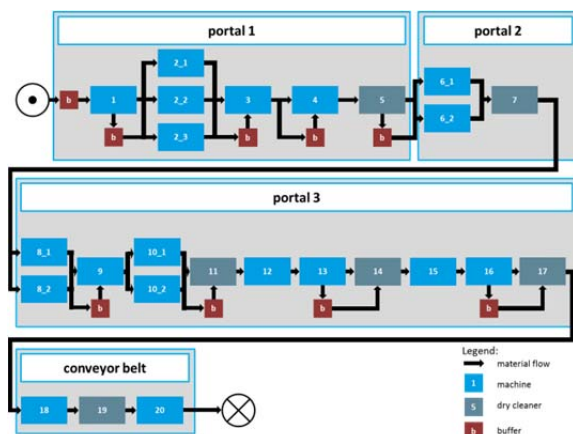


Figure 4: Detailed Conceptual Model

The most significant changes are that the black boxes were modelled in detail and buffers were implemented separately. The investigation showed that the implemented buffers led to an enormous decrease in the deviation.

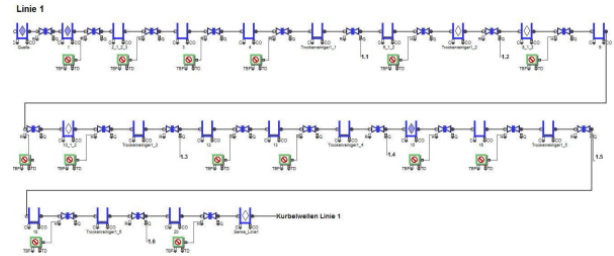


Figure 5: Detail of Final Model

The final concept (see Figure 5) of the mesoscopic model was validated and verified based on a discrete event model which illustrates the same production area. For this purpose, among others, the throughput per hour per production line, the overall compressed air and energy consumption as well as the operation of the compressor were compared (see Table 1). It can be seen that using the detailed mesoscopic approach led to almost the same results as in the DES model. So it can be stated that the detailed mesoscopic model was applicable for the investigated topic.

Table 1: Comparison of Verification and Validation Criteria

	DES Model (100 %)	Detailed Mesoscopic Model	Deviation [%]
Throughput per Line and Hour	42,3	42,0	-0,71
Compressed Air Consumption [m ³]	50.813,26	50.370,20	-0,87
Energy Consumption [kWh]	1.785,37	1.774,98	-0,58
Duty Cycle [min]	30	29	-3,33
Interrupting time [min]	56	57	+1,79

After proofing the applicability of the mesoscopic approach the optimization of the energy demand caused by the compressed air generation and distribution was examined. In several scenarios variations of the performance specification of the compressor(s) used were studied, which should lead to a reduction of the required energy demand for the generation of compressed air. Figure 6 shows a detail of the pressure curve of a scenario in which a second compressor was installed. The red curve depicts the pressure curve in the

initial situation whereas the blue curve represents the pressure curve in the scenario.

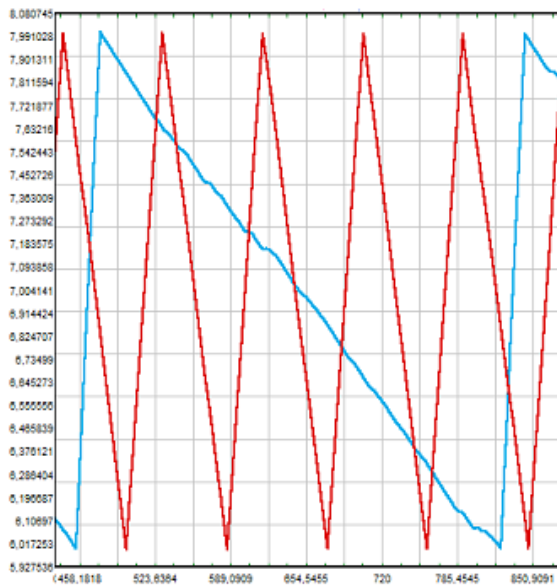


Figure 6: Detail Pressure Curve

Based on information obtained during the simulation study and by using expert knowledge of the modeller of the discrete event model, a comparison of the used simulation tools was performed. The aim of this comparison was the derivation of improvement measures for the further development of an independent mesoscopic simulator.

4. RESULTS

As a result of the work it can be stated that the mesoscopic approach is well suited for handling the analysed subject among the underlying constraints and assumptions. After initially deviations in the hourly throughput per production line of about 8.5 % were recorded, this deviation was reduced to about 0.7 % by adapting the way of modelling (see Table 2). The adapted model in this case has a significantly lowered level of abstraction compared to the originally pursued modelling approach. Based on the detailed modelling the described scenarios were investigated mesoscopically. These differ in the number and the performance of the compressors used and will contribute to improve efficiency in the production of compressed air through an optimized dimensioning.

Table 2: Comparison of the Throughputs of different modelling approaches

	DES Model	Abstract Mesoscopic Model	Detailed Mesoscopic Model
Throughput per Line and Hour	42,3	38,7	42,0

The consideration of the scenarios leads to the realization that for the investigated manufacturing area

a compressor with a significantly reduced performance should be used in comparison to the initial situation. Through the use of the compressor with the optimized performance, a saving of approximately 7.6% with respect to the compressed air related energy consumption compared to the initial situation could be realized (see Figure 7). Since the values obtained for the performance of the compressor are only slightly higher than the actual consumption of the production area, changes in the complete control logic of the pneumatic system could be considered. In the investigations of this work a discontinuous full load intermittent control of the compressor was implemented. On the basis of the results of the scenarios the control logic could be switched to a continuous, speed-controlled operation of the compressor. In order to change the control logic further investigations of the compressor and other components of the pneumatic system would be necessary.

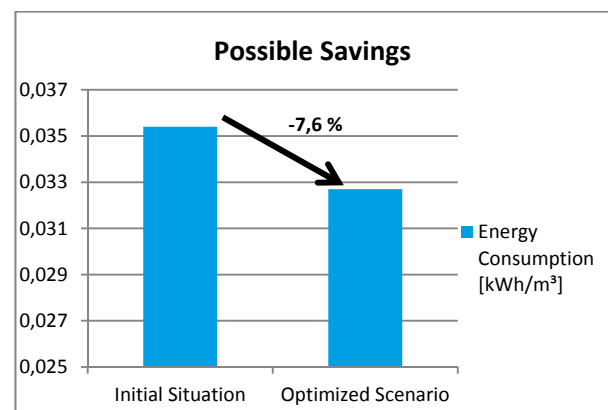


Figure 7: Possible Energy Consumption Savings

Furthermore a crucial requirement which was expressed in the development of mesoscopic simulation approach is confirmed. As required the mesoscopic approach leads to a significantly reduced simulation time for the examined subject compared to the discrete event approach. In actual applications, the mesoscopic model takes only a few seconds to perform a simulation run of the length of one day. For a run of the same observation period, the discrete event model takes about 12 minutes. With the expansion of the simulation period, the differences will be even much more pronounced in all probability. The mesoscopic approach demonstrates its better usability for very short-term, operational issues. To support this advantage even further a high user-friendliness should be ensured through appropriate input forms, in order to make quick and easy adjustments of model parameters. In addition to applications in operations other potential application areas for the mesoscopic approach were identified in the field of automobile industry. To promote the development and dissemination of the approach, it is advisable to gain practice partners from this sector to model other automotive topics mesoscopically.

Therefore implementing further mesoscopic models in the automotive industry with partners should be the next step. Particular emphasis may be placed on the consideration of supply chain networks. These networks correspond very well to the principle of the mesoscopic approach through their abstract nature, for example by using average rates for transport times or the use of handling time packages. Furthermore forwarders may also be interesting partners for such studies who want to deal with the optimization of their regional forwarding network. Another possible area directly from the automotive sector to support the development of the mesoscopic approach can be found in the simulative consideration of trades such as the press shop that are characterized by a lot size production or a small number of variants.

In the course of these investigations, the independent mesoscopic simulator should be further developed and adapted to the requirements of the customer. Another interesting aspect arises in the question whether already existing discrete event models can be converted automatically into mesoscopic models to achieve an even higher acceptance by the users. Since many production areas in the automotive industry are already present as discrete event models, another time advantage could be generated through this automated conversion of models, as an entirely new model could be omitted.

5. CONCLUSION

In summary it can be said that the mesoscopic simulation approach offers much potential for further investigation of practically relevant applications in the automotive sector but also for other industries. More interesting tasks and collaborations may arise that can proof the efficiency of the approach.

REFERENCES

- Aschauer G. J., 2013. The development of sustainable transports through the right logistics strategy – a system dynamics approach. Proceedings of the 31st International Conference of the System Dynamics Society. July 21-25, Cambridge (USA).
- Borshchev, A., Filippov, A., 2004. From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Technique, Tools. Proceedings of the 22nd International Conference of the System Dynamics Society. July 25-29, Oxford (England).
- EnEffAH, 2012. Energieeffizienz in der Produktion im Bereich Antriebs- und Handhabungstechnik. Available from: http://www.eneffah.de/EnEffAH_Broschuere.pdf [accessed 10 July 2016].
- Herrera M. M., Becerra M., Romero O. R., Orjuela J. A., 2014. Using System Dynamics and Fuzzy Logic to Assess the Implementation RFID Technology. Proceedings of the 32st International Conference of System Dynamics Society, July 20-24, Delft (Netherlands).
- Klaas, A., Klibi, K.: Ganzheitliche Supply-Chain-Optimierung mittels Kombination von mathematischer Optimierung und Simulation. In: Rabe M., Clausen U.: Simulation in Production and Logistics 2015. Stuttgart: Fraunhofer IRB Verlag, pp. 129-135.
- Krahl D., 2009. Extendsim Advanced Technology: Discrete Rate Simulation. Proceedings of the 2009 Winter Simulation Conference, pp. 333-338. December 13-16, Austin (USA).
- Law A. M., Kelton, W., 2007. Simulation Modeling and Analysis. Boston: McGraw-Hill.
- Müller-Sommer H., Strassburger S., 2009. Ausprägungen und Nutzungsgrad der Logistiksimulation im Umfeld der Automobilindustrie. Konferenzband zum 20. Symposium Simulationstechnik. pp. 353-361, September 23-25, Cottbus (Germany).
- Reggelin T., 2011. Mesoskopische Modellierung und Simulation logistischer Flusssysteme. Thesis (PhD), Otto-von-Guericke-Universität Magdeburg.
- Witthaus M., Kamphues J., Hegmanns T., 2015. Simulative Planungsassistenz für die Disposition in mehrstufigen Distributionsnetzwerken. In: Rabe M., Clausen U.: Simulation in Production and Logistics 2015. Stuttgart: Fraunhofer IRB Verlag, pp. 137-146.

AUTHORS BIOGRAPHY

Andreas Hennig, M.Sc. studied industrial engineering and management with the focus Supply Chain Management and Network and Sustainable Logistics at Otto von Guericke University Magdeburg, Germany. He received his master's degree in 2016.

Dr.-Ing. Tobias Reggelin is a member of the faculty of the Institute of Logistics and Material Handling Systems at Otto von Guericke University Magdeburg and a Research Manager at the Fraunhofer Institute for Factory Operation and Automation IFF in Magdeburg, Germany. Tobias Reggelin studied industrial engineering and management at Otto von Guericke University Magdeburg and Rose-Hulman Institute of Technology in the USA.

Dipl.-Ing. Daniel Wolff is the head of the division Production Systems and Information Processes at INPRO GmbH. He studied mechanical engineering with the focus factory planning and operation at TU Chemnitz, Germany.

MODELING AND SIMULATION OF THE EFFECTS OF THERMAL EXPANSION MISMATCH AND TEMPERATURE ON THERMAL INDUCED STRESSES IN MODERN NANO PMOS

Abderrazzak El Boukili

Al Akhawayn University in Ifrane, Morocco

Email: a.elboukili@au.ma

ABSTRACT

Stress is used to enhance dramatically the electrical and optical performances of modern nanotransistors and solar cells. The purpose of this paper is to develop and apply accurate mathematical models to simulate and analyze correctly the thermal induced stresses in nano pMOS transistors having Silicon-Germanium (SiGe) thin films as stressors. These stresses are due to the growth temperatures of the SiGe films and to the thermal mismatch between the thermal expansion coefficients (TECs) of SiGe films and Silicon substrate. In the proposed models, we are accurately taking into account the effects of the fabrication temperatures on the TECs of SiGe films and Silicon substrate to calculate the resulting thermal strains and stresses. Numerical results showing the effects of growth temperatures and thermal mismatch between TECs of SiGe and Silicon on the resulting thermal stresses will be presented, discussed and compared with literature for Intel 14 nm pMOS transistor.

Keywords: Growth temperature effects, thermal expansion mismatch effects, modeling thermal induced stresses ,Intel 14 nm nano pMOS transistors

1. INTRODUCTION

Nano pMOS and nMOS transistors are the backbone of today's economy. They have many applications in our daily life. They are used in the RAMs, CPUs, and GPUs of any new laptops, smart phones, tablets, or in solar cells. They are also used in the new generation of iPad and smart-phone touch screens, or light-emitting diode displays for televisions, mobile phones, and computer monitors. Nanotransistors provide high performance, lower power consumption, lower cost per transistor and high level of integration compared to the micro transistors used between the invention of the first transistor in 1945 and the year 2000. The first nanotransistors were mainly born in the beginning of the year 2000. For example, Intel's Core i3 and i5 CPUs, released in January 2010, were among the first mass-produced processors to use nanotransistors with 32 nm technology node.

In February 2015, Samsung announced that its flagship smartphones Galaxy S6 and Galaxy S6 Edge would use nanotransistors with 14 nm technology node. Nanotransistors based on Graphene material are the new candidates for the next generation solar cell applications and photovoltaic industry (Yin et al., 2014).

Since 2004, the leading manufacturers of nano pMOS and nMOS transistors as Intel, IBM and Samsung have been using intentionally material mismatch stress and thermal mismatch stress to enhance dramatically the mobility and speed of the new generation of nanotransistors. It has been shown in (Moroz et al., 2006) that the application of a compressive stress originating from different stressors enhanced the mobility of holes in nano pMOS transistors by 200%.

Intel technology is using epitaxy method and high temperatures T to grow Silicon-Germanium ($\text{Si}(1-x)\text{Ge}(x)$) thin film pockets in source and drain of nano pMOS to introduce intentionally a compressive extrinsic stress σ in the channel (Moroz et al., 2004, 2005, 2006, 2012, 2013; Bera et al., 2006; Chui et al., 2007). This compressive extrinsic stress σ is due mainly to an intrinsic (i.e. initial) compressive stress σ_0 which is generated in $\text{Si}(1-x)\text{Ge}(x)$ thin films after their growth on Silicon (Si) at elevated temperatures T . This intrinsic stress σ_0 is the sum of the initial stresses σ_0^{mm} and $\sigma_0^{tm}(T)$. Where, σ_0^{mm} is due to the lattice mismatch between $\text{Si}(1-x)\text{Ge}(x)$ films and Silicon substrate. And, $\sigma_0^{tm}(T)$ is the initial thermal mismatch stress due to the thermal mismatch between the TECs of $\text{Si}(1-x)\text{Ge}(x)$ thin films and Silicon substrate and to the change of temperature from the growth temperature T to the room temperature T_0 . In this paper, we are ignoring σ_0^{mm} and focusing on the study of the initial thermal mismatch stress $\sigma_0^{tm}(T)$ and the resulting extrinsic thermal mismatch stress σ .

Let $\Delta\alpha(x, T) = -\alpha_{Si(1-x)Ge(x)}(T) + \alpha_{Si}(T)$ represents the thermal expansion mismatch between the thermal expansion coefficient $\alpha_{Si(1-x)Ge(x)}(T)$ of Si(1-x)Ge(x) thin films and the thermal expansion coefficient $\alpha_{Si}(T)$ of Silicon substrate. And, let $\Delta T = T - T_0$ represents the change in temperature. T_0 is the room temperature.

We assume that Silicon, Germanium, and Si(1-x)Ge(x) are elastic materials. In this case, we can use Hooke's law to calculate the initial thermal induced stress tensor $\sigma_0^{tm}(T)$:

$$\sigma_0^{tm}(T) = (\sigma_0^{xx}(T), \sigma_0^{yy}(T), \sigma_0^{zz}(T), \sigma_0^{xy}(T), \sigma_0^{yz}(T), \sigma_0^{xz}(T))$$

in terms of the initial thermal induced strain tensor $\varepsilon_0^{tm}(T)$:

$$\varepsilon_0^{tm}(T) = (\varepsilon_0^{xx}(T), \varepsilon_0^{yy}(T), \varepsilon_0^{zz}(T), \varepsilon_0^{xy}(T), \varepsilon_0^{yz}(T), \varepsilon_0^{xz}(T))$$

as follows:

$$\sigma_0^{tm}(T) = D\varepsilon_0^{tm}(T), \quad (1)$$

where D represents the 6 by 6 stiffness matrix. It is calculated from the elastic constants c11, c12, and c44 as in (El Boukili, 2010). The terms $\sigma_0^{xx}(T), \sigma_0^{yy}(T), \sigma_0^{zz}(T)$ are the initial normal stress components. And, $\sigma_0^{xy}(T), \sigma_0^{yz}(T), \sigma_0^{xz}(T)$ are the initial shear stress components. The terms $\varepsilon_0^{xx}(T), \varepsilon_0^{yy}(T), \varepsilon_0^{zz}(T)$ are the initial normal strain components. And, $\varepsilon_0^{xy}(T), \varepsilon_0^{yz}(T), \varepsilon_0^{xz}(T)$ are the initial shear strain components.

We could see from Equation (1) that to calculate correctly and accurately the initial thermal stress $\sigma_0^{tm}(T)$ in Si(1-x)Ge(x) thin films, we need to calculate correctly and accurately the initial thermal mismatch strain $\varepsilon_0^{tm}(T)$.

The initial thermal mismatch strain $\varepsilon_0^{tm}(T)$ is calculated from the thermal expansion coefficients $\alpha_{Si(1-x)Ge(x)}(T), \alpha_{Si}(T)$, and $\alpha_{Ge}(T)$ of Si(1-

x)Ge(x) thin films, Silicon (Si), and Germanium (Ge) respectively.

There are three levels of originalities in this paper. The first originality is to propose and apply an accurate and physically based model to calculate accurately the TECs $\alpha_{Si(1-x)Ge(x)}(T), \alpha_{Si}(T)$, and $\alpha_{Ge}(T)$ under the effects of the fabrication temperatures T and Ge mole fraction x.

The second originality is to propose and apply new and accurate physically based model to calculate accurately the thermal induced initial strain $\varepsilon_0^{tm}(T)$ and stress

$\sigma_0^{tm}(T)$. The third originality is to proposed and apply new and accurate physically based model to calculate the thermal induced extrinsic stress σ from $\sigma_0^{tm}(T)$ using Finite Volumes method in the whole structure of Intel 14 nm nano pMOS transistor.

The investigation, analysis, and quantification of the effects of the thermal expansion mismatch represented by $\Delta\alpha(x, T)$ and the change in growth temperatures T on the resulting thermal extrinsic stress σ will be presented and discussed. This analysis will show to us how much thermal stress σ we will get when we change $\Delta\alpha(x, T)$ and ΔT . And, if we know how much thermal stress we get, we will know by how much the speed performance of the nano pMOS will increase. Since the speed of nano pMOS increases with increasing hole mobility. And, the hole mobility increases with decreasing compressive stress σ according to (JianLi et al., 2014; Kumar et al., 2012). For example, it has been shown in (Moroz et al., 2006) that the application of a compressive stress increased the hole mobility by 200%. So, this analysis is of great importance to the manufacturers, engineers, and designers of modern nanotransistors.

And, it is also of great importance to the researchers and designers of the modern concentrators photovoltaic (CPV) solar cells or just Silicon based (SB) solar cells working in arid regions where the temperature is high as the desert of Morocco or elsewhere. Since the thermal induced tensile stresses will reduce the efficiency of the CPV (or SB) solar cells.

This paper is organized as follows. Section 2 presents the main different sources of initial stress. Section 3 will present the proposed model to calculate the TECs $\alpha_{Si(1-x)Ge(x)}(T)$ and $\alpha_{Si}(T)$ under the effects of the fabrication temperatures T. Section 4 will outline the proposed new and physically based models to calculate the thermal induced intrinsic and extrinsic stresses ($\sigma_0^{tm}(T)$ and σ). Section 5 will present the numerical results for the resulting thermal induced

stress σ . It will present the validation and comparison of these results with experiments from literature. It will also present a quantitative and qualitative analysis of the effects of the thermal expansion mismatch $\Delta\alpha(x, T)$ and the change in the fabrication temperatures ΔT on the thermal induced stresses σ in Intel 14 nm nano pMOS transistors. Section 6 presents the concluding thoughts and future work.

2. DIFFERENT SOURCES OF INITIAL STRESS

The initial stress σ_0 in the Si(1-x)Ge(x) thin films or in the Silicon substrate depends on many different types of sources as: deposition, oxidation, etching, shallow trench isolation, diffusion, implantation, and neighboring. After deposition, two types of initial stresses will be created in the Si(1-x)Ge(x) thin films: the initial stress due to the lattice mismatch between Si(1-x)Ge(x) and Silicon materials. And, the initial stress due to thermal mismatch. The initial stress σ_0 due to lattice mismatch has been investigated in our previous work (El Boukili 2010). In this paper, we are going to focus on the modeling and analysis of the initial stress σ_0 induced by thermal mismatch.

2.1. Initial stress due to lattice mismatch

Right after deposition, thin films are either stretched or compressed to fit the substrate on which they are deposited. The film wants to be smaller if it was stretched earlier, thus creating a tensile initial stress. And, similarly, it creates a compressive initial stress if it was compressed during deposition. The initial stress generated due to this phenomenon can be quantified by Stoney's equation which relates the stress σ_0 to the substrate curvature or by other advanced models (El Boukili 2010).

2.2. Intrinsic stress due to thermal mismatch

Thermal mismatch stress depends on the mechanical properties of the materials. It occurs when two materials with different coefficients of thermal expansion are heated or cooled down. When two different materials are heated or cooled down, they will expand or contract at different rates. During thermal processing, thin film materials like Si(1-x)Ge(x), Poly-Silicon, Silicon Dioxide, or Silicon Nitride expand and contract at different rates compared to the Silicon substrate according to their thermal expansion coefficients. This creates an initial thermal mismatch strain and stress in the film and also in the substrate. The deposition takes place at elevated temperatures. When the temperature is decreased from the fabrication temperature T to the room temperature T_0 , the volumes of the grains of Si(1-x)Ge(x) films and Silicon substrate shrink and the thermal stresses in the material increase. The initial thermal stress σ_0 in the Si(1-x)Ge(x) films depend mainly on the fabrication temperature T , the TECs of Si(1-x)Ge(x), Germanium

ratio x , the substrate orientation, the doping, and the deposition technique used. For the current nano pMOS transistors, Si(1-x)Ge(x) films are grown using low pressure chemical vapor deposition (LPCVD), plasma enhanced chemical vapor deposition (PECVD), or Epitaxy. Intel company is using Epitaxial growth for Si(1-x)Ge(x) films.

3. MODELING OF TEMPERATURE EFFECTS ON TECs OF SI AND SIGE

To take into account the effects of the fabrication temperatures T on the TECs of Silicon, Germanium, and Si(1-x)Ge(x) thin films, we are proposing and applying the following models:

$$\alpha_{Si}(T) = s1 \cdot (1 - \exp(-s2 \cdot (\frac{T^2}{s3} - s4))) + s5 \cdot T \quad (2)$$

$$\alpha_{Ge}(T) = a1 \cdot (1 - \exp(-a2 \cdot (\frac{T^2}{a3} - a4))) + a5 \cdot T \quad (3)$$

$$\alpha_{Si(1-x)Ge(x)}(T) = (1-x)\alpha_{Si}(T) + x\alpha_{Ge}(T) \quad (4)$$

The parameters $s1, s2, s3, s4, s5, a1, a2, a3, a4,$ and $a5$ have been calculated, using the nonlinear least squares method, from fitting the models (2), (3), and (4) with experimental data found in (Okaji et al., 2004; Singh, 1968). The models (3)-(5) are accurate and fit excellently well with the experimental data from (Okaji et al., 2004; Singh, 1968).

4. NEW AND ACCURATE MODLS FOR THERMAL MISMATCH INDUCED INTRINSIC AND EXTRINSIC STRESSES

To take into account the effects of the fabrication temperatures T , the thermal mismatch $\Delta\alpha(x, T)$, and the Germanium content x on the thermal mismatch induced intrinsic strain $\varepsilon_0^{tm}(T)$ in the Si(1-x)Ge(x) thin films, we propose and apply the following accurate and physically based models:

$$\varepsilon_0(T) = \int_{T_0}^T \Delta\alpha(x, t) dt . \quad (5)$$

And, we take:

$$\varepsilon_0^{xx}(T) = \varepsilon_0^{zz}(T) = \varepsilon_0(T), \quad (6)$$

$$\varepsilon_0^{yy}(T) = -\varepsilon_0(T), \quad (7)$$

$$\varepsilon_0^{xy}(T) = \varepsilon_0^{yz}(T) = \varepsilon_0^{xz}(T) = 0. \quad (8)$$

We assume that the shear strains in (8) are all zero since we are using Intel technology which introduces only uniaxial strains and stresses after deposition of the Si(1-x)Ge(x) thin films. The normal intrinsic strains and

stresses introduced by Intel technology for nano pMOS transistors are compressive (i.e; negative) in x and z directions and tensile (i.e positive) in y direction. These x, y, and z directions are shown in Figure 1. From equation (5), $\varepsilon_0(T)$ is negative. Then, $\varepsilon_0^{xx}(T)$, and $\varepsilon_0^{zz}(T)$ are compressive and $\varepsilon_0^{yy}(T)$ is tensile. The models we are proposing in the equations (2)-(8) to calculate the thermal mismatch induced intrinsic strain in Si(1-x)Ge(x) thin films are original, accurate, in 3D, and physically based. Now, we use the equation (1) and the equations (2)-(8) to calculate accurately all the 3D components of the thermal mismatch induced intrinsic stress $\sigma_0^{tm}(T)$ in the Si(1-x)Ge(x) thin films as follows:

$$\begin{aligned}\sigma_0^{xx}(T) &= (c11 + c12) \cdot \varepsilon_0^{xx}(T) + c12 \cdot \varepsilon_0^{yy}(T) \\ \sigma_0^{yy}(T) &= (c11 + c12) \cdot \varepsilon_0^{xx}(T) + c11 \cdot \varepsilon_0^{yy}(T) \\ \sigma_0^{zz}(T) &= (c11 + c12) \cdot \varepsilon_0^{zz}(T) + c12 \cdot \varepsilon_0^{yy}(T) \\ \sigma_0^{xy}(T) &= \sigma_0^{yz}(T) = \sigma_0^{xz}(T) = 0\end{aligned}\quad (9)$$

Where c11, and c12 are the elastic constants of Si(1-x)Ge(x) films. We have used Vegard's Law (Denton et al., 1991) to calculate them as follows:

$$cij = (1 - x) \cdot cij(Si) + x \cdot cij(Ge).$$

Where cij(Si), and cij(Ge) are the elastic constants of Si and Ge materials at room temperatures and i,j=1,2. The model proposed here in (9) to calculate the thermal induced intrinsic stress $\sigma_0 (= \sigma_0^{tm}(T))$ in Si(1-x)Ge(x) films is also original, accurate, and physically based. It is original since:

Firstly, we are taking into account the effects of both the thermal expansion mismatch $\Delta\alpha(x, T)$, growth temperatures T, and Ge mole fraction x. Secondly, this model can be used to calculate separately the intrinsic stress σ_0 that is due only to thermal mismatch. We should know that the intrinsic stress σ_0 obtained from measurements includes both the effects of thermal mismatch and lattice mismatch and even other source of stress like doping, neighboring, layout, annealing, etching, packaging, capping layers and more. Then, we cannot use the values of σ_0 coming from measurements to investigate and quantify only the effects of the thermal mismatch or only the effects of lattice mismatch. Designers, engineers, and manufacturers would like to know for example which kind of effects are dominant for improving the performances of nano pMOS: thermal mismatch effects or lattice mismatch effects. For some devices, thermal effects are dominant. For other devices, the lattice mismatch effects are dominant. Sometimes the lattice mismatch effects are dominant for a certain range of

growth temperatures and the thermal mismatch effects become dominant for another range of temperature (Alireza et al., 2013). From the model in (9), we can quantify the amount of intrinsic stress σ_0 that results from thermal mismatch only.

As far as we know, no theoretical model has been proposed in the published literature to calculate the initial stress σ_0 in Si(1-x)Ge(x) films. To quantify the amount of intrinsic stress σ_0 originating only from the lattice mismatch, we need to use another model which is not in the scope of this paper. The model in (9) is based on Hooke's law. And, it is accurate since the models we have used to calculate the temperature dependence of the TECs of Si, Ge, and Si(1-x)Ge(x) films are in excellent agreement with measurements. The justification of the model in (5) for thermal mismatch strain is based on the definition of the thermal expansion coefficient. In general, the thermal expansion $\alpha(T)$ is defined as the rate of change of the thermal strain $\varepsilon(T)$ with respect to temperature as follows:

$$\alpha(T) = \frac{d\varepsilon(T)}{dT}. \quad (10)$$

We have applied this formula to the Si(1-x)Ge(x) films when they are free and when they are attached to the substrate to get the model in (5). We should note that the thermal strain defined in a material is different from the thermal mismatch strain which is defined at the interface between two different materials. The model we proposed in (5) is for thermal mismatch strain. Now, we are going to present the models we are using to calculate the thermal mismatch induced extrinsic stress σ which is defined in the whole structure of 14 nm pMOS transistor. The extrinsic stress σ is generated from the intrinsic stress σ_0 that exists in the SiGe films after their growth. To calculate σ , we use the following procedure:

- a) We calculate the extrinsic stress tensor σ from the extrinsic strain tensor ε and σ_0 by using Hooke's law as follows:

$$\sigma = D\varepsilon - \sigma_0 \quad (11)$$

$$\varepsilon = (\varepsilon^{xx}, \varepsilon^{yy}, \varepsilon^{zz}, \varepsilon^{xy}, \varepsilon^{yz}, \varepsilon^{xz})$$

$$\sigma = (\sigma^{xx}, \sigma^{yy}, \sigma^{zz}, \sigma^{xy}, \sigma^{yz}, \sigma^{xz})$$

- b) We calculate the components of ε from the displacements u, v, and w as described in (El Boukili, 2015).
- c) We use Finite Volumes method to calculate numerically the displacements u, v, and w by solving a system of a second order partial

differential equations described in (El Boukili, 2015).

We should note that the intrinsic stress σ_0 in Equation (11) is equal to the thermal mismatch induced stress $\sigma_0^{tm}(T)$ in Si(1-x)Ge(x) films. And, we assume that σ_0 is equal to zero in all the remaining parts of the Si substrate. This assumption is valid since the Si substrate is thick and there will be no intrinsic stress build up in it. From the proposed model in (11), we can study, analyze, and quantify the effects of the thermal expansion mismatch $\Delta\alpha(x, T)$, growth temperatures T, and Ge content x on the extrinsic stress σ . And, this will allow us to study and quantify the effects of the same parameters $\Delta\alpha(x, T)$, T, and x on the speed performances of all the modern nano pMOS transistors. The reason is that the mobility of holes increases with decreasing compressive stress σ according to (JianLi et al., 2014; Kumar et al., 2012; Moroz et al., 2013, 2012, 2006, 2004). And, we know that the speed of nano pMOS will increase if we increase the mobility of holes. The speed of nano pMOS is related to the mobility of holes (see El Boukili, 2015) and the mobility of holes is directly related to the extrinsic stress σ (see JianLi et al., 2014; Kumar et al., 2012). And, the extrinsic stress σ is directly related to the intrinsic stress σ_0 from Equation (11). And, finally the intrinsic stress σ_0 is related to $\Delta\alpha(x, T)$, T, and x as can be seen from the proposed models (2)-(9).

5. SIMULATION RESULTS, VALIDATION, AND ANALYSIS

In this section, we are going to apply, analyze, and validate the proposed models in the Equations (1) to (11) to calculate accurately the thermal mismatch induced extrinsic stress σ in 14 nm nano pMOS transistor developed by Intel in 2014. The gate length (Lg) of this device is 23 nm. Its source and drain are made up of Si(1-x)Ge(x) thin films that are used as stressors (see Figure 1). Their thicknesses are 60 nm. Their lengths are 288.5 nm. The length in x direction of the Si substrate is 300 nm. The thickness of Si substrate in y direction is 190 nm. And, the width of the Si substrate in z direction is 300 nm. The Si substrate orientation is (100). And, the channel orientation is <110>.

5.1. Simulation results and validations

In the following Figures 2, 3, and 4 the growth temperature T is 600°C.

Figure 2 shows the 3D distribution of thermal induced extrinsic stress σ^{xx} when x=20%, $\Delta\alpha(x, T) = -0.7934e-6 / ^\circ\text{C}$.

Figure 3 shows the 3D distribution of the thermal induced extrinsic stress σ^{xx} when x=30%, $\Delta\alpha(x, T) = -1.19e-6 / ^\circ\text{C}$.

Figure 4 shows the 3D distribution of the thermal induced extrinsic stress σ^{xx} when x=40%, $\Delta\alpha(x, T) = -1.586e-6 / ^\circ\text{C}$.

In the following Figures 5, and 6 the Ge content x is 20%.

Figure 5 shows the cross-section distribution of the thermal induced extrinsic stress σ^{xx} when $\Delta\alpha(x, T) = -0.847e-6 / ^\circ\text{C}$ and the fabrication temperature T=700°C.

Figure 6 shows the cross-section distribution of the thermal induced extrinsic stress σ^{xx} when $\Delta\alpha(x, T) = -1.007e-6 / ^\circ\text{C}$ and the fabrication temperature T=1000°C.

The obtained results for σ^{xx} are qualitatively in excellent agreement with the simulation and the experimental results found in literature. From Figure 2, where x=20%, and Si(80%)Ge(20%) film is the only stressor, $\sigma^{xx} = -1.117e+8$ Pa at the center of the channel. And, $\sigma^{xx} = -3.0034e+7$ Pa just below the center of the channel. The simulation results found in (Moroz et al., 2005) for a pMOS structure with Si(80%)Ge(20%) films in source and drain are: $\sigma^{xx} = -10e+8$ Pa at the center of the channel, and, $\sigma^{xx} = -5e+8$ Pa just below the channel. The shape of σ^{xx} in Figure 2 is very close to shape of σ^{xx} in (Moroz et al., 2005). So, our simulation results are qualitatively in good agreement with those of (Moroz et al., 2005). We believe that our values are smaller than those of (Moroz et al., 2005) since in our simulations we are using only one stressor (Si(1-x)Ge(x)) and our σ^{xx} is due to thermal mismatch only. The σ^{xx} in (Moroz et al., 2005) is bigger since it represents the sum of the compressive stresses coming from: thermal mismatch, lattice mismatch, layout effects, and capping layer. Our simulation results are, also, qualitatively in good agreement with the simulation results found in (Alireza et al. 2013; Moroz et al., 2012, 2006). They are also in good agreement with measurements found in (Peng et al., 2006; Heicker et al., 2007).

5.2. Discussion and analysis

From Figures 3, 4, and 5, we can see that the compressive stress σ^{xx} is decreasing with increasing Ge mole fractions and the thermal expansion mismatch $\Delta\alpha(x, T)$ when the fabrication temperature T is 600°C . According to the Figure 7 from (Packan et al., 2008), the hole mobility is increasing with decreasing compressive stress σ^{xx} . And, the speed of nano pMOS transistors is increasing with increasing hole mobility. We could, then, conclude that the hole mobility and the speed of 14 nano pMOS will increase with increasing Ge content x and with increasing thermal expansion $\Delta\alpha(x, T)$. On the other hand, we already know that the hole mobility increases with increasing Ge content x for 35 nm or above technology node of pMOS (Moroz et al., 2006). However, and as far as we know, no investigation has been done about the effects of Ge content x for 14 nm technology node for pMOS.

The most important achievement and the originality of this paper is the calculation of the amount of stress σ^{xx} that is due to solely thermal mismatch effects. From Figure 4 the amount stress σ^{xx} at the center of the channel that is due to the thermal mismatch $\Delta\alpha(x, T) = -1.586e-6/^\circ\text{C}$ for $x=40\%$ and $T=600^\circ\text{C}$ is $-1.8028e+8$ Pa. And, from the Figure 7, we could see the effects of this amount of thermal stress on the hole mobility enhancement. In published simulation results (Moroz et al., 2013, 2012, 2006, 2004; Peng et al., 2006), the calculated values of σ^{xx} involve so many effects: thermal mismatch effects, lattice mismatch effects, layout effects, doping effects, capping layers effects, and more. Therefore, from published results, we cannot quantify the contribution of solely thermal mismatch effects on the hole mobility and speed performance of nano pMOS transistors.

Another originality of this paper is the calculation of the amount of thermal stress σ^{xx} for different thermal expansion mismatch $\Delta\alpha(x, T)$ between Silicon substrate and $\text{Si}(1-x)\text{Ge}(x)$ films and different temperatures T . In practice, when we change x is like we have a different material. So, different values of x give us different materials with different mechanical properties. This analysis is then beneficial for Concentrators Photo Voltaic (CPV) Solar Cells with multi-junctions or even for Silicon based single junction solar cells working in arid regions where the temperature is high like the desert of Morocco or elsewhere. The high operating temperatures (sometimes above 100°C) cause thermal expansions of the materials making up the solar cell which may generate high thermal strains and stresses and which may reduce the efficiency by 50% or more.

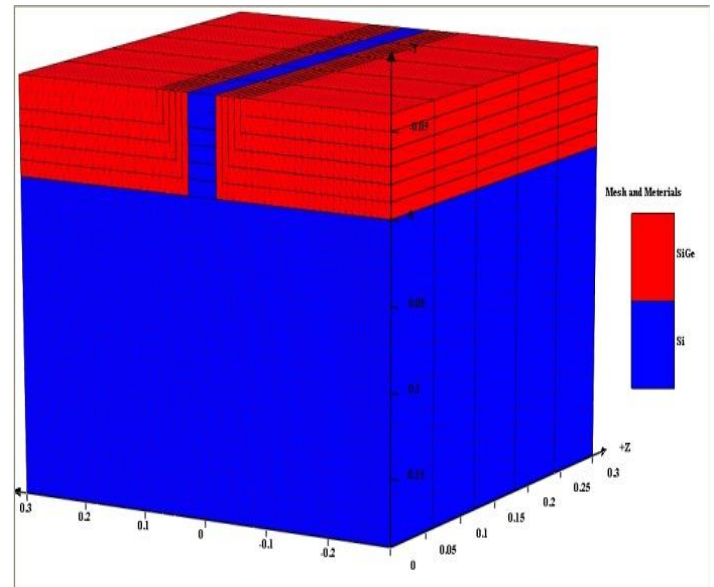


Figure 1: Materials And Mesh Of The Simulated Structure: 14 nm nano pMOS Transistor.

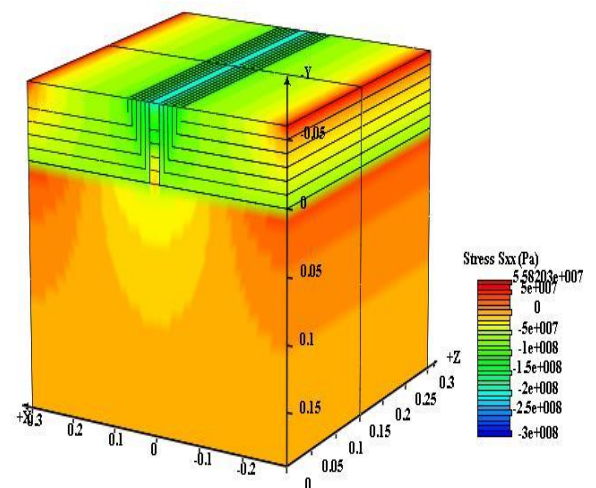


Figure 2: 3D Distribution of σ^{xx} When $x=20\%$, $\Delta\alpha(x, T) = -0.7934e-6/^\circ\text{C}$, And $T=600^\circ\text{C}$.

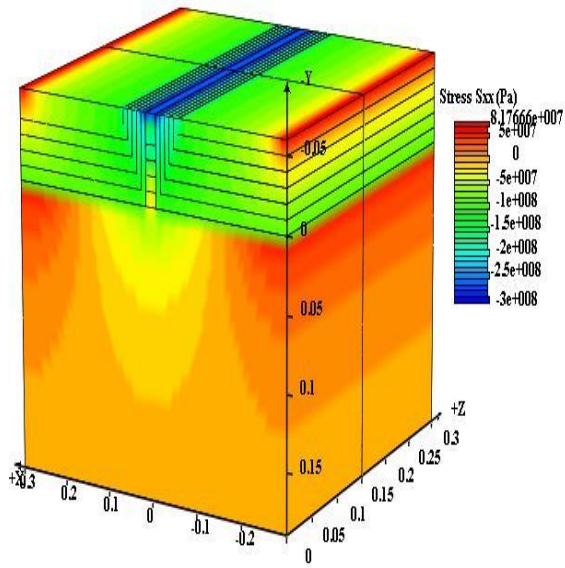


Figure 3: σ^{xx} When $x=30\%$, $\Delta\alpha(x,T) = -1.19e-6$ /°C, And $T=600^\circ\text{C}$.

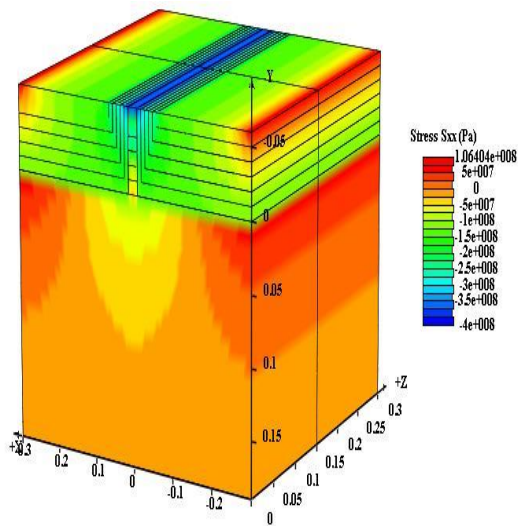


Figure 4: σ^{xx} When $x=40\%$, $\Delta\alpha(x,T) = -1.586e-6$ /°C, And $T=600^\circ\text{C}$.

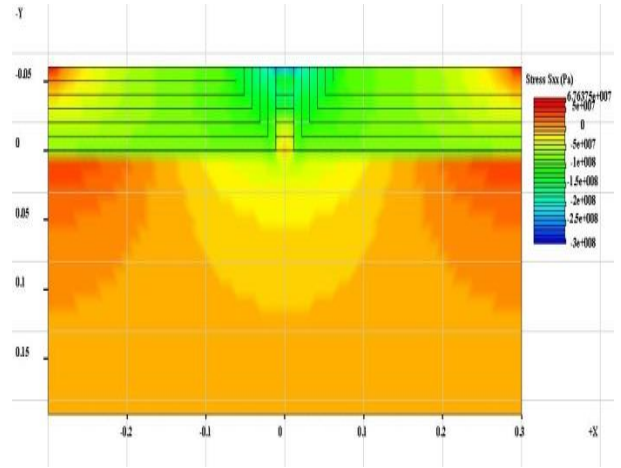


Figure 5: σ^{xx} When $x=20\%$, $\Delta\alpha(x,T) = -0.847e-6$ /°C, And $T=700^\circ\text{C}$.

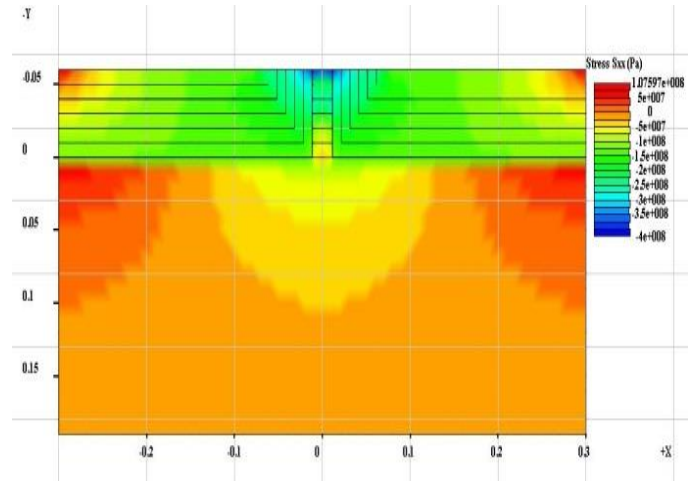


Figure 6: σ^{xx} When $x=20\%$, $\Delta\alpha(x,T) = -1.007e-6$ /°C, And $T=1000^\circ\text{C}$.

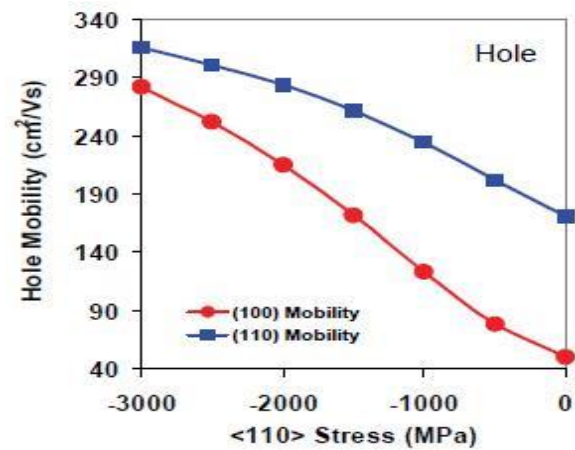


Figure 7: Compressive Stress (σ^{xx}) Enhanced Hole Mobility From (Packan et al., 2008).

6. CONCLUSIONS

In this paper, we have developed, validated, and applied original, accurate and physically based models to simulate and analyze the effects of solely thermal expansion mismatch between Silicon substrate and Si(1-x)Ge(x) films on the speed performance of 14 nm nano pMOS transistors developed by Intel in 2014. Our numerical results in Figures 2, 3, and 4 show that the amount of thermal stress obtained for different thermal expansion mismatch for x=20%, 30%, and 40% is between $-1.8028e+8$ Pa and $-1.117e+8$ Pa. This is a great finding for engineers, designers, and manufacturers. Since it shows the effects of thermal mismatch induced stress on the speed performance of 14 nm nano pMOS transistors. In the future, we will investigate the effects of both thermal mismatch and lattice mismatch of the hole mobility and the speed performances of nano pMOS devices.

REFERENCES

- Alireza, M., Haihui, R., Zhang, L., Mei, L.,(2013). Residual stresses in thin film systems: effects of lattice mismatch, thermal mismatch and interface dislocations. *Int. Jour. Of Solids*, Vol. 50, 3562-3569.
- Bera, L.K., 2006. The impact of uniform strain applied via bonding onto plastic substrate on MOSFET performance. *IEEE Electron Device Lett.*, Vol. 27, 58-61.
- Chui, K.J., 2007. n-MOSFET with silicon-carbon source/drain for enhancement of carrier transport. *IEEE Trans. Electron Devices*, Vol. 54, 249-256.
- Denton, A. R.; Ashcroft, N. W. (1991). Vegard's law. *Phys. Rev.*, A 43 (6), 3161-3164.
- El Boukili, A., 2010. 3D Stress simulations of nano transistors. In: Gunther, M., Bartel, A., Brunk, M., Schops, S., Striebel, M., eds. *Progress in Industrial Mathematics at ECMI*. Berlin Heidelberg: Springer, 85-91.
- El Boukili, A., 2015. New models for thermal stress simulation in 3D nano CMOS transistors. *Proceedings of the first international conference on electrical and information technologies*. pp. 25-27, March, 25-27, Marrakech (Morocco).
- Ghani, T. et al., 2003. A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors. *Proceedings of IEDM Technical Digest*, pp.978-980. Dec. 8-10, Washington DC (USA).
- Ghani, T., 2010. Challenges and innovations in nano-CMOS transistor scaling. *IEEE workshop on microelectronics and electron device*. April 16, Boise (Idaho USA).
- Hecker, M., Geisler, H., (2007). Strain state in silicon structures for microprocessors technology. *Materials Science-Poland*, Vol. 25, No. 1, 7-18.
- JianLi, M., Fu, Z., Liu, P., Zhang, H. (2014). Hole mobility enhancement in uniaxial stressed Ge dependence on stress and transport direction. *Science China, Phys., Mech. & Astr.*, Vol. 57, 1860-1865.
- Kumar, A., Xiu, K., Haensch, W., Robinson, R., Bajaj, M., Johnson, J., Furkay, S., Williams, R. (2012). A simple, unified 3D stress model for device design in stress-enhanced mobility technologies. *Proceedings of SISPAD*, pp. 3000-3003, Sept. 5-7, Denver (CO, USA).
- Moroz, V., 2004. Analyzing strained-silicon options for stress-engineering transistors. *Solid State Technology Magazine*. July, Vol. 47, issue 7, p49.
- Moroz, V., Eneman, G., Nouri, F., Smith, L., (2005). The impact of layout on stress-enhanced transistor performance. *Proceedings of Int. Conf. on Simulation of Semiconductor Processes and Devices*, pp. 143-146. Sept. 1-3, Tokyo (Japan).
- Moroz, V., Washington, L, Farah, N., Sunderraj, T., Eneman, G., Smith, L., Screntelkamp, R., (2006). pMOS with 200% mobility enhancement induced by multiple stressors. *IEEE Electron Dveice Letter*, Vol. 27, No. 6, 511-513
- Moroz, V., et al., 2012. 14nm FinFET stress engineering with epitaxial SiGe source/drain. *Proceeding of ISTDM*, pp. 1-2. June 4-6, Berkeley (CA, USA).
- Moroz, V., et al., 2013. A group IV solution for 7nm FinFET CMOS: Stress engineering using Si,Ge, and Sn. *Proceeding of IEDM*, pp. 26.3.1-26.3.4 Dec. 9-11, Washington DC (USA).
- Okaji, M., Yamada, N., Watanabe, H., (2004). Linear thermal expansion of Silicon from 293 to 1000 K. *International Journal of Thermodynamics*, Vol. 25, No 1, 221-236.
- Okada, Y., Tokumaro, Y., Watanabe, H., (1984). Precise determination of lattice parameter and thermal expansion coefficient of Silicon between 300 and 1500°K. *Journal of Applied Phys.*, Vol. 56, No 2, 314-320.
- Packan, P., Cea, S., Desphande, H., Ghani, T., Giles, M., Panale, P., Zawadzki, K., (2008). High performance high-K+ metal gate strain enhanced transistors on (110) silicon. *Proceedings of IEDM*, pp. 1-4, Dec. 15-17, San Francisco (CA,USA).
- Peng, Z., Andrei, A., Haifeng, H., Joel, W., Chris, N., Eric, S., (2006). Analysis of nano-scale stress in strained silicon materials and microelectronics devices by energy-filtered convergent beam electron diffraction. *ECS Transactions*, 2(2), 559-568.

- Singh, H., (1968). Determination of thermal expansion of Germanium, Rhodium, and Iridium by X-rays. *Acta Cryst.*, Vol. A 24, No 2, 469-471.
- Yin, Z., Zhu, J., He, Q., Cao, X., Tan, C., Chen, H., (2006). Graphene-based materials for solar cell applications. *Advance Energy Materials*, Vol. 4, No. 1, 1-19.

SIMULATION SYSTEM OF MINING WORKS

Aleksey N. Starodubov^(a), Vasily V. Sinoviev^(b), Victor V. Okolnishnikov^(c)

^{(a),(b)} Federal State Budget Scientific Centre «The Federal Research Center of Coal and Coal-Chemistry of Siberian Branch of the Russian Academy of Sciences», Federal State Budget Educational Institution of Higher Professional Education Kuzbass State Technical University named after T.F. Gorbachev

Kemerovo, Russia

^(c) Design Technological Institute of Digital Techniques of Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia

^(a)a.n.starodubov@gmail.com, ^(b)zv150671@gmail.com, ^(c)okoln@mail.ru

ABSTRACT

Information system for simulating of tunneling works, which was engineered by the Federal Research Center of Coal and Coal-Chemistry of Siberian Branch of the Russian Academy of Sciences (FRC CCC SB RAS), is shown in the article. The system is a supporting tool for making decisions in tunneling works projection and efficiency evaluation. After selection of geological factors, user gets list of appropriate equipment. The technology will be formed with using of this equipment. System chic is automatic creation of simulating models of tunneling works from simulating modules of technology elements, which were developed with use of special language GPSS World. Moreover, there were automatizations of processes of models start, getting results of simulating, creation of report-file with technical and economical characteristics and tunneling works schedule. Using of the system in part of carrying experiments allows finding ways for increasing of efficiency detect bottleneck and find solutions how to eliminate bottleneck, make evaluation of efficiency of mine technical systems and find ways of efficiency increasing.

Keywords: simulation, mine technical systems, CAD, decision support system

1. INTRODUCTION

Process of complex familiarization and learning of coal fields contains the next actual headings:

- Innovate technologies of familiarization and learning of coal fields systems analysis and substantiation;
- Design and development of software tools, methods and approaches for modeling of innovative technologies and complex familiarization development;
- Developing of methods and algorithms of prediction, evaluation of trustworthiness and efficiency mine-engineering system.

Accumulated experience shows that efficient tool for decision supporting for planning new high-efficient different industrial areas is simulating which displays system dynamics in PC by algorithm, which simulates system's behavior (Konyukh V. 2010; Oparin V.N. 2007; Andrei A. and Dinescu S. 2013; Hubalovsky S. 2014; Guizzi G., Murino T. and Romano E. 2009; Kleinrock L. 1979; Starodubov A.N. 2013; Sinoviev V.V., Starodubov A.N., Majorov A.E. and Kochetkov V.N. 2013; Poletayev V.A., Sinoviev V.V., Starodubov A.N. and Chicherin I.V. 2011).

2. MATERIALS

There are innovate system creation projects in Institute of Coal FRC CCC SB RAS for tunneling works simulation. Modern computer simulation language GPSS World was chosen as software implementation models tool for complex industrial systems simulating (Sinoviev V.V. 2013; Lou A.M. and Kelton A.D. 2004; Devytkov V.V. 2013; Okolnishnikov V., Rudometov S., and Zhuravlev S. 2010; Starodubov A.N., Sinoviev V.V. and Dorofeev M.U. 2010; Sinoviev V.V., Okolnishnikov V.V., Starodubov A.N. and Dorofeev M.U. 2015). The system designed on base of machine code C#4 in Microsoft Visual Studio 2012. Projects and templates is in "xml" format. Simulation model storage is in "gps" format. Information system allows analyzing the next list of tunneling works design elements:

- Equipment choice and selection of technological scheme of works;
- Simulation model building and simulating experiment leading;
- Statistical data accumulating;
- Visual display of tunnel works dynamics developing;
- Organizational works schedule;
- Technical and economic indicators calculation.

Figure 1 presents a scheme of options for using of informative tunnel works simulation model elements.

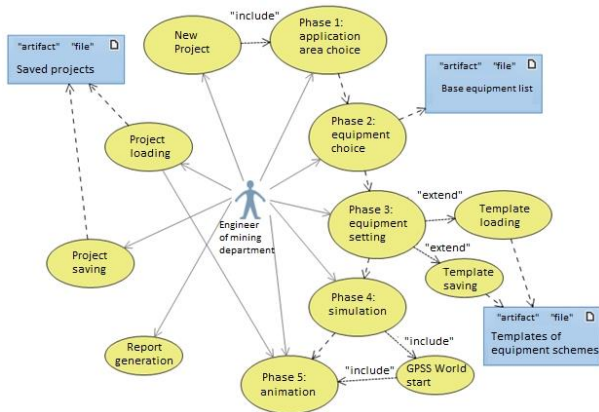


Figure 1: Diagram of options for using of informative system elements

User creates new project and choose geological factors. The next step is automatic formation of appropriate mining machines from database. User creates set of equipment from this list. User is able to correct characteristics if necessary. There is an opportunity of choice of prepared template and user is able to create new template and save it. Generation models module takes signal by button push for launching. This module forms a mine-works variant from library. The imitation module launches model and shows the result of simulation experiment. Technology animation and report files are formed automatically. There are main technical and economic indicators and tunneling works planogram in report files. User is able to save project of developed technology for modification and using in near future.

Simulation modeling system functioning is showed by example of underground drilling scheme with using of cutter-loader with attached implements and belt loader (Figure 2).



Figure 2: Technological scheme of underground drilling with using of cutter-loader (with attached implements and belt loader)

There are these works inside mining face: cross drift, engineering complex moving, belt loader enragement, rail way enlargement, pump tree enlargement, belt loader and re-loader technical service, rig two-way delivering, sinking, water pressing-in, rig work, barring, pipes and wires moving, materials delivering.

At the first stage of working with simulation system you should enter the next items: size in the clear, mine working angle of slope, amount of inclination, toughness coefficient, vertical conformance, thickness of a steam (Figure 3).

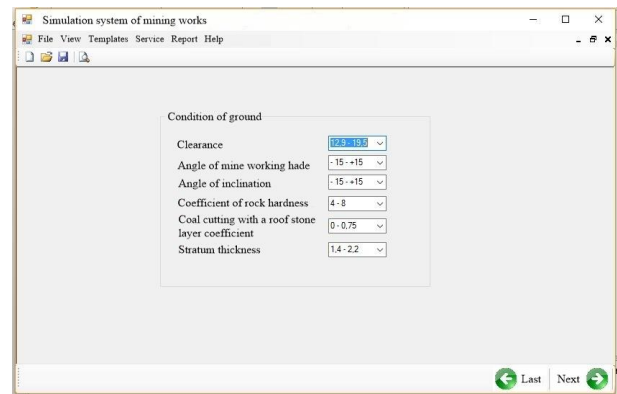


Figure 3: Geological factors selection

Transition to the next stage is due to button "NEXT". User will select equipment from list "Initial equipment" and will add it in list "Selected equipment" (Figure 4).

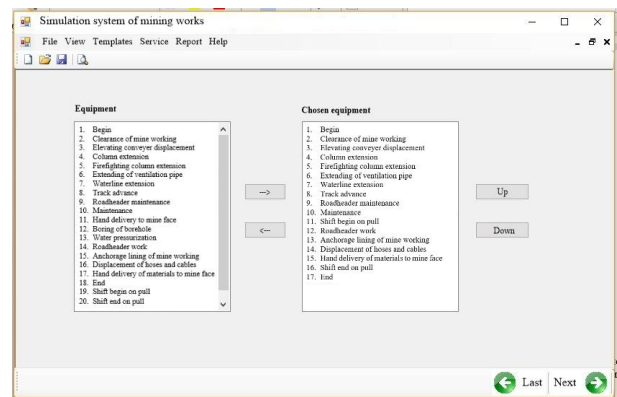


Figure 4: Equipment selection

On the next stage user will select necessary works, which will be done due to special equipment. User has to enter random duration for each work and number of workers (Figure 5).

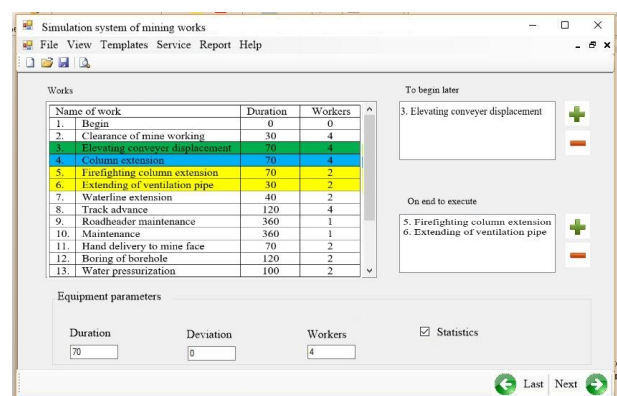


Figure 5: Works selection

User launches generation of model by prepared options, defines options, and makes simulation start (Figure 6).

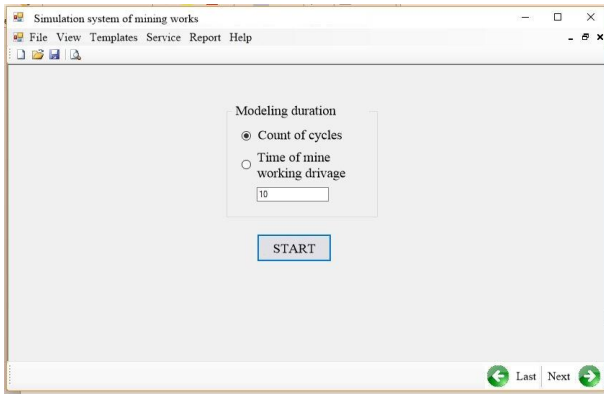


Figure 6: Simulation parameters setting

GPSS World model of developed scheme was based on the next approach. It is defined that requests for execution of works are equal with simulation model transactions. Block GENATATE was created for input to the model. Instruments in the model are works and mining equipment, which described by raw of blocks SEIZE-RELEASE and ENTER-LEAVE (for displaying of both work equipment). To model the duration of operations was used ADVANCE blocks, which delay transacts at time of execution of technological operations. TERMITATE block is created for deleting of transacts after end of the modeling. According to technological schemes, functioning of various equipment can be carried out at the same time. SPLIT block was created for separation. Block ASSIMBLE is for synchronizing of concurrent tasks. LOOP blocks are for simulation of technological operations (to the beginning of the cycle) and ASSIGN block (to specify the number of repetitions, which are stored in the parameter transact).

Model consists of the following components:

- Initialization of the model
- Performed works
- Finalization of the model
- Collecting of statistical data

During initialization of the model in GENERATE block occurs the appearance of the one transact. This transact will exist during the all simulation time and simulation will be completed after destruction of transact. Transact parameter is written in the number of full cycles (days) of simulation in the ASSIGN block. Next is a pointer to the beginning of the change of the care and synchronization transact in block ASSEMBLE. ASSIGN command is for input the number of care cycles. SPLIT block simulates beginning of parallel works. TRNSFER command moves in block, which simulates the first technological operation beginning (Figure 7).

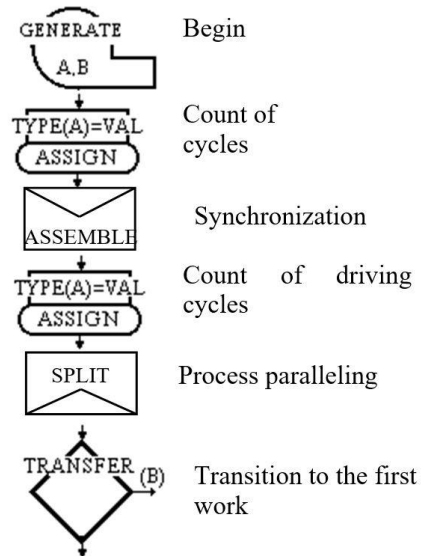


Figure 7: Model initialization flow chart

Model initialization code will be:

	GENERATE	0,,,1
	ASSIGN	2,1
START1	ASSEMBLE	1
	ASSIGN	3,9
	SPLIT	1,WORK2
	TRANSFER	,WORK11

There is inspection of task, transact equality to zero (responsible for the number of simulated cycles) after all works for finalization of the model. If the parameter is zero then there is a moving to the next block TERMINATE, which removes transact and ends simulation. If the parameter is not equal to zero then number reduced by one and there is moving to beginning of the model.

Figure 8 shows the scheme of model finalization.

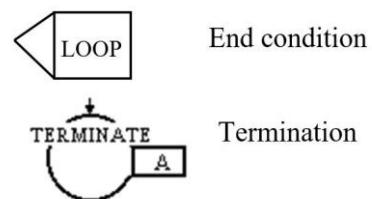


Figure 8: Scheme of model finalization

Model finalization code will be:

	LOOP	2,START1
	TERMINATE	1

Construction of work models in this way allows you to generate technological model entirely. After START button system receives process and operations GPSS World-codes from data base, which were set up by user on the works selection stage (pic 5), in addition there are reading of models parameters (temporary delays, deviation, necessity of statistical data for this operation) and formation of model of technology in appliance with indicated order of executed operations. System uses this

data to arrange size of operand in blocks SPLIT and ASSEMBLE. It is dividing of operations or ban of synchronous operations execution. Model is supplemented by initialization and finalization blocks, which have been mentioned above. The next step is moving of generated model for start in simulation environment.

Animation of scraper-works (Figure 9) and report documents with experimental results in MS Excel format will be formed after simulating ending. There are two tables with report data: technical and economic parameters for the selected scheme and schedule of works. Technical and economical parameters (Figure 10) consist of information about speed of elaboration leading with chosen technology on different time periods. For example, the number of passed meters for one day-shift, day or month. In addition, there are data in tables for mine works planning: working days number, the number and duration of shifts in day, cycles in shift and the required number of service staff.

On the schedule of works graph (Figure 11) you can see order and time of sinking cycle execution, order and combination of processes in mining face.

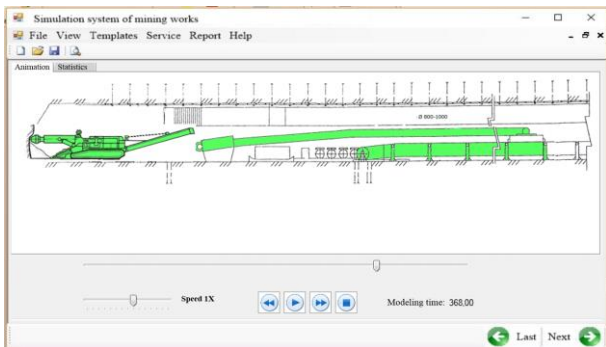


Figure 9: Tunnel works animation

Animation allows switch frames forward and backward stepwise, change positions of shown elements of control, animation speed, monitoring of the current time of technological operations. Functioning equipment is green, equipment, which is in service - yellow.

There are principal technical and economical characteristics of technology and tunnel works planogram in report file.

Technical and economical characteristics		
No	Characteristics	
1	Speed of mine working drive	
	m/month	540
	m/day	18,00
	m/shift	6,00
2	Face advance on cycle, m	2,00
3	Count of cycles	
	in shift	3,00
	in days	9
4	Count of working days in month	30
5	Count of shifts in day	4
6	Shift duration, h	6
7	Cycle duration, min.	120
8	Workers count	
	in shift	4
	on pull	8

Figure 10: Technical and economical characteristics for chosen scheme

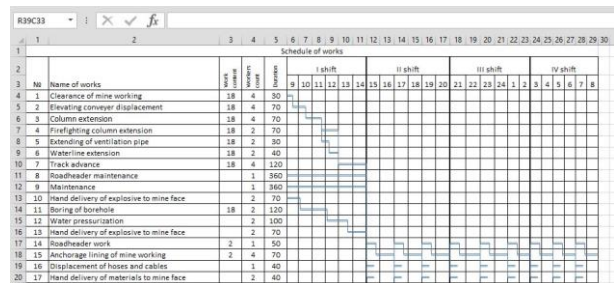


Figure 11: Schedule of works

The proposed system allows through leading of simulation experiments to define ways of increasing of productivity, to identify bottleneck and find solutions how to eliminate bottleneck, to assess the effectiveness of mine technical systems and ways of its increase. For example, experiments were conducted to identify the effectiveness of proposals for robotics of combine technology. Row of experiments on models of the original technology and modified technology after applying of robotics is performed. Histogram showing the influence of tram on degree of use cutter-loader in basic and robotic technologies is shown on Figure 12. Determined that the degree of use of cutter-loader in robotic mining technology increases to 22-50% (depends of tram capacity) compared to the conventional technology.

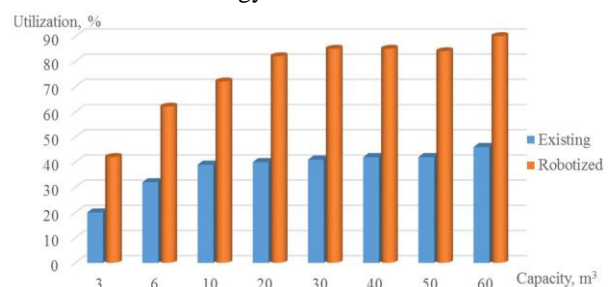


Figure 12: Impact of transport capacity on the combine utilization

3. RESULTS

As result, there are information system, which allows automatically to select equipment for given geological factors, to build simulation models of scraper works leading, to show dynamic of scraper woks leading on PC monitor, to collect statistical data about main technical and economical parameters, to build schedule works graph.

User enters geological factors into the information system. System forms list of the equipment. Simulated models will be created automatically from standard ready-made GPSS World modules, which shows the most important technological processes. On the next stage system launches models, obtains results and shows statistical data about simulating results in special report-file. Finally, a graph of scraper works schedule and the main technical and economic indicators will be shown on PC screen.

Use of the system will speed up and improve the process of designing a solid mining technology, and eliminate costly technological risks by automating the calculation of technical and institutional options of conducting mining operations. Moreover, this system can serve as a tool to support decision-making in determining ways to improve efficiency, identify and eliminate bottleneck in the technology and assess the effectiveness of mine technical system. Coal mine is a complicated technological object, so the establishment of such information systems is an important step in the development and transfer of the coal industry to the higher technological level.

ACKNOWLEDGMENTS

This research was financially supported by the Russian Foundation for Basic Research (project 16-47-420490 p_a).

REFERENCES

- Andrei A. and Dinescu S., 2013. Modeling and Simulation of the Complex Mining Production Systems for Coal Extraction. Recent Advances in Civil and Mining Engineering, WSEAS Press, pp. 248–255. Antalya, Turkey,
- Devytkov V.V., 2013. The methodology and technology of simulation studies of complex systems: current status and prospects of development. Monograph, Moscow, Russian Federation (in Russian).
- Guizzi G., Murino T., and Romano E., 2009. A Discrete Event Simulation to model Passenger Flow in the Airport Terminal, Mathematical methods and applied computing, WSEAS Press, pp. 427–434. Vouliagmeni, Athens, Greece.
- Hubalovsky S., 2014. Modeling and Computer Simulation of Static, Dynamic and Feedback Systems as Tool of Development of Logical Thinking. International Journal of Mathematics and Computers in Simulation, 8: 276–285.
- Kleinrock L., 1979. Queueing theory. Moscow, Russian Federation (in Russian).
- Konyukh V., 2010. Simulation of mining in the future. Proceedings of the IASTED International Conference on Automation, Control, and Information Technology - Control, Diagnostics, and Automation, ACIT-CDA, pp. 1–6. June 15-18, Novosibirsk, Russian Federation.
- Lou A.M. and Kelton A.D. 2004. Simulation modeling and analysis. Sankt-Petersburg, Russian Federation (in Russian).
- Okolnishnikov V., Rudometov S., and Zhuravlev S., 2010. Simulation environment for industrial and transportation systems. Proceedings of the International Conference on Modelling and Simulation, pp. 161–165. Prague, Czech Republic.
- Oparin V.N., 2007. World experience in automation of mining operations at the mines. Publishing house Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation (in Russian).
- Poletayev V.A., Sinoviev V.V., Starodubov A.N. and Chicherin I.V., 2011. Design of the computer integrated production systems. Monograph, Kemerovo, Russian Federation (in Russian).
- Sinoviev V.V., 2013. Modeling multi face gallery tunneling using a simulation approach. Mining Informational and Analytical Bulletin (Scientific and Technical Journal), Moscow, Russian Federation, SV6: 138–144 (in Russian).
- Sinoviev V.V., Okolnishnikov V.V., Starodubov A.N. and Dorofeev M.U., 2015. The Experience of Discrete Event Simulation Robotic Technology of Mining. Proceedings of the 19th International Conference on Systems (part of CSCC '15) Recent advances in systems, pp. 499–502. Zakynthos Island, Greece.
- Sinoviev V.V., Okolnishnikov V.V., Starodubov A.N. and Dorofeev M.U., 2015. Discrete Event Simulation Robotic Technology of Mining. Proceedings of the 2th International Conference on Mathematics and Computers in Sciences and Industry (part of MCSI 15), pp. 17–19. August 17-19, Sliema, Malta.
- Sinoviev V.V., Starodubov A.N., Majorov A.E. and Kochetkov V.N., 2013. Simulation approach in modeling complex in Energy coal processing. Monthly production and mass magazine Energetik, Moscow, Russian Federation, 1: 26–29 (in Russian).
- Starodubov A.N., 2013. The generalized structure of model of a power technological complex in the form of systems of mass service. Separate release of the Mining Informational and Analytical Bulletin (Scientific and Technical Journal), Moscow, Russian Federation, OB6: 145–151 (in Russian).
- Starodubov A.N., Sinoviev V.V. and Dorofeev M.U., 2010. Simulation of the technological complex for deep coal processing. Ugol, Moscow, Russian Federation, 2: 8–12 (in Russian).

INTELLIGENT SIMULATION FOR TOURISM AND CULTURAL HERITAGE MANAGEMENT

Galina Merkurjeva^(a), Vitaly Bolshakov^(b), João Martins^(c), Ricardo Gonçalves^(d), Gumersindo Bueno Benito^(e)

^{(a)(b)}Riga Technical University, Kalku Street 1, LV-1658, Riga, Latvia

^{(c)(d)}Universidade Nova de Lisboa, 2829-516 Monte de Caparica, Portugal

^(e)Fundación Santa María la Real, Monasterio de Santa María la Real, 34800, Aguilar de Campoo, Palencia – España

^(a)galina.merkurjeva@rtu.lv, ^(b)vitalijs.bolsakovs@rtu.lv, ^(c)jf.martins@fct.unl.pt, ^(d)rg@uninova.pt,
^(e)gbueno@santamarialareal.org

ABSTRACT

The paper addresses an intelligent simulation approach to tourism and cultural heritage management. Emergent information technologies, modelling approaches and simulation paradigms provide new capabilities for intelligent and interoperable management of tourism and cultural heritage. Nowadays mobile applications for tourism organisation at heritage sites have become quite popular. Smart mobile applications with embedded simulation capabilities can significantly increase intelligence of managerial decisions and improve the quality of customer services. The concept of a mobile application that for a particular tourist place suggests objects to be visited preferably at the moment is introduced in this paper. A simulation-based case study for intelligent management of pedestrian tourist flows at heritage sites is given. An agent-based modelling paradigm is applied to simulate tourists' behaviour and analyse operational solutions. Post-experimental analysis to explore behaviour of the developed simulation model is performed.

Keywords: intelligent and interoperable management, agent-based simulation, pedestrian flows, smart mobile application, heritage site

1. INTRODUCTION

All over Europe, cultural heritage, being part of the human history, traditions and creativity and having a social function while creating economic value, attracts millions of tourists. This sometimes creates difficulties for visitors to have easily access to a particular heritage site. Also, overcrowding may lead to the destruction and degradation of heritage buildings (HBs) (Lyu et al. 2016).

The overall idea of the intelligent tourism organisation to provide both more comfortable environment for visitors and less stress on the heritage is rather topical (Xiaowen et al. 2012). As computing power is growing and new IT technologies appear, many visitors can use mobile applications to gain more information from the local organisations. As the number of smart mobile devices in use is continuously growing, it would be

reasonable to provide visitors with intelligent applications that would guide them and suggest tourist objects or attractions to be visited preferably at the moment (Marie-Theres 2016).

The development of such application for a particular tourist place could require advanced forecasting and planning in order to consider all local peculiar properties and manage crowd. Nevertheless, many aspects of local spatial information and pedestrian crowd behaviour could be assessed in a simulation process (Lyu et al. 2016). Such mobile applications would not be useful for pedestrian tourists in case of large, extensive geographical areas. But they would be highly effective for heritage sites, which can be investigated within walking distances, such as small historic city centres and medieval monasteries.

This paper addresses some issues and application of an intelligent and interoperable approach to tourism logistics and cultural heritage management. The emergent information technologies, modelling approaches and simulation paradigms that enable enhancing the effectiveness of decisions in managing tourism and cultural heritage are discussed. A simulation-based case study in intelligent management of pedestrian tourist flows in a historical city centre is described. It is aimed to analyse pedestrian tourist flows and evaluate possible outcomes of application of smart technologies to control them. An agent-based modelling paradigm is employed for simulating tourists' behaviour. A structure of a simulation model and its layout are presented, and behaviour of intelligent agents is detailed. Post-experimental analysis to explore the model behaviour is performed.

2. APPROACH AND MODELLING PARADIGM

Information exchange and knowledge sharing is a prerequisite for collaborative good decision making about HBs' protection, rehabilitation and operation processes. This requires an intelligent and interoperable approach to collaborative management produced with process interoperability when human beings share a common understanding across a network, business

systems interoperate, and work processes are coordinated (Sarraiya et al. 2010).

Advanced and emergent information technologies create new capabilities for intelligent and interoperable management of cultural heritage. Particularly, they provide methods and tools for heritage data sensing, processing and management; imaging and visualization of heritage; modelling and simulation (such as 3D modelling, crowd simulation, agent-based simulation); socio-economic modelling of the environment; climate impact and risk assessment for cultural heritage; modelling for interoperability of HBs and mobile application services.

Nowadays, specific mobile applications for heritage sites are developed to amend or supplement tour guides and provide application also reducing tourist crowds users with additional information about them (Rolando et al. 2013). Examples of such heritage sites which extensively address mobile device users are Barcelona and Malta (Mobile Apps 2016, Malta Tourist Mobile Apps 2016). They provide a wide range of specific and free to use mobile applications for different types of tourists' purposes and could be grouped as follows.

The first group allows interactive orientation in a new unknown location and provide geographical and navigational information for the attractions to visit. The second one mainly focuses on providing additional textual, audio and other media information within a covered area. The third group covers applications, which include both navigational and descriptive options. The present work is focused on applications with a navigation opportunity.

However, most of such applications do not include enough information on a current tourist activity within the covered area, which could be sensitive to mass tourism destinations and the most popular tourist attractions. Mobile applications which will incorporate such information would be useful here (Li and Wang 2011). These applications could allow navigating the tourists to less-crowded sites. Also, more advanced information, such as visitor's preferences, real distance, times to reach the site and optimal routes could be considered.

To make these mobile applications more intelligent by providing suggestion on the next object to visit and taking into account various factors, different decision making techniques may be used. Usually, the evaluation of the alternatives is based on the distance and value of heritage or touristic attractions. Complementing it with the information on the current state of the object will allow the holder of such application visiting less overcrowded attractions, while avoiding visiting the objects, for which currently it is necessary to wait a long time in a queue. Due to complexity of real-life experiments, it is assumed to evaluate the alternatives by means of simulation experiments.

In practice, modelling and simulation are used at two levels of a system description (Merkuryev et al. 2014). Macroscopic modelling processes geographic objects and regions in an aggregated form based on the black-

box principle where each region is specified by prescribed average statistical indicators, such as land and heritage building use, status of objects, population, transport nets and services. The small amount of the necessary data and a small number of computational resources make macroscopic modelling one of the most widely used approaches in spatial planning.

Microscopic modelling provides the depiction of the real system in a very detailed manner and execution of simulation activities with homogeneous spatial units. Microscopic models are gaining ever broader application area in recent years due to the rapid development of information technologies and increase in computing power resulted in higher speed and accuracy of simulation processes. So, the microscopic simulation of pedestrian flows (Liu et al. 2014) has attracted increasing research attention since a reliable simulation model for pedestrian flow may greatly benefit in transportation management, as well as in urban planning and architecture. Microscopic models may have stochastic behaviour.

With regard to the use of modelling paradigms, so called System Dynamics paradigm (System Dynamics Society 2016) enables analysis of complex dynamic systems, such as population dynamics, ecological systems and economical processes. It assumes a high level of aggregation of system's objects and mainly is used for long-term strategic decisions in tourism and heritage management. Nevertheless it can be used also for short and mid-term models, where structure and dynamics of a system should be investigated. System dynamic simulation allows easier interpretation of systems structure and internal relationships. Examples of efficient tools that provide system dynamic simulation are Vensim DSS and AnyLogic simulation software. The last one allows combining system dynamics with other powerful simulation paradigms such as agent-based modelling and discrete event simulation, which allows expanding the model to higher detailing of internal processes.

The agent-based modelling provides an intelligent approach and advanced methodology for building microscopic models. It becomes the dominant approach in social simulation as worldwide social systems are complex, highly decentralised and composed of a multitude of heterogeneous objects called by agents. These agents act with some purpose and their interaction, usually through time and space, generates emergent behaviour, often at higher levels than those at which such agents operate (Crooks et al. 2008).

In traffic simulation, such agent-based models allow considering various aspects of pedestrians behaviour. Here, pedestrians are modelled as independent agents, which interact among themselves and within a defined spatial environment. This allows considering individual characteristics of pedestrians, revealing weak elements of pedestrian infrastructure, and providing highly detailed model and visualisations (Camillen et al. 2009). The case study is built on an agent-based simulation model applied to simulate tourists' behaviour and

support decision making for individual tourists on the attraction to be preferably visited next, as well as to assess possible outcomes of application of smart mobile applications to control tourist flows and crowds.

3. CASE STUDY

The cultural heritage site represents a central part of the original and historical area of the capital of Latvia, which is known as the Old Town of Riga (often called also “old city”). It is the most popular tourism area for pedestrian tourists including 5 open-air sites or tourist attractions and 3 indoor sites. The open-air sites represent either large squares with a number of heritage objects or vicinities of historic buildings with crowds of people gathered around buildings. The indoor attractions include two museums and the Riga Cathedral hall. It is assumed, that indoor attractions have limited capacity. The heritage has three entrance points, from which tourist agents come into a heritage site with a constant rate. It is assumed, that tourists would like visiting all attractions within a certain period of time. Also, incoming tourists are individual, not organised in groups, and do not have any predetermined sequence of visited attractions.

The heritage site includes separate buildings or building quarters separated by streets. Some of these streets are pedestrian and can be passed by agents at any place. But other streets that have motor car traffic can be crossed only at specific places. Outdoor attractions are located in the pedestrian street areas, or around some specific objects. The outer walls and fences of such building quarters create insurmountable obstacles for moving of tourists in the environment.

There may be two types of tourists visiting the heritage: (1) ordinary tourists without mobile applications for a tour guide; these tourists follow the closest attractions, however their behaviour can be unexpected or random; and (2) advanced tourists equipped with mobile applications, which try to adhere to the attractions proposed by mobile guides. Relative numbers of ordinary and advanced tourists may vary. Usually, all tourists visit each attraction only once.

The aim of the study is to analyse if introduction of applications would avoid tourist crowds in the old city and significantly reduce waiting times at the most popular attractions.

4. AGENT BEHAVIOUR MODELLING

In the model, tourists are simulated as intelligent agents having their own decision-making capabilities. Each agent has its speed and width, which are not deterministic and defined by uniform distributions within ranges [0.4; 0.8] and [0.4; 0.6] m/s, correspondingly. The dynamic behaviour of tourist agents is defined as follows (Bolshakov and Merkuryeva 2016). Each agent created in the model performs a sequence of actions:

1. Selects one of non-visited attractions. For each attraction the probability of being chosen is defined in section below. For ordinary tourist

agents and ones equipped with applications, the logic of the action is different.

2. Moves to the chosen attraction in the defined environment, which has walls and non-accessible areas.
3. Arrives at the attraction and stays there for a short period of time if the attraction is of an open-air type. The time period is randomly defined within the time ranges that vary between different attractions.
4. Arrives at the attraction. If the attraction is of an indoor type, first waits in a queue in case its current capacity is overloaded. When leaves the queue, stays at the attraction for a longer period of time. All time periods are randomly defined similar to ones in the previous action.
5. Leaves the heritage environment if all the attractions visited or being tired after certain time has passed.
6. Otherwise, selects the next attraction (step 1).

For each attraction the probability of being chosen is defined by the formula:

$$p_i = \frac{w_i}{\sum_{j=1}^n w_j},$$

where p_i is a probability of attraction i to be selected, w_i is attraction weight for a particular agent, and n is the total number of attractions.

For ordinary agents, the closest attraction will have higher probability of being selected, so weight w_i is calculated as follows:

$$w_i = \begin{cases} 0, & \text{visited} \\ \frac{1}{d_i}, & \text{otherwise} \end{cases}$$

where, d_i is a distance from the current agent's location to the i -th attraction.

In case of advanced agents equipped with smart mobile applications, weights are calculated considering the current load of attractions, i.e.

$$w_i = \begin{cases} 0, & \text{visited} \\ \frac{1}{d_i} \left(\max\left(\frac{s_i - l_i}{s_i}, 0\right) \right), & \text{otherwise} \end{cases}$$

where s_i defines the maximal capacity of i -th attraction, and l_i is the current number of agents at i -th attraction.

Finally, tiredness of the agent is modelled by a state chart and timeout transition which is randomly defined within the range from 1 to 4 hours.

5. SIMULATION MODEL

The layout of the agent environment (Figure 1) is built based on the city map developed by OpenStreetMap

contributors (OpenStreetMap 2016). Attractions' areas are bordered by green dotted lines. Building quarters and non-accessible areas are surrounded by walls shown as orange lines. Queues of tourists waiting at the entrance of indoor attractions are marked by dark blue lines. Agents have three entry points into the environment, i.e. on the left, right and bottom edges of the model screenshot, shown with green solid lines. The entry rates are 2500, 5000, and 3000 agents per hour correspondingly. The capacity of indoor attractions is limited by 200 visitors.



Figure 1: Layout of the Heritage Site

Figure 2 presents a screenshot of a fragment of the model visualisation in more detail. The agents are marked by red, black or blue dots. Red and black dots correspond to ordinary tourists without an application and tourists with it, while tired tourists, which exit the environment, are marked by blue dots.

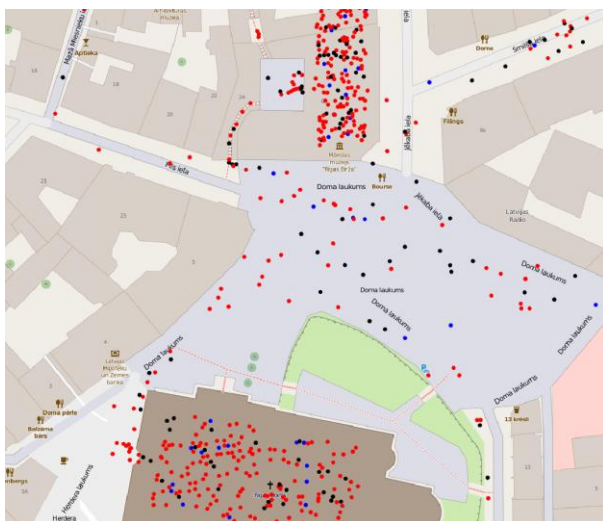


Figure 2: Fragment of the Model Screenshot

6. SIMULATION AND EXPERIMENTAL ANALYSIS

The increasing number of international tourists visiting heritage cities and capacity limits for the most heritage sites worldwide may lead to queuing of tourists at the

entrances to popular museums, art galleries, etc. In the case study, long queues appeared at the entrances of some indoor attractions, particularly, at the entrance to the cathedral (Figure 3). Thus the length of the queue is defined as the main performance indicator to validate the proposed approach and assess outcomes of the use of mobile applications by performing simulations.

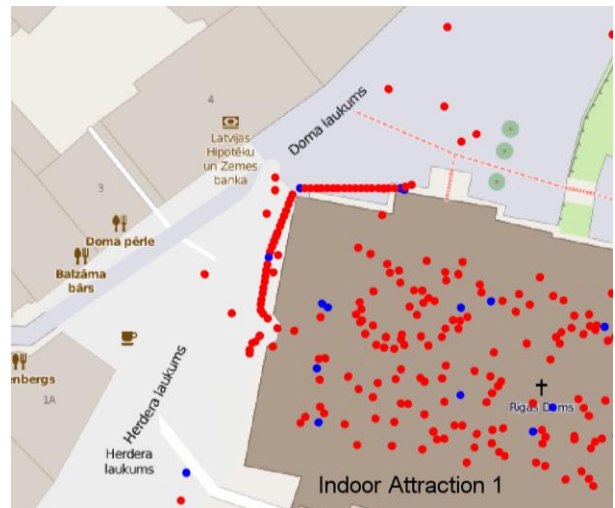


Figure 3: Queues at the Entrance of Indoor Attraction 1

A number of simulation experiments have been performed with different relative numbers tourists equipped with smart applications. Time plots for queue lengths at the entrances of indoor attractions (i.e., museums in the model) received from simulations are shown in Figures 4 and 5. In the case where all agents represent ordinary tourists (Figure 4), queue lengths tend to increase over time (scored in seconds). It can be seen that the distribution of numbers of tourists between museums is very uneven, and there is almost no queue to get to the Museum 3. In the case where about 50% of tourists use smart mobile applications to choose the next attraction (Figure 5), queue lengths are much smaller than in the previous scenario. All statistics are reset at the end of a warm-up period.

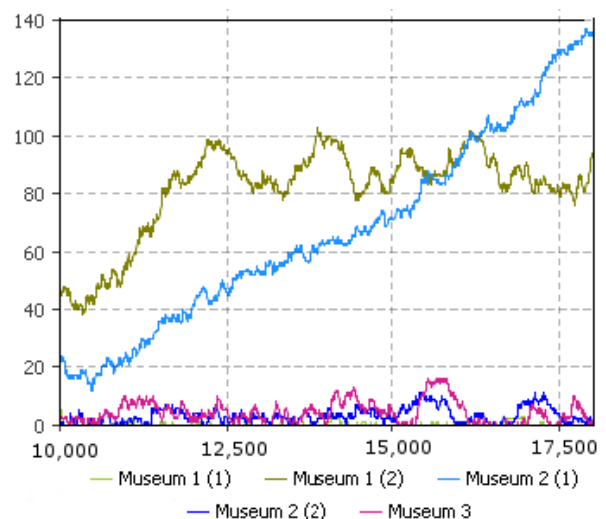


Figure 4: Queue Lengths (no application used)

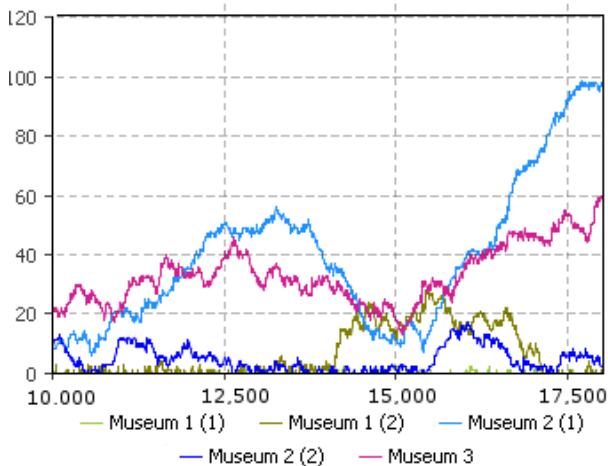


Figure 5: Queue Lengths (50% of tourists use applications)

In the case where relative numbers of advanced tourists are fairly high, the queues are still formed. This is due to a time lag. In the current logic of the model, tourists are suggested to go to the next attraction when they leave the previous one. While they are moving, the attraction can be occupied by other tourists. Thus, logic of the applications is supposed to be adjusted to deal with this time lag.

A regression metamodel is built to measure the sensitivity of queue lengths from a relative number r of tourists with the proposed application. In the model, the total queue length L is calculated as a sum of all tourists waiting in the queues to indoor attractions at the end of the simulation period and $r \in [0, 0.2, 0.4, 0.6, 0.8, 1.0]$. In each experiment, 200 simulation replications have been performed to obtain the mean value with the 95% confidence interval (Figure 6).

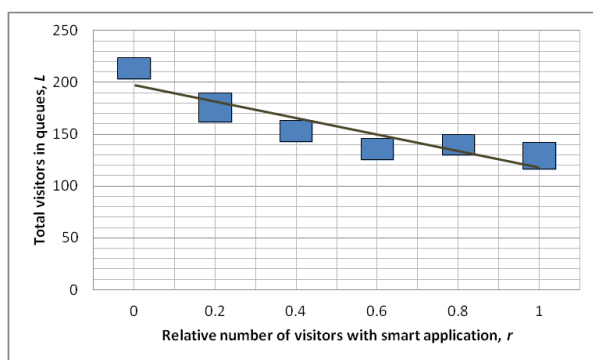


Figure 6: Confidence Intervals for Total Queue Lengths

As a result, the simulation metamodel is approximated by a linear function, i.e., $L \approx 197.9 - 78.1r$. So, an increase in the number of tourists which would use such applications can significantly reduce a number of visitors waiting in queues while managing tourist flows and overloading of heritage sites, thereby reducing their negative impacts on the heritage buildings.

7. CONCLUSIONS

Smart mobile applications with embedded simulation capabilities can significantly improve intelligence of managerial decisions and quality of customer services. The proposed concept of an intelligent touristic mobile application which integrates agent-based simulation has shown its sustainability and allowed partially unload overcrowded heritage sites. It may be advanced by a more sophisticated processing logic and decision making procedures.

ACKNOWLEDGEMENTS

The research is performed within the COST (European Cooperation in Science and Technology) Action TD1406 in "Innovation in Intelligent Management of Heritage Buildings" (i2MHB).

REFERENCES

- Bolshakov V. and Merkurjeva G., 2016. Simulation in Intelligent Management of Pedestrian Flows at Heritage sites. Proceedings of Second International Conference on Systems Informatics, Modelling and Simulation, Riga, 2016, pp. 18-24.
- Camillen F., Capri S., Garofalo C., Ignaccolo M., Inturri G., Pluchino A., Rapisarda A. and S. Tudisco, 2009. Multi agent simulation of pedestrian behavior in closed spatial environments. Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference, Toronto, ON, 2009, pp. 375-380.
- Crooks A., Castle C. and Batty M., 2008. Key challenges in agent-based modelling for geo-spatial simulation, Computers, Environment and Urban Systems, 32, 417-430.
- Li Y. and X. Wang, 2011. Tourist flows analysis and decision support system based on intelligent mobile phone. Proceedings of the IEEE International Conference on Cloud Computing and Intelligence Systems, pp. 418-422, Beijing.
- Liu S., Lo S., Ma J. and W. Wang, 2014. An Agent-Based Microscopic Pedestrian Flow Simulation Model for Pedestrian Traffic Problems, IEEE Transactions on Intelligent Transportation Systems, 15 (3), 992 - 1001.
- Lyu M.N, Yang Q.S., Yang N. and S.S. Law, 2016. Tourist number prediction of historic buildings by singular spectrum analysis, Journal of Applied Statistics, 43 (5), 827-846.
- Malta Tourist Mobile Apps, 2016. Available: <http://www.visitmalta.com/en/mobile-apps>
- Marie-Theres Albert, 2016, Perceptions of Sustainability in Heritage studies: Walter De Gruyter GmbH, Berlin-Boston.
- Merkuryev Y., Merkurjeva G., etc., Eds., 2014. Information Technologies and Support Tools for Space-Ground Monitoring of Natural and Technological Systems. Riga: RTU, 2014, 110 p.

- Mobile Apps - Visit Barcelona, 2016. Available from: <http://www.barcelonaturisme.com/wv3/en/page/1464/mobile-apps.html> [accessed 15 May 2016]
- OpenStreetMap, 2016. Available from: <http://www.openstreetmap.org/> [accessed 20 June 2016]
- Sarraipa J., Jardim-Goncalves R. and Steiger-Garcia A., 2010. MENTOR: an enabler for interoperable intelligent systems. *International Journal of General Systems*, 39 (5), 557-573.
- System Dynamics Society, 2016. Available from: <http://www.systemdynamics.org/> [accessed 15 May 2016]
- Xiaowen L., Jin Yu, Yu Z. and Lu Na, 2012. Design of Service-flow Oriented Intelligent Tourism Collaborative Service System based on MAS. *Proceedings of the 24th Chinese Control and Decision Conference*, pp. 2672-2675.

AUTHORS BIOGRAPHY

Galina Merkurjeva holds a Doctor of Engineering degree in Technical Cybernetics and Information Theory from Institute of Electronics and Computing Techniques of Latvian Academy of Sciences and Habilitated Doctor of Technical Sciences in Control of Social and Economic Systems from V.A. Trapeznikov Institute of Control Sciences, RAS. She is Professor at the Department of Modelling and Simulation and leading researcher at the Institute of Information Technologies of Riga Technical University. Her professional interests and experiences have focused in discrete-event simulation, simulation metamodeling and optimisation, simulation-based training, decision support and expert systems, and supply chain management. She is an author of more than 160 publications, including 5 books.

Vitaly Bolshakov holds a Doctor of Engineering degree in Management Information Systems from Riga Technical University, Latvia (2013). He is a lecturer at the Department of Modelling and Simulation and a leading researcher at the Institute of Information Technologies of Riga Technical University, Latvia. His main research interests and experiences are in the fields of metaheuristic optimisation, fitness landscape analysis and simulation-based optimisation.

João Martins holds a PhD degree in Electrical Engineering from Instituto Superior Técnico (IST), Technical University of Lisbon. He is an Auxiliar Professor at the Electrical Engineering Department, Faculty of Sciences and Technology of Universidade Nova de Lisboa, Portugal and a researcher at UNINOVA Institute, Portugal. João Martins has more than 30 books and journal publications. He has published more than 100 papers in conference proceedings with referees. His main interests include fault diagnosis and fault tolerant operation, alternative energies and power quality, intelligent and energy

efficient buildings, and related engineering education issues.

Ricardo Jardim-Gonçalves holds a PhD degree in Industrial Information Systems from the New University of Lisbon. He is Auxiliar Professor at the New University of Lisbon, Faculty of Sciences and Technology, and Senior Researcher at UNINOVA institute. His research activities have focused on standard-based intelligent integration frameworks for interoperability in complex systems, covering system's adaptability in evolutionary and networked environments, architectures, methodologies and toolkits to support improved development, harmonisation and implementation of standards for data exchange in industry, from design to e-business. He has been technical international project leader for more than 15 years, with more than 80 papers published in conferences, journals and books and 4 books edited.

Gumersindo Bueno Benito holds a degree in Law from the University of Cantabria and a Master degree in Project Management from the University of Valencia Spain. Management Professional (PMP) certified by Project Management Institute. Currently he is a project manager and Head of the Project Management Department at the Foundation Santa María la Real, Spain.

USING NODEJS FOR DISCRETE SIMULATION IN A WEB-BASED ENVIRONMENT

Thomas Wiedemann^(a)

^(a)University of Applied Science Dresden, Germany

^(a)wiedem@informatik-htw-dresden.de,

ABSTRACT

There is a quiet revolution in the web application area. Instead of using hundreds of heavy web-servers in expansive server farms, only one powerful server is used for the black Friday super-business at companies like Walmart. The secret behind it is the base software that is being used: a JavaScript engine called NodeJS for server applications. In contrast to heavy weight multitasking procedures of the operating systems with a maximum of about 200 processes at one time, NodeJS uses its own light weight multitasking in combination with Google's performance optimized JavaScript-V8 engine with up to 1 Million processes on the same machine.

This paper analyses the options for using this multitasking power for discrete simulation. The additional benefit of using JavaScript on a webserver is the easy integration in web-oriented environments.

Keywords: NodeJS, JavaScript (JS), light weight multitasking, web-based JavaScript environments

1. INTRODUCTION

The web community is seeing a dramatic change: nearly all web companies focus on the new JavaScript (JS) options on the server and develop new JS-frameworks for all needs like visualization, data handling and 3D-design. The first time in web history there is no programming break between the browser client and the server. By using the JS-framework NodeJS JS-code is implemented both on the browser and the client. So there is just one code base for both sides.

An additional benefit is the high performance of the whole system. In contrast to heavy weight multitasking procedures of the operating systems with a maximum of about 200 processes at one time, NodeJS uses its own light weight multitasking in combination with Google's performance optimized JS-V8 engine with up to 1 Million processes on the same machine. It was reported that Walmart's extreme black-Friday business was done by only one powerful server machine instead of a large computer farm (Pasqual 2015). Because this saves some millions of dollars, this technology will quickly emerge on all other web servers and big web solutions.

The power behind the NodeJS is the new developed V8-engine from Google. When this engine was put to work for the first time in 2010, it impressed the community with the speed of the JavaScript inside the browser.

Instead of drawing the menus of some big web application like TYPO3 step by step, all the menus and backend options were shown at the same time; it was a visible performance jump of about 1000%.

The secret behind this performance jump is the optimized V8-JS-compiler from Google (see (Daniel Clifford, 2012)).

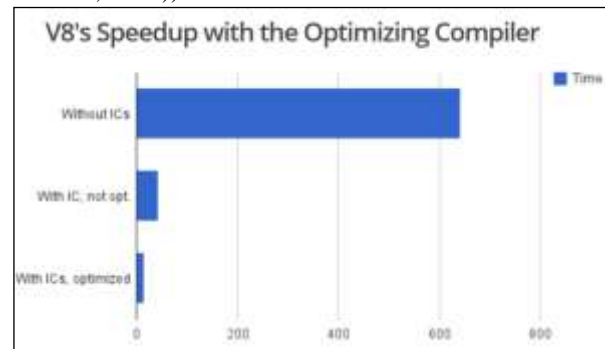


Figure 1: The performance Jump of Googles V8-engine

There is a long list of details, how to use the optimizing compiler in the correct way. In the best case with all optimizations active, the speed of the interpreted JavaScript program is nearly the same (only -17%) as a compiled C++-program. This was never achieved before and is the cause for the quiet revolution in computer science, especially in web-based apps.

2. NODEJS

The JavaScript based Node.js (here NodeJS) engine for server applications was built since 2010. NodeJS uses an event-driven, non-blocking I/O model that makes it lightweight and efficient (see (NodeJS 2016) for details). The further development of NodeJS has been managed since 2015 by the NodeJS Foundation which involves all big and important computer companies. So the future and full acknowledgement and support of NodeJS in the whole computer science community is nearly 100% guaranteed.

2.1. Code efficiency of NodeJS

NodeJS is not only powerful in the time dimension. The coding efficiency is also very high. The following code example from the Node.js homepage is all that is needed for a high-performance HTTP-webserver.

```

var http = require('http');
var static = require('node-static');

var file = new static.Server();

http.createServer(function (req, res) {
  file.serve(req, res);
}).listen(1337, '127.0.0.1');

console.log('Server running ... ');

```

Figure 2: A simple HTTP-Server under Node.js

In addition, there is a very powerful package manager for Node.js, called npm. It helps to install additional Node.js modules from the global npm registry and manages their life-cycle with updates and solving dependencies from other packages.

3. IMPLEMENTING THE NODE-SIMULATOR

3.1. Basic architecture

The main task of the Node.js-engine and framework is the implementation of the light-weight multitasking environment for a large number of concurrent processes – the same task that a discrete simulator has.

In NodeJS the task is completed by a non-blocking layer below the main programming level. While an IO-device-access will take some time, NodeJS switches automatically to the next pending process. As a result of this **event-driven programming**, which is very different from the well-known simulation procedural sequential programming, the structure of a JavaScript NodeJS-based software is very different. The main architecture of NodeJS is shown in fig. 1. (NodeJS 2016)

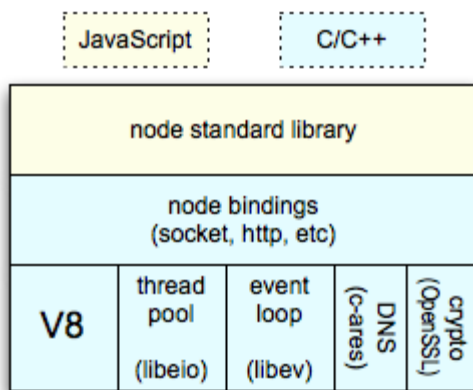


Figure 3: The NodeJS-Architecture (NodeJS 2016)

NodeJS is a combination of a number of C++-programmed basic modules, which could also be extended by new, user specific C++-modules and the JavaScript-application code above.

3.2. Typical code structures in NodeJS

As a result of the event-driven programming paradigm the structure of a simulation oriented program is different from the programming in procedural oriented environments like C++ or SLX.

According to the new event based programming style the code must be divided in two groups;

Blocking code will delay the further execution of the NodeJS-engine and should not be used in NodeJS. In PHP or C such code is used for IO-operations or longer OS-operations (e.g. waiting for user input). In a simulation language like SLX, the corresponding blocking code parts use the unconditional time delay **advance(time)** or the conditional time delay **waituntil(condition)**. Like in NodeJS, these SLX-code-parts switch the execution sequence of the simulation processes. This is not seen in the code and unexperienced SLX-programmers will not understand the main underlying function principle of SLX (this is often seen in SLX-beginners courses)

Both the typical PHP-web program and the code of a SLX-program must be converted by the following conversion rules (see details in (Howard 2013)):

1. All non-blocking code can be written in the sequence shown before. If the source language is C-style compatible, this is done in a very similar manner in JavaScript.
2. Blocking code must be converted to a combination of calling the desired blocking function with a link to a callback to a function, which is called after completion of the time consuming task. The call back function is often provided below the function call, but can be defined also at some other place before.

By following these conversion rules a standard PHP program with some calls to blocking IO-function is converted to a chain of NodeJS callbacks (fig. 4). Please note that the NodeJS is not a sequential program, but the code exits the program after calling each fs.write-function and comes back a long time later (in processor time) to the callback function below the function call.

```

// PHP program writing 3 lines of text
$fp = fopen('fp.txt', 'w');
fwrite($fp, 'line1');
fwrite($fp, 'line2');
fwrite($fp, 'line3');

// corresponding NodeJS-program
fs.open('fp.txt', 'w', 0666, function(error, fp)
{ fs.write(fp, 'line1', null, 'utf-8', function() {
  fs.write(fp, 'line2', null, 'utf-8', function() {
    fs.write(fp, 'line3', null, 'utf-8', function(){});
  });
}); // as a sequence of nested callbacks

```

Figure 4: A PHP to NodeJS conversion example

3.3. The simulation code structure

According to the typical code structure of non-blocking NodeJS code the resulting code for a simulation with delays must be structured in the same way. Every time a simulation process must advance in time – like the SLX and GPSS commands `advance(time)` or the waiting construct in SLX `wait_until(condition)`, a combination of a corresponding function call with a callback function, where the next process steps are defined, is necessary. The resulting program code for a simulation program for NodeJS is shown in fig. 5.

```
// Single simulation with 2 operation steps
var simdis = require('./simdis');//sim modul
console.log("Start SimPro " + simproclD);
// advance to start time of process
simdis.advance( simStartTime,
function() // callback to next step
{ console.log(simproclD +":Step1");
  simdis.trace(simproclD +":Step1");
  // do the job
  simdis.waituntil( simCond2(),
    function(err) // next callback
    { console.log(simproclD + ":Step3");
      // and so on ...
    }
  });
}); // end of nested function calls
```

Figure 5: A typical NodeJS simulation code

The `simdis.advance(..)` is a function call to a simulation specific function from the new `simdis`-module (see also 3.4). The second parameter of the function call is a nameless call back function, where the code for the simulation process after returning from the advance-waiting time is located.

The `simdis.waituntil()` –function is similar. The first parameter is also a function call to a predefined function (not shown in the code example), where a condition is calculated and a Boolean is returned. If the result is true, NodeJS will continue the execution to the callback function inside the second parameter.

Actually, this building of a deep nested sequence of functions is not very easy and it is very difficult to find mistakes. It would therefore be a good idea to generate such deep nested code from graphical representations of simulation models.

3.4. Implementation of the simulation core

The main task of implementing a simulation scheduler under Node.js is to change the common event based switching layer to a simulation time controlled module - the already well-known discrete future event list scheduler (see (Schriber 2013) for details). Instead of returning the first finished process after a somewhat random timed IO-operation, the new switching module

will precisely select the next process in simulation time from the event list and activates this process.

All the simulation specific code is located in a new module “`simdis`”, which is loaded in the first program line in the example shown above. This module is using the same `libev-C++-library` for scheduling events in order to switch from a blocking scheme to a non-blocking scheme defined by function call and a callback function inside. By using this basic library of NodeJS the main event loop can also be used for the simulation scheduling.

Although first performance tests were very promising, the actual main task is the implementation of very efficient future event list modules under JavaScript.

In the actual version a very simple future event list is used for first tests. In the future version a more sophisticated event list implementation will be used, in order to test the maximal possible runtime speed. This task will be done according to well-known scheduling strategies with callback-functions like in SLX for avoiding a slow polling of conditions in blocking conditions.

It must be noted that that the described scheduling technology is based on the main single threaded NodeJS-loop which runs only on one core. If there are more cores on the hardware, there is a second option for distributing processes over different cores.

3.5. Horizontal distribution of processes

According to the web orientation of NodeJS and the main focus on high-parallel HTTP-server implementations, NodeJS is also able to distribute its processes over different processor cores.

This technology is efficient if the processes do not communicate too much with each other, because the communication is relatively slow in result of the used web-interfaces with TCP/IP or also faster UDP packets. The program example is fig. 6 shows a multicore simulation, where the `simproc()` function contains code like in fig. 5.

```
// Multicore simulation
var cluster = require('cluster');
var http = require('http');
var nCPUs = require('os').cpus().length;

if(cluster.isMaster)
{ console.log("Starting Master process");
  console.log("Sim on CPUs = " + nCPUs);
  for(var i = 0; i < numCPUs; i++)
  { cluster.fork(); }
}
if(cluster.isWorker)
{ console.log("Sim " + cluster.worker.id);
  simproc(cluster.worker.id);
}
```

Figure 6: A horizontal process scheduler

One option for avoiding any communication by the distributed process is the so called Hyper computing simulation, where each single Node-process on every core executes a full version of a simulation task. Only the results of each simulation are transferred to a main controlling module on single core, which collects all results and generates the statistics and final report for all replications.

3.6. General performance issues under JS

JavaScript itself was long known as a slow interpreted language inside web browsers. In general this is still true, because JavaScript is still an interpreted language, which is not compiled before runtime. But the V8-engine for Google changed the situation. The V8-interpreter is a highly sophisticated system, which tries to start the execution of the JS-program very quickly, but during the execution checks the code more closely and generates native code for the existing processor. This native code generation depends heavily on some race conditions. The most important is the constant usage of the same type definitions over the sub-function calls. Because JS is not strongly typed, it is possible to call a function with very different variable types. In such a case the generated native code will be invalidated and the performance falls back to an interpreted very slow level. Because the web community is also interested in a very high overall performance of a NodeJS based web server, there are a lot of monitoring and testing tools for analyzing the code and its behavior (see the book “Deploying Node.js” (Pasqual 2015)).

3.7. Extending and running the simulation system

The supporting functions for a simulation environment like graphical presentation of results, database import and export interfaces, connections to other TCP/IP based interfaces like web-services are already ready to use from a large number of libraries like JQuery, AngularJs and D3.js.

NodeJS can run on nearly all OS without any special requirements. The Node.js-environment is free and perfectly prepared for a new style of web-based simulation. It could be expected that large cloud service providers like Amazon and Google will provide Node.js based cloud services in the near future. The new NodeSim will perfectly fit into this cloud environment and this will open new dimensions to the simulation community for joint development and usage of a very modern and open simulation framework.

Like the actual quiet revolution in web oriented development, this could also lead to new horizons in discrete simulation.

4. OUTLOOK TO THE FUTURE USAGE OF JS IN SIMULATION

Until now, most universal cloud providers like Microsoft’s Azure, Google and Amazon offer only execution options for well understood code for their sites in result of security limitation. The reason is that it is very critical for the cloud provider to understand native machine code delivered from unknown cloud users. They therefore allow only .NET-based code with a well-defined set of .NET commands or very similar Java-bytecode.

As a result of this security barrier no well-known complex simulation environments like SLX, Enterprise Dynamics can be executed on a common cloud system. (Of course it is possible to rent a full user-defined cloud server with full OS-access and to install this software, but this very expensive and not much better than a set of existing PC’s).

JavaScript and NodeJS will change this situation. Because JS is a interpreted language, the user must deliver the source code and the cloud provider can check this against its security rules. The V8-engine will then compile and optimize this code to native code for the available hardware, so there is no disadvantage against compiled code in the future. Future work will analyze the real parameters on cloud solutions and will try to use some big existing cloud systems.

REFERENCES

- Clifford, D., 2012. Breaking the JavaScript Speed limit with V8 <http://v8-io12.appspot.com/#2> 2012
- Howard, D.. Node.js for PHP Developers. O’Reilly Media 2013
- NodeJS 2016. <https://nodejs.org/en/>
- Pasquali,S.. Deploying Node.js. Packt Publishing, 2015
- Schriber T.,Brunner D., Smith J., 2013. Inside Discrete Event Simulation Software: How It Works and Why It Matters. Proceedings of the 2013 Winter Simulation Conference, Pages 424-438

AUTHORS BIOGRAPHY

THOMAS WIEDEMANN is a professor at the Department of Computer Science at the University of Applied Science Dresden (HTWD). He has finished a study at the Technical University Sofia and a Ph.D. study at the Humboldt-University of Berlin. His research interests include simulation methodology, tools and environments in distributed simulation and manufacturing processes, intranet solutions and database applications.
Email: wiedem@informatik.htw-dresden.de

A COMPARATIVE STUDY OF HYPERELASTIC CONSTITUTIVE MODELS FOR AN AUTOMOTIVE COMPONENT MATERIAL

Rafael Tobajas^(a), Daniel Elduque^(b), Carlos Javierre^(c), Elena Ibarz^(d), Luis Gracia^(e)

^{(a),(b),(c),(d),(e)} Department of Mechanical Engineering, University of Zaragoza, C/ María de Luna, 3, 50018 Zaragoza, Spain

^(a)rafaeltobajasalonso@gmail.com, ^(b)daniel.elduque@gmail.com, ^(c)carlos.javierre@unizar.es, ^(d)eibarz@unizar.es
^(e)lugravi@unizar.es

ABSTRACT

The use of thermoplastic elastomers has strongly increased in recent decades in order to reduce the size of components in the automotive and aeronautical industries. To design this kind of components, engineers face the challenge of reproducing the behavior of these materials by numerical simulations. This task is not always simple because these materials often have a strongly nonlinear behavior. In this paper an elastomer thermoplastic material has been analyzed and an automotive component has been studied by five numerical simulations with five material constitutive models. This study shows that a careful choice of the constitutive model should be made to obtain reliable results. Although several constitutive models fit well with the experimental data of uniaxial testing, when these are used in actual components, there are significant differences in the obtained results.

Keywords: hiperelasticity, thermoplastic elastomer, finite element method, simulation.

1. INTRODUCTION

Automotive and aeronautical industries consider a future challenge to further increase engine efficiency by decreasing fuel consumption and motor weight. In order to reach it, designers must be able to reduce the size of the engine block and all its components using alternative materials to metals, such as thermoplastic elastomeric polymers. (Drobny, J. G., 2014).

The use of these materials has strongly increased in recent decades due to their good mechanical properties (Štrumberger et al., 2012) (P. Consulting, 2014). In addition to its ease to manufacture complex-shaped components, these materials have great advantages as their high deformation capability, their ability to absorb vibrations and their low cost-weight ratio. (Malloy, 1994).

Despite all their advantages, developing an efficient and durable design with these materials is not an easy task as reproducing their mechanical behavior by simple

computational algorithms is not always possible. The constitutive relations between stresses and strains for these materials are nonlinear and time-dependent, and additionally they also experience other effects such as hysteresis and softening (Charlton et al., 1994).

Although hyperelastic constitutive models are the best models to reproduce the nonlinear behavior of these materials, several more complex constitutive models exist in the literature to also take into account other effects. Therefore, the selection of a proper constitutive model remains an engineering challenge to be solved since the behavior of the material has a good fit with any hyperelastic model, the mechanical behavior of the component is often not adapted to the actual behavior.

From these constitutive models, simulations techniques must be used to obtain the stress and strain fields to evaluate the component from an engineering point of view. Simulation is one of the most important fields into the world of engineering due to it is used for such varied sectors such as structural designs and manufacturing studies of plastic parts (Javierre, et al. 2013) (Javierre, et al. 2006) or numerous studies into the food sector (Jiménez, et al. 2014) (Latorre-Biel, et al. 2013).

Finite Element Method (FEM) is a numerical technique currently used in simulation processes for several fields. Specifically, into the mechanical engineering field, it is used for studies of developing of food packaging (Fernandez, et al. 2013), designing of led weatherproof luminaire (Javierre, et al. 2014) or modeling automotive products (Jiménez, et al. 2009) (Ruiz Argáiz, et al. 2008).

In this paper five simulations of a real component used in the automotive industry, based on five different constitutive models of material, have been performed, comparing the corresponding results. In this study, FEM simulations were carried out using the software Abaqus (Version 6.11).

2. MATERIAL DESCRIPTION

An automotive air duct is going to be analyzed in this paper. This component is made of a material called Santoprene 101-73 provided by ExxonMobil (ExxonMobil).

The nonlinear behavior of the material has been described by uniaxial tensile tests like simple tension, planar extension and simple compression. In addition, cyclic behavior information is also provided at different levels of deformation.

To characterize the material, the datasheet and Young modulus value supplied by the manufacturer have been used. This study focuses on the relationship between stress and strain for the first pull deformation, when it is the first time in deforming the material. The Young modulus recommended by the manufacturer in this deformation is $E = 29.7$ MPa (ExxonMobil). This relationship can also be obtained from the cyclical behavior information of the material, and it is represented by a red line in Figures 1 and 2. According to standards (ASTM D412-15a) (ISO 37:2011), the provided values of strain and stress are nominal values.

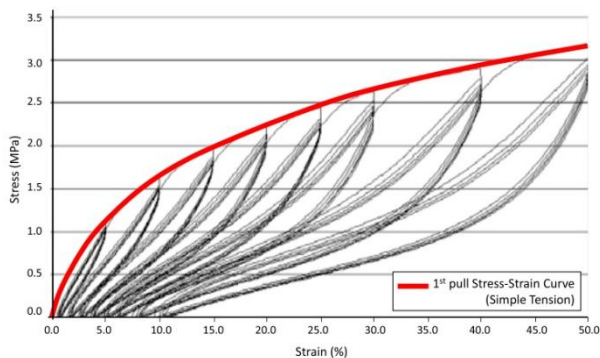


Figure 1: Strain-stress plot of Santoprene 101-73 from manufacturer datasheet (ExxonMobil) for Simple Tension.

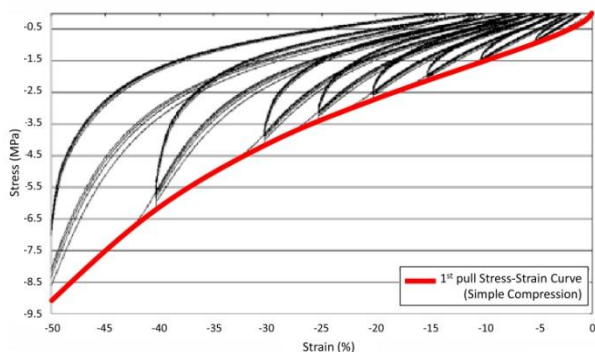


Figure 2: Strain-stress plot of Santoprene 101-73 from manufacturer datasheet (ExxonMobil) for Simple Compression.

In order to obtain values of strain energy density, strain and stress values from the first pull deformation shown

in Tables 1 and 2 are used. Strain energy density values were obtained by Equation (1).

$$W = \sum_{i,j=1}^3 \int_0^{\epsilon_{ij}} \sigma_{ij} d\epsilon_{ij} \quad (1)$$

Strain, stress and strain energy density (SED) values from the first pull deformation are shown in Tables 1 and 2.

Table 1: Strain-stress values for Santoprene 101-73

First Pull-Simple Tension		
Nominal Strain (%)	Nominal Stress (MPa)	SED (MPa)
0.000	0.000	0.000
0.900	0.345	0.001
2.000	0.615	0.006
3.200	0.859	0.014
4.700	1.084	0.027
6.900	1.338	0.054
8.700	1.516	0.080
9.900	1.623	0.098
12.20	1.792	0.137
13.80	1.913	0.165
14.90	1.981	0.188
17.60	2.122	0.241
18.90	2.193	0.269
19.90	2.232	0.290
23.00	2.374	0.359
24.20	2.422	0.388
25.00	2.451	0.406
28.40	2.585	0.493
29.90	2.631	0.530
33.90	2.768	0.638

Table 2: Strain-stress values for Santoprene 101-73

First Pull-Simple Compression		
Nominal Strain (%)	Nominal Stress (MPa)	SED (MPa)
0.00	0.000	0.0000
-0.90	-0.219	0.0010
-2.00	-0.437	0.0045
-3.20	-0.662	0.0114
-4.70	-0.884	0.0224
-6.90	-1.181	0.0454
-8.70	-1.399	0.0691
-9.90	-1.533	0.0866
-12.20	-1.783	0.1246
-13.80	-1.957	0.1543
-14.90	-2.086	0.1780
-17.60	-2.393	0.2380
-18.90	-2.550	0.2704
-19.90	-2.668	0.2955
-23.00	-3.066	0.3839
-24.20	-3.233	0.4224
-25.00	-3.338	0.4471
-28.40	-3.854	0.5715
-29.90	-4.088	0.6291

3. COMPONENT AND SIMULATIONS DESCRIPTION

The component is an air duct for a combustion engine that has a length of 100 mm and the diameter of the inlet and outlet is 24 mm. In the central area it is formed by five bellows where the maximum diameter is 29 mm and the minimum diameter is 24 mm. The wall thickness is 2 mm. Its geometry is shown in Figure 3.

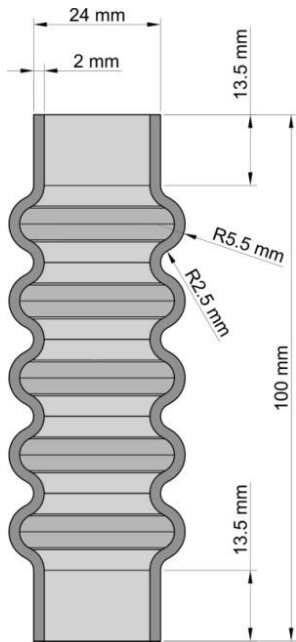


Figure 3: Geometry of the component

The simulation consists of subjecting the component to a prescribed compression deformation of 15 mm in its own axis direction (Figure 4).

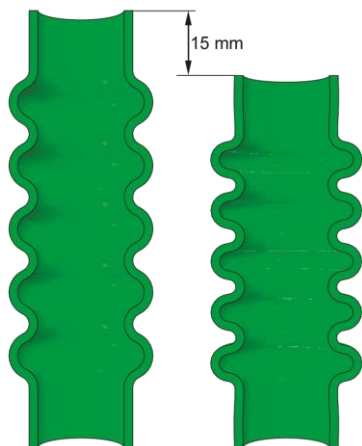


Figure 4: Prescribed deformation in the component

The finite element model was made from a 2D section of the component wall. The 2D section was meshed with 4-node quadrilateral axisymmetric element (CAX4H, Abaqus) (Abaqus Analysis User's Manual, 2011). A radial sweeping from the axis of the duct was used to generate the 3D model (Figure 5). A sensitivity

analysis, checking the strain energy convergence, was performed in order to obtain the appropriate mesh size. The mesh consists of 1148 elements and 4027 nodes.

Several simulations with different constitutive models of material are done to study the mechanical behavior of the component.

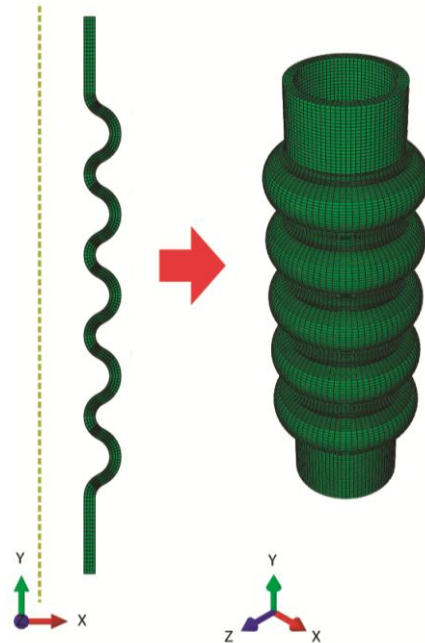


Figure 5: Finite element model of the component

Firstly, a simulation with Elastic material model with $E=29.7$ MPa was carried out (see Figure 6) to obtain the maximum and minimum strain values in the component. These values are shown in Table 3.

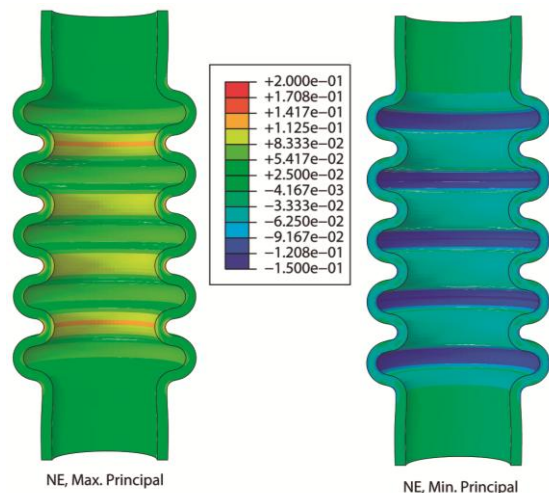


Figure 6: Simulation with Elastic material model ($E=29.7$ MPa). Strain results

Table 3: Max. and min. strains for the simulation with Elastic material model ($E=29.7$ MPa).

Max ppal strain (%)	Min. ppal. Strain (%)
19.34	-13.94

Then, the constants of hyperelastic models are going to be fitted for each model and then the simulations with these models are going to be carried out to compare the results.

4. HYPERELASTICITY

The hyperelastic constitutive models are mathematical models that attempt to simulate the behavior of materials whose stress strain relationship is nonlinear.

Usually these models are represented by strain energy density function W which is formulated as a function depending on different magnitudes associated to the strain field and the material constants, in the way:

$$W = W(\mathbf{F}) = W(\lambda_1, \lambda_2, \lambda_3) = W(I_1, I_2, I_3) \quad (2)$$

where $\lambda_1, \lambda_2, \lambda_3$ are principal stretches, and I_1, I_2, I_3 are the invariants of left Cauchy-Green strain tensor, \mathbf{B} , respectively, obtained as:

$$I_1 = \text{trace}(\mathbf{B}) \quad (3)$$

$$I_2 = \frac{1}{2}(I_1^2 - \mathbf{B} \cdot \mathbf{B}) \quad (4)$$

$$I_3 = \det(\mathbf{B}) \quad (5)$$

In hyperelastic models the Cauchy stresses are derived by differentiating the strain energy density function as follows:

$$\sigma = \frac{2}{J} \cdot \mathbf{B} \cdot \frac{\partial W}{\partial \mathbf{B}} = \frac{1}{J} \cdot \frac{\partial W}{\partial \mathbf{F}} \cdot \mathbf{F}^T \quad (6)$$

where \mathbf{F} is the strain gradient tensor and $J = \det(\mathbf{F})$.

In this paper, five different constitutive models have been used:

- Linear elastic model
- Hyperelastic Neo Hookean model
- Hyperelastic Mooney-Rivlin model
- Hyperelastic Ogden model
- Hyperelastic Marlow model

4.1. Linear elastic model

In this model the material returns to its original shape when the loads are removed, and the unloading path is the same as the loading path. This model is known as Hooke's Law. The stress is proportional to the strain and the constant of proportionality is the Young's modulus (E).

$$\sigma = E \cdot \varepsilon \quad (7)$$

The strain energy density functions is defined as:

$$W = \frac{1}{2} \cdot \sigma_{ij} \cdot \varepsilon_{ij} \quad (8)$$

4.2. Hyperelastic Neo Hookean model

This model was proposed by Treloar in 1943 (Treloar, L. R.G., 1943). In this model, the strain energy density function is based only on the first strain invariant:

$$W = C_1(I_1 - 3) + \frac{1}{D}(J - 1)^2 \quad (9)$$

where C_1 is a material constant, J is the determinant of the strain gradient tensor \mathbf{F} and D is a material constant related to the bulk modulus.

4.3. Hyperelastic Mooney-Rivlin model

The Mooney-Rivlin model (Mooney, 1940) (Rivlin, R. S. and Saunders, D. W., 1951), proposed in 1951, is one of the most used hyperelastic models in the literature. Although there are various versions of this model, the most general is based on the first and second strain invariants. The strain energy density function is defined as follows:

$$W = C_{10}(I_1 - 3) + C_{01}(I_2 - 3) + \frac{1}{D}(J - 1)^2 \quad (10)$$

where C_{ij} are material constants, J is the determinant of the strain gradient tensor \mathbf{F} and D is a material constant related to the bulk modulus. Neo Hookean is a particular case of the two parameters Mooney-Rivlin model.

4.4. Hyperelastic Ogden model

The Ogden hyperelastic model (1972) (Ogden, R. W., 1972) is possibly the most extended model after Mooney-Rivlin model. This model is based on the three principal stretches ($\lambda_1, \lambda_2, \lambda_3$) and $2N$ material constants, where N is the number of polynomials that constitute the strain energy density function, defined as:

$$W = \sum_{i=1}^N \frac{\mu_i}{\alpha_i} (\lambda_1^{\alpha_i} + \lambda_2^{\alpha_i} + \lambda_3^{\alpha_i} - 3) + \sum_{k=1}^N \frac{1}{D} (J - 1)^{2k} \quad (11)$$

where μ_i y α_i are material constants, J is the determinant of the strain gradient tensor \mathbf{F} and D is a material constant related to the bulk modulus.

4.5. Hyperelastic Marlow model

The Marlow hyperelastic model (2003) (Marlow, R. S., 2003) means the strain energy density is independent of the second invariant, and a single test such as a uniaxial

tension test is necessary to determine the material response. The strain energy density function is defined as:

$$W = \int_0^{\lambda_T-1} T(\varepsilon)d\varepsilon \quad (12)$$

where $T(\varepsilon)$ is the nominal uniaxial stress and λ_T is the uniaxial stretch.

4.6. Fitting parameters of material models

For all the previously presented models, the optimal constant values were obtained by means of an optimization algorithm to try to faithfully reproduce the material behaviour provided by the manufacturer. A least squares algorithm was used to minimize this variable. The error function is defined as:

$$\text{Error} = \sum_{i=1}^m (\text{Data}_{\text{real}} - \text{Data}_{\text{model}})^2 \quad (13)$$

where m is the number of points on the chart provided by the manufacturer.

Since the minimum principal strain provided on the component is about -14% and the maximum principal strain is 20%, the models have been adjusted to this range of strain.

Once the optimal constant values were obtained, for each model, the strain energy density and the strain-stress curves were obtained and they were compared with the actual material curves. To determine the quality of each of the models, the R^2 correlation coefficient was calculated for the strain-stress plot (Figure 7).

The values of material parameters and R^2 values for the models are shown in Table 4.

Table 4: Parameters and R^2 values of models

Model	Parameter	Optimal Value	R^2
Linear elastic	E	14.602	0.978
Hyperelastic Neo Hookean	C_1	2.416	0.977
Hyperelastic Mooney Rivlin	C_{10}	3.190	0.975
	C_{01}	-0.818	
Hyperelastic Ogden N=2	μ_1	25.560	0.977
	α_1	0.065	
	μ_2	76.579	
	α_2	0.108	

In view of the obtained results, all presented models fit well with actual values of the material datasheet because the R^2 correlation coefficient is always higher than 0.960. Therefore any of the studied models should provide good results in the simulation of the component.

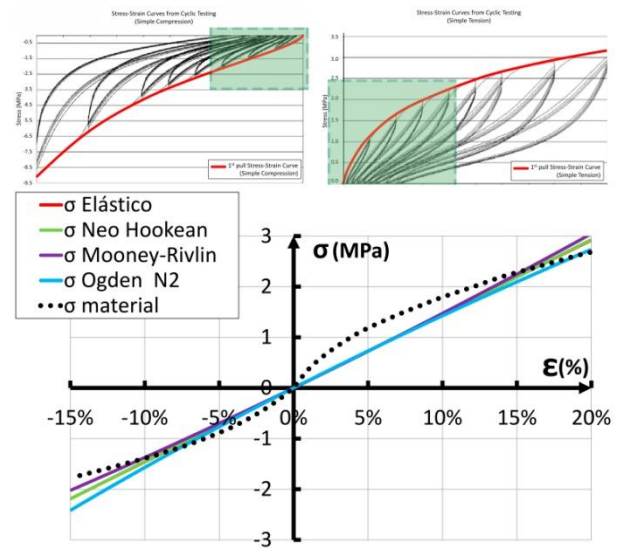


Figure 7: Strain-stress plot for different models between -14% and 20% strain

5. SIMULATIONS AND RESULTS

In order to check the mechanical behavior of the component depending on the constitutive model used, five finite element simulations of the duct compression were performed. Each of them uses one of the different material constitutive models introduced in Section 3.

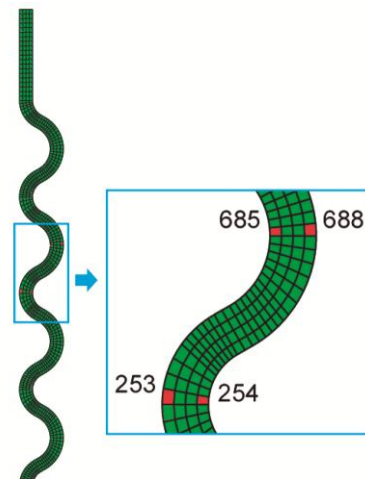


Figure 8: Four characteristic elements of the model

The values of the variables most commonly used in the engineering field, like stresses, strains and energy, have been obtained from integration points of four characteristic elements of the model (Figure 8).

In several studies maximum principal strain and the strain energy density were used to evaluate the life of component when a cyclic load was applied to it. (Mars, W. V., & Fatemi, A., 2002). Other investigations about fatigue use maximum principal stresses to calculate the life of component. (André, N., Cailletaud, G., & Piques, R., 1999). Finally, the Von Mises comparison stress is commonly used to know if the material is able to resist a static load or displacement.

So, concerning the strain field, the maximum and the minimum principal strains, respectively, were studied (Figures 9 and 10). On the other hand, for the stress field, the maximum and minimum principal stresses, respectively, (Figures 11 and 12) and the von Mises stress were analyzed (Figure 13) at each of the integration points and these values were referred at centroid of elements. And finally, the strain energy density was also studied (Figure 14).

6. DISCUSSION

The obtained results show that the mechanical behavior of Santoprene 101-73 can be accurately adjusted by using the different models of hyperelastic behavior considered with the appropriate material constants, with a good agreement with the data supplied by the manufacturer for the uniaxial stress-strain curve. A least squares algorithm allows obtaining the material constant reaching values of R^2 over 0.975 in all studied models when applied for stress-strain uniaxial curve.

However, important differences were found in stresses (up to 40%) and the strain energy density (up to 90%) when the models were applied to a real component in a wide range of strain as occurs in the normal operation of that kind of pieces.

To this respect, the pure elastic, the NeoHookean, the Mooney-Rivlin and the Ogden models provide similar results as much in principal strains as in principal stresses. A similar trend is observed for von Mises stress and strain energy density.

A more realistic behavior is obtained by using Marlow model and important divergences were found respect the other models. The divergence is more pronounced for stresses and strain energy density, whereas it is considerably lesser for strains.

So, a careful election of the appropriate constitutive model must be done in order to obtain realistic simulations of real components, in such a way that the results corresponding to strains and stresses would be reliable enough, allowing the strength and functionality verifications, considering that those magnitudes are

used to predict mechanical performance, fatigue life, etc., and then can determine important features of the component design.

7. CONCLUSIONS

The aim of this paper was to determine the best constitutive model for reproducing the mechanical behavior of Santoprene 101-73, material used in automotive industry. To reach it, four different hyperelastic models and the commonly used linear elastic model have been studied in order to obtain the material constants of each of them. An optimization least squares algorithm were used to fit the best values of material constants to each of them. In order to conclude which of them best represents the actual behavior of the material, the R^2 correlation coefficient for stress-strain relationship has been used.

In view of the obtained results, it can be concluded that all the considered models fit the actual material behavior with enough accuracy, being the Marlow model the most accurate model to reproduce the mechanical behavior of Santoprene 101-73, considering the stress-strain uniaxial curve.

However, the study has demonstrated that despite using constitutive models that fit correctly to the actual behavior of the material in standardized uniaxial tests, the different models show large differences concerning the realistic behavior when they are applied to real components, in which complex stress/strain fields can arise. So, more information from biaxial and planar tests is needed in order to perform accurate simulations for real components.

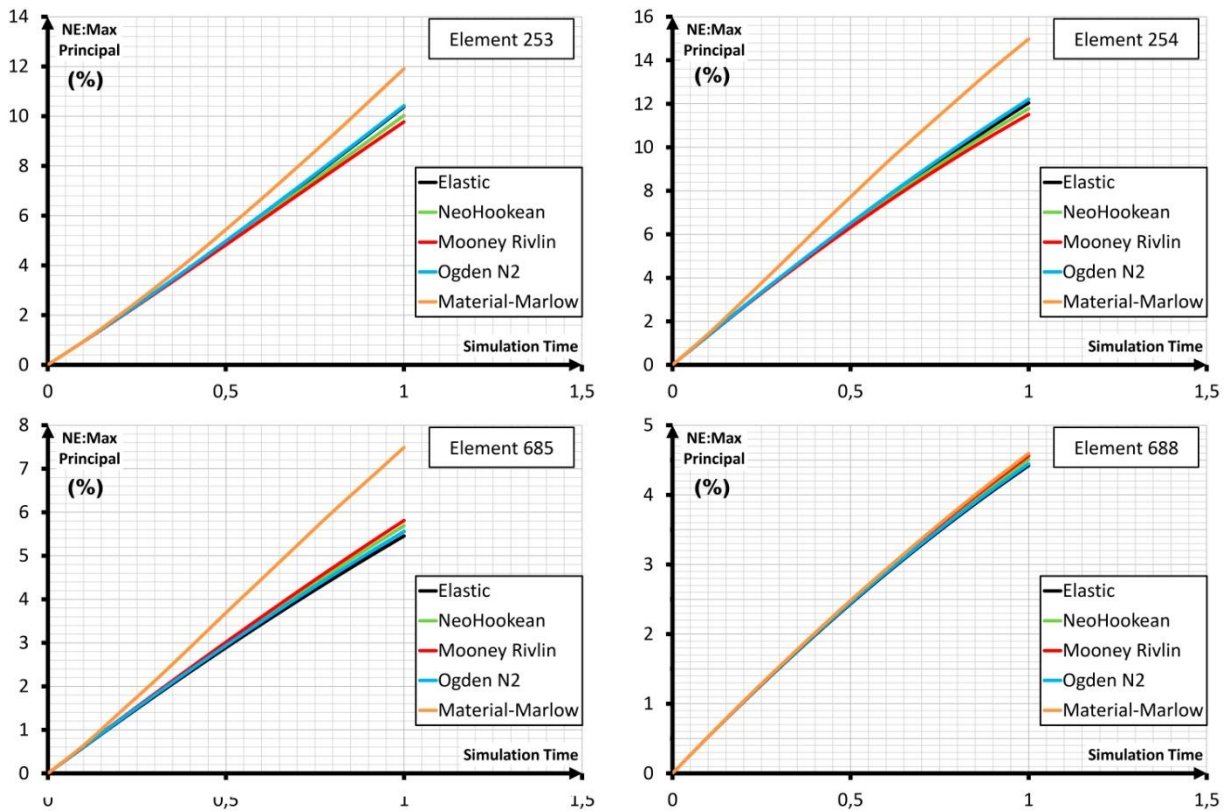


Figure 9: Maximum principal strain for different models calculated at each of the integration points and referred at centroid of four characteristic elements.

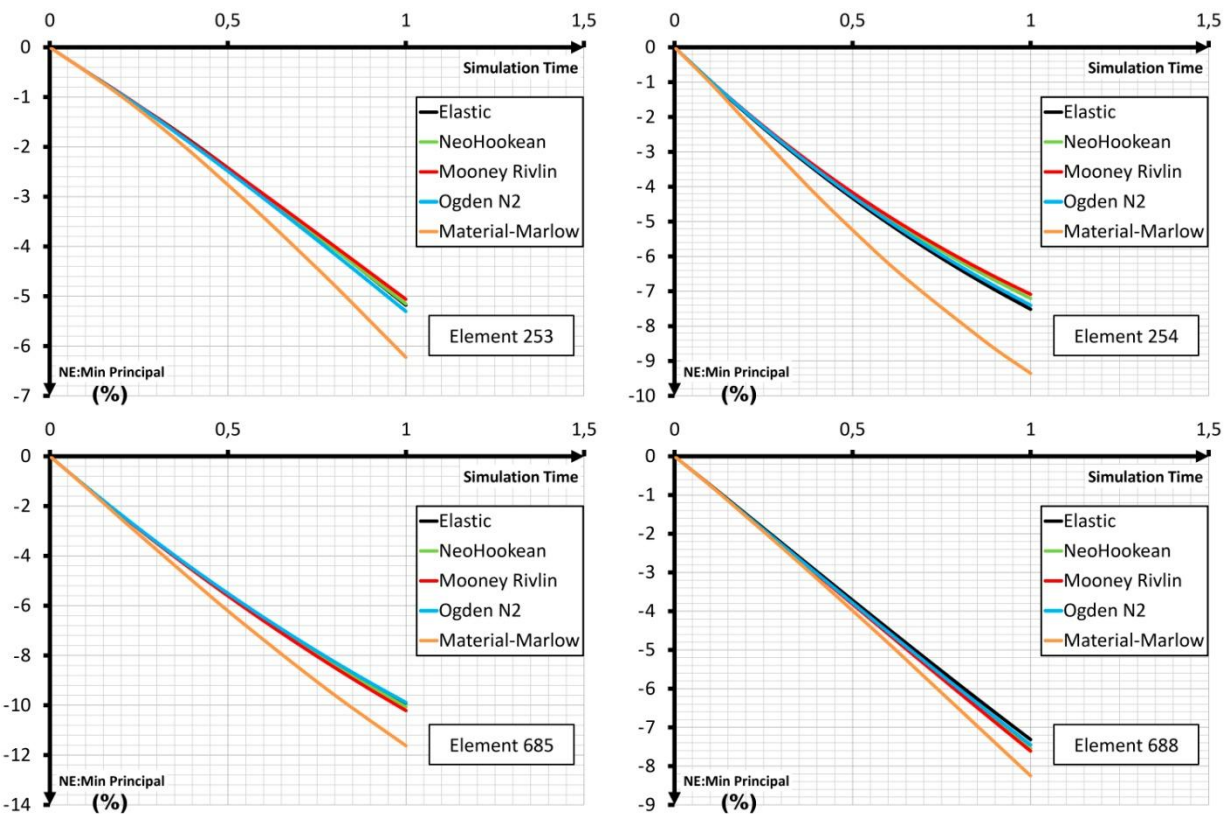


Figure 10: Minimum principal strain for different models calculated at each of the integration points and referred at centroid of four characteristic elements.

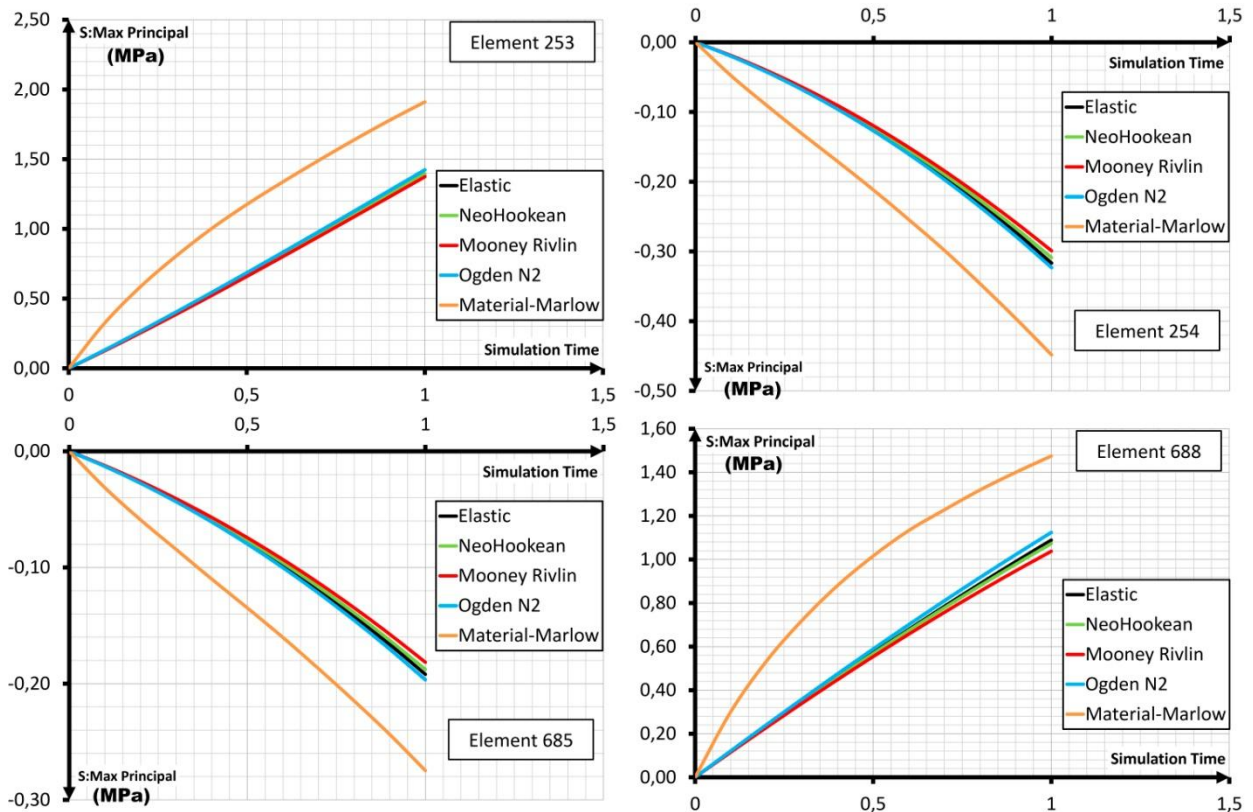


Figure 11: Maximum principal stress for different models calculated at each of the integration points and referred at centroid of four characteristic elements.

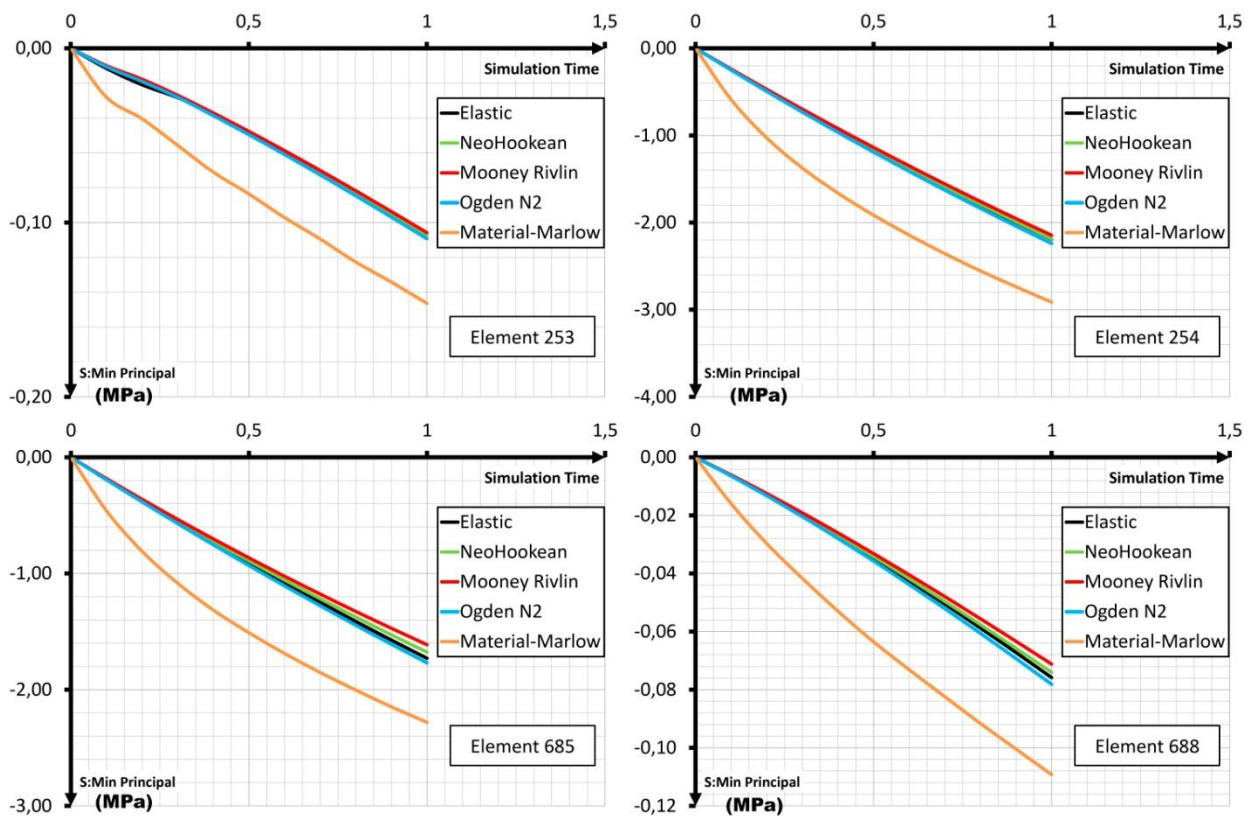


Figure 12: Minimum principal stress for different models calculated at each of the integration points and referred at centroid of four characteristic elements.

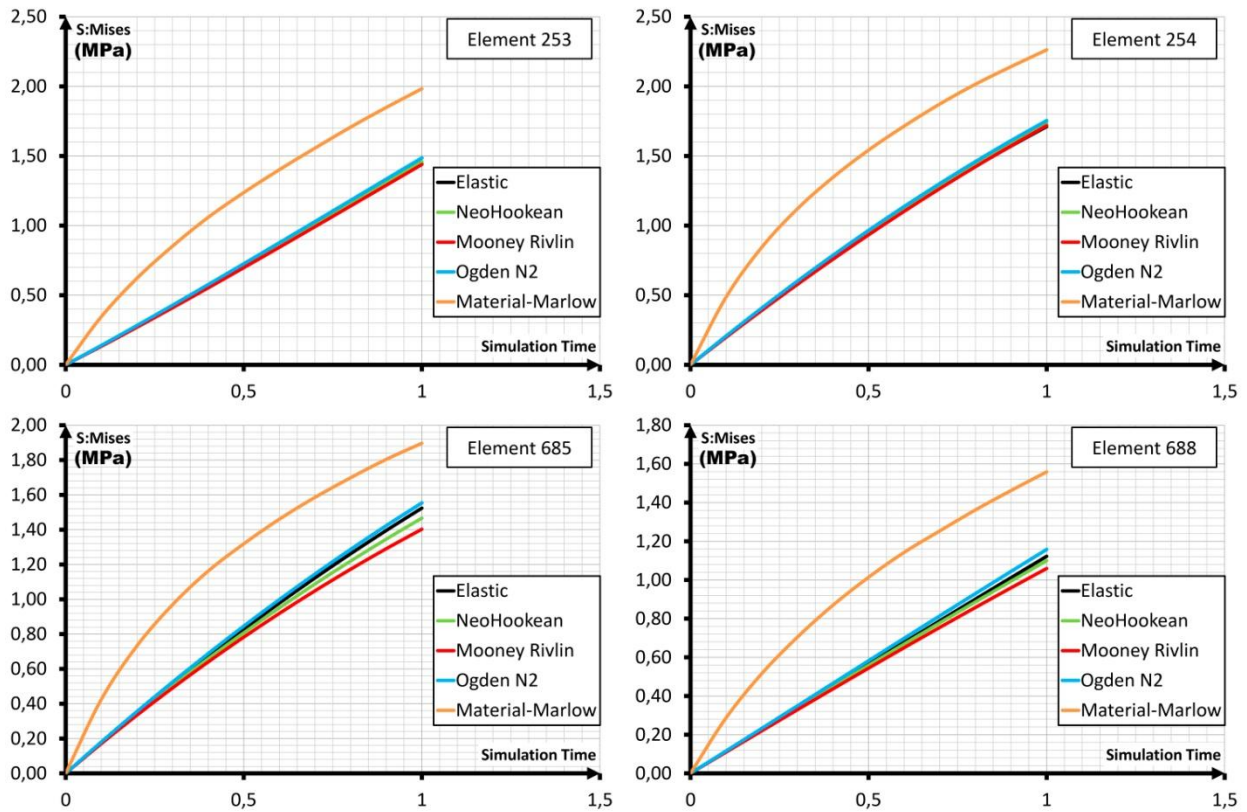


Figure 13: Von Mises comparison stress for different models calculated at each of the integration points and referred at centroid of four characteristic elements.

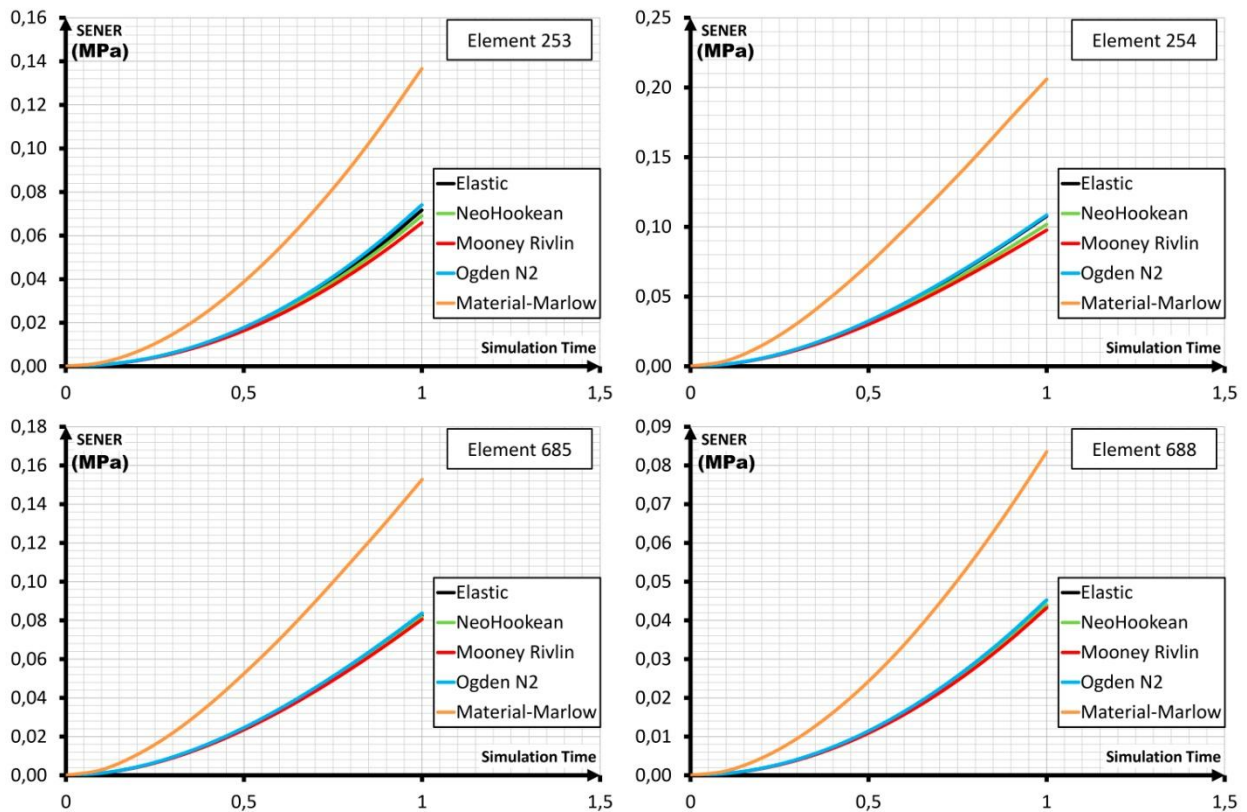


Figure 14: Strain energy density (SED) for different models calculated at each of the integration points and referred at centroid of four characteristic elements.

REFERENCES

- Abaqus Analysis User's Manual, Abaqus 6.11 Online Documentation, Dassault Systèmes, 2011.
- André, N., Cailletaud, G., & Piques, R. (1999). Haigh diagram for fatigue crack initiation prediction of natural rubber components. *Kautschuk Gummi Kunststoffe*, 52(2), 120-123.
- ASTM D412-15a, Standard Test Methods for Vulcanized Rubber and Thermoplastic Elastomers—Tension, ASTM International, West Conshohocken, PA, 2015, www.astm.org
- Charlton, D. J., Yang, J., The, K. K., A Review of Methods to Characterize Rubber Elastic Behavior for Use in Finite Element Analysis, *Rubber Chemistry and Technology* (3) 481–503.1994
- Drobny, J. G. (2014). *Handbook of thermoplastic elastomers*. Elsevier.
- Dufton, P. W. (2001). *Thermoplastic Elastomers Market*. iSmithers Rapra Publishing.
- ExxonMobil, Santoprene 101-73 Thermoplastic Vulcanizate.
- Fernández, A., Javierre, C., González, J. & Elduque, D., 2013. Development of thermoplastic material food packaging considering technical, economic and environmental criteria. *Journal of Biobased Materials and Bioenergy*, 7(2), 176-183.
- ISO 37:2011. Rubber, vulcanized or thermoplastic – determination of tensile stress–strain properties. ISO Standard; 2011.
- Javierre, C., Abad-Blasco, J., Camañes & V., F. D., 2013. Redesign of metallic parts with thermoplastic materials: Application example | Rediseño de componentes metálicos En materiales termoplásticos: Ejemplo de aplicación. *DYNA*, 88(2), 197-205.
- Javierre, C., Elduque, D., Camañes, V. & Franch, D. (2014). Simulation and experimental analysis of led weather proof luminaire thermal performance. *Paper presented at the 26th European Modeling and Simulation Symposium, EMSS 2014*.
- Javierre, C., Fernández, A., Aísa, J. & Clavería, I., 2006. Criteria on feeding system design: Conventional and sequential injection moulding. *Journal of Materials Processing Technology*, 171(3), 373-384.
- Jiménez, E. et al. 2014. Methodological approach towards sustainability by integration of environmental impact in production system models through life cycle analysis: Application to the Rioja wine sector. *SIMULATION*, 90(2), pp.143-161.
- Jiménez, E., Ruiz, I., Blanco, J., & Pérez, M. (2009). Design and simulation of production of injection pieces in automobile industry. *International Journal of Simulation: Systems, Science and Technology*, 10(3), 23-30.
- Kutz, M. (Ed.). (2011). *Applied plastics engineering handbook: processing and materials*. William Andrew.
- Latorre-Biel, J. I., Jiménez-Macías, E., Blanco-Fernández, J., & Sáenz-Díez, J. C. (2013b). Optimal design, based on simulation, of an olive oil mill. *Paper presented at the 25th European Modeling and Simulation Symposium, EMSS 2013*.
- Mars, W. V., & Fatemi, A. (2002). A literature survey on fatigue analysis approaches for rubber. *International Journal of Fatigue*, 24(9), 949-961.
- Marlow, R. S. (2003). A general first-invariant hyperelastic constitutive model. *Constitutive Models for Rubber*, 157-160.
- Mooney, M. (1940). A theory of large elastic deformation. *Journal of applied physics*, 11(9), 582-592.
- Ogden, R. W. (1972, February). Large deformation isotropic elasticity-on the correlation of theory and experiment for incompressible rubberlike solids. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (Vol. 326, No. 1567, pp. 565-584). The Royal Society.
- P. Consulting, *Global elastomeric polyolefins markets, technologies trends* (2014) 2014–2020.
- Rivlin, R. S., & Saunders, D. W. (1951). Large elastic deformations of isotropic materials. VII. Experiments on the deformation of rubber. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 243(865), 251-288.
- Robert A. Malloy. (1994). *Plastic part design for injection molding: an introduction*. Hanser Publishers.
- Ruiz Argáiz, I., Jiménez Macías, E., Blanco Fernández, J., & Pérez de la Parte, M. (2008). Design and simulation of production of injection pieces in automobile industry. Paper presented at the *Proceedings -EMS 2008, European Modelling Symposium, 2nd UKSim European Symposium on Computer Modelling and Simulation*.
- Štrumberger, N., Gospočić, A., & Bartulić, Č. (2012). Polymeric Materials in Automobiles. *PROMET-Traffic&Transportation*, 17(3), 149-160.
- Treloar, L. R. G. (1943). The elasticity of a network of long-chain molecules—II. *Transactions of the Faraday Society*, 39, 241-246.

AN INTELLIGENT, INTERACTIVE AND INTEROPERABLE PLATFORM FOR THE 21ST CENTURY MUSEUMS

Francesco Longo^(a), Letizia Nicoletti^(b), Antonio Padovano^(c), Adriano Solis^(d), Arianna Monzon^(e)

^{(a),(c)} DIMEG, University of Calabria, Italy

^{(b),(e)} Cal-Tek Srl, Italy

^(d) School of Administrative Studies, York University, Canada

^(a) f.longo@unical.it, ^(b) l.nicoletti@cal-tek.eu, ^(c) antonio.padovano@unical.it, ^(d) sasdir@yorku.ca, ^(e) arianna.monzon@cal-tek.eu

ABSTRACT

This article presents a completely innovative comprehensive social platform that includes Intelligent, Interactive and Interoperable applications that will bring beyond the way cultural digital resources are created, disseminated, preserved and re-used based on new ICT solutions. The I3-CPM (Intelligent, Interactive and Interoperable Comprehensive social Platform for the 21st century Museums) is intended to demonstrate the museum's capability to foster an active participation of large number of stakeholders to create a European Cultural Digital Space where geographic distances and time are meaningless (any time and any place approach). To this end, the I3-CPM has been designed to be interoperable, modular, scalable and portable by interconnecting History, Archaeology, and Social Sciences with cutting-edge multi-disciplinary technologies, such as Virtual Environments, Augmented Reality, Holographic Displays, Intelligent Agents, Mobile technologies, Immersive Caves, and Gestures Sensors. Digital cultural contents and information are received by users according to a new transmedia, user-driven, adaptive, interactive, intelligent and meaningful experience where History becomes a guide for life.

Keywords: Digital Museums, Virtual Environment, Augmented Reality, Intelligent Cultural Heritage Fruition

1. INTRODUCTION

ICT changed the way cultural digital resources are created, disseminated, preserved and (re)used. Art galleries, archaeological sites and exhibitions all around the world have started investing efforts and money in extending their borders: consider for example the swift success of the museums' mobile applications available on the main app marketplaces. However, mobile applications as well as traditional "brochure" websites are only meant to help the user navigating the information, but do not deliver any educational content about the exhibit. Digital technologies have been introduced in the cultural heritage world by Malraux in 1947 with the concept of "Virtual Museum". Starting from then, a number of different types of "museums"

were identified (ICOM News, 3, 2004) in order to categorize different level of user's engagement. Several heterogeneous methodologies and technologies, based on virtual (VR) and augmented reality (AR), have been used in the latest years to make the user experience more engaging (see Hall et al., 2001), as showed by a sound state of the art proposed by Styliani et al. (2009). Unlike traditional websites or applications, VR and AR are able to convey information not in a mediated way (for example by using language) but intuitively and naturally by using sensorial feedback (images, sounds, etc.), thus conveying cultural content in an effective way. VR is capable to increase drastically the slope of the learning curve because cultural contents and information are absorbed spontaneously (Osberg, 1995) by the virtual visitors who are engaged in an "experience of learning" or "learning for fun" (Packer, 2006).

Significant immersive cultural experiences and projects have been proposed and developed in the field of advanced fruition of cultural heritage. Examples are the ReaCTor at the Foundation of Hellenic World (Gaitatzes et al., 2000), the ARCO framework (White et al., 2004), the PEACH project (Stock et al., 2007), the MINERVA project (Amigoni e Schiaffonati, 2009), the V-MUST (Virtual Museum Transnational Network, www.v-must.net) project, the 3D-COFORM project (Arnold, 2009), just to name a few. In order to enhance the cultural heritage experience, all these projects seek to take advantage of synergies among different technologies (multi-agent architecture, audio-visual technologies, 3D visualization, etc.). The most common solutions range from simple interactive screens to full-immersive systems that may include projection systems, 3D stereoscopic glasses or CAVE-like environments (Gaitatzes et al., 2001; Geller, 2006; Sooi et al., 2016). Although the preservation and fruition of cultural heritage through modelling techniques, graphics and virtual/augmented reality is widespread and has been widely documented (Addison, 2000), researchers and research communities are today concerned about how the content should be made more appealing. For example, as shown by Anderson et al. (2010), the use of SGs in the cultural heritage for teaching/learning

historical contents has been instead little considered. According to the taxonomy adopted by Mortara et al. (2014), SGs can have various purposes including cultural awareness, historical reconstruction, heritage awareness and can vary according to the learning objectives, genre (e.g. adventure, simulation) and application context.

Indeed, apart from the tools and systems that can be used thanks to the most recent technological advances, it is needed to design an innovative, intelligent and interactive cultural heritage experience based on the concept of experiential learning and “edutaining”.

To this end, the innovative aspect of this research work was the design and preliminary implementation of an Intelligent, Interactive and Interoperable Comprehensive social Platform for the 21st century Museums (I3-CPM) which leverages and benefits of cutting-edge multi-disciplinary technologies, such as Virtual Environments, Augmented Reality, Holographic Displays, Intelligent Agents, Mobile technologies, Immersive Caves, and Gestures Sensors with History, Archaeology, and Social Sciences. It is worth mentioning that some of these technologies have been already successfully integrated in other application areas (e.g. Industry and Defense) as shown by articles already published (Bruzzone et al., 2010; Bruzzone et al., 2013). The I3-CPM aims at creating a European Cultural Digital Space where geographic distances and time are meaningless (any time and any place approach).

Digital cultural contents and information are received by users according to a new transmedia, user-driven, adaptive, interactive, intelligent and meaningful experience – where History becomes a guide for life – supported by an advanced Personal Assistant and Cultural Navigator (called PHRONESIS) able to intelligently interact with the users by questions/answers mechanisms. This paper describes the main components of the I3-CPM architecture in the next section and the first results of the prototype developed (Section 3). The ultimate goal is to show how the project objectives have been addressed through:

- The creation of a Virtual Interoperable Environment (configured as a serious game) where multiple museums and multiple users may engage with multiple digital cultural resources at any time and in any place, being at home, in the museum site, off-line or connected on the web.
- The creation of an Augmented Interoperable Museums where multiple users can experience augmented reality contents and receive additional pieces of information by using their mobile technologies on the museums sites.
- The creation of an Immersive Interoperable Museum to bring beyond the immersive virtual reality experience and provide the users with the possibility to fully and intelligently interact with the virtual environments by means of helmets, immersive caves, gestures sensors and touch

technologies.

- The creation of the Holographic Interoperable Museums, to be deployed in museums sites that aims at creating an advanced kiosk solution.
- The creation of an “open” social platform that can be joined also by other projects in the field of cultural heritage, such as Big and Open Data initiatives. To guarantee the interoperability with other initiatives, a gateway called NON I3-CPM Applications Bridge will be designed, conceptualized and developed. The NON I3-CPM Applications Bridge will be a way to allow a gradual and continuous stakeholders involvement within the I3-CPM platform even after the completion of the project.

2. THE I3-CPM ARCHITECTURE: A MODULAR FRAMEWORK FOR MULTI-SIDED CULTURAL HERITAGE FRUITION

The I3-CPM (Intelligent, Interactive and Interoperable Comprehensive social Platform for the 21st century Museums) is an innovative and comprehensive platform that aims at bringing beyond the way cultural digital resources are created, disseminated, preserved, re-used and re-purposed. The main ideas, models and assumptions needed to create the I3-CPM platform are summarized in the architecture overview depicted in Figure 1, which illustrates the basic features of the I3-CPM platform and shows the most important segments underpinning the I3-CPM platform.

The I3-CPM Architecture (see figure 1) includes six main parts, a description of which is provided in the next sub-sections:

- The Virtual Interoperable Museums (VIM);
- The Augmented Interoperable Museums (AIM);
- The Immersive Interoperable Museums (IIM);
- The Holographic Interoperable Museums (HIM);
- The NON I3-CPM Applications Bridge (NAB);
- A Common Infrastructure (CI) underlying the whole system.

The *Virtual Interoperable Museums (VIM)* is one of the I3-CPM platform modules which is able to interoperate with the other parts through the I3-CPM Common Infrastructure as in Figure 1. Multiple Museums (*museum 1, museum 2..., museum n*) can create and share their cultural digital resources and knowledge contents as part of a common 3D Virtual Environment (3D VE) with the aim of recreating a European Cultural Space.

The *Augmented Interoperable Museums (AIM)* is an augmented reality based tool supported by the PHRONESIS system for storytelling. The AIM can be accessed by multiple users that, by using their own mobile technologies (or mobile technologies provided by museums), e.g. smartphone, tablet, glasses, etc., can receive augmented contents while interacting with the PHRONESIS Personal Assistant and Cultural Navigator.

What we said so far for the VIM module can be applied to the *Immersive Interoperable Museums (IIM)*. Indeed,

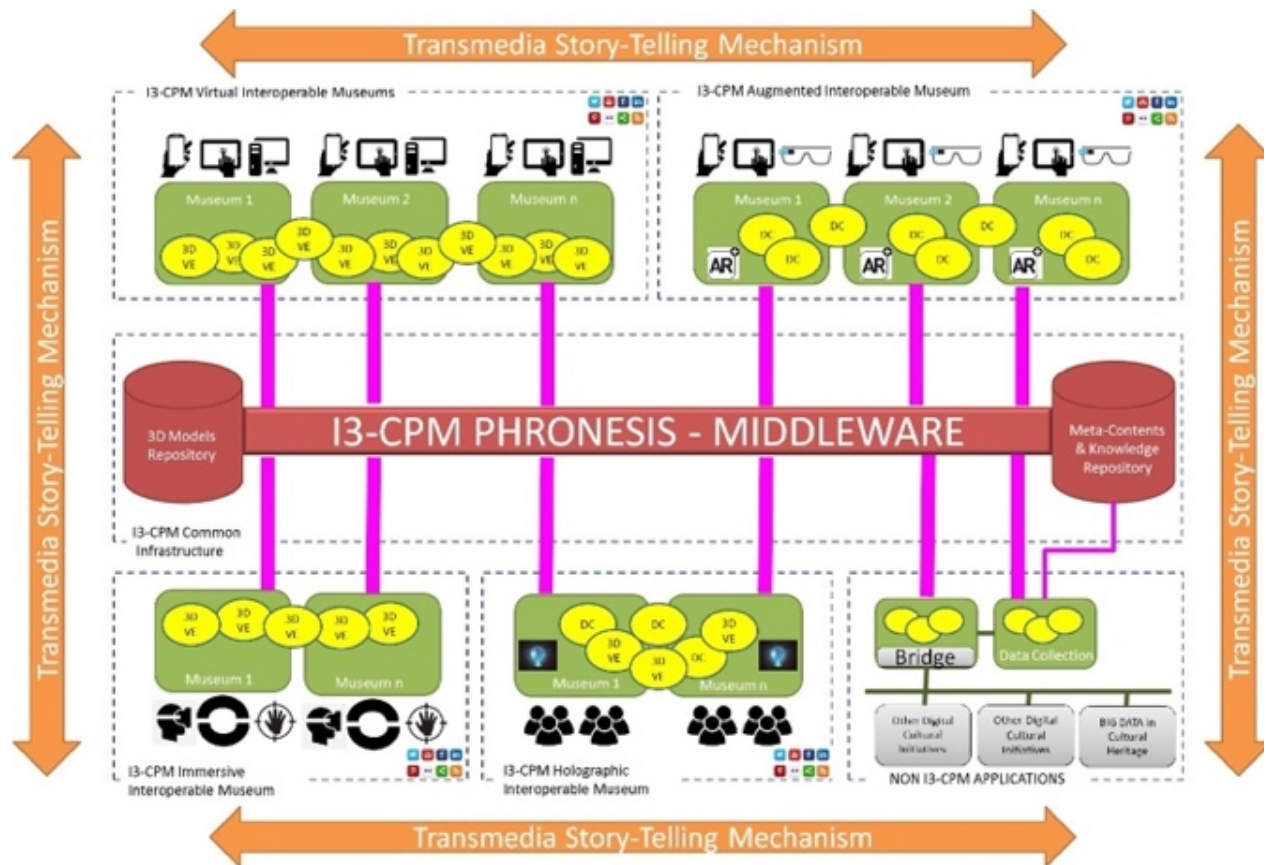


Figure 1 - I3-CPM Architecture Overview

VM are technologically demanding in terms of Virtual Reality and 3D experiences. While the VIM's 3D VE is experienced by using mobile technologies, within the IIM the experience becomes fully immersive thanks to virtual caves, helmets and gestures technologies.

The main idea of the *Holographic Interoperable Museums (HIM)* is to integrate a holographic display system able to recreate a human hologram (as well as other holographic images) with the PHRONESIS Personal Assistant and Knowledge Navigator. The result of this integration will be embedded in a kiosk solution to be deployed in the museums site to create a quasi-real museums assistant able to interact with the user, provide the user with suggestions according to his/her preferences, show multiple digital contents according to user requests.

The *NON I3-CPM Applications Bridge* allows the integration of I3-CPM applications with NON I3-CPM applications therefore extending the interoperability of the I3-CPM platform to other projects and initiatives in the field of cultural digital contents and virtual museums, such as Open Big Data Initiatives in the cultural heritage sector.

Finally, the *I3-CPM Common Infrastructure (CI)* is made up by the Middleware, the PHRONESIS system, the 3D Models Repository and the Meta-Content and Knowledge Repository.

3. I3-CPM EXPECTED IMPACTS

In terms of impact and benefits for the General Public, the I3-CPM will remove the barriers between the general public and digital cultural resources by implementing a Virtual Interoperable Environment that will allow the general public to access many digital cultural repositories as a semantically linked and interactively accessed continuum of data. To this end, I3-CPM will provide an effective content access and retrieval that will be intuitive and easily used by the general public, as well as personalized to individual users. Furthermore I3-CPM Virtual Interoperable Museums will allow both customizable single user and group experiences, online digital encounters, access from social media tools and will be portable and scalable according to the end-users needs. I3-CPM promotes familiarization of younger generations with cultural heritage through the use of media sharing and social networking technologies and a gamification context as well as it promotes the notion of shared cultural heritage thus creating a sense of community where people of various nationalities visit the same exhibition and/or share experiences. This approach also promotes the reduction of the level of social exclusion of the elderly, people with disabilities and those financially stressed by facilitating the access to content and places of cultural interest that would otherwise be

impossible to be accessed/visited by elders due to mobility or financial constraints.

In terms of benefits for educators, digital cultural resources allow educators to present material to their students in familiar formats that can be used to engage and involve the students, making culture come alive. Users are engaged in a specific self-reinforced context in which they are motivated to learn and educated through a serious game based mechanisms.

There are also benefits for researchers such as the creation of new personalizing algorithms and mechanisms for intelligent user-machine interaction such as the combination of Augmented Reality with PHRONESIS, available through mobile platforms, the development of new ways to use virtual helmets (or immersive caves), gestures sensors (e.g. MYO) and touch technologies and the creation of new solutions putting together hologram technologies with PHRONESIS.

Finally, there are also benefits for museums, institutions, curators and for the creative industry secured by dedicated and customized access to the digital content repositories, which will allow users access only to approved contents. The interactive and always available nature of mobile apps will foster deeper engagement of users with the available digital cultural material. To this end, I3-CPM provides a high potential for access, popularizing and advertising of cultural heritage through its applications using Augmented Reality, Serious Games and Holograms with PHRONESIS. The functionalities enabled by the I3-CPM will promote new interactions between the institutions and their target audiences. For example, cultural tourists or student groups might decide to visit a museum after discovering items of interest through interactions with the digital content.

4. I3-CPM PROTOTYPE AND PILOT

In order to provide evidence on the relevance of the I3-CPM and to show how all the objectives can be achieved, a fully functional prototype of the platform is going to be developed as continuation and extension of the SG-ICT research project, Serious Games at Increased Impact on Culture and Tourism (Longo et al., 2014; Longo et al., 2015). The prototype will be developed with and for users in a living lab approach fulfilling the open innovation paradigm, with the contribution of final users and all the stakeholders, centered in user needs and directly focused on market application.

A generic virtual museum has been implemented in order to shift later in time the choice of a specific real museum to use as partner in the pilot. The VE can be enjoyed through a very simple workstation (depicted in Figure 2) made up by three large screens, which give the user a very basic feeling of “immersion”, and a touchscreen to interact with the environment.



Figure 2 – Generic Virtual Museum for the I3-CPM

The deployment of one prototype for each museum requires the utilization of all the technologies and hardware needed to let users experience all the platform functionalities (VIM, AIM, IIM and HIM). Examples of already available and tested technologies are the Oculus Rift and the HTC Vive as VR Helmets (see Figure 3), the MYO Gesture Control Armband for hands/arms motion Tracking and the Virtuix Omni.



Figure 3 – Testing the VE with VR Helmets in the IIM

An initial set-up of an advanced visualization system made up of eight large touchscreens and a high-quality surround-sound system has been designed as in Figure 4 to be employed both in VIM and IIM applications and installed in museums.

As far as the AIM is concerned, some initial prototypes of a mobile application have been developed with the aim to enjoy cultural content about a specific artefact by simply pointing at a two-dimensional visual code (or tag) placed next to the real artefact as represented in Figure 5. This way augmented content, accurate 3D models of the artefact (see Figure 6), audio and video content can be enjoyed through an interactive experience which benefit also of the PHRONESIS system to ask for more detailed information.



Figure 4 – Set-up of an advanced visualization system for VIM or IIM applications



Figure 5 - Tag recognition for augmented cultural content on AIM

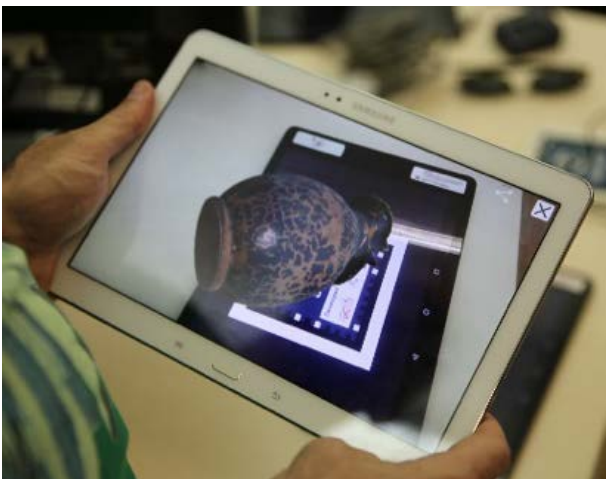


Figure 6 - Navigating a 3D model of an artefact on AIM

The prototype under development aims at providing a very intuitive and immediate environment for museums, institutions, galleries etc. that desire customizing the I3-CPM on their needs.

5. CONCLUSIONS

The ambition of this research work is to set the scene for a Cultural Digital Space that will allow the users to navigate a social platform where cultural digital contents and information are accessible to everyone any time and any place. This can be regarded as the 21st century museum concept that is expected to turn upside down the traditional museum visits. Promoting the cultural heritage through interactive and full-immersive experiences and setting new learning patterns by exploiting the informative and educational potentials of new methodologies and technologies (e.g. Serious Games) will bring beyond the way cultural digital resources are created, disseminated, preserved and re-used and will represent a “trait d’union” between past, present and future.

Since I3-CPM explores unconventional ways to deliver cultural knowledge contents (Virtual and Augmented Reality, Holography, Immersive technologies, Simulation, Serious Games, Knowledge based Systems, vocal interaction technologies, etc.), it is expected to provide users with new learning experience by:

- An interoperable and modular platform able to interconnect multiple technologies with the aim of creating Interoperable Museums where geographic and cultural distances are cut down and where physical and digital worlds are strictly linked each other.
- A scalable and portable platform that can be easily re-adapted to different types of collections (e.g. museums, archaeological sites, art galleries, etc.) and able to run using different types of technologies (Desktop Computers, Mobile devices, Caves, VR Helmets, Holograms).
- A platform able to encourage an active and gradual participation of large number of stakeholders also providing an easy way to connect with already existing and future platforms and Digital Heritage Initiatives.

Holographic visualization and the PHRONESIS system open new ways for the museums to think at the museums tour. The I3-CPM is indeed envisaged to be of benefit to General Public, Educators, Researchers, Memory Institutions, Curator Community, Creative Industry as summarized in Table 1.

ACKNOWLEDGMENTS

The research presented in this paper is part of an ongoing research project: “*SG-ICT: Serious Game at Increased Impact on culture and Tourism*” financed by the Italian Ministry of Education, University and Research under the Bando startup initiative. The main project partners include the University of Calabria and CAL-TEK SRL.

REFERENCES

- Addison A. (2000). Emerging trends in virtual heritage. IEEE Multimedia, Special Issue on Virtual Heritage, 7 (2): pp. 22–25.

- Amigoni, F., Schiaffonati, V. (2009). The Minerva System: A Step Toward Automatically Created Virtual Museums. *Applied Artificial Intelligence* 23(3): pp. 204-232.
- Anderson, E.F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., and de Freitas, S. (2010). Developing Serious Games for Cultural Heritage: A State-of-the-Art Review. *Virtual Reality* 14 (4): pp. 255-275.
- Arnold, D. (2009). 3D-COFORM: Tools and expertise for 3D collection formation. In *Proceedings of EVA*, pp. 94-99.
- Bruzzone A.G., Fancello, G., Fadda, P., Bocca E., D'Errico G., Massei M., (2010). Virtual world and biometrics as strongholds for the development of innovative port interoperable simulators for supporting both training and R&D. *International Journal of Simulation and Process Modelling*, Volume 6, Issue 1, Pages 89-102.
- Bruzzone, A.G. (2013). Intelligent agent-based simulation for supporting operational planning in country reconstruction . *International Journal of Simulation and Process Modelling*, 8 (2-3), pp. 145-159.
- Gaitatzes, A., Christopoulos, D., Roussou, M. (2001). Reviving the past: Cultural Heritage meets Virtual Reality. In: *Proceedings of Virtual Reality, Archeology, and Cultural Heritage (VAST)*: pp. 103-109.
- Geller, T. (2006). Interactive tabletop exhibits in museums and galleries. *IEEE Computer Graphics and Applications*. Volume 26, Issue 5, pp. 6-11.
- Hall T., Ciolfi L., Bannon L., Fraser M., Benford S., Bowers J., Greenhalgh C., Hellström S.-O., Izadi S., Schnädelbach H., Flintham M. (2001). The Visitor as Virtual Archaeologist: Using Mixed Reality Technology to Enhance Education and Social Interaction in the Museum. In: *Proceedings of VAST 2001: Virtual Reality, Archaeology and Cultural Heritage*, ACM SIGGRAPH, Glyfada, Greece, pp. 91-96.
- ICOM News, no. 3, 2004, available at: http://icom.museum/fileadmin/user_upload/pdf/ICOM_News/2004-3/ENG/p3_2004-3.pdf.
- Longo, F., Nicoletti, L., Vena, S., Padovano, A. (2014). Serious games at increased impact on culture and tourism *Proceedings of the 26th European Modeling and Simulation Symposium*, pp. 641-648.
- Longo, F., Nicoletti, L., Florio, G., (...), Bruno, L., Caputi, L. (2015). Inside virtual: A new app for interactive and intelligent cultural heritage fruition. *Proceedings of the 27th European Modeling and Simulation Symposium*, pp. 471-478.
- Malraux A., *La Musée imaginaire*, Gallimard, Paris, 1996 [orig. 1947].
- Mortara M., Catalano C.E., Bellotti F., Fiucci G., Houry-Panchetti M., Petridis P. (2014). Learning Cultural Heritage by Serious Games. *Journal of Cultural Heritage*.
- Osberg K.M. (1995). Virtual reality and education: where imagination and experience meet, *VR in the Schools* 1 (2), pp. 1-3.
- Packer J. (2006). Learning for fun: The unique contribution of educational leisure experiences, *Curator: The Museum Journal* 49 (3), pp. 329-344.
- Sooai, A.G., Sumpeno, S., Purnomo, M.H. (2016). User perception on 3D stereoscopic cultural heritage ancient collection. In *Proceedings of the 2nd International Human Computer Interaction and User Experience Conference*: pp. 112-119.
- Stock O., Zancanaro M., Busetta P., Callaway C., Krüger A., Kruppa M., Kuflik T., Not E., Rocchi C. (2007). Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User Modeling and User-Adapted Interaction*. Volume 17, Issue 3, July 2007, Pages 257-304.
- Styliani, S., Fotis, L., Kostas, K., Petros, P. (2009). Virtual museums, a survey and some issues for consideration. *Journal of Cultural Heritage*. Volume 10, Issue 4, pp. 520-528.
- White M., Mourkoussis N., Darcy J., Petridis P., Liarokapis F., Lister P., Walczak K., Wojciechowski R., Cellary W., Chmielewski J., Stawniak M., Wiza W., Patel M., Stevenson J., Manley J., Giorgini F., Sayd P., Gaspard F. (2004). ARCO - An architecture for digitization, management and presentation of virtual exhibitions. *Proceedings of Computer Graphics International Conference, CGI*, pp. 622-625.

AUTHORS BIOGRAPHY

Francesco Longo is currently Assistant Professor at the Department of Mechanical, Energy and Management Engineering where he teaches "Industrial Plants and Facilities" and "Industrial Plant Management" for Master Degree Students in Mechanical and Management Engineering. His research interests include Modeling & Simulation for production systems design and supply chain management. He is Director of the Modeling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES), a research laboratory operating at the University of Calabria, which is member of the MS&Net (McLeod Modeling & Simulation Network). Starting from 2016, he is also serving as CEO of CAL-TEK Srl, a Spin-off company of University of Calabria. He has published more than 150 scientific papers on international conferences and journals participating as speaker and chairman to different international conferences. He actively cooperates with many research institutions all over the world (DIPTM, University of Genoa, Kennedy Space Center, NASA, Rutgers University among others). He is editor of several journals and journal's special issues and over the last years, he has strongly supported the organization of conferences in the area of simulation applied to Industry and supply chain as member of the

International Program Committees. Starting from 2003 he worked on private research projects involving manufacturing and logistics systems operating in Italy. He has acquired experiences in consulting for Business Process Re-engineering and Logistics in different areas.

Letizia Nicoletti was CEO of Cal-tek Srl from 2012 to 2014 where she is currently Senior Manager. She obtained her Master Degree in Management Engineering, Summa cum Laude as well as her PhD in Mechanical Engineering from University of Calabria, Italy. She has followed as Scientific Responsible many research projects in different areas including logistics and distribution, Defense and Cultural Heritage. She was also the main Responsible for all the services provided by CAL-TEK to NATO STO CMRE.

She carried out several work experiences travelling in Europe and United States working as CEO of Cal-tek Srl and attending Conferences and Workshops in the M&S area.

Since 2009 she has acquired a strong experiences in software development and Modeling & Simulation (M&S) including High Level Architecture (HLA) and Distributed Real Time Simulation. Her skills include requirements definition and development, conceptual modelling, model simplification, data collection and analysis, representing unpredictable variability and selecting statistical distributions, models and software coding, experimentation, verification and validation, serious games development, visualization, simulation software in different areas, to name a few, defense, healthcare, complex systems and tourism.

She is author of more than 50 scientific papers and technical reports. In 2011, she successfully completed the NATO CAX Operator Certification Course.

Antonio Padovano is PhD student at the University of Calabria where he received his degree in Management Engineering summa cum laude.

His work focuses on the development of interoperable discrete event and agent-based simulations for decision support in complex systems in the area of Industry, Logistics and Defense with a particular attention to Human Modeling & Simulation to reproduce the main functions of human behavior in order to assess the impact of human choices, decisions and actions. He participated to different research projects both in cooperation with MSC-LES and CAL-TEK Srl (e.g. T-ESEDRAS, SMARTDISTR, SG-ICT).

He participated to different international conferences (EMSS, MAS, HMS, China-Italy Science, Technology and Innovation Week, etc) and he also served as Advisor for University of Calabria students at ICAMES 2016 (Istanbul, Turkey).

He was visiting student at the Business School at Rutgers University (The State University of New Jersey, USA) between June and September 2015 working in the field of simulation modeling and development aimed at analyzing and improving the performance of supply chains and industrial systems, in

particular, container terminals as well as implementing a software tool whose main objective is to support decision making processes within complex industrial and logistic systems in order to evaluate green management practices in terms of both environmental impact.

He teaches as tutor for undergraduate and graduate students enrolled in the Degree in Mechanical Engineering at the University of Calabria in the classes "Industrial Systems" and "Industrial and mechanical systems".

Adriano Solis teaches and conducts research in the areas of operations/logistics/supply chain management and management science. More recently he also started to extend his research areas to other topics including the application of simulation based solutions to different domains (including cultural heritage). He served for different years as General and Program Chair of the International Conference on Applied Modeling & Simulation therefore providing a strong contribution to the organization of the I3M multiconference. He is currently Director of the School of Administrative Studies at York University, Canada

Arianna Monzon is currently doing researches in the area of virtual environments for mobile applications with particular attention to Cultural Heritage area. She worked as programmer in CAL-TEK srl where she firstly acquired experience in the development of serious games by using graphic and physics engines and where she was also involved in the SG-ICT project.

AUGMENTED REALITY AND MOBILE TECHNOLOGIES FOR MAINTENANCE, SECURITY AND OPERATIONS IN INDUSTRIAL FACILITIES

Agostino Bruzzone^(a), Francesco Longo^(b), Letizia Nicoletti^(c), Marco Vetrano^(d), Luigi Bruno^(e), Alessandro Chiurco^(f), Caterina Fusto^(g), Giuseppe Vignali^(h)

^(a) DIME University of Genova, Via Opera Pia 15, Genova, Italy

^{(b),(e),(f)} DIMEG, University of Calabria, Via Ponte Pietro Bucci, Cube 45C, Arcavacata, Rende, Italy

^{(c),(d),(g)} CAL-TEK SRL, Via Spagna 240/242, Rende, Italy

^(h) Department of Industrial Engineering, University of Parma, viale G.P.Usberti 181/A, Parma, Italy

^(a) agostino@itim.unige.it, ^(b) f.longo@unical.it, ^(c) l.nicoletti@cal-tek.eu

^(d) m.vetrano@cal-tek.eu, ^(e) l.bruno@msc-les.org, ^(f) a.chiurco@unical.it, ^(g) c.fusto@cal-tek.eu,

^(h) giuseppe.vignali@unipr.it

ABSTRACT

Safety issues are a crucial aspect in industrial plants management due to regulatory, operational as well as strategic aspects. It calls for tools that could be easily embedded into everyday practices and able to combine complex methodologies with high usability requirements. In this perspective, the proposed research work is focused on the design and development of a practical solution able to integrate augmented reality and wearable technologies for operators' support in complex man-machine interactions. After establishing both functional and non-functional requirements, a structured design strategy has been adopted. The main outcomes of the aforementioned strategy include a multi-layered modular solution (developed at MSC-LES lab of University of Calabria) whose potentials are investigated through two different case studies and technical equipments.

Keywords: safety, security, maintenance, industrial plants, augmented reality, mobile technologies, wearable systems

1. INTRODUCTION

Safety and security issues are one of the major concerns in industrial plants. Due to the greater public exposure and to the increasing attention of current regulations, plant managers are concerned with workers' health and safety more than ever before. As a matter of facts, workplace accidents are likely to receive a great deal of media coverage as well as to activate legal investigations into companies' liabilities. Both ways, irrespective of actual faults, the reputation of the company involved may be damaged and misjudgment/misrepresentation could seriously damage the brand image. These are the reasons, among others, why the management is usually prone to adopt all those countermeasures that can contribute to risk factors minimization and control. In addition to safety and security, also maintenance operations play a critical role for the efficiency and productivity levels and they are often connected to safety and security issues.

In this framework, the main research contributions already available in literature fall under the following categories: models and methodologies, methods and tools. What has been done in terms of models and methodologies include mainly risk estimation (Okabe and Ohtani, 2009; Ahmad et al., 2014; Sengupta et al. 2016) and risk analysis (Tixier et al., 2002; Reniers, 2009). Besides, a state of the art review on risk assessment methodologies and analysis tools is proposed by Khan et al. (2015) while uncertainty in risk assessment for high-consequence technologies is treated in Zio and Aven (2013).

As for methods, the main effort is toward practical approaches for risk reduction taking into account even economic and regulatory aspects (Segawa et al., 2016). Basic requirements for a successful risk reduction strategy can be found in Summers (2009) while process safety management has been covered by Knegtering and Pasman (2009) and critical challenges in implementing safety programs are addressed by Qi et al. (2012) without neglecting the capability to recover from disturbances (Dinh et al. 2012).

In terms of tools, it is worth mentioning the contribution provided by Shariff et al. (2006) where risk assessment is integrated into process design utilities.

This quick and short analysis of the literature shows that there is a gap in the development of operational tools. To bridge the gap, this research work proposes a new approach where augmented reality and wearable smart technologies are seamlessly integrated for safety & security enhancement as well as to support maintenance operations. The basic idea is to support operators in complex, real time man-machine interactions. In this perspective augmented reality can provide visual, self-explanatory information on how to execute a specific task/procedure as well as demonstrate the best interaction patterns.

However, it goes with the need to preserve operational efficiency without hindering operators' workability. For this reason, the proposed approach involves smart wearable technologies aimed at ensuring great flexibility in terms of freedom of movement. For instance, using smart glasses operators can have their

hands free for executing their tasks while they access the system functionalities. Therefore, the many possibilities offered from wearable technologies have been investigated to detect the most suitable system configurations and to ensure a successful deployment at operational levels.

The reminder of the paper is organized as follows: section 2 illustrates the system functional requirements, section 3 presents the design approach, section 4 gives an application example and lastly conclusions are drawn.

2. SYSTEM OVERVIEW AND REQUIREMENTS

The system is built upon the need to ensure the following capabilities:

- Provide operators with real time feedbacks and augmented reality contents on task/procedures execution so as to minimize the risk of accidents and support training.
- Monitor men-machine interactions with augmented reality for safety enhancement.

Other non-functional requirements of the proposed system include: usability, interoperability with legacy systems, manageability, and regulatory compliance.

To fulfill such requirements the architecture depicted in figure 1 has been designed.

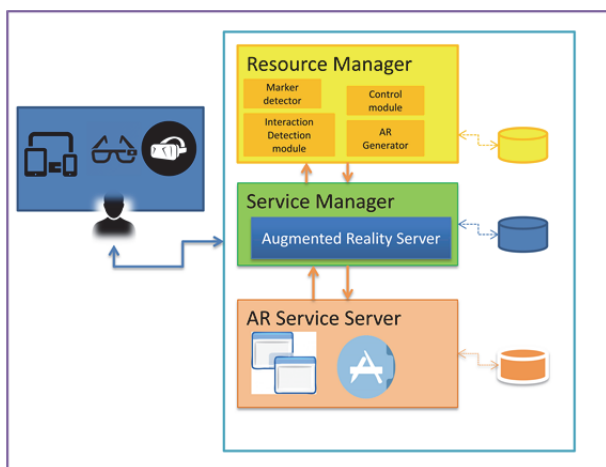


Figure 1: Architecture overview

Basically, this architecture is scalable and modular in nature since the underpinning principle is the separation of the service infrastructure from the contents base. The service infrastructure includes the service manager implementing the interactions with the Augmented reality server and the AR Service Server that fulfills users' requests sent via mobile or wearable technologies.

On the other side, the Resource manager is responsible for contents and metadata.

Furthermore, considering that the tool is envisaged for deployment at operational level, a user-centered design approach has been pursued. Bearing in mind that end-users may not have advanced technical skills, usability issues have been widely explored.

In particular, along the development cycle, usability tests have been executed over the main interaction components to detect and fix usability problems. From a conceptual point of view the architecture depicted in figure results in the following functionalities for its intended users:

- let operators acquire advanced contents based on augmented reality about man-machine interaction procedures compliant with safety standards and principles.
- preliminary training for tasks with high risk factors.
- support operators providing information that is usually not available in the workplace (i.e. machine productivity, expected maintenance operations).
- send warning messages about the outcomes of improper operations (i.e. what happens if a maintenance operation is not performed, if the operators fails, etc).

3. SYSTEM DESIGN AND IMPLEMENTATION

To conceptualize the system requirements, for analysis, design and implementation purposes, UML diagrams have been drawn up.

As a matter of facts UML diagrams allow depicting both the static structure of the system as well as its dynamical behavior over time. Structure diagrams have supported the identification of the system components and relations at different levels of abstraction. In particular the structure diagrams that have been drawn up include the class diagram and the model diagram depicted in figures 2 and 3 respectively.

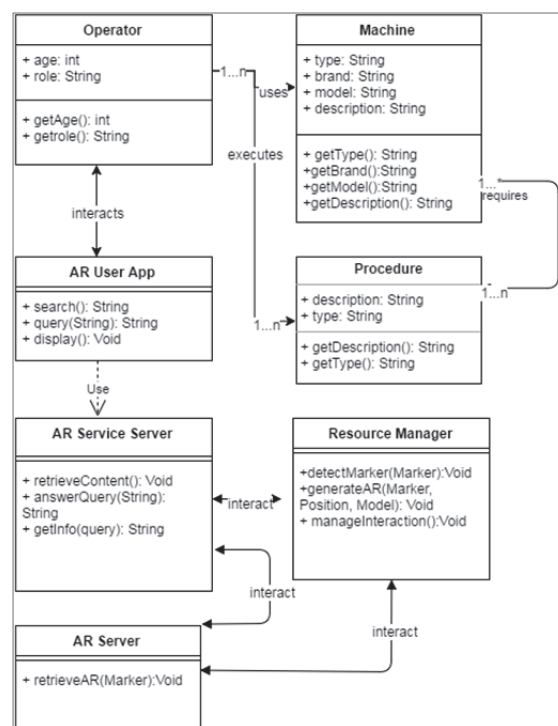


Figure 2: Class Diagram

The class diagram highlights the structure of the system in terms of classes, interfaces and related constraints, features and relationships.

Therefore it shows and describes the application domain of the system: operators with specific roles are called to interact with one or more machines executing one or more procedures; in doing so they can be supported by the system that is required to provide AR contents as well materials about machines and procedures.

To this end, the system front-end application interacts with an AR Service Server that activates a set of methods needed to collect/retrieve contents. The AR Service Server, in turn, interacts with the AR Server that implements AR algorithms and with the Resource Manager that is responsible for contents (data, metadata, models, multimedia, etc) handling. Furthermore, a complementary view of the system is given in the model diagram that is quite useful to give new insights about architectural, logical and behavioral aspects. The model diagram related to the system is represented in figure 3.

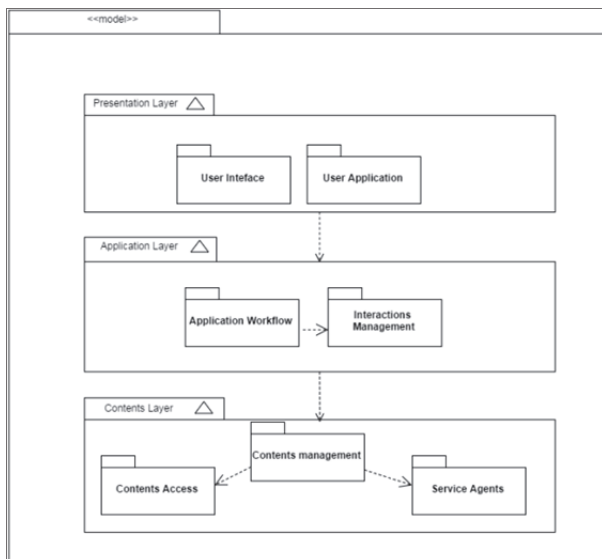


Figure 3: Model Diagram

Here, the overall model has been broken down into three main layers. The presentation layer includes interfaces and applications on the user side and is devoted to ensure that the contents passing through are in the appropriate form for the recipient providing also interaction mechanisms (i.e. access to video, 3d representations of a particular machine, documents, etc). The application layer implements the core functionalities of the system and the logics in order to met both functional and non-functional requirements. The contents layer, instead, provides access to all those contents (information, multimedia, geometric models, metadata etc) that are needed to let the system deliver meaningful information to its intended users. As mentioned before, the system evolution over time is represented by behavioral models, namely the use case diagram, the state machine diagram and the sequence diagram. As shown in figure 4, the use case diagram

highlights what the system is supposed to do (requirements) and what the system can do (functionalities).

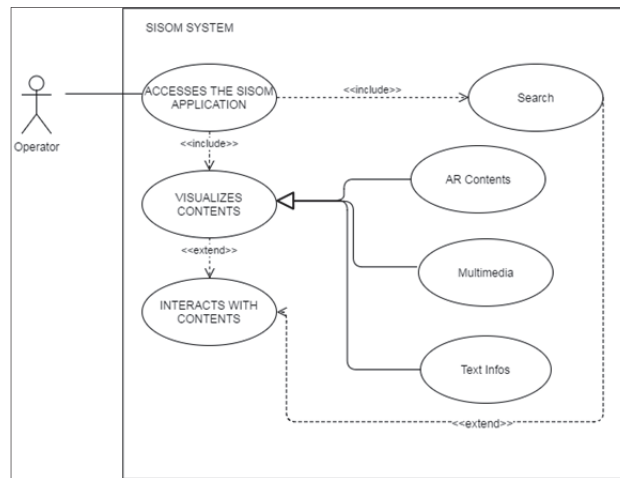


Figure 4: Use case diagram

In particular, the set of actions that the operator can do while accessing the system front-end application include: contents visualization (i.e. pointing his device toward the machine he is supposed to interact with) and/or search among the available contents.

After contents are displayed the operator can freely interact with such contents through his own device. Contents may be information, augmented contents, multimedia or a combination of the aforementioned content types.

Systems transitions and states, instead, are considered within the state machine diagram (Figure 5) that is particularly useful to understand how interaction patterns can evolve. Since transitions and states are mainly due to the user choices it is possible to ascertain that the system is event-driven in nature.

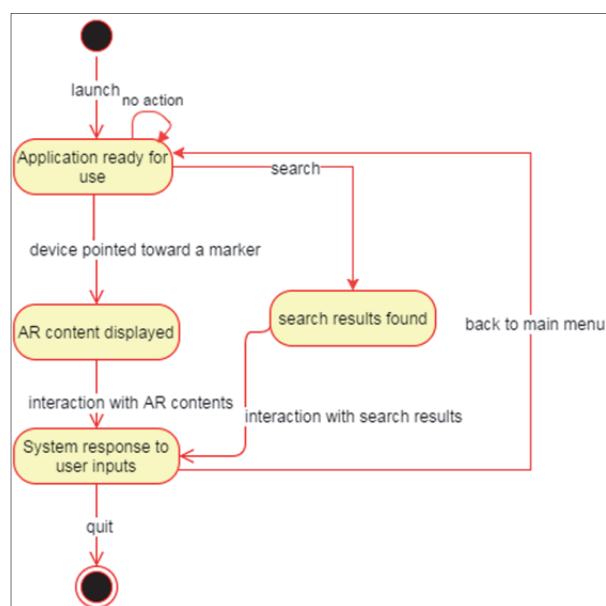


Figure 5: State machine diagram

As for messages exchange, the sequence diagram has been drawn. It is a useful representation of how the main system objects interact as well as of the communication exchange showing that it is mainly synchronous in nature.

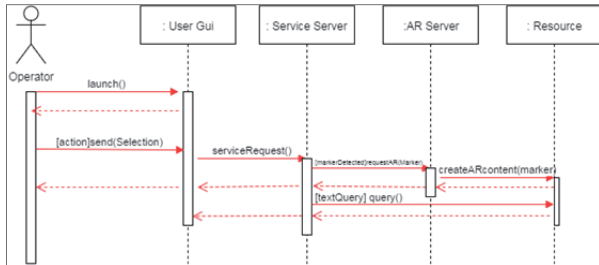


Figure 6: Sequence Diagram

Therefore the system design has been carried out according to a structured approach based on well-organized elements of solution adopting a “divide and conquer” strategy where the overall solution puts together multiple solution modules communicating each other.

4. APPLICATION EXAMPLES

The MSC-LES lab (University of Calabria) lab has a long experience in developing system and solutions based on 3D Virtual Environment and Simulation in different areas such as Industry and Logistics (Longo 2013; Longo et al. 2015). Therefore the step ahead in this project was the integration of new methodologies together with mobile and wearable technologies.

To test and show the potentials of the proposed solution the system has been deployed in two different use cases: one is about the use of a control panel and the other is about the installation of an hydraulic pump. For each use case, different fruition experience have been investigated. Considering the control panel use case, regardless of the device used by the operator, the front-end application is able to display augmented reality contents superimposing different levels of information.

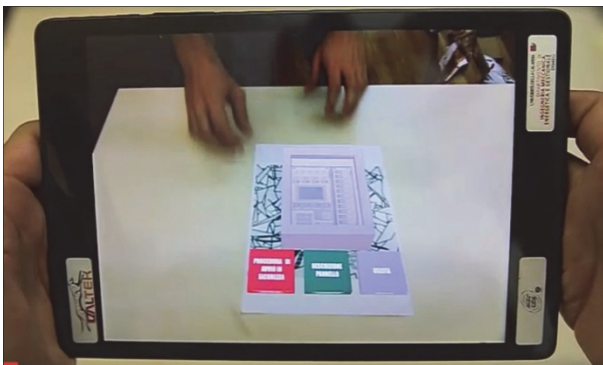


Figure 7: Control Panel use case

A first example is shown in figure 7, where augmented reality contents are displayed through a tablet over a

marker and include a 3D representation of the control panel as well as three buttons giving access to :

- The security procedure for operations starting;
- The description of the panel commands;
- The function to go back to the application main menu.

It is worth noticing that the same functionalities can be also accessed using different technologies.

As a matter of facts, instead of using a tablet, the operator can get access to the same functionalities using a headset (i.e the Samsung Gear VR) for contents display and a gesture control system like Myo Armbands for interaction purposes such as pressing the button to display the control panel commands description (as shown in figure 8).



Figure 8: Solution deployed through headset and armbands

In addition, another solution for the system fruition, include an helmet with a monocular eyewear and a gesture control system interfaced with an interactive whiteboard (Figure 9).

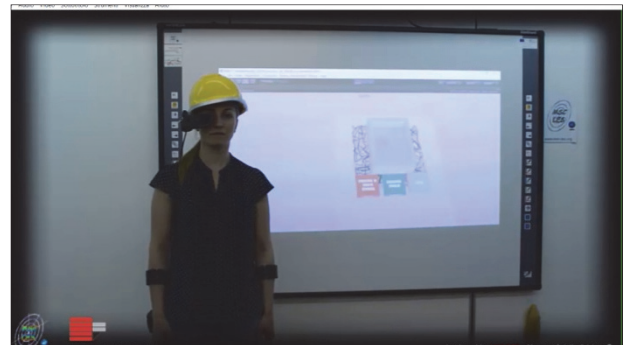


Figure 9: Solution deployed through helmet, monocular eyewear, armbands and interactive whiteboard

As shown in figure 9, AR contents are projected over the interactive whiteboard and interactions occur through a gesture recognition system implemented within armbands.

In this configuration the system is suitable for training purposes such as teaching security procedures to newcomers or let experienced operators get into new safer procedures.

For the same purpose, a lighter solution where the helmet is replaced by smart glasses has been set up (Figure 10).

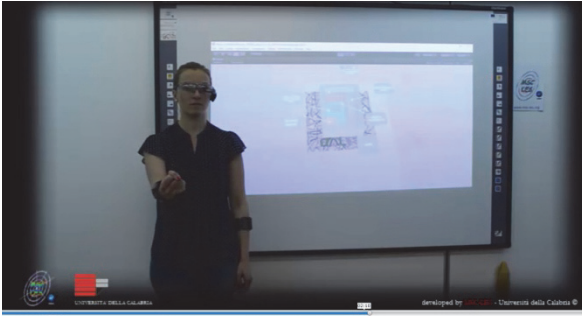


Figure 10: Solution deployed through smart glasses, armbands and interactive whiteboard

After all, to take the solution a step further, an additional effort has been made to search for an overall more convenient configuration in terms of deployment at operational level (see Figure 11). To this end, firstly a solution to use the panel as marker for its own augmented contents is under investigation and afterwards, an ergonomic analysis will be performed over the basic configurations described above to detect the most suitable one so as to preserve both the operator comfort as well as his operational efficiency.



Figure 11: Exemplw of deployment in place

As for the second use case, it is about the installation procedure of an hydraulic pump (Figure 12). The whole procedure has been broken down into four steps and each step has been faithfully recreated in the AR environment.



Figure 12: Hydraulic Pump use case

Upon interaction between the user and the AR commands, specific functions are activated to allow the virtual reality animation of the basic operations required to fulfill each step (Figure 12).



Figure 13: Detach/Attach the main pump body

Hence, using the app, the operator can view how to perform each step pressing the related AR button (see Figures 13 and 14).

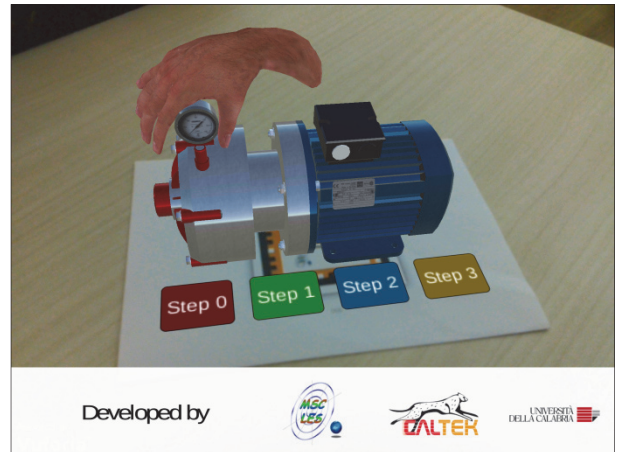


Figure 14: Detach/Attach the pump manometer

5. CONCLUSIONS

The proposed research discusses about the conceptualization and development of an advanced tool for safety enhancement in industrial plants. The proposed tool is meant to support operational processes with a particular focus on those tasks affected by high risk factors. To this end, special attention has been paid to the design process carried out according to a structured strategy. Structured and behavioral diagrams, have been drawn up to specify the main requirements and provide a multi-layered representation of the solution. Furthermore, after implementing the technical architecture, the solution deployment has been investigated in two application examples where

different configurations, in terms of technical equipment for contents fruitions, are considered. As a result, the proposed tool is able to deliver real-time augmented reality contents to support men-machine interactions as well as complex operational procedures. Two main use patterns have been envisaged:

- a daily tool for support at operational level
- a training tool for newcomers or for experienced operators that have to learn new safety procedures.

REFERENCES

- A. Sengupta, D. Bandyopadhyay, C.J. van Westen, A. van der Veen, An evaluation of risk assessment framework for industrial accidents in India, *Journal of Loss Prevention in the Process Industries*, Volume 41, May 2016, Pages 295-302, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2015.12.012>
- Akiko Segawa, Satoshi Yoshikawa, Takayuki Toyama, Hayao Nakanishi, Emi Kikuchi-Uehara, Masahiko Hirao, Hirokazu Sugiyama, Method for reducing environmental, health, and safety risks in active pharmaceutical ingredient manufacturing based on multiobjective evaluation, *Process Safety and Environmental Protection*, Available online 15 September 2016, ISSN 0957-5820, <http://dx.doi.org/10.1016/j.psep.2016.09.005>.
- Angela E. Summers, Continuous improvement and existing safety systems, *Journal of Loss Prevention in the Process Industries*, Volume 22, Issue 6, November 2009, Pages 685-688, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2009.08.005>.
- Azmi Mohd Shariff, Risza Rusli, Chan T. Leong, V.R. Radhakrishnan, Azizul Buang, Inherent safety tool for explosion consequences study, *Journal of Loss Prevention in the Process Industries*, Volume 19, Issue 5, September 2006, Pages 409-418, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2005.10.008>
- B. Knegeting, H.J. Pasma, Safety of the process industries in the 21st century: A changing need of process safety management for a changing industry, *Journal of Loss Prevention in the Process Industries*, Volume 22, Issue 2, March 2009, Pages 162-168, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2008.11.005>.
- Enrico Zio, Terje Aven, Industrial disasters: Extreme events, extremely rare. Some reflections on the treatment of uncertainties in the assessment of the associated risks, *Process Safety and Environmental Protection*, Volume 91, Issues 1–2, January–March 2013, Pages 31-45, ISSN 0957-5820, <http://dx.doi.org/10.1016/j.psep.2012.01.004>.
- Faisal Khan, Samith Rathnayaka, Salim Ahmed, Methods and models in process safety and risk management: Past, present and future, *Process Safety and Environmental Protection*, Volume 98, November 2015, Pages 116-147, ISSN 0957-5820, <http://dx.doi.org/10.1016/j.psep.2015.07.005>.
- Genserik Reniers, An optimizing hazard/risk analysis review planning (HARP) framework for complex chemical plants, *Journal of Loss Prevention in the Process Industries*, Volume 22, Issue 2, March 2009, Pages 133-139, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2008.10.005>.
- J. Tixier, G. Dusserre, O. Salvi, D. Gaston, Review of 62 risk analysis methodologies of industrial plants, *Journal of Loss Prevention in the Process Industries*, Volume 15, Issue 4, July 2002, Pages 291-303, ISSN 0950-4230, [http://dx.doi.org/10.1016/S0950-4230\(02\)00008-6](http://dx.doi.org/10.1016/S0950-4230(02)00008-6).
- Linh T.T. Dinh, Hans Pasman, Xiaodan Gao, M. Sam Mannan, Resilience engineering of industrial processes: Principles and contributing factors, *Journal of Loss Prevention in the Process Industries*, Volume 25, Issue 2, March 2012, Pages 233-241, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2011.09.003>.
- Longo, F. (2013). On the short period production planning in industrial plants: A real case study. *International Journal of Simulation and Process Modelling*, 8 (1), pp. 17-28.
- Longo, F., Chiurco, A., Musmanno, R., Nicoletti, L. (2015). Operative and procedural cooperative training in marine ports. *Journal of Computational Science*, 10, pp. 97-107.
- Masaaki Okabe, Hideo Ohtani, Risk estimation for industrial safety in raw materials manufacturing, *Journal of Loss Prevention in the Process Industries*, Volume 22, Issue 2, March 2009, Pages 176-181, ISSN 0950-4230, <http://dx.doi.org/10.1016/j.jlp.2008.12.002>.
- Ruifeng Qi, Katherine P. Prem, Dedy Ng, Morshed A. Rana, Geunwoong Yun, M. Sam Mannan, Challenges and needs for process safety in the new millennium, *Process Safety and Environmental Protection*, Volume 90, Issue 2, March 2012, Pages 91-100, ISSN 0957-5820, <http://dx.doi.org/10.1016/j.psep.2011.08.002>.
- Syaza I. Ahmad, Haslenda Hashim, Mimi H. Hassim, Numerical Descriptive Inherent Safety Technique (NuDIST) for inherent safety assessment in petrochemical industry, *Process Safety and Environmental Protection*, Volume 92, Issue 5, September 2014, Pages 379-389, ISSN 0957-5820, <http://dx.doi.org/10.1016/j.psep.2014.03.009>.

VIRTUAL AND AUGMENTED REALITY AS ENABLERS FOR IMPROVING THE SERVICE ON DISTRIBUTED ASSETS

Agostino G. Bruzzone^(a), Marina Massei^(b), Gianluca Maglione^(c),

Matteo Agresta^(d), Giulio Franzinetti^(e), Antonio Padovano^(f)

^{(a),(b),(c)} DIME University of Genoa, Italy, www.itim.unige.it

^{(c),(d),(e)} Simulation Team, www.simulationteam.com

^(f) DIMEG, University of Calabria, Italy, www.msc-les.org

^(a) agostino@itim.unige.it, ^(b) massei@itim.unige.it, ^(c) maglione@simulationteam.com,

^(d) agresta@simulationteam.com, ^(e) giulio.franzinetti@simulationteam.com, ^(f) antonio.padovano@unical.it

ABSTRACT

The evolution of Augment and Virtual Reality is enabling new solutions. This paper addresses creation of applications to support service and maintenance of distributed systems. Indeed this approach could be applied to devices provided as service for industrial and individual use and could introduce new capabilities in terms of training for operators, control and remote service support. The paper presents a case study devoted to lead the introduction of these innovative solutions in industrial and health care system.

Keywords: Augmented Reality, Virtual Reality, Simulation, Maintenance, Service, Distributed Systems

1. INTRODUCTION

Different kind of equipment and systems are extensively outsourced in Industry as well as in Public Sector. One of major aims promoting the diffusion of this approach is related to the possibility to simplify the management and to concentrate on the core activities by outsourcing the other elements affecting the processes. In addition to this aspects, outsourcing services is expected to guarantee fixed costs and to keep quality under control from sponsor point of view . The basic idea is that the external companies taking care of the outsourcing become specialized on these service and evolve to critical masses able to provide the new services as a competitive costs; basic good examples are cleaning services, energy, gas, etc.. Indeed, even if this approach is currently growing also due to popular tendencies in enterprises, it is evident that the new technologies are enabling new opportunities for outsourcing being able to address and solve specific challenges (e.g. remote control). In particular, while dealing with outsourcing related to provide and operate (or guarantee operations) of distributed systems and equipment, the tracking and maintenance of these elements becomes a pretty crucial factor. In facts, the final customers are very interested to the reliability of the systems, but being distributed such activities could result challenging and could require specific approaches. In facts, these aspects were very well addressed in ICT sector, where most of all the new hardware are remotely monitored and specific updates and maintenance procedure are performed through distributed service (e.g. mobile phones, servers, etc.).

The developments of the IoT (Internet of Things) is further reinforcing this capability and provides new solutions to integrate distributed systems and components in networks; within these architecture it could be possible to develop innovative solutions for monitoring, tracking and maintenance. This paper addresses specifically the case of equipment, systems and components that are geographically distributed and require to conduct even physical maintenance and service. A good example is related to the reload process for gas tanks, or the check of n medical equipment deployed in patient houses or hospitals. In these cases it is not possible just to operate remotely the software: it is necessary to develop a local and responsive capability to provide the service on site. In addition, in these cases, the monitoring and tracking of the equipment is also very important to increase availability, reduce losses and also to have an updated situation awareness. Indeed, another not negligible aspect is that one related to robberies and misuse of the equipment that could lead to reduced expected life and increase in costs. Due to these aspects, the authors are evaluating the development of combined approach based on Augmented Reality (AR) and Virtual Reality (VR) to support training and remote supervision to operations and maintenance of distributed systems.

2. SIMULATION, VIRTUAL AND AUGMENTED REALITY FOR SERVICE

The use of simulation for supporting education program in Industrial Plant and Components is a very consolidated approach (Ferrington et al.1992; Mosca et. Al. 1995, Del Rio et al., 2013) and even the use of Virtual Reality has been extensively applied in several sectors for training (procedures and equipment for remote operators) (Pspotka 1995; Mosca et al. 1996, 1997; Moline 1997; Wilson et al 1997; Stone 2001, Longo et al., 2013).

Indeed, once innovative virtual frameworks are created, it becomes possible to use them also to improve maintenance and service in different ways, such as optimization of the service, logistics as well as preventive analysis on the context (De Sa et al. 1999; Bluemel et al.2003). Indeed the innovative use of Simulation to support service of distributed systems has been effectively applied to aerospace and energy industries (Bruzzone & Simeoni 2002; Haritos 2005).



Figure 1: Concept of Virtual Mock Up for an Educational and Training Aid for AUV Service

Augmented and Virtual Reality could support many aspects and the service and maintenance are among the most promising since the introduction of these techniques (Azuma 1997). Also the use of mobile technologies is very interesting when we move to address distributed systems; indeed the mobile solutions for simulation based training in external logistics have been demonstrated very successful in several contexts; in fact these achievements confirm the potential of mobile training concept (Bruzzone et al. 2004a, 2004b; Monahan et al. 2008; Ally 2009; Lee 2011). Indeed, the interactive Virtual Worlds are effectively enabling new opportunities in training applied to procedures and operations in many sectors (Bruzzone 2009; Raybourn 2014). The authors are currently considering to apply the use of AR/VR to health care for supporting equipment deployed in hospitals or in patient houses; in fact the applications in Health Care of Simulation and Virtual Reality are pretty popular since many years and demonstrated to be effective (Giribone et al. 1996; Zajtchuk Satava 1997; Gaba 2004). In this paper it is proposed to investigate the use of web technologies and cloud approach to support services in conjunction with AR, VR and M&S. For instance, applications to industrial plants and components based on web technologies and virtual reality are popular from almost two decades and are continuously evolving (Bruzzone 1999; Monahan et al. 2008). Therefore even if these concepts have been investigated since many years, the recent evolution in web technologies is enabling new opportunities in service and maintenance (Bruzzone et al. 1999; Vora et al. 2002). In fact, another potential in this sector is to provide the remote users with devices able to evaluate the status of the operator in order to optimize the effectiveness of training and of operation. In this sense, the use of innovative solutions based on the capability to capture physiological parameters remotely (e.g. EEG, muscular tone, cardio frequency) has a great potential in supporting operator and user supervision; from this point of view these researches are very consolidated (Orlansky 1994; Brookings 1996) and are leading to new solutions (De Crescenzo et al. 2011; Bruzzone et al. 2016a).

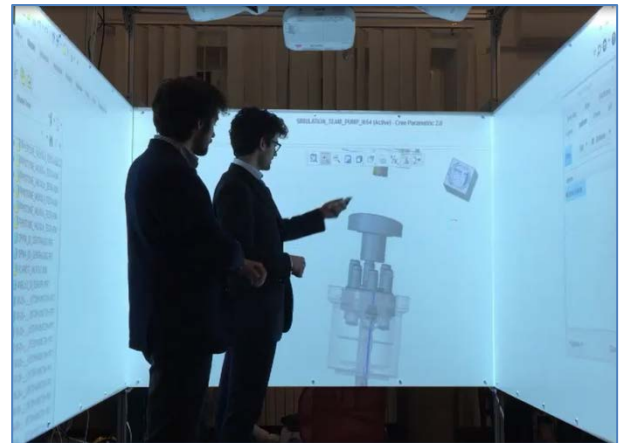


Figure 2: SPIDER: the Virtual Immersive Interoperable Interactive CAVE for Virtual Training and Engineering

The authors propose for this research the use of the innovative MS2G paradigm (Modeling, Interoperable Simulation & Serious Games) that integrates M&S (Modeling and Simulation) and Serious Games (Bruzzone et al. 2014). Indeed in our case the combined use of AR and VR strongly benefits of this new paradigm.

3. THE PROPOSED SOLUTION

The author studied in the past the potential to use AR/VR to support maintenance of complex systems; for instance studies were conducted on using these technologies on autonomous systems: for instance on AUV (Autonomous Underwater Vehicles) and USV (Unmanned Surface Vehicles) to support the service as proposed in figure 1 (Bruzzone 2013). Moreover, the Simulation Team developed an innovative mobile solution, named SPIDER (Simulation Practical Immersive Dynamic Environment for Reengineering) that allows to experiment different combinations of real equipment and virtual immersive framework (Bruzzone et al. 2016b). The SPIDER is an innovative compact and scalable CAVE (Cave Automatic Virtual Environment) based on modular units designed to fit within a standard 20'x40' High Cube Container (just 2m x 2m x 2.6m) as proposed in figure 2. The SPIDER solution is easy to be federated with other interoperable simulator and the surface of the faces are interacting by touch screen and/or tracking pencil, guaranteeing fully immersive experience (e.g. sound, vision, vibrations and acceleration, touch).

Along last months, the SPIDER has been used to support R&D in defense, marine applications as well as transportation and logistics. In this last case, the BBBUS Simulator (Box Bull Bus Simulator), developed in close cooperation with Central Labs University of Cagliari and designed to support training, demonstrated its capabilities to support also procedural training by using the same virtual reality framework as happens in other crane simulators (Bruzzone et al. 2011).



Figure 3: Head Mounted Display in Virtual Prototyping

During the last years, also other solutions have been experimented by Simulation Team Partners for supporting AR and VR in multiple applications from Virtual Prototyping of Cranes to Electrical Boards and Marine Solutions (Bruzzzone et al 2010).

Figures 3, 4, 5, 6 show different combinations of AR/VR devices during experimentation carried out by Simulation Team Members (e.g. DIME, Liophant, MAST, MSC-LES, CAL-TEK). They include Head Mounted Display (HDM), VR Googles and Glasses, AR Tablets and integrated HDM/CAVE

4. SERVICE DISTRIBUTED SYSTEMS

The researches carried out by Simulation Team identified a specific case study in service for distributed systems. In particular the analysis addressed tanks, containers and equipment used to provide O₂ to industrial operators, Health Care Infrastructures and individual patients at home. In this case, the components/systems moves from simple O₂ tanks to cryogenic containers and to respiratory devices. In addition the application field include industries, public institutions and consumers.

4.1. Gas Container Service

Indeed the Gas Container business sector involves big enterprises and include management and logistics of several hundred thousand containers of different type and size. This issue is even more crucial in Healthcare Sector. Given the average cost of a single unit, the container stock represent a capital for the companies; furthermore the containers, in particular gas cylinders, are easily lost or stolen and require to be controlled over wide regional distribution. Currently, it results pretty interesting from major companies active in providing these service to investigate new solutions and architectures to track cylinder locations in the gas container business. This aims to reducing cylinder losses, improving containers managements and stock controls. In addition, the producers of medical ventilators result interested in developing new capabilities to remotely assist patients increasing direct supervision and reducing the manning required to maintain the apparatus.



Figure 4: VR based on Oculus and Game devices

Respect old solutions, currently the use of innovative technologies based on web applications, cloud approach and enabling technologies in terms of markers and trackers allow to create really new and ad hoc architectures supporting good scalability and flexibility, as well as reusability in a wide spectrum of applications. In this sense the use of AR and VR creates synthetic environments to represent the current situation for management as well as training and educational activities. The authors are currently investigating combined solutions to propose a specific configuration and a road map for developing a prototype and a pilot experimentation on a set of gas containers.

The research aims to finalize a feasibility study on configurations and architectures able to track and monitor medical gas containers and ventilators. This research is focused on creating a tailored solution to guarantee overall performance and benefits, taking into account constraints and reliability of the adopted elements. Indeed the design of the configuration will be based on the following functions: capability estimation, costs and development times for virtual prototyping, tailoring and engineering, integration and implementation, verification, validation and testing.

The proposed solution is currently based on the integration, within a cloud, of different technologies such as AR, VR, Markers, RFIDs, Trackers, Webcams, Cloud Connectivity, Smart Phones and Web Applications to provide distributed services, tracking and training for medical ventilators, gas containers such as cylinders and cryogenic containers. The solution is expect to be available to support logistics and reduce risk of losses as well as to assist the patients remotely, limiting the need of qualified operators and supporting planned and preventive maintenance. In addition this solution will be flexible for being extended to other sectors where the containers are used (e.g. industrial application for O₂)

One major achievement is using the Augmented and Virtual Reality for remote support of the service, training and supervision of the equipment within the proposed architecture. This further increases the benefits of the architecture and prepare a large number of operators through distributed training.



Figure 5: Solution based on tabled and AR



Figure 6: AR solution based on web application

5. VV&T

It is fundamental to conduct integrated Verification, Validation and Testing (VV&T) of the proposed solution, in order to guarantee its effective application. From this point of view, it is proposed to use the AR and VR environments to conduct tests with Subject Matter Experts (SME) and to evaluate reliability, effectiveness and efficiency. Also, it is proposed to adopt different techniques during the development of the proposed solution (e.g. walk through, face validation, virtual examples). The dynamic testing on the virtual simulator and the use of Augmented Reality, on real system and mockups, guarantees to perform early-phase tests, supporting the adoption of Virtual Prototyping approach. The authors are currently working on developing an framework to be reused over different systems and component for conducting multiple experimentations. Dynamic validation on the experimental case is based on the application of Design of Experiments to identify critical factors and reliability of the proposed solution (Montgomery 2008). Moreover, the authors are considering the use of smart phones and tablets as main support for AR/VR on the field and specific testing activities are on going for identifying the hardware/software platforms to be used

by the operators. The dynamic performance measurements on the proposed architecture is conducted, until now, mostly in laboratory, waiting to move to an external pilot test in synergy with a major industrial operator. Indeed a specific complex element to be evaluated is the capability of this approach to be resilient respect non-cooperative behaviors, carried out by external people and final users, that possibly damage or alter the distributed systems to be maintained and its components.

CONCLUSIONS

The paper proposes a preliminary research about the use of Virtual and Augmented Reality for the remote service of components and systems. The proposed architecture guarantees the possibility to support local operators as well as remote supervision. In addition, the approach provides a very effective framework for training and education for staff to be involved in service both remotely and in main control rooms. This approach is expected to reduce the cost of the distributed components/systems and to increase their reliability. This will lead to further reinforce the outsource services and the competitiveness of operators. The specific case related to the O₂ Containers and equipment is a good example where the use of this technologies demonstrated to be very promising. The authors are working to develop a first pilot to demonstrate and promote this research track on the field in cooperation with a leading industry of the sector.

REFERENCES

- Ally, M. (2009) "Mobile learning: Transforming the delivery of education and training", Athabasca University Press, Canada
- Azuma, R. T. (1997) "A survey of augmented reality", in *Presence: Teleoperators and virtual environments*, MIT Press, 6 (4), 355-385
- Bluemel, E., Hintze, A., Schulz, T., Schumann, M., & Stuering, S. (2003) "Perspectives on simulation in education and training: virtual environments for the training of maintenance and service tasks", *Proceedings of the 35th conference on Winter simulation*, New Orleans, USA, December, pp. 2001-2007
- Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996) "Psychophysiological responses to changes in workload during simulated air traffic control", *Biological psychology*, 42(3), 361-377
- Bruzzone A.G., E.Page, A.Uhrmacher (1999) "Web-based Modelling & Simulation", SCS International, San Francisco, ISBN 1-56555-156-7
- Bruzzone A.G. (1999) "Interactive Web-Based Tools for Education in Industrial Management", *Proceeding of ITEC99*, The Hague (NL), April
- Bruzzone A.G., Simeoni S. (2002) "Cougar Concept and New Approach to Service Management By Using Simulation", *Proc. of ESM2002*, Darmstad Germany June 3-5

- Bruzzone A.G., Revetria R., Massei M., Simeoni S., et al. (2004a) "Models for the Introduction of Mobile Technologies in External Logistics", Proc. of ASTC2004, Arlington VA, April
- Bruzzone A.G., Viazzo S., B.M. (2004b) "Massive Training based on Virtual Reality Equipment Applied to Logistics and Heavy Haul Tracking", Proc. of SCSC2004, San Jose', CA, July
- Bruzzone A.G. (2009) "Serious Games for Training and Education on Defense Against Terrorism", Invited Speech at NATO M&S Group - MSG-069, Bruxelles, Belgium October 15, 16
- Bruzzone, A. G., Fancello, G., Fadda, P., Bocca, E., D'Errico, G., & Massei, M. (2010) "Virtual world and biometrics as strongholds for the development of innovative port interoperable simulators for supporting both training and R&D" *Int.Journal of Simulation and Process Modelling*, 6(1), 89-102.
- Bruzzone A.G., Fadda P, Fancello G., Massei M., Bocca E., Tremori A., Tarone F., D'Errico G. (2011) "Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes", *Simulation* October 2011, vol. 87 no. 10, p. 857-874, ISSN: 857-874
- Bruzzone A.G. (2013) "New Challenges for Modelling & Simulation in Maritime Domain", Keynote Speech at SpringSim2013, San Diego, CA, April
- Bruzzone A.G., Massei M., Tremori A., Longo F., Nicoletti L., Poggi S., Bartolucci C., Picco E., Poggio G. (2014) "MS2G: simulation as a service for data mining and crowd sourcing in vulnerability reduction", Proc. of WAMS, Istanbul, September
- Bruzzone A.G., Massei M., Cianci R., Longo F., Agresta M., Di Matteo R., Maglione G.L., Murino G., Sburlati R. (2016a) "Innovative Simulation for Scenario Analysis and Operational Planning", Proc. of SCSC, Pasadena, CA, USA, July
- Bruzzone A.G., Massei M., Agresta M., Di Matteo R., Maglione M., Murino G., Turi M., Cappelletti C. (2016b) "Modeling and Simulation as Support for Development of Human Health Space Exploration Projects", Proc. of EUROSIM, Oulu, Finland, September 12-16
- De Crescenzo, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., & Salti, S. (2011). Augmented reality for aircraft maintenance training and operations support. *IEEE Computer Graphics and Applications*, 31(1), 96-101.
- Del Rio Vilas, D., Longo, F., Monteil, N.R. (2013). A general framework for the manufacturing workstation design optimization: A combined ergonomic and operational approach. *Simulation*, 89 (3), pp. 306-329.
- De Sa, A. G., & Zachmann, G. (1999). Virtual reality as a tool for verification of assembly and maintenance processes. *Computers & Graphics*, 23(3), 389-403
- Ferrington, G., & Loge, K. (1992). Virtual Reality: A New Learning Environment. *Computing Teacher*, 19(7), 16-19
- Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and safety in Health care*, 13 (suppl. 1), i2-i10
- Giribone P., Bruzzone A.G., Devoti A., Cuneo R., Folco G., Testa S. (1996) "Surgery Planning by Using Simulation : a Step Forward to Built-Up a General DSS for Hospital Management", Proc. of ESS'96, Genoa, October 24-26, Italy
- Haritos, T., & Macchiarella, N. D. (2005). A mobile application of augmented reality for aerospace maintenance training" Proc. of 24th IEEE Digital Avionics Systems Conference (Vol. 1, pp. 5-B)
- Lee, K. (2012). Augmented reality in education and training. *TechTrends*, 56(2), 13-21.
- Longo, F., Nicoletti, L., Chiurco, A. (2013). Cooperative training for ships and tugboats pilots based on interoperable simulation. *Proceedings of the 25th European Modeling and Simulation Symposium, EMSS 2013*, pp. 695-703
- Moline, J. (1997). Virtual reality for health care: a survey. *Studies in health technology and informatics*, 3-34
- Monahan, T., McArdle, G., & Bertolotto, M. (2008). Virtual reality for collaborative e-learning. *Computers & Education*, 50(4), 1339-1353
- Mosca R., Giribone P. & A.G.Bruzzone (1995) "Simulator Object-Oriented Modeling for Educational purposes in the Industrial Plant Sector", Proc. of WMC95, Las Vegas, Nevada (USA), January 15-18
- Mosca R., Giribone P. & Bruzzone A.G. (1996) "Virtual Reality as a Support for Designing Fire-Fighting Standing Operating Procedures", Proc. of Simulation for Emergency Management, New Orleans LA, April 8-11
- Mosca R., P.Giribone, A.G. Bruzzone (1997) "Information Technology and Artificial Intelligence to Support Production Management in Industrial Plants", Proc. of AI'97, Innsbruck, Austria, February 18-20
- Orlansky, J., Dahlman, C. J., Hammon, C. P., Metzko, J., & Taylor, H. L. (1994) "The Value of Simulation for Training", Technical Report of Institute for Defense Analyses Alexandria VA, No. IDA-P-2982
- Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional science*, 23(5-6), 405-431
- Raybourn, E.M. (2012) "Beyond Serious Games: Transmedia for more Effective Training & Education", Proc. DHSS2012, Rome, Italy
- Stone, R. (2001). Virtual reality for interactive training: an industrial practitioner's viewpoint. *International Journal of Human-Computer Studies*, 55(4), pp. 699-711
- Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using

virtual reality technology for aircraft visual inspection training: presence and comparison studies. *Applied Ergonomics*, 33(6), 559-570

Wilson, P. N., Foreman, N., & Stanton, D. (1997). Virtual reality, disability and rehabilitation. *Disability and rehabilitation*, 19(6), 213-220

Zajchuk, R., & Satava, R. M. (1997). Medical applications of virtual reality. *Communications of the ACM*, 40(9), 63-64

Author's Index

AbouRizk	39		
Accorsi	277		
Achkar	204		
Affenzeller	107	117	56
Agresta	361		
Ali	32		
Altuntaş	162		
Andreas	56		
Aviles	87		
Axelsson	249		
Baltruks	10		
Bär	204		
BarrosdeOliveira	94		
Bautista-Martínez	224		
Belkin	141		
Bogner	256		
Bolshakov	328		
Borges	296		
Bottani	277		
Bruno	355		
Bruzzone	148	355	361
BuenoBenito	328		
Capel	1		
Carrasco-Limón	1		
Cebral-Fernandez	191	197	
Centobelli	277		
ChavesAlmanza	263		
Chiurco	355		
Cornú	204		
CorreiaNeto	94		
Crespo-Pereira	191	197	
CruzMoreira	94		
daCunhaSantos	94		
DeFelice	148		
deJesúsPortillo-Villasana	224	242	
Delahaye	67		
Dias	231		
Diviš	100		
ElBoukili	314		
Elduque	338		
Elmir	249		
Falcone	148		
Fikejz	128		
Fleck	107		
Flores	288		

FloresDeLaMota	282			
FloresMonroy	122			
Fokaides	183			
Franzinetti	361			
Fusto	355			
Garcia-del-Valle	191	197		
Ginters	10			
Glasa	155			
Golovin	141			
GomesCunha	94			
Gonçalves	328			
González-Hernández	168			
GonzálezManzanares	87			
Gracia	338			
GuerreroCoronel	87			
Günthner	15			
Hapoğlu	162			
Hennig	309			
Hofmann	134			
HuertaBarrientos	282			
Huerta-Barrientos	168	175	224	242
Iakovlev	141			
Ibarz	338			
Javierre	338			
Karder	117			
Kartak	78			
Kavicka	25			
Kavička	100			
Kim	39			
Kommenda	107			
Kylili	183			
Lang	134			
Langer	134			
LaradePaz	288			
Li	48			
Lienert	15			
Lin	48			
Liu	48			
Longo	148	355	348	
Lopez-Rojas	249			
Ma	48			
Macarulla	296			
Maglione	361			
Martins	328			
Martins P.	231			
Massei	361			
Méndez	204			

Merkuryev	10	
Merkuryeva	328	
Merta	128	
Mohamed	32	39
Montanari	277	
Monzon	348	
MujicaMota	67	
Mušič	216	
Nicoletti	348	355
Novotny	25	
Oborník	128	
Okolnishnikov	212	323
Oliveira	231	
Padovano	348	361
Pereira	231	
Petrillo	148	
Pluhacek	303	
Prante	107	
Raggl	56	
Raska	62	
Reggelin	134	309
Rinaldi	277	
Rios-Prado	191	
Rossainz-López	1	
Rouco-Couzo	197	191
Rudometov	212	
Ruiz	296	
Sadauskiene	183	
Sakne	10	
SameiroCarvalho	231	
Sánchez-Rinza	1	
Scala	67	
Scheibenpflug	117	
Seduikyte	183	
Senkerik	303	
Sigov	141	
Sinoviev	212	323
Sipkova	155	
Solis	348	
Sorrosal	296	
Sosa	270	
Starodubov	323	
Tobajas	338	
Tricoire	56	
Tunca	256	
Tyschuk	141	
Ulrych	62	

Vera-Morales	175	
Vetrano	355	
Vieira	231	
Vignali	355	
Viktorin	303	
Voracek	73	
Wagner	117	
Weigert	134	
Weisenpacher	155	
Wellens	270	263
Wiedemann	334	
Wiesinger	256	
Wolff	309	
Yang	48	
Zelinka	303	
Zhuravlev	212	
Zong	39	